Determination of headways and vehicle capacities in public transport networks using a dynamic assignment model

Master Thesis
Stefan Glück
August 2017

TU Delft

Determination of headways and vehicle capacities in public transport networks using a dynamic assignment model

By

Stefan Manfred Glück

In partial fulfillment of the requirements for the degree of

Master of Science

At the Delft University of Technology,

Faculty of Civil Engineering and Geosciences,

Department 'Transport and Planning'

Examination date: 22 August 2017

Student number: 4518152

Graduation committee:

Prof. Dr. Ir. Bart van Arem	Chairman, TU Delft (CEG)
Dr. Oded Cats	Daily supervisor, TU Delft (CEG)
Dr. Ir. Bruno F. Santos	Supervisor, TU Delft (AE)
Dr. Henk Post	External supervisor, Transdev/Connexxion NL

An electronic version of this thesis is available at http://repository.tudelft.nl/.

PREFACE

This thesis is the result of an eight-month research project at the Delft University of Technology. With this work, I finalize my masters program in civil engineering with the specialization in ‘Transport and Planning’. The two years in Delft and especially the final thesis project were both challenging but also very rewarding. I am happy and proud that it is now coming to an end. But I also have to say that without the help and support of certain persons – both at the personal and professional level – I could not have reached this point. Therefore, I would like to thank all people that have directly or indirectly contributed to this.

First and foremost, I want to express my gratitude to my daily supervisor, Oded Cats, for his kind, supportive and highly appreciated advice and guidance throughout the entire project. Without his expertise, innovative ideas, inputs and resources, this work would surely be of a lower quality. Also, his flexible and open attitude towards means of communication and organization regardless of time and place provided me with a lot of freedom which I really appreciated.

Next, I would like to thank Henk Post for providing me with valuable data for the case study. He put a lot of work in preprocessing and organizing the data for me which made it way easier for me to use them. Also, I really appreciated his advices regarding mathematical issues.

Furthermore, I want to thank Bruno Santos for his valuable feedback and comments on my work. As a researcher in the development of analytical decision tools for airline operations and planning, he could contribute a lot this project especially regarding optimization methods. I hope that this thesis will also benefit his scientific work and maybe evoke some fruitful thoughts.

Of course, I also want to thank the chairman of the graduation committee, Professor Bart van Arem, for his guidance and valuable feedback during the committee meetings.

Finally, I would like to thank my family, friends and partner for accompanying me during this process, listening to my thoughts and worries and giving me advice, but also for taking me out of myself and helping me to relax and forget about work. To my fellow students who have accompanied me during the last two years and became good friends: Thank you guys for the great support and the amazing time!

Stefan Glück, August 2017

EXECUTIVE SUMMARY

Background, research motivation and objectives

The provision of efficient and attractive public transport services is a possible sustainable solution to mitigate existing problems resulting from excessive motorized traffic especially in densely populated urban environments.

A crucial decision that needs to be made when planning public transport services is the determination of frequencies (or headways) and vehicle capacities deployed on certain lines. That is, it needs to be decided how often a service runs per hour and what kind of vehicles are used. This tactical decision is usually made before any detailed planning at the operational level such as vehicle/crew scheduling occurs. Since demand conditions may change in the longer term, a revision of these general decisions may be required on a regular basis.

In practice, authorities or operators typically use predefined service standards such as maximum vehicle occupancy rates as the basis for setting frequencies and vehicle capacities, while combining this action with experience, judgement and passenger counts. Recent advancements in research and computation power, however, lead to the development of advanced optimization models that can consider passengers' response to a new supply setting by incorporating public transport assignment models which can simulate and forecast the behavior of travelers. These methods enable a more anticipatory planning and dimensioning of supply than if service was merely adjusted to prevailing demand conditions. Yet, all methods developed so far use static assignment approaches. That is, travelers are assumed to make decisions based on average supply conditions and performance indicators can directly be computed from the given supply and passenger flows without taking into account the dynamic interaction between demand and supply. However, especially in highly-utilized networks, these dynamic interactions may cause severe issues regarding the reliability and overall performance of a service.

Dynamic public transport assignment models make use of simulation techniques that enable a detailed emulation of dynamic interactions between demand and supply. In contrast to conventional static models, they can capture congestion effects such as denied boarding, deteriorating comfort onboard a crowded vehicle as well as service headway fluctuations resulting from riding and dwell time variations.

To the best of the author's knowledge, no study has so far investigated the use of dynamic assignment models as part of a decision tool for tactical service planning. Although some scientific works on frequency and/or vehicle capacity optimization included certain congestion effects, none of them considered the implications of crowding on overall service reliability in terms of supply variations. That is, running times of vehicles are assumed to be constant and headways perfectly regular. The use of a stochastic public transport simulation model can overcome these shortcomings and thus fill a gap in current research.

Given this motivation, the objective of this thesis is the development of a model as tactical decision tool for frequency and vehicle capacity determination that is able to fully capture the dynamic behavior of demand and supply components in public transport networks. This objective is reflected in the following main research question:

How can dynamic public transport assignment models and search algorithms be combined in order to find optimal network supply conditions in terms of line frequencies and vehicle capacities given the objectives of passengers and operators?

The second objective of the thesis is to test and show the practical applicability and implications of the developed model. To this end, it also needs to be investigated to what extent the model can yield benefits when applied to problems of real scale. Thus, the second key research question is:

What are the practical implications and benefits of the proposed model when applied to a case study?

Methodology

The modelling framework of the proposed headway and vehicle capacity determination model consists of three main components (see Figure 0.1): (1) a dynamic public transport assignment model , (2) an evaluation model that determines the performance of a potential solution with respect to an objective function using the output of the assignment model and (3) a search algorithm that selects feasible solutions to be provided as input to the assignment model for further evaluation. Thus, the entire process of selecting and evaluating possible solutions is iteratively executed until a certain termination criterion is fulfilled, meaning a final solution was found. The algorithm is initialized by a starting point in terms of an initial feasible solution.

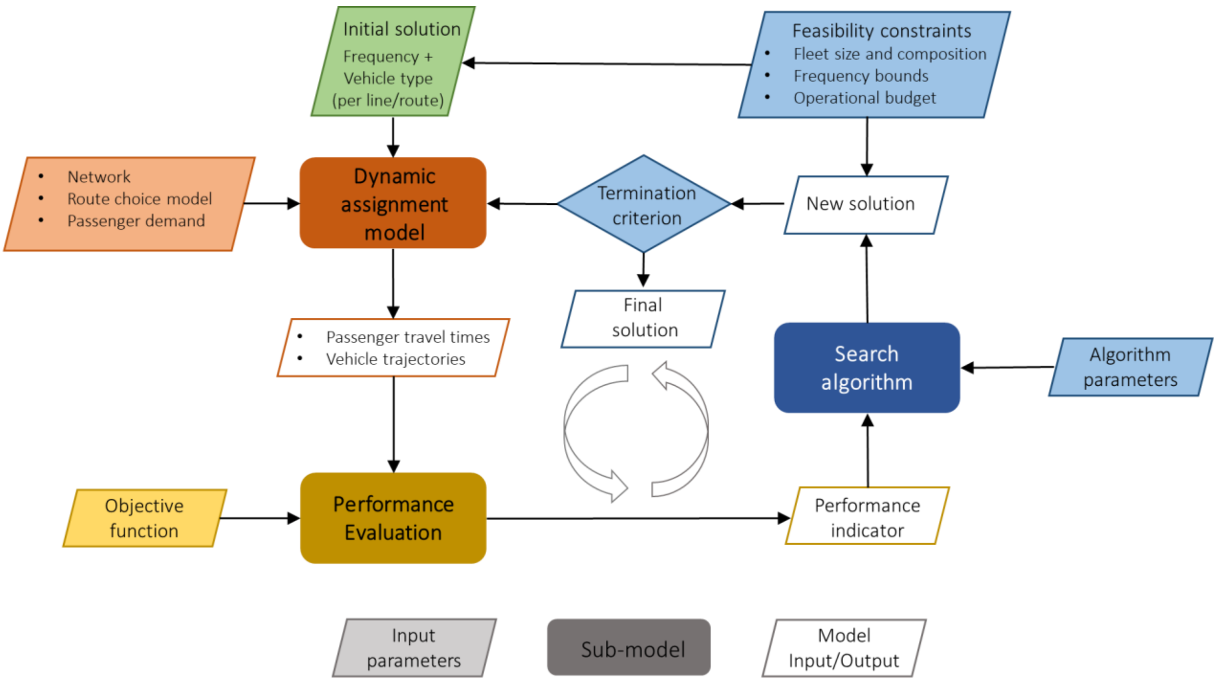


Figure 0.1: Basic framework of the headway and vehicle size determination model.

BusMezzo is a dynamic public transport operations and assignment tool designed to support the analysis and evaluation of Advanced Public Transport Systems in terms of operation, planning and control. The mutual interactions of vehicles and passengers in BusMezzo are explicitly modelled using an agent-based simulation approach. Therefore, all dynamic effects on assignment results related to congestion and service reliability can be explicitly considered by BusMezzo which makes it an ideal implementation of a dynamic assignment model. Passengers are modelled as individual decision makers who can dynamically choose their routes through the network depending on current conditions. The generation of passenger agents at stops is done stochastically, following a Poisson process, and mean generation rates are assumed to be constant (i.e. inelastic towards changes in supply). That is, it is assumed that passengers arrive randomly at a stop and do not adjust their arrival times based on a known schedule. This behavior is commonly seen in public transport networks offering high-frequency services. Moreover, an iterative network loading can be performed in the simulation that accounts for the day-to-day learning of passengers who may adjust their routes based on experienced service reliability attributes in terms of waiting times and on-board crowding levels for all tried paths. That is, experienced travel times and comfort levels may cause passengers to switch to alternative routes until those travel attributes converge to constant values. Using this feature results in network-wide steady-state conditions which can be regarded as an equivalent to the congested user equilibrium computed in conventional static assignment models.

On the supply side, BusMezzo simulates the movements of each individual vehicle through the network. The dynamic properties of stochastic running time variations, demand-dependent dwell times and onboard vehicle occupancy levels can be fully considered by the simulation model and enable the modelling of crowding and congestion effects. In each iteration of the optimization algorithm, a potential supply setting in terms of line frequencies and vehicle capacities is provided as an input to BusMezzo for simulation. That is, the headway of successive vehicle departures at a line's departure terminal and the type of vehicle in terms of seating and total capacity used on that line is defined. A line may refer to both directions of a route or each direction may be treated separately.

Each potential solution needs to be evaluated with respect to the objectives of both passengers and operator. Therefore, the objective function computes the total system costs as the sum of costs to be borne by the users and the operator of the system. The former costs are computed by the value of time and the total generalized travel time resulting from a certain supply setting. Travel time components such as waiting and in-vehicle times are given by the output of BusMezzo. Operational costs are estimated based on the total number of vehicles needed per type and the total travel distance covered by the vehicles. The model also allows defining alternative objective functions, for instance the minimization of user costs subject to an operational budget constraint.

A solution in terms of line frequency and vehicle capacity is generated by selecting each of the two decision variables from a finite discrete set of predefined values. This allows to constrain the problem and also enables the optional consideration of one type of decision variable only in case the other set contains only one value. The number of possible solutions increases exponentially with the number of lines (or route variants), hence resulting with a problem of large combinatorial complexity. Since a closed analytical formulation describing the mathematical relation between the decision variables and the objective function value is not available, conventional optimization methods such as gradient-based algorithms cannot be employed. Therefore, two versions of a search algorithm that solely use the objective function value are developed for solving this specific problem.

A local search algorithm is developed that finds locally optimal solutions by moving from one candidate solution to the other in the solution space following a descent path in terms of the objective function value. Thereby, the algorithm only considers the local feasible neighborhood of a solution and selects the best performing neighbor as the next candidate solution. A final solution is found when all feasible neighboring solutions cannot improve the current candidate, implying local optimality. However, it cannot be guaranteed that a global optimum is found since different starting solutions may lead to different locally optimal solutions due to the topology of the solution space.

A second search algorithm is developed based on the theoretical methodology of simulated annealing, a probabilistic metaheuristic. By sometimes accepting solutions that worsen the current objective function value, a trapping in locally optimal regions of the solution space might be avoided. In this way, local maxima can be overcome and the solution space can be examined more thoroughly, thereby increasing the probability of finding the globally optimal regions. The acceptance of an uphill move is stochastically simulated based on a probability value which is dependent on the relative difference between the current and the new solution as well as a parameter value (temperature) which monotonically decreases throughout the runtime of the algorithm. As execution time progresses, this acceptance probability gradually decreases which causes the algorithm to finally converge to a solution. This procedure, however, does not guarantee finding the global optimum.

Application

In the next step, the developed model is applied to two different case studies in order to investigate the model's behavior towards changes in certain input parameters and prove its practical applicability as well as identify benefits resulting from the findings obtained.

First, the model is applied to a hypothetical medium-sized public transport network consisting of four lines. Three tests are executed that aim at investigating the influence of certain input parameters on the search algorithms, these include: the starting solution defined to initialize the algorithm, the shape of the decreasing temperature parameter function in the model based on simulated annealing and the relative weighing of the waiting time component in the objective function.

Results indicate that the nature of the initial solution in terms of decision variable values provided as a starting point to the model has a significant influence on the characteristics of the final solution obtained by the search algorithm based on local search (see Figure 0.2). In case the method based on simulated annealing is used, no dependency can be observed. The parameter value influencing the shape of the cooling function in the latter algorithm has a direct effect on overall runtime of the model. It is clearly observable that a longer run time can lead to a qualitative improvement in terms of objective function value of the respective solutions found (see Figure 0.3). Finally, it can be shown that the relative weighing of certain terms in the objective function can significantly affect the final solution. In this specific case, an increase of the relative importance associated with the disutility due to the waiting time leads to solutions that decrease the average waiting time per passenger at the expense of increased operational costs.

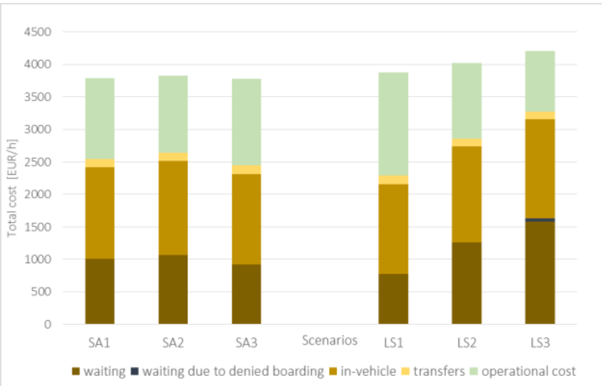


Figure 0.2: Total cost of the final solutions found for different scenarios (SA = simulated annealing; LS = Local Search; 1, 2, 3 = high, medium, low capacity starting solution)

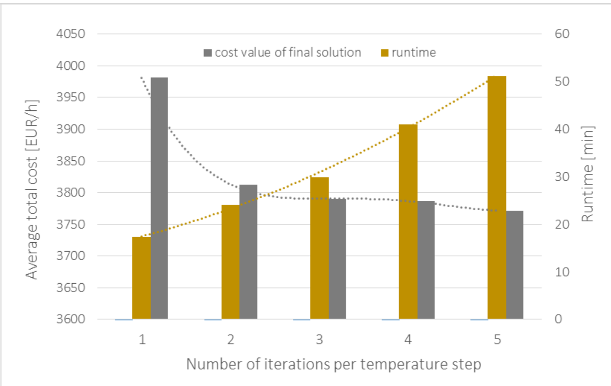


Figure 0.3: Cost value of final solution vs. runtime for different parameter settings of the SA algorithm.

In a second step, the developed model based on the simulated Annealing algorithm is applied to a real case study located in the 'Zaanstreek' area north of Amsterdam. The bus network consists of 6 different route variants (lines) and two different demand settings in the morning and the evening peak hours are analyzed. The model is applied to multiple scenarios involving different combinations regarding the formulation of the objective function, constraints on the decision variables and the considered passenger demand.

The results indicate that there exists a clear difference between the two demand periods examined regarding both the found decision variables and objective function values. A minimization of total costs (TC) can yield significant reductions of the total system costs by up to 1.6% compared to the current situation in the morning peak period (AM), whereas in the evening peak (PM), no significant reductions can be achieved. Moreover, in the former period, all solutions tend to reduce passenger-related costs at the expense of increased operational costs, while in the latter case two of the found solutions are close to the current total cost composition and one solution suggests to decrease operational costs which implies higher user costs. In all cases, the major factor which causes changes in user costs is a reduction in waiting time. A mere minimization of total travel costs (UC) subject to a current operational budget constraint yields significant passenger benefits in both peak periods as supply is increased up to the budget limit. The separate determination of line frequencies per direction of a route variant (ASYM) yields lower user costs compared to a conventional symmetric setting of frequencies (SYM) irrespective of the objective considered. A simultaneous determination of both vehicle capacities and line frequencies (VEHCAP) can even yield larger user benefits. Figure 0.4 below shows the found solutions in terms of user and operational costs for all scenarios investigated.

All in all, the obtained results are in line with the expectations on the outcomes of the scenarios. Using small vehicles can decrease the average operational costs per vehicle which allows to increase overall network capacity and thus reduce passenger-related costs. It proved clearly evident that lower user costs are generated at the same amount of operational expenses when determining frequencies separately per line direction than if setting them equally in both directions. The fact that the potential of improvement of the current supply provisions is greater in the morning than in the evening peak is also in line with previous expectations. Notwithstanding, results emphasize the quality of the current service provision in the evening peak from a total cost point of view and this also proves that the model finds realistic solutions.

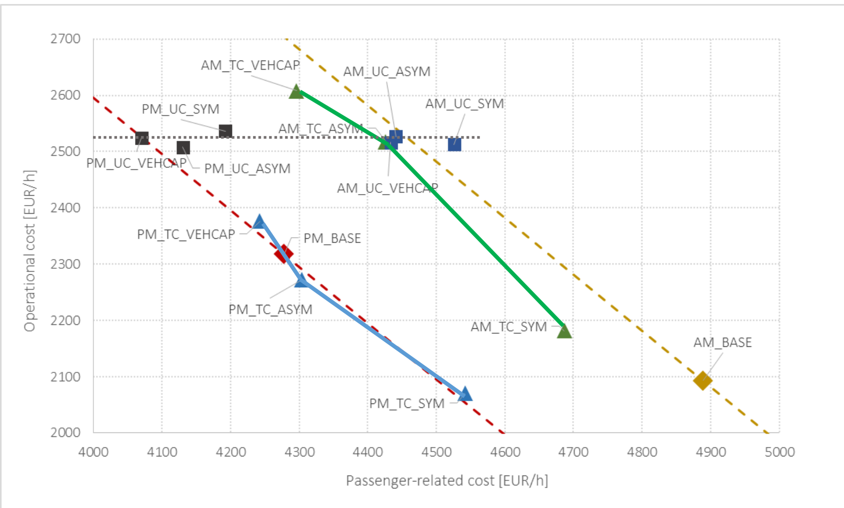


Figure 0.4: Overview of the performance of all solutions found for the different scenarios in terms of associated passenger-related and operational costs.

Conclusions and recommendations

The main scientific contribution of this study is the successful implementation of a dynamic public transport assignment model within the framework of a tactical decision tool for the simultaneous determination of line frequencies and vehicle capacities. To this end, a search algorithm based on the method of simulated annealing and combined with BusMezzo, an agent-based assignment model, was developed. The integration of BusMezzo allows to fully consider the dynamic interaction between demand and a potential supply setting and the resulting consequences on overall system performance. From a practical point of view, an application of the proposed model at the tactical planning level may lead to a better operational performance of the transport system than if a conventional method was used since operational issues such as service reliability are already explicitly taken into account during this early planning phase. This makes an application of the tool particularly beneficial in case of highly-utilized and potentially crowded and congested public transport networks.

The tests on certain model input parameters have clearly proved the superiority and advantages of the simulated annealing method over the method based on local search. The results indicate the presence of multiple locally optimal solutions in the specific problem and thus confirm and justify the suitability of the application of the SA methodology. Moreover, it was found that a longer execution time of the SA algorithm, meaning a more intensive search, can lead to an increase in the quality of found solutions. A third test revealed that the relative importance attached to a certain aspect in the objective function can significantly affect the final solution found by the model. Hence, it is advised to pay special attention towards the formulation of the objective function when defining the model input parameters.

The application of the model to a case study of real size proved its practical applicability and revealed solutions that can beneficially improve the current situation. Overall results indicate that the potential of improvement of the current supply provision is largest for the considered morning peak hour. During this period, significant travel cost savings can be generated by a change of the current supply and it is thus advised to the incumbent operator to increase overall supply provision during this period. In the evening peak, however, a change of the current situation is not necessary from a total system costs' point of view. This result confirms the quality and optimality of the current situation given the prevailing demand conditions.

Furthermore, the results clearly highlight the advantages of asymmetric service provision during periods of directed passenger demand which is currently present in the regarded network. The use of asymmetric frequency settings can lead to a more effective satisfaction of the present demand and also confirms the suitability of the current asymmetric supply setting in the evening peak. Moreover, a simultaneous optimization of vehicle capacities and line frequencies proved the benefits of deploying a mixed vehicle fleet on the regarded network in both periods considered.

Next to the practical benefits, also some limitations and shortcomings of the developed model were identified. A major limitation of the developed tool is the negligence of vehicle scheduling since each vehicle is simulated for one trip on a specific line only. Consequently, potential delays being present at the destination terminal of a line cannot affect the punctuality of the following departure from the terminal which implies that effects of the propagation of delays and degraded service reliability among multiple lines and route directions are not accounted for. Moreover, operational costs are computed based on an estimated fleet size rather than an exact fleet size value resulting from vehicle scheduling. Hence, in order to improve those limitations and further increase the practical utility of the developed tool, future research should examine the integration of a vehicle scheduling model into the present modelling framework.

Another limitation of the proposed model relates to the disregarding of passenger demand elasticity towards changes in supply which lead to changes in generalized travel costs. In the present model, demand is assumed to be a constant input in terms of (mean) passenger flow rates per OD pair. In reality, however, this factor may be influenced by changes in supply provision as those may cause effects related to modal shift and/or induced demand. To account for this, an additional feedback loop needs to be implemented into the modelling framework that adjusts the OD matrix based on the relative changes of generalized travel times per OD pair and a demand elasticity function. This would allow to investigate crucial practical topics such as forecasting ridership growth and resulting additional revenues or supply optimization given the objective of operator profit maximization or subsidy minimization.

Areas of research in which the model could be applied beyond the tactical level are the strategic network design and tactical supply determination during special events. In the former case, the model could be applied to a network consisting of all potential lines. Those lines resulting in zero or very low frequencies could then be removed from the set of attractive routes. Running the model on a modified network or special demand configurations in case of special circumstances such as construction works or big events can create valuable outputs which can be used as a tactical basis for predefined service plans. Future research using the model as a methodological basis could investigate the capabilities and suitability of the developed framework for the design of automated public transport services. In this context, aspects that differ from conventional public transport services and thus require further investigation may include the negligence of on-board staff costs and changes in passengers' perceptions as a result of vehicle automation. Furthermore, future research should quantify and validate the added value of the proposed model by conducting a comparative analysis between the developed tool and a conventional public transport supply optimization model using a static assignment approach. In this way, the practical advantages and disadvantages of the proposed model compared to the state of the art could be identified on a generalizable level as well.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

This chapter provides an introduction to the present study, starting with a section describing the background and scope of the topic as well as the research motivation. The following sections elaborate on objectives and research questions and present the approach that is followed to meet the objectives. Finally, the chapter closes by presenting the structure of the thesis report.

1.1 Background and research motivation

Especially in densely populated urban environments, issues caused by motorized traffic such as congestion, noise and air pollution and conflicts with vulnerable road users have been increasing in many places all over the world for the past decades. One way to mitigate these issues is the provision of public transport systems offering high levels of service to its users and thus making people switch from private motorized modes of transport to more sustainable public transport modes. Therefore, public authorities and operators should promote the development of high-quality public transport systems in order to improve people's overall quality of life.

Providing attractive public transport services is a challenging task which involves decision making on various supply elements ranging from short-term day-to-day operational strategies to long-term network planning. The final goal is the provision of an attractive, reliable and comfortable service that meets users' needs and is still economically efficient in terms of operation. The frequency setting and fleet size determination problem can be regarded as a tactical planning decision at the medium term horizon of a few weeks up to months or years. In order to maintain a good level of service or reduce operational cost by cutting excessive and underutilized supply it is necessary to regularly adjust frequencies and vehicle capacities on public transport lines to demand variations along different seasons of the year or times of day.

In practice, public transport agencies typically use service standards such as crowding levels, allowed maximum standees, and upper (policy) and lower limits on headways as the basis for setting frequencies and vehicle capacities, while combining this action with experience, judgement and passenger counts (Furth & Wilson, 1981). Recent advancements in research and computation power, however, lead to the development of advanced optimization models that can consider passengers' response to a new supply setting by incorporating assignment models which can simulate and forecast the behavior of travelers. Using this approach, a more anticipatory planning and dimensioning of supply would be possible than if service was merely adjusted to prevailing demand conditions. Yet, there is still room for improvements of these new models since all methods developed so far use static assignment approaches. That is, travelers are assumed to make decisions based on average supply conditions and performance indicators can directly be computed from the given supply and passenger flows without taking into account the dynamic interaction between demand and supply. The impact of these effects on travel time reliability is not accounted for as average waiting and in-vehicle times are computed based on perfectly regular headways and average travel times of vehicles. In practice, some static assignment models can partly consider crowding and congestion effects by computing user equilibrium conditions in the network based on experienced in-vehicle times that are adjusted by crowding factors. However, especially in highly-utilized networks, the dynamic interactions between demand and supply may not only affect onboard crowding levels but also cause severe issues regarding the reliability and overall performance of a service for instance resulting from passenger flow-dependent dwell times. Therefore, the inclusion of dynamic system behavior is desirable when making tactical decisions such as frequency and capacity determination. To the best of the author's knowledge, no study has investigated the use of dynamic public transport assignment models for tactical planning purposes so far.

1.2 Objectives and research questions

Given the afore-mentioned research motivation, the main objective of this thesis is the formulation of a model as tactical decision tool for frequency and vehicle capacity determination which is able to capture the dynamic behavior of demand and supply components in public transport networks. This objective is reflected in the following research question:

How can dynamic public transport assignment models and search algorithms be combined in order to find optimal network supply conditions in terms of line frequencies and vehicle capacities given the objectives of passengers and operators?

A number of sub-questions can be formulated which help to gain supportive knowledge to answer the main question and structure the research approach:

- *How can potential solutions neighboring a given solution and satisfying certain feasibility constraints be generated?*
- *Which components should be included in the objective function for performance evaluation of a potential solution to incorporate both the interests of travelers and operators?*
- *How can this problem be solved using an efficient search algorithm that uses outputs from the simulation model and finds a feasible and well-performing solution?*

The second objective of the thesis is to test and show the practical applicability and benefits of the proposed model. Thus, the second research question is:

What are the practical implications and benefits of the proposed model when applied to a case study?

The related sub-questions are:

- *What is the effect of different starting solutions and search algorithm parameters on the quality of the final solution obtained?*
- *To what extent do the solutions and resulting performance indicators obtained by the model differ from a reference situation when applied to a real case study?*
- *How robust/sensitive are the model outputs against varying demand conditions?*
- *What are the practical limitations of the model?*

1.3 Approach

This section describes the approach chosen in order to meet the objective and answer the research questions. The research approach consists of three stages: a literature review, the model development and its application.

A thorough literature review facilitates the development of the model by gaining insight into the approaches proposed in previous studies. First, the superordinate topic of public transport service planning which forms a basis for the actual topic is briefly introduced. Thereafter, various older and more recent works on headway and vehicle size optimization are presented and classified according to their approaches. The basic strategies identified in the literature study can serve as a good basis for the present model. Finally, a comparative analysis and synthesis of the reviewed literature is executed in order to identify those basic strategies, review the historical development of modelling approaches within this specific domain, and also highlight concepts and elements which have been neglected or rarely treated so far and which the present study intends to address.

In the next stage, the actual model development takes place, which means finding an answer to the first research question. Figure 1.1 shows the basic methodology of the proposed model consisting of three main components which constitute the decision tool: A dynamic public transport assignment model that considers the dynamic interaction between demand and supply and its potential impacts on service reliability, a model that evaluates the performance of a potential solution by transforming the outputs of the assignment model into a suitable performance indicator, and a search algorithm that selects potential solutions. The optimization model aims at finding a well-performing solution by iterative search. Thus, the formulation of the optimization problem leads to the adoption of a suitable search algorithm that selects appropriate solutions in the neighborhood of a given solution. Special attention needs to be paid to the generation of a feasible set of solutions respecting the defined constraints. The performance of a solution is evaluated using an objective function which reflects both the interests of travelers and operators. After the model has been implemented, it needs to be checked whether it works as expected and produces reasonable results. Using a small example case study for model verification purposes facilitates the tracing of potential errors and bugs and overall interpretation of the produced results. Moreover, the effects of assumptions on certain model inputs can be easily tested and compared using a small network which allows for an exhaustive search, i.e. the enumeration of all possible solutions.

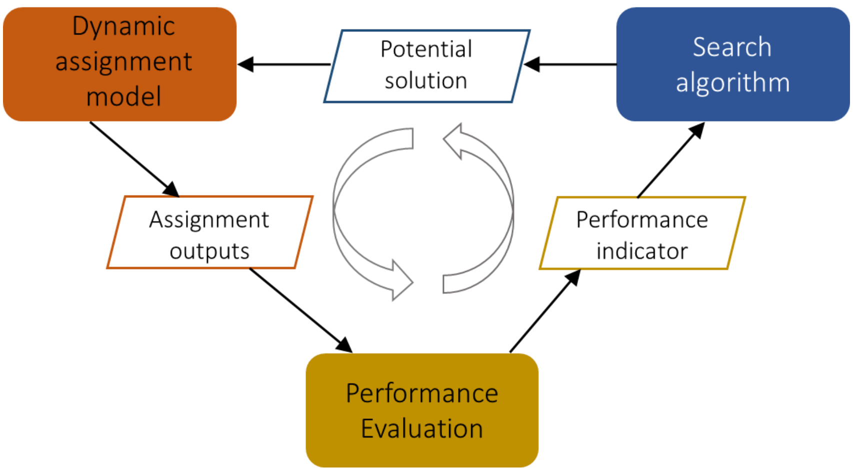


Figure 1.1: Basic components of the proposed model.

Finally, the proposed model is applied to case studies in order to show its practical implications. Since the behavior of the proposed model may depend on several external input parameters, different scenarios need to be executed under different assumptions on constraints, model input parameters and demand conditions. A real-world case study is used to compare the performance of an optimized supply with that of the current situation. The analysis of the model results for different cases and scenarios helps to identify the effects of different assumptions on model inputs as well as the practical limitations and finally provides an answer to the second research question. Based on the analysis of the obtained results, conclusions can be drawn and recommendations for practice and research can be made.

1.4 Thesis structure

The report of this study is structured as follows: Chapter 2 provides a comprehensive review on literature dealing with headway and vehicle size optimization approaches. It closes with a summarizing synthesis including a positioning of the present study within the current research. Chapter 3 gives a detailed explanation of the selected methods used to develop the model. The theoretical formulation of the model is introduced step-by-step and, finally, model verification results are presented. Chapter 4 presents practical applications of the developed model to case studies including discussions of the obtained results. Finally, Chapter 5 summarizes the findings of this thesis by giving answers to the research questions and providing recommendations for public transport authorities and operators as well as for future research projects. The final chapter concludes by a personal reflection on the work process and lessons learned during the thesis project.

2 LITERATURE REVIEW

This chapter provides a comprehensive review on literature which is relevant in the context of the present study. It starts with giving a brief introduction on the more generic and superordinate topic of optimal planning of public transport services. Thereafter, various studies presenting models and approaches for determining and optimizing vehicle capacities and headways in public transport networks are reviewed. The chapter concludes with summarizing the most important findings of the literature review and pointing out missing elements that the present study can address.

2.1 Public transport service planning

Designing and operating public transport networks is a very complex task involving various decisions on different supply elements which range from short-term operational strategies to long-term network planning. The overall goal is the provision of a public transport system that offers an attractive, reliable and comfortable service to its users and is still economically efficient in terms of operation.

In literature, the entire public transport service planning problem is usually subdivided in four phases or sub-problems (Ceder & Wilson, 1986 ; Ceder, 2007):

1. Public transport network design: Comprises the strategic determination of the locations of interchange, terminals and intermediate stops on a line as well as the line's route. Independent inputs and constraints may include land-use characteristics, authority constraints such as budget limits and current patronage.

2. Frequency setting and timetabling: Includes the tactical determination of service frequencies for certain periods of operation as well as the definition of arrival and departure times of vehicles at all stops. These decisions result in a required fleet size to operate the timetable based on known route travel time distributions.

3. Vehicle scheduling: In this phase, vehicles of a certain type are assigned to the trips determined in the previous phases. This results in an operational vehicle schedule and a required fleet size per type of vehicle.

4. Crew scheduling: Finally, drivers are assigned to planned vehicle trips and thus driver roosters are generated taking into account for example working regulations, payment structures and individual preferences of drivers.

All four phases constitute the public transport service planning problem and are highly interdependent since outputs obtained at one stage are provided as inputs to the following stage. A decision made at the frequency setting stage, for instance, may strongly influence the subsequent phases. Therefore, it would be desirable to simultaneously solve all stages of the problem in one step. However, due to the high complexity of each phase, an integral approach is difficult to realize in terms of computational time given the current state of computing power of standard PCs and algorithms available. Instead, each phase or a combination of multiple phases is solved separately and the entire problem can finally be solved using an iterative procedure which uses outputs in terms of a solution obtained for one problem as inputs to the subsequent problem. Ibarra-Rojas et al. (2015) provide an extensive review of scientific papers dealing with all sub-problems of the public transport service planning problem and also real-time control strategies. Especially in recent years, this research domain has gained substantial progress in terms of new types of approaches and models being developed.

2.2 Headway and vehicle size determination

This section provides a comprehensive review on scientific literature both dealing with headway and vehicle size determination in public transport networks. It starts with reviewing analytical models using exact optimization approaches which were mainly developed between the years 1980 and 2000. Thereafter, recent works on the issue incorporating sophisticated heuristic optimization methods are presented and elaborated on.

2.2.1 Exact optimization approaches

The network design, frequency setting and timetabling stages of the public transport service planning methodology described in Section 2.1 usually considers only one type of vehicle having a certain capacity. In practice, however, various types of vehicles having different capacities such as minibuses, articulated and double-decker buses or standard buses may be employed. Note that the decision variables frequency and vehicle capacity are strongly interrelated as a line segment's capacity in terms of maximum number of passengers transportable per hour is the product of both variables. That is, for a given line frequency and maximum load segment and desired average vehicle load factor a minimum vehicle capacity can be directly computed. The same holds for a given vehicle capacity and the resulting frequency.

The following two sections elaborate on analytical models developed to determine optimal headways and vehicle sizes using exact optimization approaches. Most of the models are only applicable to a single line without taking into account network effects.

2.2.1.1 Optimal headway

This section presents analytical models and optimization approaches for determining the optimal headway of a single public transport line. Note that all these methods only aim at determining a suitable supply in terms of frequencies and thus assume a given vehicle capacity. Models for determining the vehicle size are discussed in the following section (2.2.1.2).

One of the earliest models was proposed by Newell (1971) who derived an analytical expression for the difference in time (headway) between two consecutive vehicle departures by minimizing passengers' waiting time and a constant operating cost per vehicle for a single line. According to this formula, the frequency and the number of passengers per vehicle is proportional to the square root of the passenger arrival rate. The model is formulated for the case of vehicles with unlimited capacity and also for the case where a capacity constraint is present.

The work of Newell was later extended and refined by Salzborn (1972) who developed a model for determining a complete vehicle schedule resulting from an optimized frequency setting by splitting up the problem into two objectives. The primary objective is to determine the minimum fleet size for a single transportation route given a known passenger arrival rate and the line's cycle time. During peak it is assumed that the minimum fleet size/maximum headway results from a maximum load factor constraint: i.e. during peak, busses have to be loaded at maximum. In a second step, the actual bus departures are determined by minimizing passengers' waiting time subject to the fleet size constraint. In contrast to Newell (1971), Salzborn considers that busses can be used for more than one trip (trip chaining). The optimization problem is solved with the calculus of variations. Practical implications of the model are illustrated by means of a real-world metropolitan bus route in Adelaide, Australia, using observed passenger counts. The number of required vehicles and their departure times at certain control points are determined by the model. Moreover, it is shown how the developed model can be easily extended to optimize a pair of connected lines.

Ceder (1984) describes a method for the practical frequency determination of a single line using passenger count data and desired occupancy levels. In total, four methods are proposed which use two kinds of data: the maximum load at selected stops (point check) or a complete load profile of the entire route (ride check). A criterion is developed that helps to decide which type of method should be used and thus which type of data should be collected. The criterion uses the measure of the load profile density, which is the observed measure of total passenger-km (i.e. total ridership over the route), divided by the product of the length of the route and its maximum load. Lower profile densities indicating significant load variability among the stops suggest the use of the ride check procedure whereas the point check method would be appropriate in the case of a flatter load profile (high density).

All studies reviewed previously present a methodology for determining optimal headways for a single public transport line given an observed passenger demand. That is, the supply is optimally adjusted in response to prevailing demand conditions. However, in reality, demand is not independent from supply but rather distributes across a network depending on the supply characteristics of certain lines in the network. The study of Han and Wilson (1982) was one of the first considering these network effects, yet in a very simplified fashion. They proposed a model for allocating busses to a network with significant overlap between routes given passenger demand in form of an OD-matrix. The overall objective is to minimize waiting time and crowding levels subject to fleet size and demand satisfaction, i.e. capacity should be sufficient on any line segment. A simplified assignment model is proposed that assigns passenger flows to parallel overlapping lines proportional to the respective frequencies. The solution of the problem is decomposed into a base allocation procedure and a subsequent surplus allocation algorithm: The base allocation increases the frequency on each line until all constraints are satisfied, i.e. generating the supply condition with the minimum number of busses required; the surplus allocation procedure increases the frequencies on each line separately by minimizing overall crowding levels until the maximum number of available vehicles is reached. Practical application is illustrated by a simple network (a three-lines part of the bus system in Cairo) for which the algorithm was applied in a complete manual procedure.

2.2.1.2 Optimal vehicle capacity

This section presents exact optimization approaches for determining the optimal vehicle size of a public transport service. Note that a selected vehicle capacity has direct implications on the headway since both measures show a proportional relation assuming that a constant line capacity and an average vehicle load factor is given. That is, an optimally determined vehicle type directly results in an associated headway.

The general problem when determining the optimal vehicle size lies in the trade-off between the costs to be borne by the user and the operator of the system. That is, the overall aim is to minimize the social costs of the public transport system consisting of the generalized travel costs of passengers and the operating costs associated with a certain supply. This contrasts with the previously presented models for setting optimal frequencies since those are based on demand satisfaction rather than cost minimization. All of the analytical models mentioned below are derived in a similar fashion. The equation for optimal vehicle size computation is derived by first formulating a social cost equation and then setting the first derivative with respect to the vehicle size to zero. Depending on their theoretical assumptions the presented models differ in terms of complexity and amount of data required.

A very simple analytical formulation to compute the optimal vehicle size was proposed by Walters (1980), who claims that the optimal size which minimizes passengers' waiting time costs and wage vehicle costs is indirectly proportional to the total number of passengers. This implies that for increasing passenger volumes, vehicles sizes and corresponding headways decrease. In contrast to that, Gwilliam et al. (1985) attack this idea by establishing a similar expression which entails a direct proportional relationship between passenger demand and vehicle size. By assigning common values to the input variables, the authors show that operating much smaller buses than those commonly used is not justified.

A far more sophisticated model was developed by Oldfield and Bly (1988). Their assumptions are very detailed comprising various features and components. In contrast to most other models, demand is assumed to be elastic towards changes in the supply in terms of passenger trip costs according to a constant generalized cost elasticity value. Moreover, the influence of different vehicle sizes and resulting frequencies (i.e. flow of vehicles) on road congestion is modelled explicitly. In order to account for the influence of different demand levels on dwell times and thus operating speed, passengers' in-vehicle time is assumed to be dependent on the occupancy of the vehicles. The probability of denied boarding as busses become full is explicitly accounted for as well. Considering all these components, the model is capable of replicating the entire public transport system in great detail compared to other models within this domain. However, this comes at the price of an increased effort when gathering and calibrating required input data.

Unlike all models presented so far, Shih and Mahmassani (1994) proposed an approach that is applicable to a network of lines and not only to a single line. Demand in terms of total passenger trips and maximum link volume per line is not assumed to be constant but a result of the supply offered and determined by a simple public transport assignment model only taking into account line frequencies, travel times and number of transfers. The model aims at computing optimal frequencies and vehicle sizes by using an iterative procedure. First, an initial set of preliminary route frequencies is assigned to the network and a demand matrix defined at network level is assigned. Then, the optimal vehicle size per line is computed using the maximum link flow resulting from the passenger assignment and a predefined maximum load factor. An updated frequency resulting from the vehicle size and load factor is computed and reassigned to the network. The procedure is repeated until frequencies converge. Note that this approach does not aim at minimizing overall system costs but rather optimizes the vehicle sizes for each line separately. A real-world case study in Austin, Texas, is conducted to illustrate the practical implementation of the model.

An early work by Jansson (1980) paid special attention towards determining the optimal vehicle size to be used both in peak and off-peak periods of the day. By incorporating passenger peak to off-peak flow ratios and considering the mean flow rate over the whole day the author developed an expression that determines the optimal bus size to be used for the entire daylight service. Using that model, he showed that particularly in off-peak, social costs can be significantly reduced by employing smaller vehicles at higher frequency. Similar to the previous work, Lee et al. (1995) also developed a model that takes into account different periods of the day. The model is able to optimize vehicle sizes on multiple routes as well, i.e. it considers the entire network. The model also aims at analyzing the advantages of using different vehicle sizes on a line during the day by taking into account additional capital cost resulting from a mixed vehicle fleet. Numerical results show that operating a network with two different vehicle sizes instead of one single size is advantageous in cases where the demand in peak hours is approximately two times bigger than in off-peak periods. Using a mixed-fleet may lead to a lower overall fleet size than if a single vehicle size was used.

Gronau (2000) tackles the vehicle size determination problem from a different perspective by evaluating the viability of an unexplored option to offer different types of vehicles on the same route for travelers having different values of time. That is, small vehicles operating at high frequency should serve high-value-of-time passengers whereas larger vehicles can offer a supply to low-value-of-time travelers. He shows that the optimal decision whether to use one or two types of vehicles depends on the length of the route and the distribution of passengers' value of time. The longer the route and the more dispersed the distribution of the population, the greater becomes the tendency to use two types of vehicles rather than one. The optimal number of runs associated with each service concept is used as a simple decision criterion, i.e. the concept requiring the fewer runs is preferable over the other one.

Another different viewpoint of the problem was examined by Tisato (2000) who optimized public transport subsidy levels given multiple constrained scenarios concerning desired load factor and bus size. Results show that the optimal unit subsidy declines as patronage increases but the decline was less pronounced in unconstrained cases where both load factor and bus size are optimizable decision variables. That is, if both variables are optimized simultaneously the change of bus sizes due to increased demand will be smaller than if only one variable is optimized because marginal optimization conditions can be better met by increasing both load factor and vehicle size.

Several studies have adopted vehicle size optimization approaches for designing and evaluating public transport feeder services, i.e. a feeder line that connects a remote area with an adjacent major transportation hub or interchange station. Analytical models were developed to compare fixed route conventional services over flexible route subscription services for deterministic (Chang & Schonfeld, 1991) and stochastic demand (Chien et al., 2001). Results show that flexible services require smaller vehicles at higher frequencies than conventional services do. This implies higher operating costs and lower user costs for the former and the opposite constellation for the latter case. Moreover, a threshold value of the demand density in the area connected by the feeder service was established which can be used to make a decision about which service concept to use. Another study dealing with the design of a bus feeder service was done by Chien (2005) who optimized headway, vehicle size and the route of the feeder line simultaneously given a discrete set of possible vehicle sizes and routes having associated lengths and travel times. The procedure computes the optimal headway by an analytical expression for each combination of vehicle size and route and finally selects the optimal solution which results in the minimum total cost. The model was applied to a real-world case study in New Jersey, USA.

2.2.2 Heuristic optimization approaches

Most of the analytical approaches presented in the previous section assume a fixed demand-line assignment, i.e. constant rates of passengers boarding and alighting at the stops along a route independent from the supply offered. In practice, however, when an entire public transport network consisting of multiple lines needs to be optimized, this assumption does not hold anymore since passengers choose their route through the network from origin to destination based on the supply characteristics of (potential) multiple routes. These demand-supply interactions and network effects introduce non-linear components to the optimization problem such as the inverse relationship between waiting time and headway. Exact and efficient solution methods are rarely available for such highly non-linear problems. In order to solve these complex problems, heuristic methods that find approximate near optimal solutions are often applied.

The following two subsections present heuristic optimization approaches for frequency and capacity determination. First, models for determining optimal line frequencies in public transport networks assuming given vehicle capacities are reviewed. Thereafter, approaches are presented that simultaneously optimize headways and vehicle capacities using heuristic optimization methods.

2.2.2.1 Frequency determination assuming given vehicle capacities

In order to account for the route choice implications of passengers, recent scientific works on public transport supply optimization incorporate bi-level optimization models. These models typically consist of two hierarchically structured optimization procedures: An upper level supply optimization model and a lower level assignment model computing equilibrium passenger flows resulting from a certain supply given by the upper level model (Ibarra-Rojas et al., 2015).

An early work on bi-level frequency optimization in public transport networks was done by Constantin and Florian (1995). At the upper level, the total travel time of passengers is minimized subject to a fixed fleet size constraint and a lower frequency bound. Assignment of passengers at the lower level is done using optimal strategies, a linear optimization problem which can be solved by an algorithm of polynomial complexity, e.g. a label-setting algorithm (Spiess & Florian, 1989). The upper level problem is solved by a projected sub-gradient algorithm which uses a (sub)gradient obtained from the assignment model in each iteration. The model is tested on three real life public transport networks.

Yu et al. (2010) use the same upper level objective and lower level assignment approach in their bi-level model as in the previous study. Vehicle capacities and fleet sizes are a constraint per set of lines belonging to a specific bus company. In the assignment model, multiplicative crowding factors reflecting the degree of comfort are used when computing in-vehicle times. Frequencies at the upper level model are optimized using a genetic algorithm. The proposed model is tested on a real life network in China. Results reveal that travel times can be reduced compared to the current supply situation. Moreover, it is shown that an integrated organization of bus companies sharing their vehicles can lead to an even more efficient allocation of resources in the entire network.

Another approach is examined by Yoo et al. (2010) who set the upper level objective of maximizing demand in response to a certain frequency setting. Constraints include frequency bounds as well as a fixed given vehicle fleet. Passenger flows are modelled using a stochastic user equilibrium assignment model considering transfer delays between lines at the lower level. Moreover, capacity constraints for line segments as well as increased waiting times because of denied boarding are implemented. Frequencies are iteratively optimized using a gradient projection method. Information on the gradient is obtained by the assignment model in each iteration. The model is tested on a hypothetical network and numerical experiments show that the algorithm converges well to an optimal point.

Huang et al. (2013) proposed a bi-level formulation that is able to handle uncertain demand in terms of a mean trip matrix and associated variance. The objective of the upper level frequency optimization model is to minimize the weighted sum of operator costs and travel time variance. The lower level assignment model applies the theory of optimal strategies and computes the mean and variance of passenger flows through all links. A genetic algorithm is applied to the upper level model, which optimizes frequencies constrained by a fixed fleet size and the requirement of all lines being served by at least one vehicle. Practical implementation of the proposed algorithm to a network in China shows that costs resulting from current operations can be reduced by changing the supply characteristics in terms of frequencies.

Next to the previously reviewed studies which used a bi-level structure in their proposed optimization models, there are other works proposing models for frequency optimization that cannot be classified according to a bi-level approach. That is, the problem is solved in one optimization procedure rather than using two hierarchical models.

Martínez et al. (2014) propose an alternative formulation of the model proposed by Constantin and Florian (1995). By linearizing the bi-level structure, they are able to reformulate the problem into a mixed integer linear program (MILP) for which dedicated techniques can be applied to solve the problem to optimality. The problem considers discrete frequencies, unlimited capacity of vehicles, and limited fleet size. Although the proposed MILP formulation enables to compute optimal solutions to the problem, largely-sized instances are expected to be hard to solve. Therefore, a metaheuristic solution algorithm based on Tabu search is proposed by the authors as an alternative solution method. Both approaches are tested on real-life cases and numerical results show that current operations can be improved.

Verbas et al. (2015) tested the influence of different demand elasticities corresponding to various levels of disaggregation on the solution of the frequency allocation problem by extending the model presented by Furth and Wilson (1981). The problem is formulated with a non-linear program which minimizes the weighted sum of ridership and wait time savings subject to constraints such as budget, fleet size, headway bounds for each line, and bounds for load factors. Route choice of passengers was assumed as fixed and estimated based on boarding and alighting data. Numerical results indicate that elasticities based only on temporal aggregation result in an underestimation of the potential improvements as compared to elasticities which account for some spatial characteristics, such as land use or the opportunity to transfer.

A comprehensive model considering various components and interactions influencing the performance of a public transport system was proposed by Yu et al. (2011). In their approach, frequencies are optimized for an entire network allowing for a differentiation between line directions. Moreover, cost components related to crowding such as increased waiting times due to denied boarding, onboard comfort levels and the influence of crowding on dwell times are explicitly taken into account. For the sake of simplicity, route choice decisions of passengers are neglected and demand is assumed as constant boarding/alighting rates. The objective of minimizing total system cost is optimized using a parallel approach of genetic algorithm in combination with local Tabu search. Application of the model to a real-life public transport network shows that the current overall level of service can be improved by reallocating available resources.

2.2.2.2 Simultaneous determination of frequency and vehicle capacity

All of the approaches presented in the previous section aim at optimizing supply in terms of choosing optimal line frequencies and thereby assuming vehicle size/capacity as given exogenous variable. However, there are two papers by Dell'Olio et al. (2012) and Ruisanchez et al. (2012) which optimize both decision variables at the same time using a bi-level approach. Line frequencies are modelled as continuous decision variables whereas bus sizes are assumed being discrete and constrained by a maximum number of available vehicles per type. At the upper level, the objective is to minimize the sum of operator and user costs. At the lower level, commercial software solving the public transport user equilibrium assignment is used. A Hook-Jeeves pattern search algorithm (Hook & Jeeves, 1961) proposes frequencies at the upper level and resulting vehicle sizes satisfying a demand constraint are computed. The proposed model is applied to a real-world bus network consisting of 15 lines in the city of Santander, Spain. Results indicate that using a mixed vehicle fleet in terms of bus capacities may lead to lower overall system costs than if a homogenous fleet was used. Ruisanchez et al. (2012) compared the results obtained by the Hook-Jeeves algorithm (Dell'Olio et al.) with those resulting from the application of a Tabu search algorithm. It turned out that Tabu search converges almost 50% faster than the other algorithm while the final optimal results are quite similar. Hence, Tabu search is more attractive if there is a need to solve the problem multiple times and for large networks.

Similar to the previous studies, Canca et al. (2016) propose an optimization model which simultaneously determines line frequencies and vehicle capacities in dense railway rapid transit networks. Line frequencies are modelled as discrete integer decision variables and the capacity of a train unit can be modified by varying the number of carriages deployed between two self-propelling units at the front and the rear of the train. The two decision variables are constrained by defining a finite set of admissible frequencies and a lower and upper bound for the number of carriages on a train respectively. Due to the technical differences of rail-bound traffic in terms of safety and rail capacity requirements compared to road traffic, additional constraints are introduced that ensure that these requirements are met given a potential supply setting. The objective function used in the model considers both the operator and user points of view by minimizing total operational and generalized travel costs. In case capacity on a segment is not sufficient to serve demand, the objective function value of the respective solution is penalized by adding a high value to it. Passengers are assigned to line segments by solving a capitated minimum cost flow problem taking into account the k-shortest paths between each OD pair. Since the model is formulated as a Mixed Integer Non-Linear Programming model, it can be solved by dedicated methods. The authors use the Extended Cutting Plane method (Still & Westerlund, 2001), which decomposes the problem into a set of sub-problems by discarding the non-linear constraints and then generating a set of linear cuts from the non-linear constraints for each solution that is infeasible in terms of these constraints. This set is added to the sub-problem and resolved again until an integer solution is found. The practical applicability of the proposed model is demonstrated on a simplified version of the metropolitan railway network in Madrid consisting of 7 lines. Using this case study, the model's sensitivity on certain input parameters such as overall demand and the specification of the value of time is tested indicating overall consistency among the resulting solutions and tendencies.

2.3 Summary and synthesis

The literature review on headway and vehicle size determination showed that early studies focused mainly on analytical models that can be solved using exact optimization methods or closed analytical expressions. Most of these models neglected or strongly simplified network effects, service reliability and passenger route choice decisions. Many recent studies, however, aim at searching for a system-wide optimum solution which takes into account the before mentioned aspects. Nonlinear correlations between the demand and supply model highly increases the complexity of the entire decision problem. In order to solve such problems efficiently, heuristic solution methods that produce approximate results of sufficient quality are needed.

Most of the studies dealing with models for vehicle capacity determination were published between the years 1980 to 2000. Table 2.1 presents a comprehensive characterization of all reviewed papers according to assumptions and determining factors considered in the model. There is no consistency among the models concerning which trip time elements influence optimal vehicle size. Some of the approaches consider various elements in a detailed fashion whereas others only take into account some basic elements. It is also worth mentioning that different models show opposite viewpoints on the relationship between demand and vehicle size: in some, a high passenger flow will increase the vehicle capacity needed; in others, it will justify using smaller vehicles. The representation of dynamic effects such as increased waiting times due to denied boarding or onboard comfort levels are very limited among all the models.

Recent models for public transport supply optimization are quite complex and require advanced solution methods. Table 2.2 presents a classification of the reviewed models according to selected characteristics including the present study. Most of the models aim at determining optimal supply conditions with respect to a certain objective in terms of line frequencies. Only very few studies tackle the simultaneous optimization of vehicle size and headway. In most of the works, the composition of the vehicle fleet is assumed as an exogenous input to the model. Most of the solution algorithms can be classified as gradient-based or metaheuristic algorithms such as genetic algorithms or Tabu search. The underlying public transport assignment models are mostly based on average and static supply conditions and do rarely consider effects related to crowding when determining passenger route choice decisions. There are some studies, however, that do explicitly account for certain crowding effects and dynamic demand-supply interactions. In all models, service is considered to be perfectly reliable in terms of stochastic supply variations. That is, running times of vehicles are assumed to be constant and headways perfectly regular.

The present study can capture the missing or rarely considered features of earlier studies by including the following aspects and thereby filling certain gaps. Using a public transport simulation model for dynamic assignment of passengers will enable a detailed representation of demand and supply including stochastic behavior and issues related to reliability. In contrast to current models, the consideration of dynamic supply-demand interactions and resulting issues such as denied boarding, discomfort due to crowding, and headway regularity can be implicitly captured by the simulation model. A detailed output on demand and supply performance data produced by the simulation model will enable a targeted optimization which can consider various potential objectives. Finally, a simultaneous determination of headway and vehicle size is possible to implement but each decision variable can be easily optimized separately as well by assuming an exogenous input for the other one. This makes the model very flexible since a reformulation of the decision problem is not necessary.

Table 2.1: Characteristics of reviewed analytical models for vehicle size determination.

	Jansson (1980)	Walters (1982)	Gwilliam et al. (1985)	Oldfield & Bly (1988)	Shih & Mahmassani (1994)	Lee et al. (1995)	Gronau (2000)	Tisato (2000)	Chang & Schonfeld (1991)	Chien et al. (2001)	Chien (2004)
Assumptions											
Elastic/variable route demand	No	No	No	Yes	Yes	No	No	No	No	Any demand function	No
Consideration of road congestion	No	No	No	Yes	No	No	No	No	No	No	No
In-vehicle time dependency on bus occupancy	No	No	No	Yes	No	No	Yes	No	No	No	No
Consideration of denied boarding	No	No	No	Yes	No	No	No	No	No	No	No
Multiple periods of the day	Yes	No	No	No	No	Yes	No	No	No	No	No
Determining factors											
Subsidy	No	No	No	Yes	No	No	No	No	No	No	No
Fare	No	No	No	Yes	No	No	No	No	No	No	No
Waiting time	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Access/egress time	No	No	No	Yes	Yes	No	No	No	Yes/No	Yes/No	Yes
In-vehicle time	Yes	No	No	Yes	Yes	Yes	Yes	No	Yes/No	Yes	Yes
Maximum vehicle load factor	Yes	No	No	No	Yes	No	No	Yes	No	No	No
Constant operating cost component per vehicle	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	per vehicle type

	Jansson (1980)	Walters (1982)	Gwilliam et al. (1985)	Oldfield & Bly (1988)	Shih & Mahmassani (1994)	Lee et al. (1995)	Gronau (2000)	Tisato (2000)	Chang & Schonfeld (1991)	Chien et al. (2001)	Chien (2004)
Vehicle size dependent operating cost component	Yes	No	No	No	No	No	Yes	No	Yes/No	Yes	Fixed cost per vehicle type
Route length or travel time	Yes	No	No	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes

Table 2.2: Characteristics of reviewed models for public transport supply optimization including present study.

	Decision variable Frequency	Decision variable Capacity	Objective	Solution method	Fleet size	Other constraints	Assignment method	Consideration of crowding	Stochastic supply
Constantin & Florian (1995)	Continuous	exogenous	Min total travel time	Projected sub-gradient algorithm	Fixed	Lower frequency bound	Optimal strategies	No (but model can be extended)	No
Yu et al. (2010)	Continuous	Fixed per line/company	Min total travel time	Genetic Algorithm	Fixed	Upper and lower frequency bounds	Optimal strategies	Comfort level onboard	No
Dell'Olio et al. (2012)	Continuous	Discrete	Min total users' and operator's costs	Hook-Jeeves pattern search algorithm	Fixed per vehicle type	Demand satisfaction	Commercial software (User equilibrium)	No	No
Ruisanchez et al. (2012)	Continuous	Discrete	Min total users' and operator's costs	Tabu search algorithm	Fixed per vehicle type	Demand satisfaction	Commercial software	No	No

	Decision variable Frequency	Decision variable Capacity	Objective	Solution method	Fleet size	Other constraints	Assignment method	Consideration of crowding	Stochastic supply
Martínez et al. (2014)	Discrete	exogenous	Min total travel time	CPLEX solver/ Tabu search algorithm	Fixed	Demand satisfaction , frequency bounds	Optimal strategies	No	No
Yoo et al. (2010)	Continuous	exogenous	Max demand	Gradient projection algorithm	Fixed	Demand satisfaction , frequency bounds	Stochastic user equilibrium	Increased waiting time because of denied boarding	No
Huang et al. (2013)	Continuous	exogenous	Min operator's costs and travel time variance	Genetic algorithm	Fixed	At least one bus per line	Optimal strategies	No	No
Verbas et al. (2015)	Continuous	exogenous	Max ridership and waiting time savings	KNITRO solver (commercial software)	Fixed, different bus sizes	Subsidy, policy headway, max bus capacity	Fixed route choice, based on boarding/alighting data	No	No
Yu et al (2011)	Continuous (per line direction)	exogenous	Min total users' and operator's costs	Parallel genetic algorithm in combination with local Tabu search	Fixed	Lower headway bound	Fixed route choice, based on boarding/alighting data	Denied boarding, onboard comfort, crowding coefficient influencing dwell times	No

	Decision variable Frequency	Decision variable Capacity	Objective	Solution method	Fleet size	Other constraints	Assignment method	Consideration of crowding	Stochastic supply
Canca et al. (2016)	Discrete	Discrete	Min total users' and operator's costs	Extended Cutting Plane Method	Not constrained	Finite set of discrete headways , Track segment compatibility, bounds for number of carriages per train	Solving capacitated minimum cost flow problem	Penalization of unserved demand,	No
Present study	Discrete (setting per line direction possible)	Discrete	Min total users' and operator's costs	Simulated Annealing	Fixed per vehicle type/not constrained	Finite set of discrete headways	Dynamic / by simulation	Denied boarding, onboard comfort, demand-dependent dwell time functions	Stochastic running and dwell time components

3 METHODOLOGY AND MODEL DEVELOPMENT

This chapter presents the methodologies adopted to develop the proposed decision tool and introduces the theoretical formulations and concepts of the model step by step. First, a general overview of the modelling framework is provided in order to highlight each of the components the model is built of. Thereafter, each of the components is explained in detail. The chapter closes by applying the developed model to a simple example case for model verification.

3.1 Modelling framework

The basic framework of the model consists of three sub-models which together form the integrated headway and vehicle size determination model. Figure 3.1 depicts the model framework including the sub-models, inputs and outputs as well as input parameters. A dynamic public transport assignment model needs to be initialized with parameters corresponding to the situation to be modelled and analyzed. These external parameters include for instance information on the network such as routes or travel times and demand-specific parameters such as the specification of the route choice model or an OD-matrix. A solution in terms of headway and vehicle type per line is provided as an input to the simulation model. Feasibility of potential solutions is checked by user-specified constraints. These comprise network-specific features such as scheduled cycle times and line lengths and conditions regarding the availability of supply such as the number of vehicles per type available or an operational budget constraint. Moreover, headway constraints in terms of an upper and a lower bound can be defined as well. Outputs produced by the assignment model related to passenger travel times and vehicle movements are evaluated by another sub-model which is specified by an objective function that computes a performance measure for the given supply condition. Using the relative performance of potential solutions, a search algorithm computes new solutions which are again provided to the assignment model as an input. The procedure is repeated until a user-specified stopping criterion that is checked in every iteration is fulfilled and thus a final well-performing solution is found.

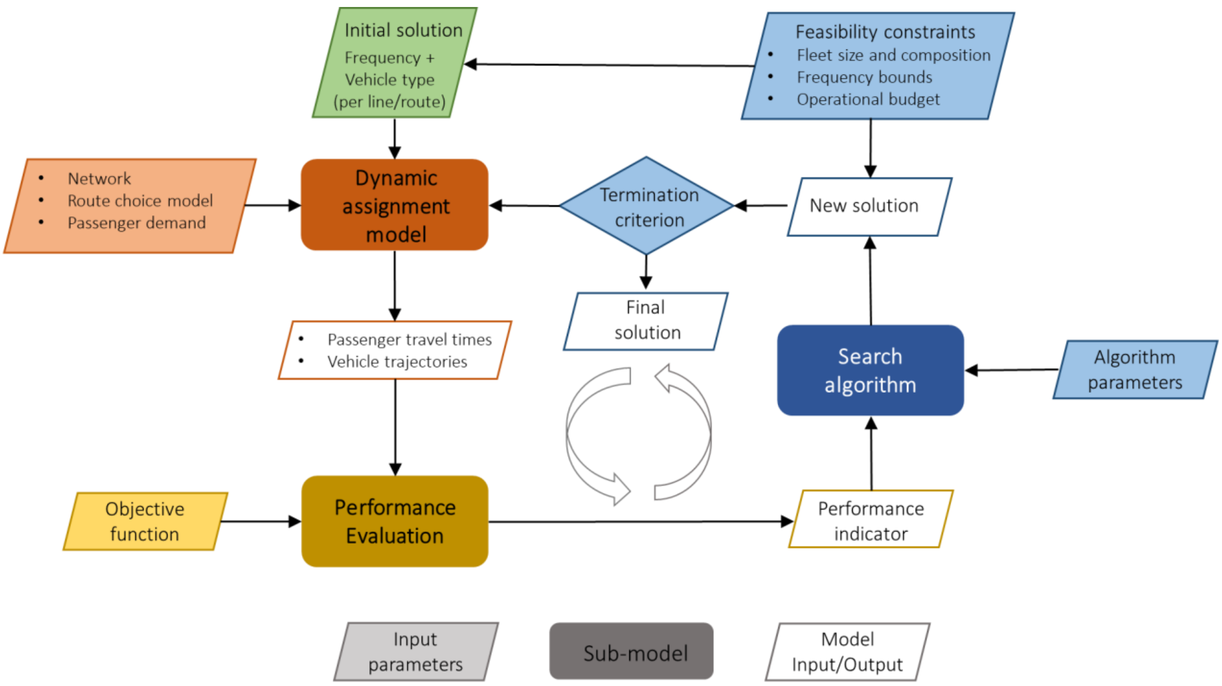


Figure 3.1: Basic framework of the headway and vehicle size determination model.

3.2 Dynamic public transport assignment model

This section describes the simulation model used for the dynamic assignment of passengers in this study. First, approaches and capabilities related to the modelling of demand and supply in BusMezzo are presented. Finally, it is explained how the simulation model is implemented and embedded within the model to be developed.

BusMezzo is a dynamic public transport operations and assignment tool designed to support the analysis and evaluation of Advanced Public Transport Systems in terms of operation, planning and control (Cats et al., 2010). BusMezzo is built within the platform of Mezzo, an event-based mesoscopic traffic simulation model developed by Burghout (2004). This model replicates individual vehicles and queues without explicitly modelling microscopic elements such as lane changes or acceleration. In contrast to time-based models which update the network status at constant and discrete time steps, Mezzo is an event-based simulation model in which subsequent time steps are defined based on a chronological list of booked events. The mutual interactions of vehicles and passengers in BusMezzo are explicitly modelled using an agent-based approach. Previous studies using BusMezzo have shown that the model can replicate effects resulting from dynamic demand-supply interactions in public transport networks such as congestion effects including variations in onboard crowding levels and denied boarding (Cats et al., 2016) or the well-known bunching phenomenon which can be seen as a degradation of service reliability along a line (Cats et al., 2010). Therefore, BusMezzo was chosen as a suitable implementation of a dynamic public transport assignment model.

3.2.1 Demand modelling

There are multiple ways of defining travel demand as an input to BusMezzo. In general, boarding and alighting rates can either be defined as a fixed input on the line level or as a network-wide OD-matrix indicating the number of trips per pair of stops. The latter one requires the dynamic modelling of passengers' route choice decisions whereas in the former one all paths are predefined. Since the present study aims at properly considering the relation between demand and supply, a complete network-wide modelling of route choice decisions and implications will be adopted. Hence, passenger demand is modelled using demand rates per OD-pair and a dynamic path choice model. During the simulation, passengers are generated randomly at the origins following Poisson-distributed and independent arrival processes.

The generic method of passenger demand modelling in BusMezzo follows a two-stage modelling approach consisting of a choice-set generation and a dynamic path choice model. The former one produces a background path set for each pair of stops in the network at the beginning of the simulation. During the simulation, travelers might adapt their choices and consider paths that were not part of their choice-set in the beginning. The choice set is generated by an algorithm considering logical constraints such as no loops and no abrupt travel legs and behavioral filtering rules such as the maximum number of extra transfers or the maximum additional in-vehicle time of alternative paths (Cats, 2011). Overlapping paths are finally merged into hyperpaths. Note that the initial background path set generation process is solely based on the static topology of the network and thus independent from the supply offered in terms of line headways and vehicle capacities. Additional time-dependent filtering rules are applied during the simulation and may further filter the initially generated path set. Hence, the initial path set generation procedure only needs to be executed once for a given network. This approach is advantageous for the entire optimization in terms of running time since the path set generation process may consume a significant amount of computation time depending on the size of the network.

In contrast to static assignment models, passengers simulated in BusMezzo do not make a single decision in terms of path choice in the network. Their final route through the network is rather an outcome of individual successive decisions based on the current situation in the network. Therefore, a dynamic path choice model using the theory of random utility discrete choice models is implemented in BusMezzo. The core of the choice model are the three subordinate decision models for connection, boarding and alighting actions. Along the journey, each alternative action is evaluated by the choice model based on the joint utility of all alternative paths that may result from that action using the commonly applied logsum expression over the respective path utilities (Cats, 2011). Path utilities are dynamically updated throughout the simulation making travelers able to reconsider their decisions. Passengers' expectations on downstream alternatives (e.g. expected travel time) may for instance be updated according to the experience onboard (e.g. experienced travel times considering vehicle-specific passenger loads as function of demand and supply variations) and hence may lead to a reconsideration of the alighting decision. Note that also the optional provision of real-time information may have an effect on travelers' behavior and is therefore incorporated within the choice model. Several levels of spatial aggregation of real-time information provision can be implemented in BusMezzo. Moreover, there is an additional feature implemented in BusMezzo which allows analyzing passengers' learning process and adaption from day to day with respect to waiting time, onboard crowding levels uncertainty and travel information. That is done by the execution of multiple simulation runs (representing within-day dynamics) while iteratively updating the accumulated memory of each passenger by remembering the travel time attributes of all tried paths. In this way, it is possible to analyze the behavioral adaption of passengers as well as the credibility of provided real-time information and generate network-wide steady-state conditions (Cats & Gkioulou, 2014) which can be seen as an equivalent to the congested user equilibrium in static assignment models. This so-called day-to-day learning feature will be used in the application presented in Chapter 4.

3.2.2 Supply modelling

BusMezzo simulates individual vehicle movements taking into account the key components of the supply side: riding and dwell times, timetables and vehicle schedules. In addition to that, disruptions and control strategies can be modelled explicitly as well. Hence, it enables the dynamic representation of public transport operations in its entirety. Note that since BusMezzo is implemented within the platform of Mezzo, a mesoscopic traffic simulation model (Burghout, 2004), the dynamic interaction between public transport and private traffic can be modelled as well. However, for the sake of simplicity, private traffic will not be modelled in this study. The influence of traffic congestion on the performance of public transport services can be partly considered by including stochastic riding time variations.

Timetables in Busmezzo can be defined according to various formats dependent on the way travel times between stops and headways are defined. Both characteristics can be defined for each trip on a line separately. That is, travel times and/or headways may differ from trip to trip. Alternatively, both parameters remain constant on each line throughout the entire simulation. That is, a line is associated with a certain scheduled headway and travel times.

Vehicle schedules or driving rosters, i.e. the assignment of trips (defined in the timetable) to specific vehicles, need to be defined in BusMezzo as well. In this way, trip chains can be defined as a sequence of trips executed by one vehicle. It also allows to model the potential propagation of delays from trip to trip. Note that the construction of vehicle schedules needs to be performed as preprocessing and may require the implementation of layover and recovery times at terminal stops.

Riding times (i.e. travel times between successive stops) in BusMezzo can be modelled deterministically or stochastically. That is, vehicles travel at constant speed between successive stops or riding times are subject to stochastic fluctuations respectively. Stochastic riding times can be modelled using so-called node servers that model the discharging rates of individual vehicles as a stochastic process having a mean and a standard deviation (Burghout, 2004). These server values can be estimated using field observations such as automatic vehicle location data (AVL). Next to riding times, dwell times constitute the second part of public transport trip travel times. It is the additional time needed to serve a stop. It includes the time needed to get off traffic and enter the stop, opening the doors, boarding and alighting of passengers, closing the doors and getting back to traffic. Due to its strong dependency on boarding and alighting passenger flows, dwell time is an important and dynamic determinant of service reliability when modelling travel times of passengers in a public transport network. In BusMezzo dwell times are modelled explicitly for each stop in the network using a predefined set of given dwell time functions which can be associated with different vehicle types. The general form of the dwell time function contains a constant delay, the passenger service time and a stochastic error term. The constant may vary across vehicle types as well as stop types. A bay stop for instance imposes an extra time for the vehicle to get back to traffic. Due to technical constraints such as door opening and closing mechanisms, the constant part of the dwell time depends on the type of vehicle as well. Therefore, BusMezzo offers the possibility to define different values depending on the vehicle and stop type. Moreover, it is possible to define an extra delay which is added to the constant delay in case the stop is already occupied when a vehicle is arriving. There are various models mentioned in literature to compute the passenger service time. Most of them take into account the number of boarding and alighting passengers at a stop and the service time needed for each boarding and alighting passenger respectively. The service time is the marginal contribution of each boarding/alighting passenger to the total passenger service time. Other models go more into detail and also consider vehicle occupation levels (crowding) when approaching a stop, door configurations as well as boarding/alighting regimes. Currently, there are six different forms of dwell time functions implemented in BusMezzo. The function coefficients can be estimated based on observed data per vehicle type or common values provided by literature can be adopted. All in all, it can be said that the representation of vehicle movements in BusMezzo can be modelled in a highly detailed fashion allowing for capturing the effect of dynamic demand-supply interactions on service reliability and overall system performance.

Besides the strategic analysis of daily public transport operations, BusMezzo can also be used to evaluate the implementation of real-time control strategies and to analyze system behavior in case of disruptions. Travel time disruption scenarios can be modelled implicitly by the exogenous travel time distribution on the respective servers or explicitly by introducing a traffic incident on a defined link in the network. In terms of operational control, various holding strategies at stops can be implemented and thus their effects on selected performance indicators can be evaluated by the simulation model.

3.2.3 Implementation

BusMezzo is implemented within the present model as it simulates the dynamics of a public transport network given a solution input in terms of line frequencies and vehicle capacities. Some of the outputs of BusMezzo are further processed to evaluate the relative performance of a potential solution. Figure 3.2 depicts the input and output components which are relevant in the context of this study.

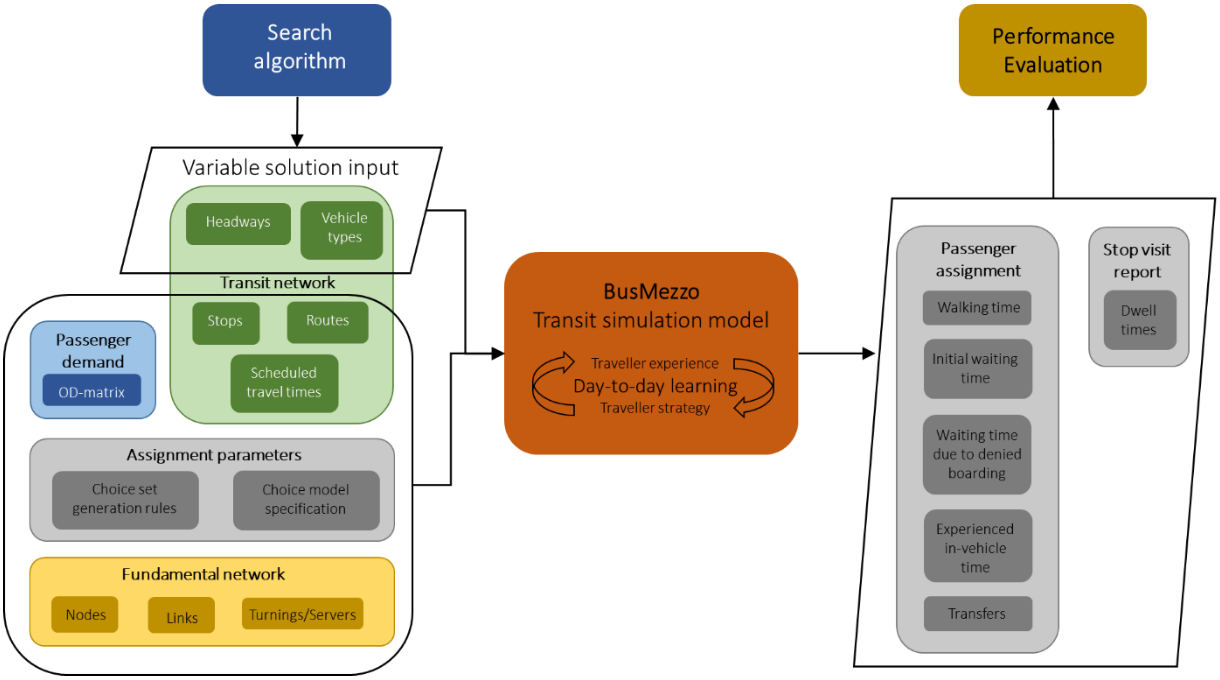


Figure 3.2: Input and output data items used by BusMezzo within the optimization model.

Input elements to the simulation model can be grouped according to fixed and constant input parameters as well as variable solution inputs. The latter ones are provided by the search algorithm and may change in each iteration of the algorithm. That is, headways and vehicle types are assigned to each line according to the given input that satisfies all feasibility constraints. Note that each available vehicle type having associated properties such as dwell time functions and capacity values is defined within BusMezzo.

For the sake of simplicity, vehicle scheduling in terms of trip chaining is not considered in the simulation. Instead, a vehicle is only assigned to one trip starting and ending at the respective terminals of a line. This assumption facilitates overall implementation since the vehicle scheduling problem can be very complex itself and would thus strongly increase the complexity of the entire model. Furthermore, in the tactical planning workflow, the vehicle and crew scheduling problems follow the frequency determination phase and are designed to find the minimum cost solution for a given service frequency. As a consequence, however, propagation of delays between trips and trip chaining across multiple lines are not replicated by the model. It is assumed that recovery times at terminals are set large enough to settle most of the delays accumulated along a line’s itinerary. Note that the disregarding of scheduling considerations also allows to set line frequencies differently for each direction of a line.

Inputs related to the public transport network that are assumed to stay constant throughout the entire procedure include the specification of stops along lines. Various features such as length, overtaking possibility and additional dwell times can be associated to a stop in BusMezzo. Moreover, walking distances between stops allowing for transfers can be defined as well. The routes of lines are defined as a sequence of traversed links. In a periodic timetable assuming constant headways, scheduled dispatching times at terminal as well as travel times between stops need to be defined once for each line in order to constitute a complete timetable. Hence, the actual departure times at stops result from a given headway. Passenger demand is assumed to be inelastic and is given in from of an OD-matrix indicating the hourly passenger generation rates per OD pair of stops. Behavioral parameters specifying the dynamic path choice model described in Section 3.2.1 include threshold values acting as filtering rules for the choice set generation process and the coefficients of the utility function which reflect the relative costs associated with certain trip elements. Finally, the fundamental network consisting of links and nodes needs to be specified as a constant input to BusMezzo as well. So-called turning movements having associated stochastic servers need to be specified at node level for each pair of connected links. In this way, the stochastic processing of vehicles passing a node is modelled enabling the emulation of riding time distributions as described in Section 3.2.2.

Outputs produced by BusMezzo and used for further evaluation include items related to the assignment of passengers, i.e. the realized travel path of each individually simulated traveler, and to the dynamic behavior of supply, i.e. the movement of vehicles. The dwell times of vehicles are considered as a dynamic component of supply which influences operational cost. Costs to be borne by passengers include walking, waiting, the time spent inside the vehicle and the disutility associated with transferring. The detailed use of the afore-mentioned figures and components when evaluating a potential solution is described in the following section. Note that BusMezzo produces various other outputs related to the performance of the public transport system including for instance detailed vehicle trajectories and performance measures such as headway and dwell time variability. The latter ones affect service reliability, which may itself have an impact on the fleet size requirement since larger travel time variations might require more vehicles to operate a certain schedule. However, due to the simplification made regarding vehicle scheduling, this impact will be disregarded in this analysis.

The day-to-day learning feature available in BusMezzo is used by the decision tool since it is necessary to investigate passengers' iterative behavioral adaptions towards a certain supply offered when making tactical planning decisions such as frequency and vehicle capacity determination. Multiple simulation runs representing successive days are simulated until travelers' expectations and associated selected strategies converge with their actual experiences. In this way, network-wide steady-state conditions are generated which may be seen as an analogy to congested network equilibrium conditions in traditional static assignment models.

Regarding the simulation of individual vehicles and passengers, the generation periods of demand and supply should be adjusted to ensure that the full supply is present as soon as the first passengers start their journey and all passengers can reach their final destination within the simulation period. Hence, the simulation of public transport supply should start earlier than the demand generation in order to fill the network and avoid unreasonable high waiting times (warm-up phase). On the other hand, supply should last long enough to provide enough capacity to carry all passengers and should thus last longer than the period of demand generation (cool-down phase).

Since BusMezzo is a stochastic model, a certain number of simulation replications is required in order to obtain statistically significant averages. A commonly used formula to estimate the number of samples needed to calculate the population mean with a predefined allowable error is:

$$N(m) = \left(\frac{S(m) \cdot t_{\alpha/2}}{\bar{X}(m) \cdot \varepsilon}\right)^2 \tag{3.1}$$

Where $N(m)$ is the number of required samples

$\bar{X}(m)$ is the mean value of the measured variable based on m initial samples

$S(m)$ is the standard deviation of the measured variable based on m initial samples

$t_{\alpha/2}$ is the value of the student t-distribution for level of significance α

ε is the allowable relative standard error to estimate the population mean

Note that equation (3.1) can also be used to calculate the resulting allowable standard error in case a desired number of samples is known.

In this study, the required number of simulation runs, in order to assess the quality of a potential solution, is estimated based on the objective function value. In order to do so, an initial sample of independent simulation runs having different random seeds is produced for each scenario and the mean and standard deviation of the resulting objective function values are computed. Based on these values, the sample size or the allowable error can be computed respectively.

3.3 Performance evaluation

In order to evaluate a potential solution in terms of its relative performance with respect to an objective, outputs obtained from the simulation model need to be transformed into a performance indicator. Therefore, an objective function is proposed. The objective is to minimize the total cost TC of the system which should both contain users' and operator's interest. These are reflected in the costs UC and OC respectively. Hence, the general objective function to be minimized is:

$$MIN\ TC = UC + OC \tag{3.2}$$

The total user costs UC can be computed from the simulation output as follows:

$$UC = \theta_{w,initial}TWT_{initial} + \theta_{w,denied}TWT_{denied} + \theta_{v}TIVT_{crowded} + \theta_{t}TT + \theta_{wlk}TWLKT \tag{3.3}$$

Where

$TWT_{initial}$: total initial waiting time [hours]; $\theta_{w,initial}$: value of initial waiting time [EUR/hour]

TWT_{denied} : total waiting time due to denied boarding [hours]; $\theta_{w,denied}$: value of waiting time due to denied boarding [EUR/hour]

$TIVT_{crowded}$: total in-vehicle time multiplied by crowding factor reflecting discomfort [hours]; θ_{v}: value of in-vehicle time [EUR/hour]

TT : total number of transfers ; θ_{t} : monetized value of one transfer [EUR]

$TWLKT$: total walking time [hours]; θ_{wlk}: value of walking time [EUR/hour]

The relative weights of each trip component can be estimated using stated or revealed preference or travel patterns surveys. Note that an extra waiting time due to denied boarding is a penalty to the passenger and thus may impose a higher cost than usual (initial) waiting.

The total operational costs can be computed as the sum of direct and indirect costs. Usually, indirect costs (e.g. administration costs etc.) can be calculated as a fixed share of the direct costs (α). The latter ones consist of driving costs (kilometers covered) CD, hourly costs due to standing still with the engine running while dwelling CS, personnel costs CP, and unit fixed costs CF. Hence, the total operational costs are:

$$OC = (1 + \alpha)(CD + CS + CP + CF) \tag{3.4}$$

The total driving costs per hour can be computed as:

$$CD = \sum_{l\in L}\sum_{c\in C} L_l \cdot f_l \cdot CD_c \cdot \delta_{l,c} \tag{3.5}$$

Where L_l is the length of line (or route variant) l in km , f_l is the determined frequency on line l and CD_c are the unit costs per kilometer covered by bus size c. $\delta_{l,c}$ is a binary auxiliary variable indicating if bus capacity c was assigned to line l. The set of all lines and vehicle capacities available is denoted as L and C, respectively.

The cost of the buses standing still with the engine running is:

$$CS = \sum_{l\in L}\sum_{c\in C} TDT_l \cdot CS_c \cdot \delta_{l,c} \tag{3.6}$$

Where TDT_l is the total dwell time on line l within one hour summed over all busses and CS_c is the unit cost per hour of bus size c standing still with engine running.

The personnel cost is considered as the cost of staff who are actually working on the busses:

$$CP = C_p \cdot \sum_{l\in L} \left\lceil \frac{T_l}{H_l} \right\rceil \tag{3.7}$$

Where C_p is the hourly cost of personnel, T_l is the minimum cycle time of line l and H_l the determined headway. The ratio of both variables is the theoretical number of vehicles needed and needs to be rounded to the next larger integer value. Note that these costs are independent from the type of vehicle used.

Finally, the fixed costs are calculated considering the vehicles that are actually circulating.

$$CF = \sum_{l\in L}\sum_{c\in C} \left\lceil \frac{T_l}{H_l} \right\rceil \cdot CF_c \cdot \delta_{l,c} \tag{3.8}$$

Where CF_c is the unit fixed cost per hour of bus type c.

The complete objective function of the problem can finally be formulated as:

$$\begin{aligned} MIN\ TC = {} & \theta_{w,initial} TWT_{initial} + \theta_{w,denied} TWT_{denied} + \theta_v TIVT_{crowded} + \theta_t TT + \theta_{wlk} TWLKT \\ & + (1+\alpha) \cdot \left[\sum_{l\in L}\sum_{c\in C} L_l \cdot f_l \cdot CD_c \cdot \delta_{l,c} + \sum_{l\in L}\sum_{c\in C} TDT_l \cdot CS_c \cdot \delta_{l,c} + C_p \cdot \sum_{l\in L} \left\lceil \frac{T_l}{H_l} \right\rceil \right. \\ & \left. + \sum_{l\in L}\sum_{c\in C} \left\lceil \frac{T_l}{H_l} \right\rceil \cdot CF_c \cdot \delta_{l,c} \right] \end{aligned} \tag{3.9}$$

Not all components of the objective function can be directly computed using a solution in terms of selected headways and vehicle capacities per line. However, most of the operational costs can be computed directly from a given solution except for the costs due to standing still with the engine running since the dwell times are an output of the simulation model. All components related to user costs need to be computed using output data from the simulation model. Summing the actual trip times experienced by each individual traveler in the simulation yields the respective trip time elements needed for the computation of the total user costs in (3.3).

As an alternative to the minimization of total system costs, the objective could just consider either the costs to be borne by the users or the operator. In the first case, operational constraints in terms of an available vehicle fleet or budget constraint need to be defined to limit the maximum quantity of supply provided. In the latter case, a constraint ensuring that demand is served satisfactorily needs to be introduced. This could for instance be the condition that denied boarding should not occur on any line segment, i.e. provided capacity always suffices demand.

3.4 Search algorithm

This section presents two versions of a search algorithm that were implemented to find a solution to the present headway and vehicle size determination problem. A simple local descent search procedure is introduced as a basis for a more complex search algorithm that follows the theoretical foundations of simulated annealing. The functioning and structures of both algorithms with respect to the present model are explained in detail. Finally, it is introduced how feasible solutions as well as neighboring solutions are generated and represented by the model.

3.4.1 Local search

Local search is a heuristic method for solving computationally hard optimization problems. This type of algorithm aims at finding a locally optimal solution by moving from one candidate solution to the other in the solution space. Thereby, the algorithm only considers the local neighborhood of a solution when searching for the next direction to move. In contrast to gradient-based optimization approaches, local search does not need any information on the gradient, i.e. derivative, of the objective function but solely the function value of any solution. This makes the method particularly suitable for problems where the objective function is not differentiable or a closed analytical expression describing the mathematical relation between the decision variables and the objective function value is not present. The latter condition holds for the present study as a potential solution is not evaluated using a classical mathematical function but by applying a stochastic simulation model.

In the present study, the solution space consists of all feasible combinations of headway and vehicle type per line. The generation of feasible solutions and the associated neighborhood is further discussed in Section 3.4.3. Figure 3.3 shows the schematic flowchart of the local search algorithm designed as a simple descent search for the present study. The algorithm is initialized by defining a starting point in terms of a solution respecting all feasibility constraints. This initial solution can be defined by the user or generated at random. In the next step, a set of feasible solutions in the direct neighborhood of the current solution is generated. Each solution, i.e. the starting solution as well as all neighboring solutions, are then evaluated by the simulation model and respective objective function values are computed using equation (3.9). The best solution among all neighboring solutions, i.e. the one having the lowest cost value, is selected as the next solution point if it is better than the current one. Thereafter, the procedure of direct neighborhood evaluation is repeated. The algorithm is terminated if the current solution cannot be improved by any solution in its direct neighborhood. That is, the algorithm follows a descent path in the solution space until a point is reached which has no lower cost solutions around it. Hence, it can be guaranteed that the algorithm will always find a locally optimal solution.

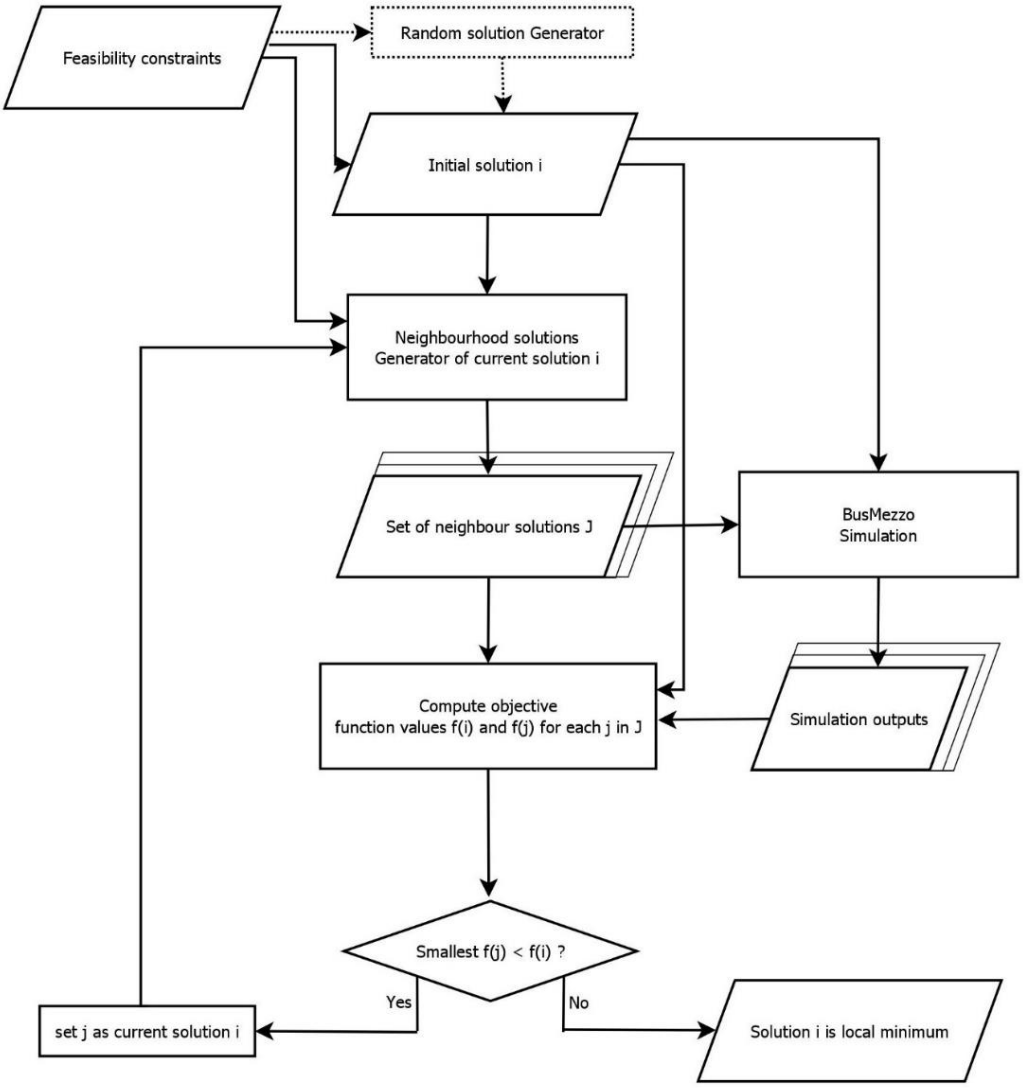


Figure 3.3: Flowchart of the local descent search algorithm.

However, it cannot be guaranteed that a global optimum is found since different starting solutions may lead to different locally optimal solutions due to topology of the solution space. Figure 3.4 shows a simple example of a two-dimensional optimization problem having a decision variable x defined on the feasible interval $x \in [x1, x2]$ and an objective function F(x). As can be seen from the function plot, the optimization problem has several local minima (points D, E, C and F) and a global minimum at point C. It is assumed that the direct neighbors of any point $x \in [x1, x2]$ are defined as $x + \Delta x$ and $x - \Delta x$., where Δx is the step size. From the function plot, it becomes obvious that the global optimum C can definitely be found by the local search algorithm if the starting solution is defined within the range $[xA + \Delta x, xB - \Delta x]$. In case the initial solution is not defined within this interval, the algorithm may converge to any of the local optima (D, E or F). Hence, if the problem at hand possesses several local optima, a simple local search algorithm following the decent direction of the solution space may not be able to find the global optimum with respect to the entire solution space.

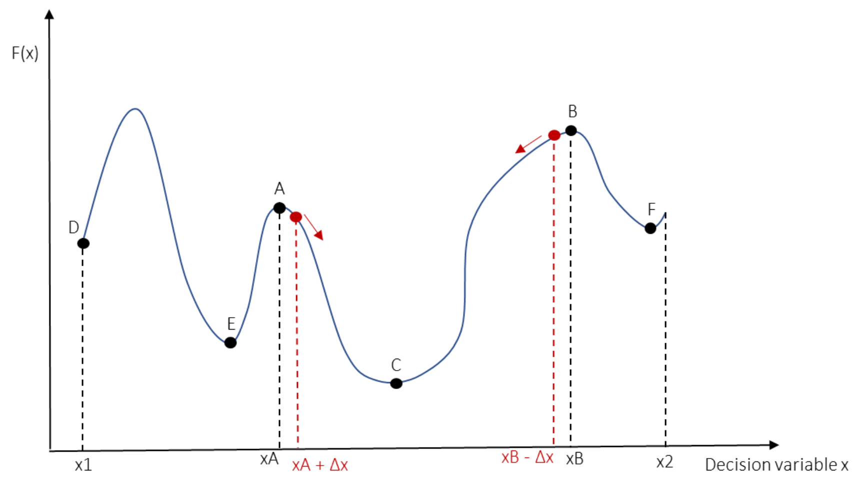


Figure 3.4: Example of a simple optimization problem with multiple local minima.

3.4.2 Simulated annealing

Simulated annealing (SA) is a probabilistic metaheuristic to approximate the global optimum in large search spaces of combinatorial and continuous optimization problems. Although SA can also be applied to continuous optimization problems, i.e. where the set of potential decision variables is continuous, this section will present a SA algorithm that is applicable to discrete problems only.

The name and inspiration of SA comes from the physical annealing of solids, which is the process of finding low energy states of a solid by initially melting the substance, and then lowering the temperature slowly and in a controlled way. The minimum energy configuration will have a particular structure, for instance as seen in a crystal. If the cooling is not done slowly, the resulting solid will not attain the ground state, but will be frozen into a metastable, locally optimal structure, such as a glass or a crystal with several defects in the structure. In the analogy, the different states of the substance correspond to the different feasible solutions to the optimization problem, and the energy of the system corresponds to the function to be minimized.

Kirckpatrick et al. (1983) and Cerny (1985) were the first who, independently from each other, showed that a stochastic Monte Carlo method for simulating the annealing of solids as proposed by Metropolis et al. (1953) could be used for solving large combinatorial optimization problems such as the travelling salesman problem. Since then, SA has been applied to various problem domains including research in the field of public transport. Fan and Machemehl (2006), for instance, showed how a SA algorithm can be applied to solve the public transport route network design problem, while Zhao & Zeng (2006) used SA in combination with a genetic algorithm to solve this specific problem.

SA is in fact a modification of the local search algorithm that tries to avoid the disadvantage of getting trapped in local optima by sometimes accepting neighborhood moves that worsen the current objective function value. In this way, local maxima can be overcome and the solution space can be examined more thoroughly, thereby increasing the probability of finding the global optimum. Figure 3.5 shows the SA algorithm in pseudo-code. Suppose that S is the solution space, i.e. the finite set of all feasible solutions, and f(x) is the cost/objective function defined on the members x ∈ S. Moreover, a temperature cooling function T(t) as well as an expression giving the number of iterations at each temperature N(t) need to be defined.

```
Select an initial state i ∈ S;
Select an initial temperature T > 0;
Set temperature change counter t = 0;
Repeat
       Set repetition counter n = 0;
       Repeat
              Generate state j, a random neighbor of i;
              Calculate δ = f(j) - f(i) ;
              If δ < 0 then i := j;
              else if random(0, 1) < exp(-δ/T) then i := j;
              n:=n+l;
       until n = N(t).
       t := t + l ;
       T := T(t), T ∈ ℝ;
until stopping criterion true.
```

Figure 3.5: Simulated annealing algorithm in pseudo-code.

The algorithm tries to avoid becoming trapped in local optima by sometimes accepting a neighborhood move which increases the cost function. For this sake, an acceptance function is used which computes the selection probability of a new (worse) solution considering the cost difference of the two neighboring solutions $f(new) - f(old)$ as well as the current temperature $T(t)$.

$$p(t) = e^{-\left(\frac{f(new)-f(old)}{T(t)}\right)} \quad (3.10)$$

That is, the smaller the difference between the old (better) solution and the new (worse) solution is the more likely is it that the new solution is accepted as the next step. It also implies that when T is high most moves will be accepted, but as T approaches zero most uphill moves will be rejected. So, the SA algorithm is started with a relatively high value of T, to avoid being prematurely trapped in a local optimum, and the temperature is gradually decreased. The acceptance probability $p(t)$ is modelled by the stochastic sampling technique of drawing a random number between 0 and 1.

The underlying assumptions of the stochastic approach in SA is the theory of Markov chains. If the temperature parameter is kept constant, then the transition matrix P_{ij} representing the probability of moving from state i to state j is independent of the iteration number, which corresponds to a homogenous Markov chain (Eglese, 1990).

It has been shown that the procedure of SA can converge to globally optimal solutions or regions. When applying any variety of the SA algorithm four generic choices regarding the parameters of the algorithm have to be made. According to Eglese (1990) these choices comprise:

(i) the initial value of the temperature parameter T_0

(ii) a decreasing cooling function T(t) that determines how the temperature is changed over time

(iii) the number of iterations N(t) to be performed at each temperature

(iv) a suitable stopping criterion to terminate the algorithm

The starting temperature T_0 should be high enough such that during the initial phase of the algorithm, virtually all possible solutions are accepted. This corresponds to the physical analogy of heating up a substance until all the particles are randomly arranged in a liquid. Laarhoven and Aarts (1987) proposed the following equation for determining the initial temperature T_0:

$$\chi_0 = e^{\frac{-\overline{\Delta F^+}}{T_0}} \tag{3.11}$$

Where $\overline{\Delta F^+}$ is the average positive change of the objective function for a series of random transitions in the solution space. Using a value of χ_0 close to one (e.g. $\chi_0 = 0.9$) ensures a high acceptance rate for the solutions during the initial phase of the algorithm. Hence, the starting temperature value can be determined by sampling a series of random neighborhood transitions for the problem at hand and is thus a problem-specific parameter.

The cooling schedule, that is, the decreasing cooling function of the temperature over time, is crucial to the efficiency of simulated annealing. If the temperature is reduced too rapidly, a premature convergence to a local minimum may occur. In contrast, if cooling works too slowly, the algorithm is also very slow to converge and may pass the globally optimal solution several times. In order to guarantee Boltzmann annealing, the theoretical foundation of simulated annealing, to converge to the global minimum with probability one, the temperature value needs to decrease logarithmically with time which is practically way too slow. Therefore, a faster schedule based on the decreasing exponential function $T(t+1) = \alpha T(t)$ with $0.85 \leq \alpha \leq 0.96$ is commonly used in practice to achieve a (sub)optimal solution (Du & Swamy, 2016).

The number of iterations at each temperature N(t) can be determined by a fixed minimum number of neighborhood transitions being accepted subject to a constant upper bound (Kirpatrick et al., 1983). Simpler approaches may keep N(t) constant throughout the entire procedure and make it proportional to the size of the solution space or the size of the neighborhood defined (Eglese, 1990). Note that the determination of the cooling function and those of N(t) are two interdependent choices as they determine the length and the difference between each homogenous Markov chain respectively. Increasing the relative magnitude of the temperature drop requires, for instance, longer Markov chains, i.e. a larger value for N(t) (Herman, n.d.).

The SA algorithm is finally terminated when the obtained solution is unaltered for a defined number of consecutive iterations. The final state then corresponds to the optimal (i.e. best-performing) solution or the 'frozen' state (Eglese, 1990).

Figure 3.6 shows the flowchart of the SA algorithm as it is implemented within the framework of the present headway and vehicle size determination model. The initialization of the algorithm works the same as in the simple descent search algorithm presented previously. However, in contrast to the local search approach, not all neighbors of a solution are investigated during SA but only a randomly chosen one. Note that if a solution was selected from the set of neighbors as a new solution by the algorithm it is also removed from the set. Although this modification is not considered in the classical version of SA, it has been argued that it can increase overall efficiency of the algorithm (Eglese, 1990). Especially in the latter part of an SA run, when the temperature parameters gets low, the probability of accepting worse solutions than the current one becomes small. Thus, in regions close to a local optimum, most of the computer time is spent rejecting worse solutions. If there are only few moves which give improved solutions, the basic SA algorithm may take a lot of time to find them. Therefore, solutions are removed from the set of neighbors once they have been evaluated and a new neighborhood of solutions is only generated if a new solution is accepted by the algorithm or all neighboring solutions have been investigated, i.e. the current set of neighbors is empty. A solution is accepted as a new move when it improves the objective function value. When a solution is worse than the current solution in terms of the objective, it is only accepted if a randomly drawn number between zero and one is smaller than the acceptance probability according to equation (3.10). Every time a solution is not accepted, a variable counting the number of successive rejections is increased by one and reset to zero if a new solution is accepted. After each change in temperature, it is checked whether the counter value is below a certain threshold value. The temperature is changed according to the cooling function after N iterations. N(t) is assumed to be constant throughout the entire search procedure. The algorithm is terminated once the number of successively rejected solutions exceeds this threshold criterion. Note that this threshold value should be equal to or larger than the maximum size of the neighborhood in terms of number of solutions in order to make sure that all neighboring solutions have been checked by the algorithm before it is terminated. Also note that the best solution found during the run of the algorithm is stored as it is possible in any single SA run for the final solution to be worse than the best solution found during the run. Moreover, Glover and Greenberg (1989) argue that with this modification, there is less need for the SA algorithm to rely on a strong stabilizing effect over time.

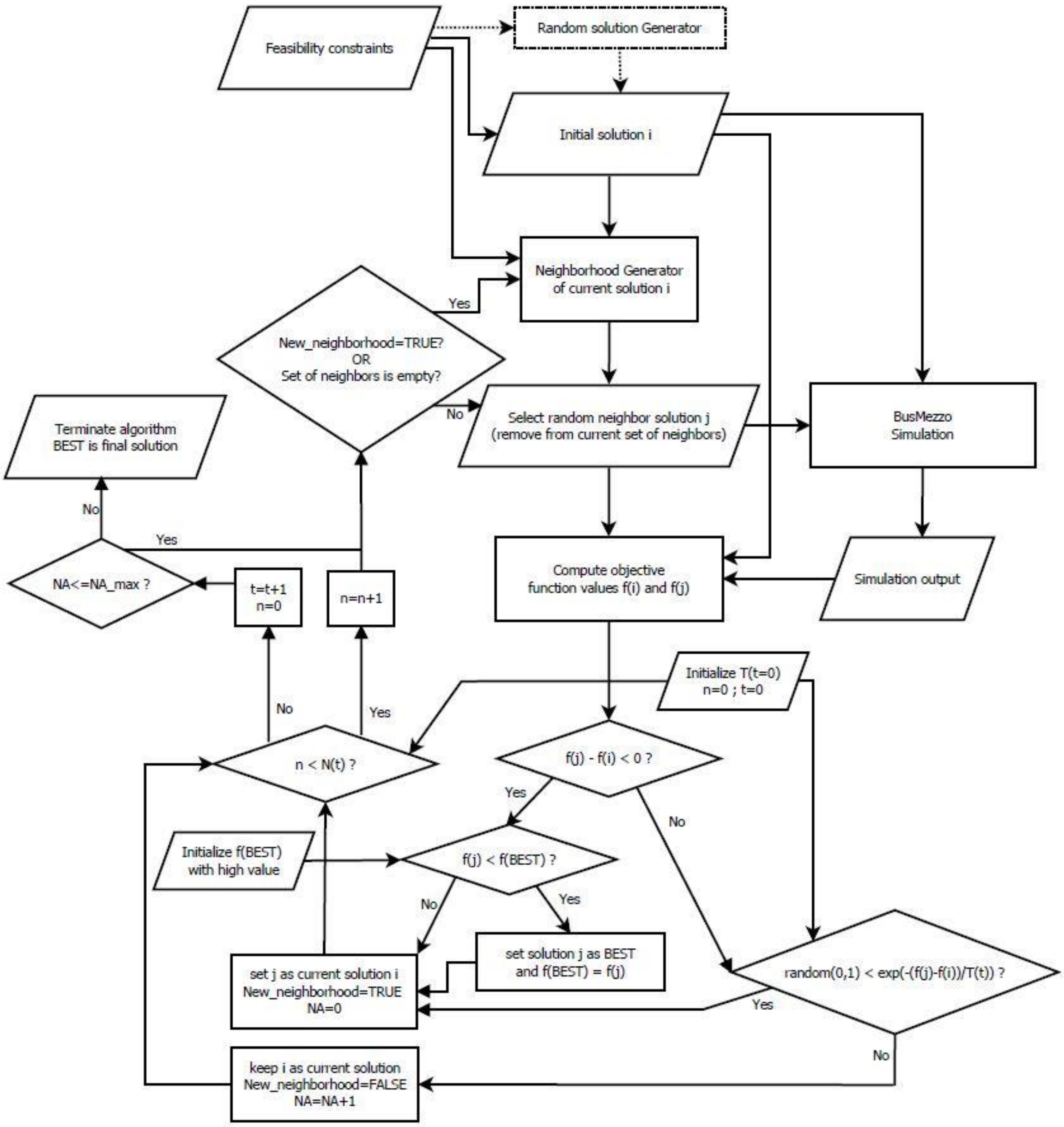


Figure 3.6: Flowchart of the developed algorithm based on simulated annealing.

3.4.3 Generation of feasible neighborhood moves

The neighbors of a state are new states of the problem that are produced after altering a given state in some well-defined way. The well-defined way in which the states are altered in order to find neighboring states is called a "move" and different moves give different sets of neighboring states. These moves usually result in minimal alterations of the last state in order to help the algorithm keep the better parts of the solution and change only the worse performing parts.

For the frequency and vehicle capacity determination problem, suppose we have the following discrete solution space arranged in an ascending order:

Headways: $H = \{h_1, h_2, \dots, h_{N_H}\}$ with: $h_1 < h_2 < \dots < h_{N_H}$

and

Vehicle capacities: $C = \{c_1, c_2, \dots, c_{N_C}\}$ with: $c_1 < c_2 < \dots < c_{N_C}$

In order to make sure that the timetable resulting from a certain frequency is periodic, i.e. scheduled minutes of arrival/departure repeat in every hour with constant headway, the ratio of one hour (3600 seconds) divided by potential headways in seconds needs to be an integer value. Table 3.1. shows all potential headways satisfying this condition between one hour and two minutes. Note that frequency and headway show an inversely proportional and non-linear relationship.

$$\frac{3600}{H} \in \mathbb{Z} \tag{3.12}$$

Table 3.1: Potential discrete headways and associated frequencies.

headway	seconds	3600	1800	1200	900	720	600	514	450	400	360	327	300	277	257	240	225	212	200	189	180	171	164	157	150	144	138	133	129	124	120
	minutes	60.0	30.0	20.0	15.0	12.0	10.0	8.6	7.5	6.7	6.0	5.5	5.0	4.6	4.3	4.0	3.8	3.5	3.3	3.2	3.0	2.9	2.7	2.6	2.5	2.4	2.3	2.2	2.1	2.1	2.0
frequency	veh/hour	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30

Assuming that all potential solutions in terms of headway-vehicle combinations per line are feasible, the total number of possible solutions can be computed as:

$$(N_H * N_C)^L \tag{3.13}$$

Where L is the number of lines of the public transport network to be optimized and N_H and N_C are the number of elements in the set of discrete headways and vehicle capacities respectively. From equation (3.13), it becomes obvious that the size of the solution space increases exponentially with the number of lines. For 8 different headways, 3 vehicle types and 5 lines, the number of possible solutions becomes almost 8 million. For 7 lines, there are around 4.6 billion of different combinations. Hence, with increasing network size and operational possibilities in terms of headways and vehicle types, the number of possible solutions will explode. In order to keep the size of the solution space as small as possible and thereby reduce complexity, it is favorable to limit the set of potential headways to reasonable values only. This can be done, for instance, by only considering headways of integer minutes and excluding certain headways a priori from the set that are operationally infeasible or unfavorable because of any reason (e.g. policy guidelines).

A solution can be coded as a $L\ x\ (N_H + N_C)$ binary matrix. For an example case of 5 lines, 5 possible headways and 3 possible vehicle types, a possible solution could be coded as shown in Table 3.2. Line 3, for instance, is operated by a vehicle capacity of 100 and a headway of 720 seconds (12 minutes).

For each possible solution, there exists only one unique binary matrix. Note that such a matrix can be easily randomly generated by a computer while respecting the condition that each line is only associated with one headway and vehicle type.

Table 3.2: Binary coding of an example solution.

	headway					vehicle capacity		
Line	300	450	600	720	900	30	60	100
1	0	0	1	0	0	0	1	0
2	0	1	0	0	0	1	0	0
3	0	0	0	1	0	0	0	1
4	1	0	0	0	0	0	0	1
5	0	0	0	0	1	0	1	0

A neighbor of a specific solution can now be generated by altering one characteristic of a certain line in a well-defined way and thereby keeping all other variables unchanged. This is done by changing the headway or capacity of a selected line to the next smaller or larger values, i.e. just flipping two neighboring bits. In that way, a set of neighboring solutions can be generated for any current solution. The maximum number of neighboring solutions is:

$$N_{max} = (2 + 2) \cdot L \tag{3.14}$$

That is, either headway or capacity can be change to the next higher or lower value, yielding 4 possible changes per line. Note that a specific solution can have less neighbors than computed in equation (3.14) if it contains boundary values, i.e. the minimum or maximum headway/capacity is selected on any line, because there is only one possibility to switch from a boundary value to the next higher or lower value.

Finally, it is worth mentioning how feasibility constraints are incorporated in the model when generating solutions. Not all solutions can be technically feasible since there are often vehicle availability, fleet size or budget constraints. By considering a discrete set of potential headways to be used in a solution, one implicitly incorporates lower and upper bounds for headways. Given the number of available vehicles of size c called N_c and the cycle time of a line T_l one can easily determine if a potential solution is feasible by

$$\sum_{l \in L} \frac{T_l}{h_l} \delta_{l,c} \leq N_c \quad \forall\, c \in C \tag{3.15}$$

Where h_l is the chosen headway on line l, C is the set of all vehicle capacities and L the set of all lines considered and $\delta_{l,c}$ is a binary variable:

$$\delta_{l,c} = \begin{cases} 1, & \text{if capacity c is used on line l} \\ 0, & \text{otherwise} \end{cases} \tag{3.16}$$

Hence, inequality (3.15) checks whether a solution can be operated given the available number of vehicles per type and the assignment of headways and vehicle types to lines which have a certain cycle time. The latter one is used to estimate the number of vehicles needed on each line to operate the selected frequency. Note that, for the sake of simplicity, it is assumed that each line can solely be operated by one vehicle type and that vehicles are not circulating among multiple lines during the defined interval of operation.

When generating random initial or neighboring solutions, an immediate feasibility check can be done using inequality 3.15 and thus infeasible potential solutions can be rejected or excluded from the list of neighboring solutions on beforehand. Depending on the case of application, the size of the solution space may be significantly reduced due to the infeasibility of certain solutions.

3.5 Model verification

The previously described model components were implemented in a computer program. MATLAB was used as a platform to code the search algorithms and coupled with the simulation model BusMezzo to obtain output data for solution evaluation. This required a mutual exchange of data between MATLAB and the simulation model during runtime.

In order to confirm the proper functioning and plausibility of results of the model, a simple network consisting of two lines and 5 stops was designed. The small scale of this example case allows to examine and visualize the entire solution space of the discrete decision problem as well as the trajectories of the search algorithms. Figure 3.7 shows the network configuration and a given passenger demand matrix. There are a total of 800 passengers per hour travelling through the network. Note that there is only one possible route available per OD pair and therefore route choice effects do not play a role in this case. The cycletime on each line was assumed being 1800 seconds (30 minutes) and the line length in both directions was set to 10 kilometers for both lines.

Table 3.3 shows the parameters and coefficients assumed for the verification of the model. Cost components such as value of time or operating cost components per vehicle type were reasonably assumed based on commonly known and used values in literature. Note that this analysis uses a purely hypothetical example that aims at investigating the topology of the solution space and the trajectory of search algorithms and not at finding optimal supply conditions. Therefore, knowing exact parameters and coefficients representing real-world conditions is not necessary in this case.

Table 3.3: Assumed parameters and coefficients for model verification.

User cost components		Operational cost components [EUR/h]	Vehicle capacities [pax/veh]			Parameters SA algorithm	
			10	20	30		
VOT	6.75 EUR/h	Driving (CD_c)	0.1	0.2	0.3	Initial temp. T_0	40.000
$\theta_{w,initial}$	2	Standing (CS_c)	0.02			Cooling function	T(t+1)=αT(t)
$\theta_{w,denied}$	2*3.5	Personnel (C_p)	30			Factor α	0.9
$\theta_{transfer}$	5	Fixed (CF_c)	10	12	14	Iterations N(t)	2 (constant)
θ_{walk}	(2)	Share indirect costs	0.3				

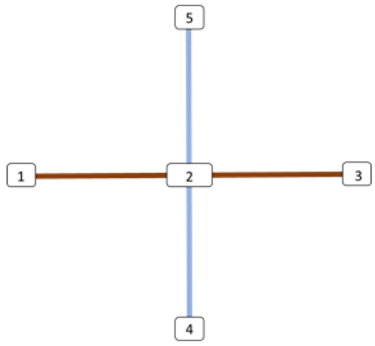
The number of required simulation runs (replications) was set to 10 yielding a maximum allowable relative error of 1.9% of the objective function value at a level of significance of 95%. The set of discrete headways in seconds was defined as:

$$H = [180, 240, 300, 360, 450, 600, 720, 900]$$

Moreover, three possible vehicles having different capacities in persons per vehicle are available:

$$C = [10, 20, 30]$$

The initial temperature T_0 for the simulated annealing algorithm was computed as 38642 according to equation (3.11) and rounded to 40000. The factor in the exponential decreasing cooling function was set to 0.9 and 2 iterations per temperature step were assumed.



	1	2	3	4	5
1	-	100	40	20	30
2	40	-	30	20	20
3	20	100	-	30	10
4	30	60	10	-	20
5	40	60	20	40	-

Figure 3.7: Simple test network and demand matrix in passengers per hour.

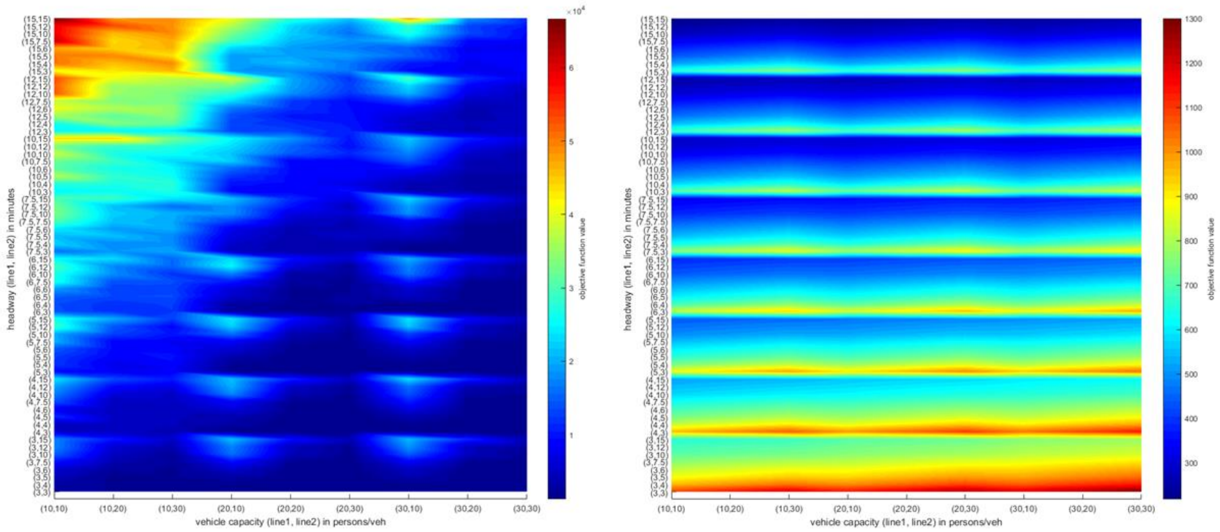


Figure 3.8: User cost (left) and operating cost (right) for the entire solution space.

8 different headways and 3 possible vehicle types yield a total of 576 solutions in the case of two lines provided that all solutions are feasible. In order to achieve this, the fleet size constraint was relaxed by setting the number of available vehicles per type to a sufficiently high number. Figure 3.8 shows the computed user and operating cost for the entire solution space. It can be observed that costs to be borne by the passengers increase heavily with decreasing vehicle sizes and line frequencies as supply in terms of transportable passengers per line segment per hour decreases. On the other hand, costs to be borne by the operator show an inverse relationship as they decrease with smaller vehicles and lower frequencies.

Adding up the cost values depicted in Figure 3.8 yields the total system cost and thus the objective function value to be optimized. Figure 3.9 shows these values for the entire solution space. Note that for small values there is no clear differentiation in terms of colours and contour lines since the absolute difference between the largest and the smallest cost value is very large compared to the values inside the 'valleys' of the solution topology. Looking at the trajectories of both the local search and the simulated annealing algorithm (SA), one can nicely observe the behavior of both procedures. Starting at the same initial solution (headways 15 and 7.5 minutes / vehicle capacities 20 and 30 persons for line 1 and 2 respectively), the local search algorithm directly proceeds towards regions of low cost values and reaches the minimum solution after 6 iterations. In contrast to that, SA first explores regions of higher cost values (upper left area) and then intensifies the search in low cost regions. After 108 iterations, the final value is found. Hence, the developed model behaves as expected. The behavior of the SA algorithm may suggest an inefficiency regarding the fast convergence towards optimal regions of the solution space compared to local search in this example. However, in case of larger solutions spaces having a more complex topology with multiple altering high- and low-cost regions, the final solution obtained by SA may be significantly better than those obtained by LS and thus outperform the latter one.

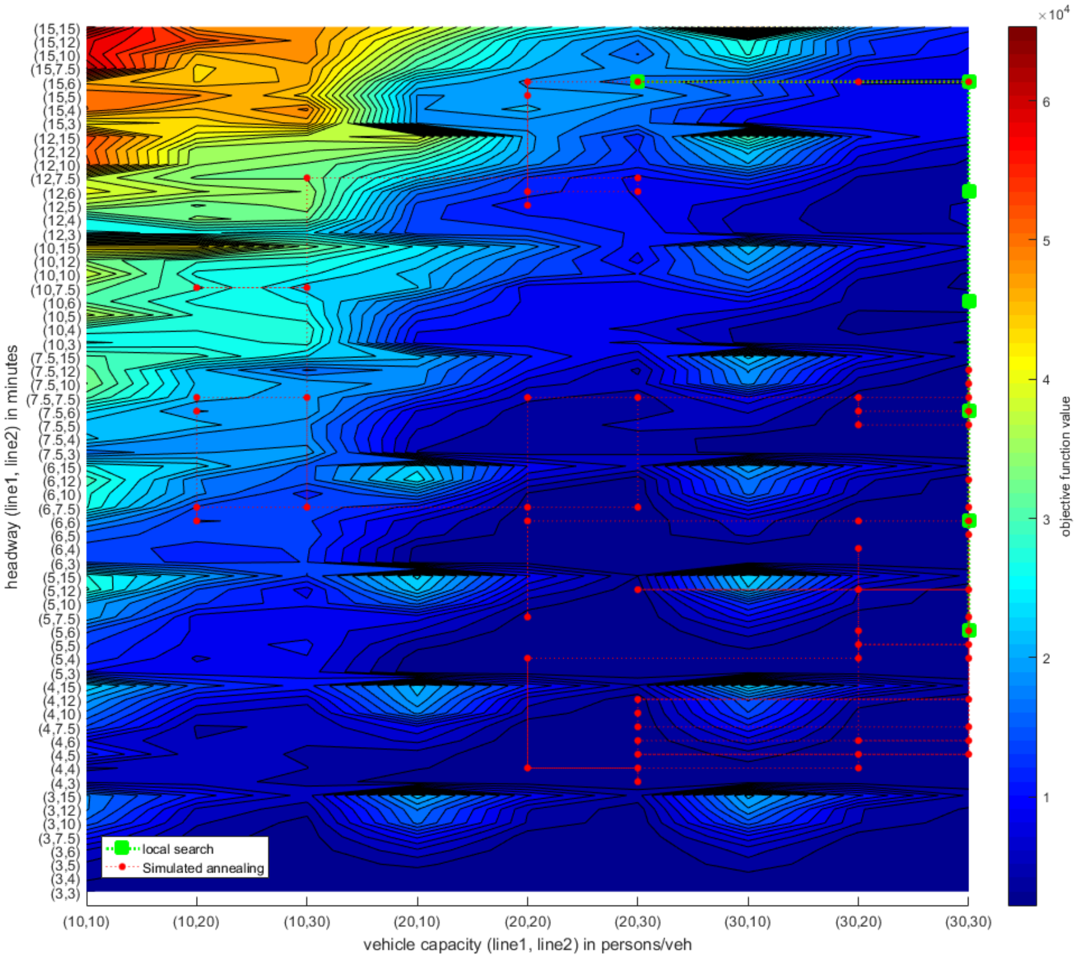


Figure 3.9: Total system cost for the entire solution space including search trajectories of local descent search and simulated annealing algorithm.

4 APPLICATION

This chapter presents two applications of the proposed model. The first case study investigates the model's behavior towards changes in certain input parameters using a hypothetical simple public transport network. In the second case study, the practical benefits and limitations of the model are identified using a real-world scenario of a bus network in the North of Amsterdam.

4.1 Network by Spiess and Florian

In this section, the first application of the developed model on a hypothetical case study is presented. As a network for numerical tests, the one proposed by Spiess and Florian (1989) was used. Figure 4.1 shows a schematic representation of the network including the riding times between stops on the respective lines. This network offers interesting trade-offs regarding passengers' route choices and is yet relative simple. A hypothetical symmetric passenger demand matrix is depicted in Table 4.1. In total, there are 1000 passengers using the network per hour. For the four lines, cycle times were set to 60, 36, 26 and 18 minutes respectively (assuming 5 minutes of turn-around and slack at each terminal). Note that in the original network presented by Spiess and Florian, lines were assumed to operate in one direction only. The line lengths were computed based on the assumption of an average riding speed of 30km/h and the given travel times. Like in the application for model verification, cost components as an input to the objective function were reasonably assumed based on commonly used values in literature. The number of required simulation runs was set to 5 yielding a maximum allowable relative error of 1.0% of the objective function value at a level of significance of 95%. The day-to-day learning feature of BusMezzo was used in this case in order to mimic steady-state conditions in terms of passengers' route choice adaptions towards supply conditions. The same set of 8 possible line headways as in the example for model verification was used. Possible vehicle capacities were chosen as 60, 90 and 120 persons per vehicle, i.e. 3 types of vehicles are available. Note that the number of available vehicles per type was set high enough such that all theoretical combinations of headways and capacities are practically feasible. This yields a total of 331.776 feasible potential solutions.

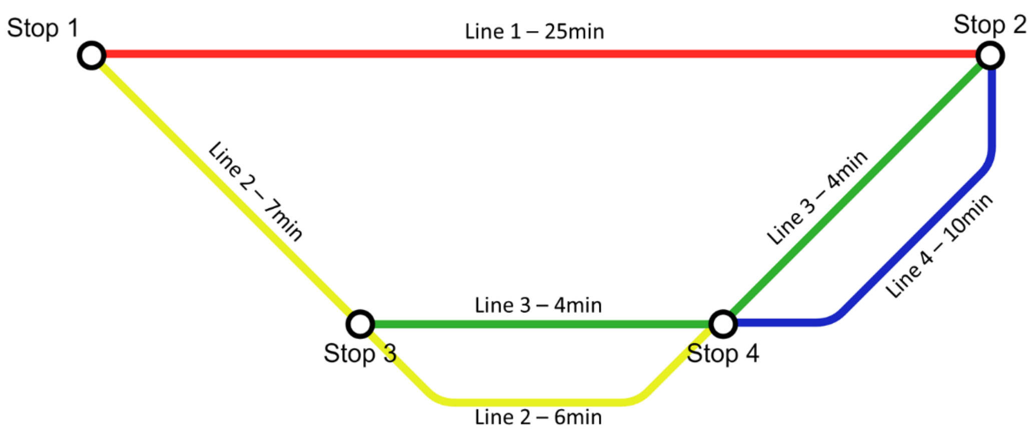


Figure 4.1: Hypothetical public transport network proposed by Spiess and Florian (1989) including travel times on line segments.

Table 4.1: Hypothetical demand matrix in passengers per hour.

	Stop 1	Stop 2	Stop 3	Stop 4
Stop 1	0	150	50	100
Stop 2	150	0	100	50
Stop 3	50	100	0	50
Stop 4	100	50	50	0

4.1.1 Scenario design

In order to test the model's behavior against various parameter settings, three test scenarios were designed. Table 4.2 gives an overview of the three test cases and associated parameter settings.

Table 4.2: Summary of the test scenarios performed.

Test 1: Initial solution						Test 2: SA parameter					Test 3: Waiting time sensitivity			
High capacity		Medium capacity		Low capacity		Number of iterations per temperature step					Cost ratio waiting/in-vehicle time			
SA	LS	SA	LS	SA	LS	N=1	N=2	N=3	N=4	N=5	1	1.5	2	2.5

Test one aims at analyzing the sensitivity of the initial solution provided to the search algorithm on the final solution. Therefore, three significantly different solutions are provided to the model as a starting condition. First, a high-capacity situation is assumed, meaning all lines are served at the highest frequency (headway of 180 seconds, i.e. one vehicle every three minutes) and large vehicles (120 person/vehicle). Second, a medium-capacity setting is provided with headways of 7.5 minutes and a capacity of 90 persons per vehicle on all lines. Finally, all lines operate at the lowest frequency and the smallest vehicle, i.e. 15 minutes headway and 60 persons per vehicle. For each of the scenarios, the final solution is computed using the local descent search algorithm (LS) presented in Section 3.4.1 as well as the simulated annealing (SA) procedure explained in Section 3.4.2. SA has a higher probability of finding a global optimum than LS since the structure of the SA algorithm allows for escaping local optima, whereas LS can only follow a descent path in the solution space that leads to the closest local optimum. Therefore, the results of both algorithms may differ depending on the initial solution selected. SA may arrive at similar final solutions independent from the starting point provided that all parameters of the algorithm are set appropriately. On the other hand, it is expected that solutions found by LS clearly differ from each other and show a similarity with or closeness to the respective starting solution in terms of the decision variables.

The second test case examines the influence of different parameter settings in the SA algorithm on the quality of the final solution found. The model is applied for different numbers of iterations to be executed at each temperature (N), which influences overall running time of the algorithm and may also affect the final solution found. Since more iterations will lead to a more intensified search and exploration of the solution space, the probability that SA will visit areas close to the global optimum increases. Therefore, it is expected that more iterations, i.e. larger values of N, may result in better solutions with respect to the objective. However, this improvement comes at the price of longer computation time.

Finally, a third test case was designed which investigates how the weighting of waiting time in the objective function and the route choice model affects the final solution found by the model. The model is applied to 4 different cases in which the relative weighting of waiting time in the objective function and passengers' route choice factors differ accordingly. Hence, this scenario tests how the relative importance of waiting as perceived by the passenger affects the optimal design of the supply. As the relative importance of waiting increases, also the waiting costs will increase and lead to an overall increase of the total costs. However, when supply in terms of frequencies is increased, nominal waiting times will decrease and cause the waiting costs to decrease. The question is whether the increase in operational costs associated with the increase in frequencies is larger or smaller than the savings in waiting costs. In the latter case, it is worth to increase the supply since overall costs will be less than if frequencies were not changed.

4.1.2 Results

Test 1

Table 4.3 shows the final solutions in terms of frequencies and vehicle capacities per line as well as the average total costs obtained for different initial solutions and search algorithms applied. Note that due to the stochastic nature of BusMezzo the relative standard deviation of each average cost value is about 1%. In Figure 4.2, the total cost associated with each solution found per scenario is depicted and divided according to cost components related to travelers and operator.

From the table and the figure above it is obvious that solutions found by the SA algorithm do not differ significantly with respect to the frequency and capacity combinations found and the associated cost components. The final solutions yield low frequencies on lines 1 and 4 and relatively high frequencies on lines 2 and 3. Solutions obtained by the local search (LS) algorithm, however, differ significantly especially with respect to the costs related to waiting time and operation. It is obvious that the final solutions are affected by the initial solution since optimal waiting costs increase as the capacity of supply of the starting solution is reduced. The same holds inversely for the operational costs. Note that the final solution generated by LS and starting from a low-capacity solution even contains denied boarding. That is, LS cannot find a solution that completely satisfies demand if initialized from a very low capacity point in this case study. Hence, it can be concluded that final solutions found by the

Table 4.3: Results obtained for different initial solutions and search algorithms.

Initial solution scenario	Line	Simulated annealing			Local search		
		frequency	capacity	total cost	frequency	capacity	total cost
1: high-capacity	1	4	60	3785.41	4	60	3874.82
	2	10	60		12	60	
	3	10	60		12	60	
	4	4	60		12	60	
2: medium-capacity	1	4	60	3832.31	4	60	4020.59
	2	8	60		6	90	
	3	10	60		8	60	
	4	6	60		8	90	
3: low-capacity	1	4	60	3774.25	4	60	4209.75
	2	10	60		6	90	
	3	12	60		5	60	
	4	5	60		4	60	
coeff. of variation				0.81%			4.16%

proposed LS algorithm are dependent on the starting solution whereas SA finds similar solutions independent from the starting solution and is thus robust against this parameter.

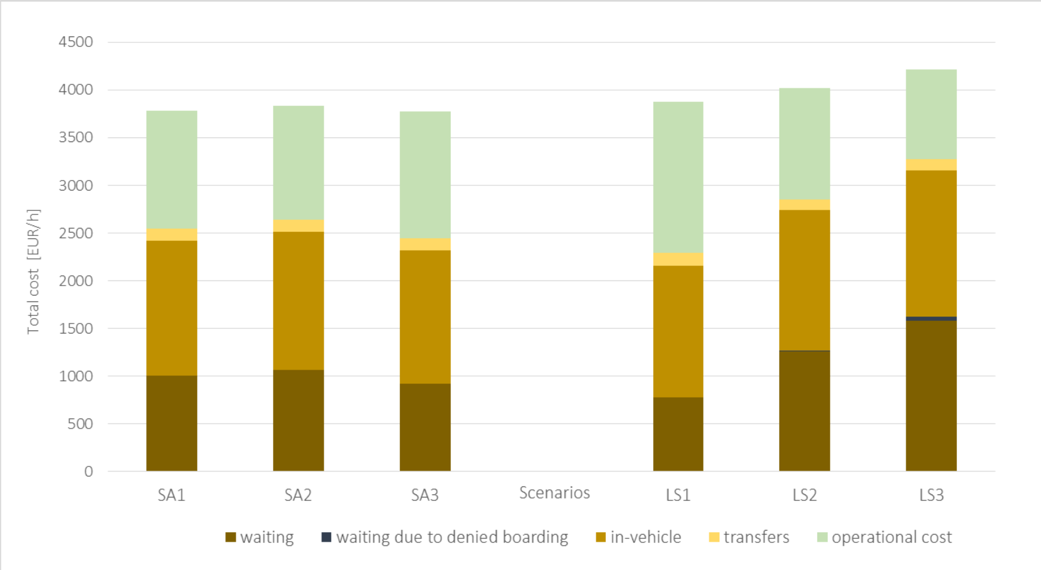


Figure 4.2: Total cost of the final solutions found for different scenarios.

Test 2

Figure 4.3 shows the evolution of the SA algorithm in terms of the objective function value of visited solutions for different values of N, i.e. the number of iterations performed at each temperature value, as well as the current temperature at each iteration. The cooling function is defined by the exponentially decreasing function $T(t+1) = \alpha T(t)$ as described in Section 3.4.2. The parameter α was set to 0.9 and an initial temperature was estimated as 900 based on equation (3.11). Starting from the same initial solution, the algorithm gradually converges to a minimum cost value. During the initial phase of the search progress, the objective function value of visited solutions fluctuates a lot as high temperatures increase the probability that worse solutions are accepted. Since a higher value of N causes the temperature to decrease more slowly, the fluctuations and thus the time needed to converge takes longer for larger N than for smaller.

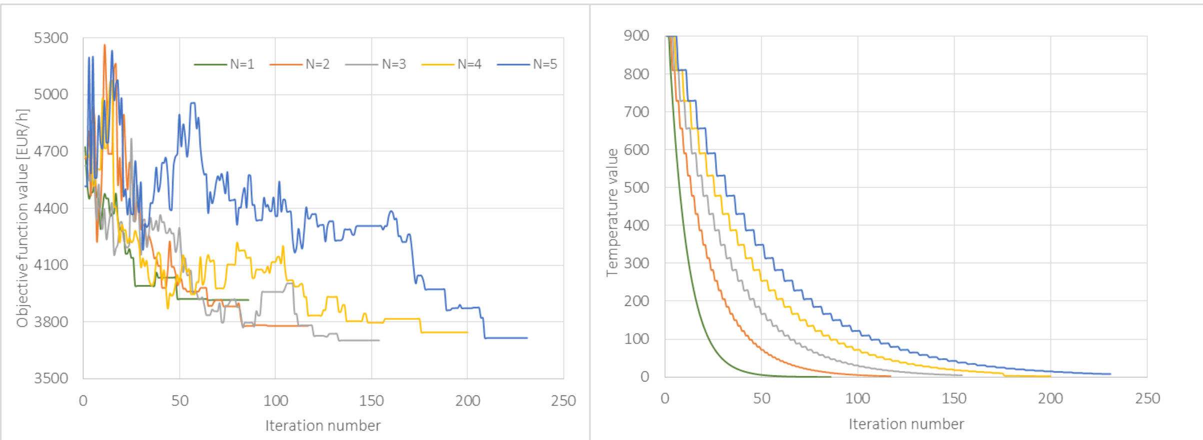


Figure 4.3: Evolution (left) and temperature values (right) of the SA algorithm for different parameter settings.

Figure 4.4 clearly indicates that an increase in parameter N leads to an increase in the overall running time of the algorithm. This finding confirms the previous analysis on algorithm search progress since the number of iterations needed to converge is directly linked to the shape of the cooling function. Figure 4.4 shows that the average cost value of the final solution found is significantly reduced when making two iterations per temperature step instead of one. Hence, a longer runtime of the model, i.e. a more intensive search, can lead to better performing solutions. However, significant improvements are limited as the application of a student's t-test shows no significant differences (at a level of confidence of 95%) among the mean cost values obtained for 3, 4 and 5 iterations. Thus, once a certain runtime is exceeded, the quality of the found solution cannot be further improved.

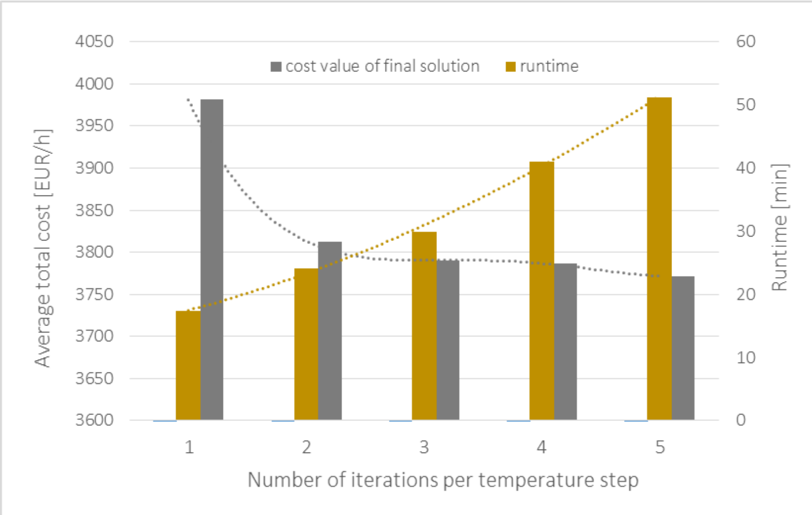


Figure 4.4: Cost value of final solution vs. runtime for different parameter settings of the SA algorithm.

Test 3

In this test, the influence of different weight ratios of waiting to in-vehicle time on the final solution found by the SA algorithm was investigated. Table 4.4 shows the final results obtained for different weight ratios. A ratio of one implies, for instance, that waiting time is perceived and evaluated the same as the time spent on board a vehicle, whereas a ratio of two means that one minute of waiting is perceived the same as two minutes of in-vehicle time by the passenger and the cost value of waiting time is also twice as high. The results clearly show that with increasing weight associated with waiting, overall headways decrease since shorter headways result in smaller waiting times.

Table 4.4: Final results obtained for different weight ratios of waiting to in-vehicle time.

	Weight ratio waiting / in-vehicle time							
	1		1.5		2		2.5	
	H [sec]	C [Pax/veh]	H [sec]	C [Pax/veh]	H [sec]	C [Pax/veh]	H [sec]	C [Pax/veh]
Line 1	900	60	900	60	900	60	900	60
Line 2	600	60	720	120	360	60	300	60
Line 3	600	60	360	60	360	60	300	90
Line 4	900	60	900	60	900	60	720	90

Figure 4.5 shows the objective function value and its composition as well as the average waiting time per passenger for the different weight ratios. It is clearly visible that the overall cost of a solution increase with the relative importance associated with waiting. However, it is also interesting to observe that only an increase in operational cost leads to an overall raise of the costs as those components belonging to the users remain relatively constant. The increase in operational cost can clearly be explained by the change in supply towards smaller headways, which implies higher frequencies and thus more vehicles to operate. The cost of waiting remains more or less constant and thus independent from the increasing importance associated with waiting. Total weighted passenger waiting time remains at the same level because the increase in waiting time coefficient is fully compensated by the decrease in average waiting time per passenger. The objective function value associated with user costs remains relatively stable while the solution itself is quite sensitive. However, as the relative weighting of waiting time increases, an overall raise of the system's cost is caused by an increase on the supply side, that is, higher operational cost. Hence, in this configuration, the absolute savings in waiting time costs are larger than the additional operational costs resulting from the increased supply. Note that in case resources are limited, the model may behave differently since supply is subject to an upper bound and cannot be increased arbitrarily.

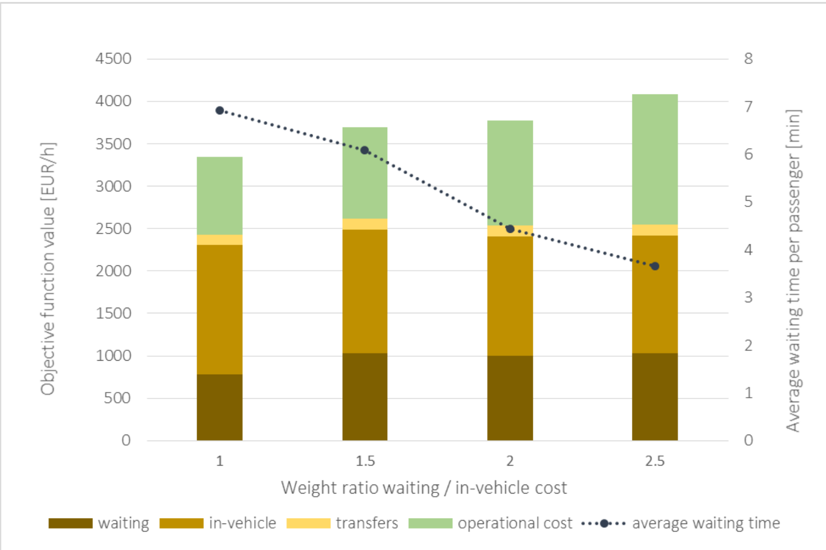


Figure 4.5: Cost values and average waiting time associated with the optimal solution for different weights of waiting time.

4.2 Real-world case study: R-net 'Zaanstreek' concession

This section presents the second application of the developed model to a large-scale case study in order to explore and demonstrate its practical applicability, benefits and limitations. It starts with a description of the case study set-up followed by the presentation of the scenario design developed to explore several features of the model in real settings. Finally, the results obtained by the model for the different scenarios examined are presented and discussed.

4.2.1 Case study description

The case study deals with a bus network to the north of Amsterdam (The Netherlands) which is operated by Connexxion, a Dutch transportation company. The so-called 'Zaanstreek'-concession is named after the area in the north of the Amsterdam metropolitan area which includes several municipalities in the province of 'Noord-Holland'. This concession is currently operated by Connexxion and comprises a total of 14 different bus lines of which two are only operated at night. The entire network can be divided into two parts: 5 lines connecting the central locations of the 'Zaanstreek' area with different locations in Amsterdam and the remaining 7 lines connecting the more remote parts of the northern municipalities. The former group of lines is part of the R-net (or 'Randstadnet'), which is a cooperation of local authorities and operators in the metropolitan area of the Randstad aiming to provide high-quality public transport services. The latter group of lines can be regarded as a local and feeder service to the R-net and is currently operated at low frequencies of 1 to 2 vehicles per hour and line. In the remainder of this chapter, these lines are referred to as local lines. Passengers using these lines usually consult the timetable before starting their journey in order to avoid long waiting times at their stop of origin. The developed model, however, is designed for high-frequency public transport services assuming a random arrival process of passengers at stops. Therefore, the case study will only consider the 5 lines belonging to the R-net, but without ignoring passenger demand using the local lines. More details about passenger demand modelling will follow later in Section 4.2.2.

Figure 4.6 shows a geographical as well as a schematic representation of the considered bus network. In the former representation, some of the local lines are indicated in grey. The entire network belonging to the R-net consists of 5 lines (associated with different colors) and 62 stops in total. Most of the stops are served by multiple lines increasing the amount of supply offered at these stops. Moreover, multiple routes are available for the majority of OD pairs making this network especially interesting in terms of passenger route choice effects and considerations. All five lines are serving the stop 'De Vlinder', which can be regarded as a central transfer hub. Table 4.5 provides a tabular overview of the 5 lines including static properties such as number of stops, line length and origin/destination stops as well as dynamic properties such as period-dependent scheduled riding times and line frequencies during different times of the day. The three lines 391, 392 and 394 are connecting Amsterdam Central Station with the railway station in Zaandam and a location situated further north called Zaanse Schans. Although lines 392 and 394 do have the same origin and destination terminals, their routes significantly differ. Line 392 is slightly longer and has more stops offering the only connection for a segment in the eastern part of the network. Line 395 connects the railway station Amsterdam Sloterdijk with a park-and-ride facility located at the highway A7 (exit 'Wormerland') in the northeast of Zaandam. Note that line 395 is operated with varying terminals during the day. That is, half of the service already ends at the hospital in Zaandam (Zaans Medisch Centrum) and thus the line frequency between the northern branch Zaans MC – P+R A7 is lowered. Due to the formulation of the model, these two varieties of line 395 are treated as two separate lines. More information on the model specification will be given in Section 4.2.2. Line 398 connects the southern business district in Amsterdam with the hospital in Zaandam. Note that this line is only operated in one direction depending on the period. In the morning service runs southbound towards Amsterdam, in the evening busses run northbound. Also note that this line is solely operated at peak times in the morning and the evening and not in operation during other times of the day and at night.

According to a practical expert in the field of public transport service planning in The Netherlands, the frequency determination is usually done in a simplified fashion using norm capacity values. This value is defined as the maximum capacity a vehicle may reach, which is the seating capacity plus a certain share of the standing capacity. The frequency on a line is then determined by dividing the total number of passengers on the busiest line segment in the network by the defined norm capacity. This approach implies a reactive adjustment of supply towards (observed) demand conditions and does not consider demand-supply interactions which need to be computed iteratively using assignment models. In case an assignment model is used, however, crowding and congestion issues are hardly taken into account. If so, crowding is only considered as decreasing comfort onboard a vehicle by incorporating crowding factors and iteratively computing network equilibrium conditions in terms of (perceived) travel times. Hence, issues resulting from dynamic effects such as travel time reliability cannot be taken into account by current static assignment models. By applying the proposed model, these aspect can be addressed.

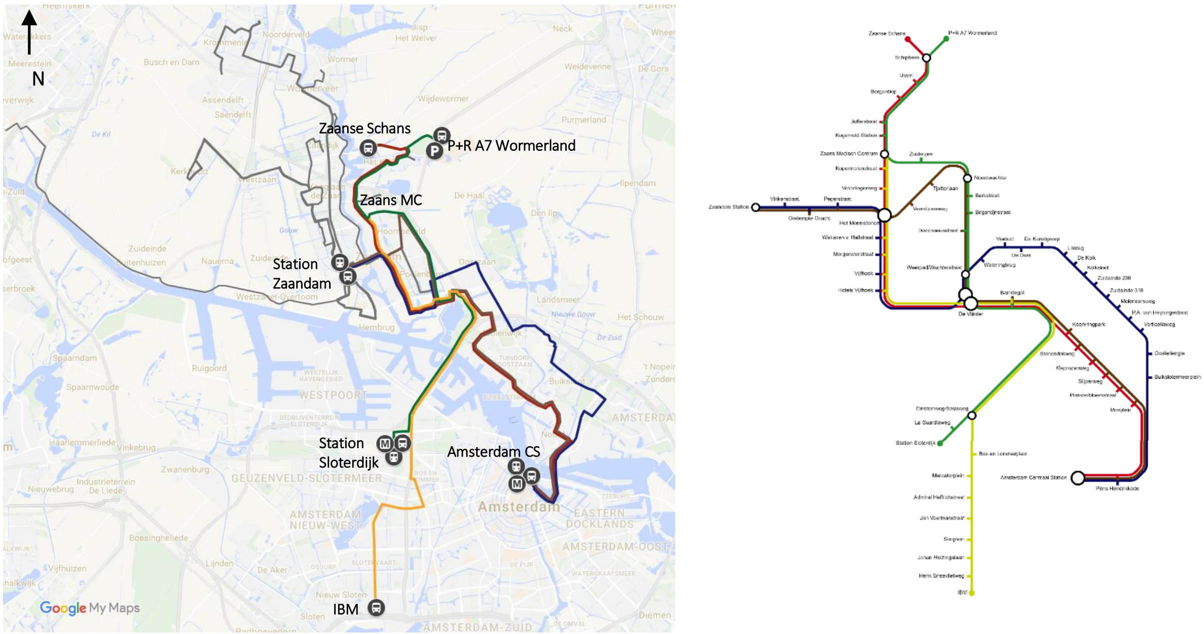


Figure 4.6: Geographical (left) and schematic representation (right) of the 'R-net' bus network.

Table 4.5: Overview of the line-specific characteristics.

Line	Origin	Destination	Frequency [veh/h]		# stops	Length [km]	Scheduled riding time AM peak/PM peak [min]
			During the day 5:00 – 19:00	At night 19:00 - 1:00			
391a	Centraal Station, Amsterdam	Zaanse Schans, Zaandam	4	2	24	21.5	45/48
391b	Zaanse Schans, Zaandam	Centraal Station, Amsterdam	4	2	24	21.1	49/45
392a	Centraal Station, Amsterdam	Station, Zaandam	4/8	2	27	19.2	44.5/49
392b	Station, Zaandam	Centraal Station, Amsterdam	4/8	2	27	19.2	49.5/44
394a	Centraal Station, Amsterdam	Station, Zaandam	4	2	21	17.9	40/43
394b	Station, Zaandam	Centraal Station, Amsterdam	4	2	21	18.1	44/40
395a1	Station Sloterdijk, Amsterdam	P+R A7, afrit 2 Wormerland	4	2	18	17.2	32.5/35
395b1	P+R A7, afrit 2 Wormerland	Station Sloterdijk, Amsterdam	4	2	17	16.4	36/33
395a2	Station Sloterdijk, Amsterdam	Zaans Medisch Centrum, Zaandam	4	0	12	12.2	21/25
395b2	Zaans Medisch Centrum, Zaandam	Station Sloterdijk, Amsterdam	4	0	11	11.9	25/22
398a	IBM, Amsterdam	Zaans Medisch Centrum, Zaandam	PM:3	0	19	18.1	42/40
389b	Zaans Medisch Centrum, Zaandam	IBM, Amsterdam	AM:3	0	19	18.0	39/41

Table 4.6 shows the current supply in terms of line frequencies offered on the R-net during the two peak periods considered. The selected morning and evening peak hours are the result of a passenger demand analysis which will be elaborated on in the Section 4.2.2. It can be seen that almost every line is operated at a frequency of 4 vehicles per hour (15 minutes headway) and line 398 at 3 vehicles per hour in both periods considered. Depending on the presence of multiple lines on a segment, this can add up to a frequency of up to 19 vehicles per hour (between stops 'de Vlinder' and 'Barndegat'). Note that in the evening peak, the frequency setting on line 392 is asymmetrical. That is, there are 8 vehicles per hour on the northbound direction of line 392 and the usual 4 vehicles per hour on the southbound direction. According to information provided by Connexxion, there are currently 3 types of vehicles used to operate the Zaanstreek concession. The majority of the fleet (75%) consists of a standard (non-articulated, 12m) bus type (see Figure 4.7) having 42 seats and an overall capacity of 83 passengers per

vehicle, i.e. 41 standing places. In addition to that, a smaller (10m, 34 seats) and a larger (13m, 52 seats) type of vehicle is available. Since there is no information available on the assignment of vehicle types to lines and the relative differences in terms of capacity are marginal among the different vehicle types, it is assumed, for the sake of simplicity, that the standard bus type is currently deployed on all lines of the R-net. Note that this assumption might involve potential consequences for the assessment of benefits of an optimized supply setting. The magnitude of a reduction in crowding levels, for instance, would turn out to be lower in case a larger vehicle than the standard one was currently used on a specific line. Potential benefits in terms of reductions in waiting times, however, should not be significantly affected by this assumption.

Table 4.6: Current supply in terms of line frequencies for both peak periods considered (base cases).

Line	Frequency (vehicles/hour)	
	AM peak (08:00-09:00)	PM peak (17:00-18:00)
391a	4	4
391b	4	4
392a	4	8
392b	4	4
394a	4	4
394b	4	4
395a1	4	4
395b1	4	4
395a2	4	4
395b2	4	4
398a	0	3
389b	3	0

Figure 4.7: Standard bus type currently deployed on the R-net.

In order to be able to determine the performance of potential solutions in terms of the supply characteristics line frequencies and vehicle capacities, the network was coded in the simulation tool BusMezzo. The fundamental network in BusMezzo consists of 150 nodes and 158 links. Each stop was explicitly modelled per line direction. At transfer hubs such as 'De Vlinder', walking distances between multiple stops served by different lines are introduced in order to allow for transfers. These distances and the average walking speed were reasonably specified with average values of 50 meters at transfer stops and a speed of 50 meters per minute, respectively.

Riding times between stops can be modelled deterministically or stochastically in BusMezzo using predefined distributions. In order to fully capture the dynamic effects of riding time variability on overall public transport reliability and performance, stochastic node servers were implemented in the model. The running time distribution of bus services is characterized by a log-normal distribution (Mazloumi et al., 2010). In Busmezzo, each node server modelling the running time RT_{ij} between stops i and j needs to be specified by a constant delay $DELAY_{ij}$ resulting from the free-flow travel time and a variable term VAR_{ij} which is sampled from a log-normal distribution defined by a mean $MEAN_{ij}$ and a standard deviation SD_{ij}.

$$RT_{ij} = DELAY_{ij} + VAR_{ij}(MEAN_{ij}, SD_{ij}) \tag{4.1}$$

The constant delay is computed using the distance $DIST_{ij}$ to be travelled between stops i and j as well as the average riding speed $\overline{SPEED}_{ij}$ which is dependent on the type of road and traffic (e.g. highway with dedicated bus lane vs. mixed traffic in urban area):

$$DELAY_{ij} = \frac{DIST_{ij}}{\overline{SPEED}_{ij}} \tag{4.2}$$

The mean of the remaining variable part of the riding time ($MEAN_{ij}$) can be computed as the difference between the scheduled travel time TT_{ij} and the sum of free flow travel time and average dwell time $\overline{DT}$ which was set to 7 seconds in this case.

$$MEAN_{ij} = \ TT_{ij} - \left(DELAY_{ij} + \ \overline{DT}\right) \tag{4.3}$$

Based on a study investigating the determinants of bus riding time variations using automatic vehicle location records (Cats, 2017), the coefficient of variation of the variable part of the running time (i.e. the size of the standard deviation relative to the mean) was set to 0.2. Hence, SD_{ij} is always $0.2 * MEAN_{ij}$.

Next to stochastic variations in riding times, BusMezzo is able to simulate the effect of passenger flows on service reliability which is mainly manifested through the dwell time. In BusMezzo dwell times are modelled explicitly for each stop in the network using a predefined set of given dwell time functions. According to Cats (2011), the general form of the dwell time function of vehicle type f on trip k of line l at stop s is:

$$D_{s,l}^{k,f} = lost\_time_s^f + \ PST_{s,l}^{k,f} + v_{s,l}^{k,f} \tag{4.4}$$

Where: $lost\_time_s^f$ is a constant delay associated with stop s and vehicle type f
$PST_{s,l}^{k,f}$ is the total passenger service time
$v_{s,l}^{k,f}$ is a stochastic error term

There are various models mentioned in literature to compute the passenger service time. Most of them take into account the number of boarding and alighting passengers at a stop and the service time needed for each boarding and alighting passenger respectively. The service time is the marginal contribution of each boarding/alighting passenger to the total passenger service time. Other models go more into detail and also consider vehicle occupation levels (crowding) when approaching a stop, door configurations as well as boarding/alighting regimes. In this case study, a formulation for the calculation of the total passenger service time proposed by Weidmann (1994) is used.

$$PST_{s,l}^{k,f} = max\{\beta_a^f \cdot \ A_{s,l}^k \,, \beta_b^f \cdot \ B_{s,l}^k\} \cdot \left[1 + \frac{3}{4}\left(max\left\{0, \frac{L_{s,l}^k - seats^f}{cap^f - seats^f}\right\}\right)^2\right] \tag{4.5}$$

Where β_a^f is the service time per alighting passenger from vehicle type f
$A_{s,l}^k$ is the number of alighting passengers at stop s on trip k of line l
β_b^f is the service time per boarding passenger from vehicle type f
$B_{s,l}^k$ is the number of boarding passengers at stop s on trip k of line l
$L_{s,l}^k$ is the passenger load on trip k of line l when approaching stop s
$seats^f$ is the total number of seats available in vehicle type f
cap^f is the total capacity of vehicle type f

This model implies that not only the number of boarding and alighting passengers contribute to the passenger service time but also the crowding level inside the vehicle. Note that the extra dwell time due to crowding increases non-linearly (quadratic) with the number of standing passengers. The service time coefficients β_a^f and β_b^f are specified for different public transport services depending on the respective boarding regime and number of doors.

Although BusMezzo is able to simulate vehicle scheduling, i.e. the movement of vehicles executing multiple trips on one or multiple lines, this feature is disregarded in this application as already explained in Section 3.2.3. Instead, each vehicle is assigned to one trip on a line only, implying that dynamic effects related to trip chaining such as the propagation of vehicle delays over multiple trips are not taken into account. The supply for each line is specified by the dispatching time of the first vehicle run and the line's headway which is assumed to be constant throughout the analyzed period. Recovery times at terminals for fleet size estimations and operational cost computations are assumed as 5 minutes on all lines. The dispatching times are based on the current timetable (Connexxion, 2016) and kept constant for all potential headway settings per line. That is, planning decisions related to timetabling in terms of the definition of vehicle arrival/departure times are not taken into account. The only decisions variables are headway, i.e. the frequency of vehicle arrival, and the vehicle capacities on the line level. Although the current timetable does not always imply perfectly equally spaced headways for the departure times at origin terminals of a line, departure headways are assumed to be homogeneous throughout the simulation.

4.2.2 Passenger demand analysis

In this case study, passenger demand is assumed to be given as a constant input in form of an OD matrix specifying the average flow of passengers per hour between each pairs of stops in the network. Using this input, the random arrival of passengers at stops is stochastically simulated in BusMezzo according to a process following a Poisson distribution. Note that a feedback loop between demand and supply in terms of demand elasticity is not taken into account in this model.

Figure 4.8 shows the distribution of the average number of trips aggregated for a time interval of one hour during an average working day (Monday to Friday). The analysis is based on a large sample of recent smartcard transaction records for the entire month of February 2017 consisting of approximately 439.000 records. Each transaction record contains the time and date of check-in and check-out as well as the initial boarding and final alighting stop. Moreover, the line used for the first boarding and the final alighting is recorded as well. Note that this kind of demand description does not allow for an exact replication of the routes taken by passengers through the network as intermediate transfer stations and lines are not recorded. Yet, route choice is explicitly modelled by BusMezzo, hence allowing passengers to react to changes in supply. Figure 4.8 clearly indicates that the morning and evening peaks occur in the time between 08:00-09:00 AM and 05:00-06:00 PM, respectively. In the morning peak approximately 1.300 trips are performed on average whereas during the evening peak this values is slightly less with an average of 1.200 trips.

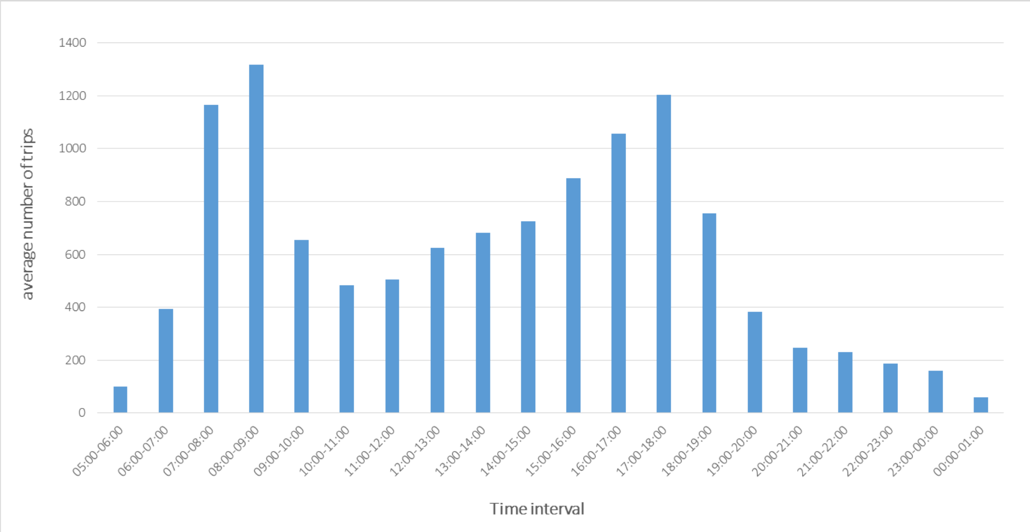


Figure 4.8: Distribution of average number of trips on the network during a working day according to time intervals.

Figure 4.9 shows the spatial distribution of boardings and alightings in terms of locations of demand generation and attraction on the network during the both peak hours highlighting the four busiest stops. It is clearly visible that the general spatial distribution of origins in the morning matches those of the destinations in the evening. The same holds vice-versa for origins in the evening and destinations in the morning. This observation suggests a typical commuting pattern with a tour starting and ending at the same location in the morning and the evening respectively. However, if one takes a closer look at Figure 4.9 one can see that there is not a perfect symmetry between the patterns in the morning and the evening. In the morning, for instance, Amsterdam Central Station is the busiest stops in terms of origins whereas in the evening destinations are more balanced between Central Station and Prins Hendrikkade. This asymmetry becomes even more obvious when Figure 4.10 is observed. Although there are some similar patterns visible when comparing passenger OD flows in the two periods, there are also significant differences. Especially the flow patterns originating/ending at Amsterdam Central Station significantly differ among the two periods since this stop far busier in the evening than in the morning. Other asymmetries can be observed at the stops of Station Sloterdijk and Zaandam. All in all, these observations suggest that some passengers tend to use the services on the R-net only in one direction of their commute and might thus use alternative modes such as train/metro/tram for the other direction or simply use different routes in terms of different origin/destination stop depending on the period.

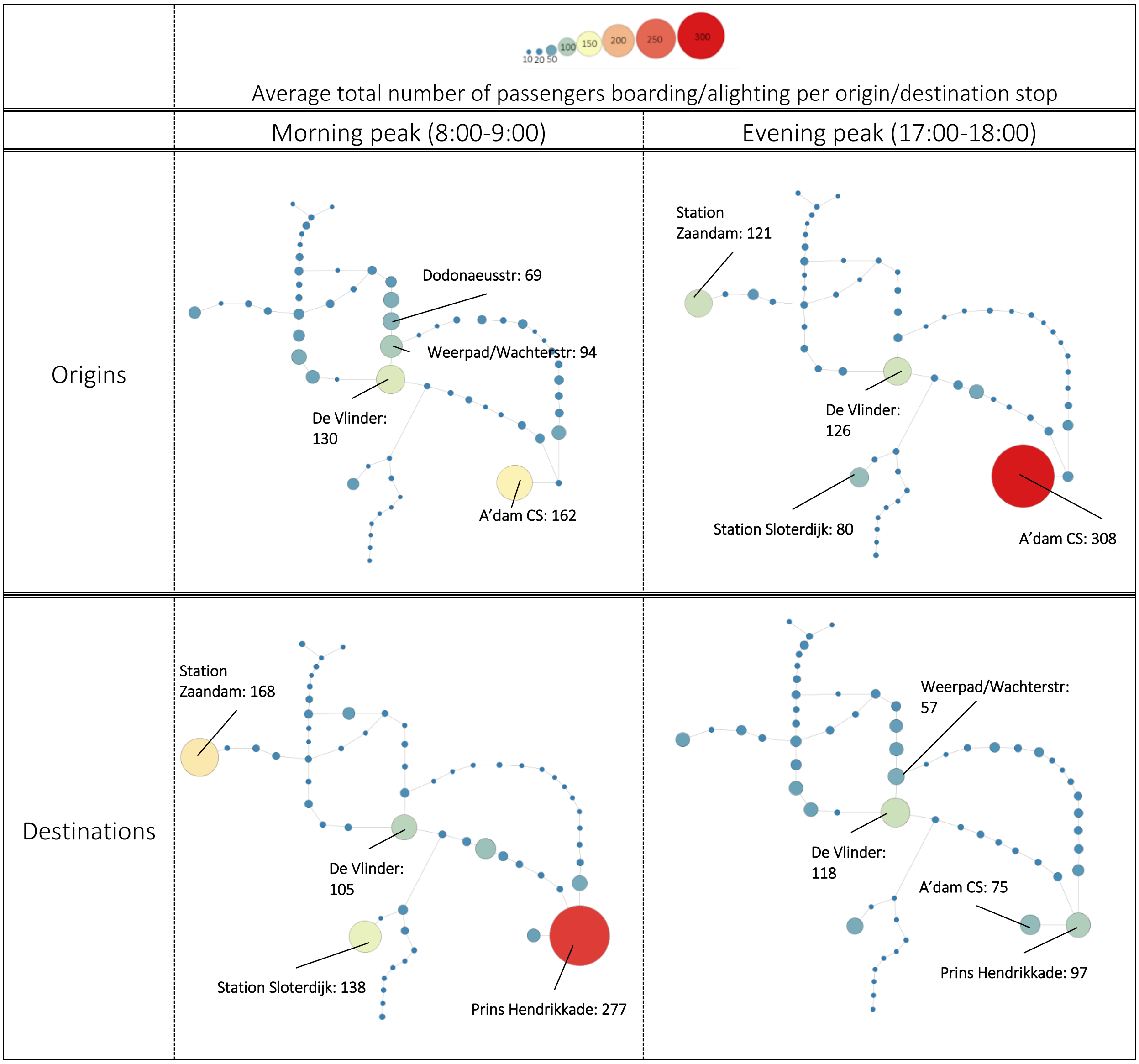


Figure 4.9: Average total number of passengers boarding/alighting per origin/destination stop during the morning and the evening peak hour including the four busiest stops.

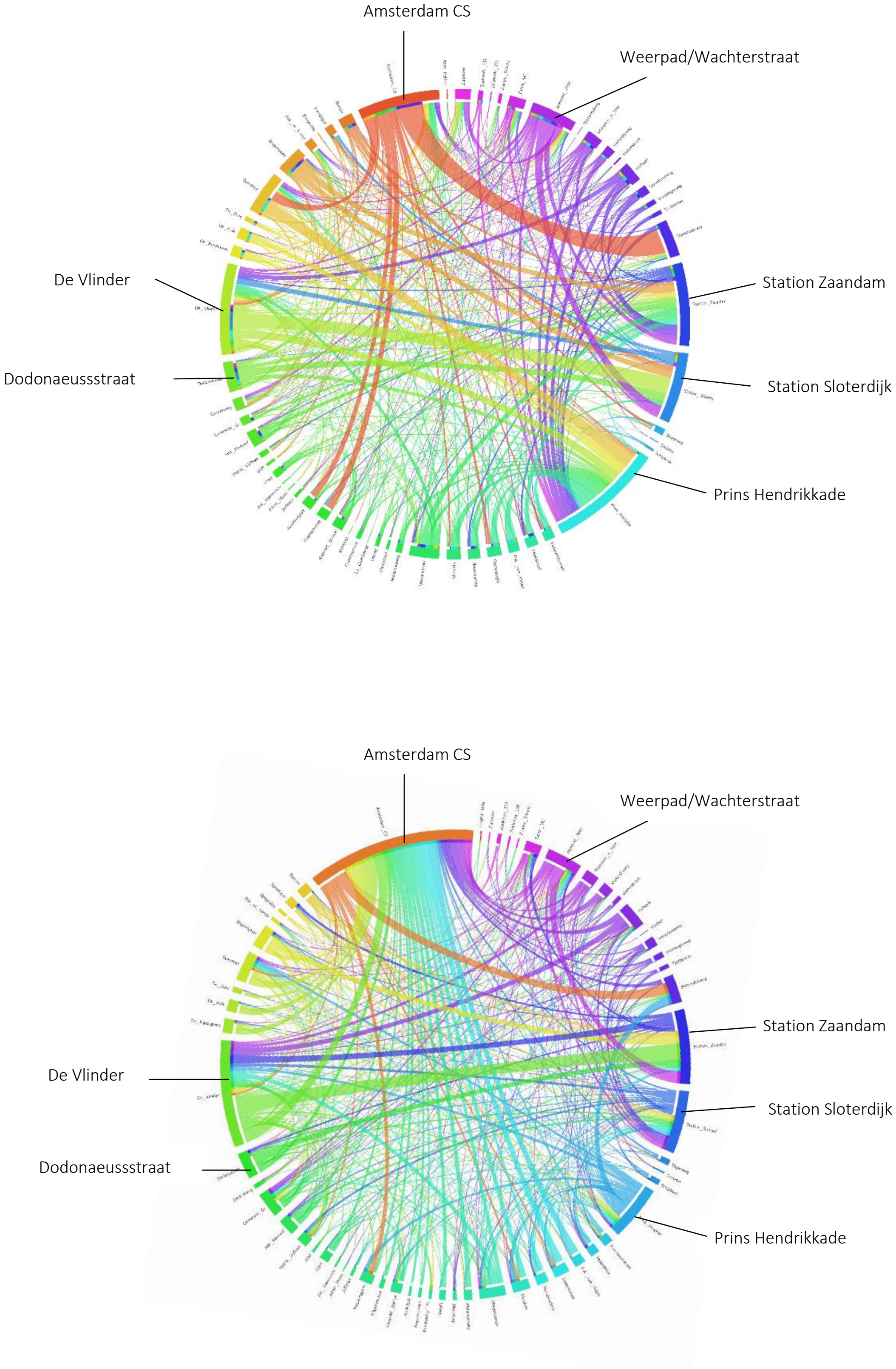


Figure 4.10: Trips made on the network during the morning peak (top) and evening peak (bottom), colored by station of origin (top) and station of destination (bottom) respectively.

Figure 4.12 shows the results of a passenger assignment to the network using the BusMezzo simulation tool. Note that the day-to-day learning feature was used in order to create steady-state conditions. The thickness of the network links indicates the total number of passengers traversing that link within the considered peak hour. The link color indicates the average seat occupancy level OCL_{ij} on the line segment between stops i and j which is computed by the passenger flow on the link PL_{ij}, the seating capacity C_L of a vehicle deployed on line L traversing the link and the line's frequency f_L. Note that the total seating capacity of a link is added over all lines traversing the link.

$$OCL_{ij} = \frac{PL_{ij}}{\sum_{L\in(ij)} C_L * f_L} \tag{4.6}$$

As can be clearly observed from Figure 4.12, average occupancy levels are higher in the morning than in the evening peak. On the southbound direction of line 392, maximum occupancy levels of almost 100% are reached in the morning peak. That is, on average all seats are occupied on the respective segments of line 392. Note that this also implies that total vehicle capacities are (on average) never exceeded in any period and any point in the network.

Moreover, a clear direction of flows depending on the period can be observed as flows are not always balanced between the two directions of a line segment during a given time period. Figure 4.11 shows the average passenger flow on a line segment per line, peak period and direction. Moreover, the absolute difference in passenger flows between the two directions of a line is depicted for both periods. This difference is slightly larger on lines 394 and 395 in the evening than in the morning. On line 392, this difference is most pronounced as in the morning, flows are almost equally balanced in both direction and in the evening the directional difference is more than doubled.

Another way of quantifying the spatial balance of demand in a network is to examine the symmetry of the OD matrix. Since a symmetric matrix OD is equal to its transposed matrix OD^T, a measure quantifying the degree of symmetry of a matric can be defined as (Mathforum, 2017):

$$Sym(OD) = \frac{|(OD + OD^T)/2|}{|OD|} \tag{4.7}$$

Where $|\ |$ denotes the norm of a matrix which is the square root of the sum of the squared matrix entries. $Sym(OD)$ is normalized and ranges over $[0,1]$ with $Sym(OD) = 1$ meaning perfect symmetry and $Sym(OD) = 0$ total asymmetry. Using equation 4.7, this value yields 0.5541 for the morning peak and 0.5522 for the evening peak. Hence, the OD matrix in the morning peak is slightly more symmetric and balanced than in the evening peak which is in line with the results depicted in Figure 4.11.

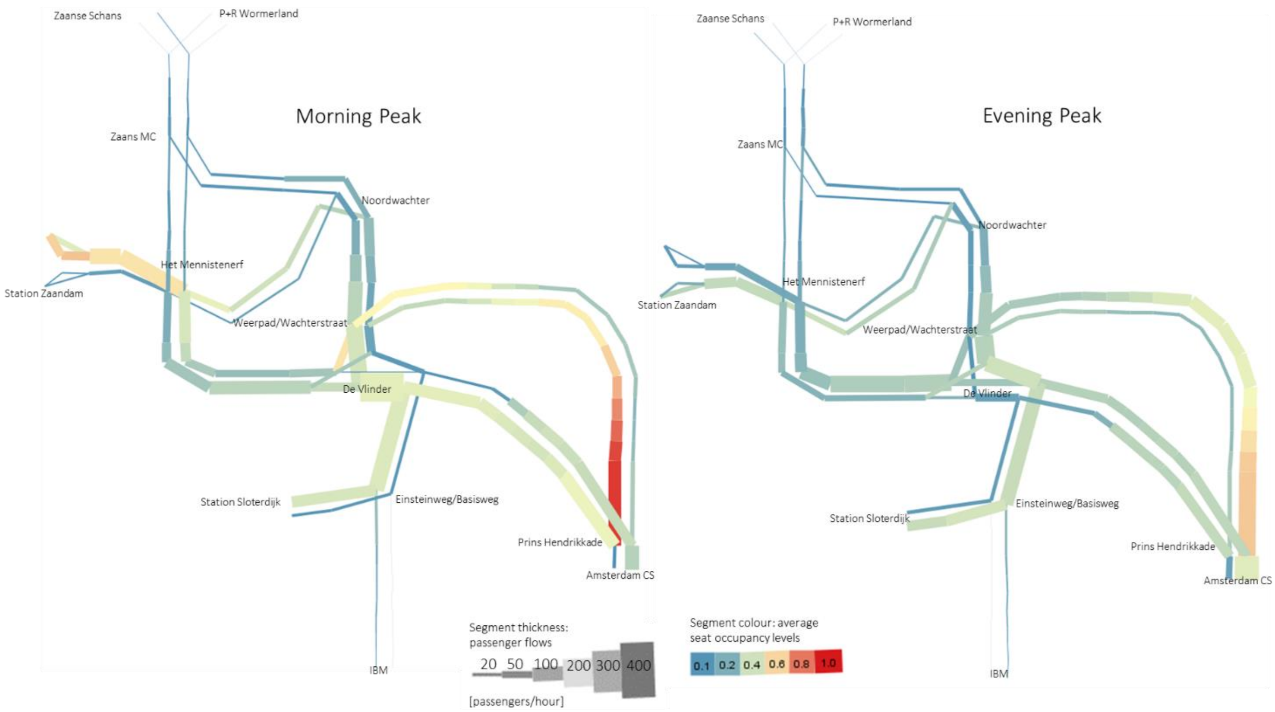


Figure 4.12: Total passenger flows on the network during the morning and the evening peak hour including average seat occupancy levels.

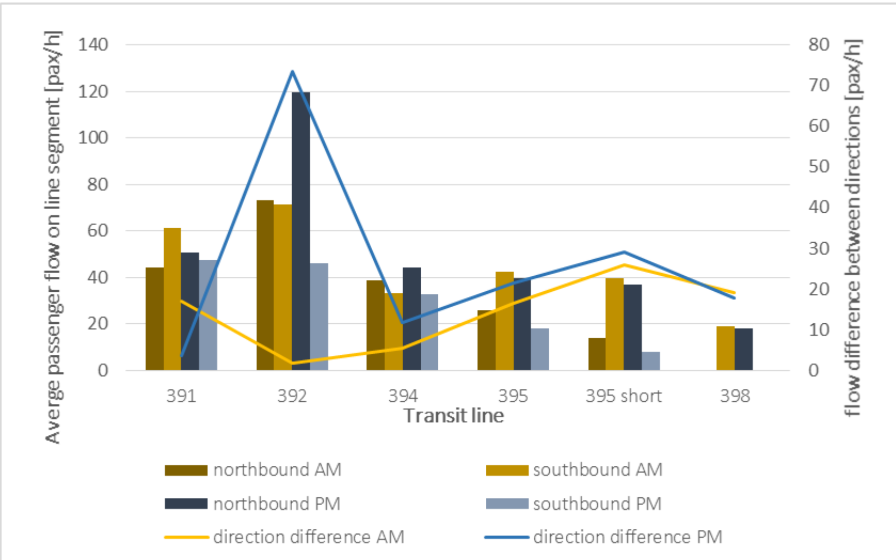


Figure 4.11: Average passenger flow on a line segment (between two successive stops) for different line directions and periods including the absolute flow difference between the directions of each line per peak period.

4.2.3 Scenario design

In order to examine the effect of different assumptions regarding input parameters and constraints on the model outputs different scenarios have been designed. All of them can be categorized according to three characteristics regarding the specification of decision variables, the formulation of the objective function and the passenger demand setting. Table 4.7 lists an overview of all 12 scenarios categorized according to these characteristics.

The model aims at finding a good solution in terms of line frequencies and/or vehicle capacities. Since the two decision variables can be determined simultaneously or separately, it is interesting to investigate the effect of the assumption on these settings on the final solution. Therefore, some scenarios do not take into account the decision variable vehicle capacity but consider a given homogenous vehicle type as an exogenous input variable. Other scenarios determine vehicle capacities as well as line frequencies by selecting appropriate types of vehicles from a given heterogeneous vehicle fleet. In that way, the influence of deploying different vehicle capacities in the network can be tested. Using small vehicles on lines with low demand can, for instance, save operational costs that can instead be used to increase capacity on highly-utilized lines.

As presented in the previous section, the case study is characterized by significant differences regarding the passenger demand during different periods of the day. Also, currently provisioned supply is different during these periods. Therefore, it is interesting to make a further distinction between the input parameter passenger demand in order to explore how the structure of different OD matrices may affect the final result. Moreover, it is worth exploring and comparing the potential of improvement for the two base situations. Since the overall quantity of supply provided in the morning peak is currently less than in the evening there might be a greater potential for improvement from a passenger's point of view.

Given the asymmetric distribution of demand within the network it is moreover worthwhile investigating how different assumptions on the decision variable line frequency can affect the quality of the final result. Next to the standard way of setting frequencies equally in both directions of a line, other scenarios are investigated in which the frequency is determined separately per line direction. This relaxation of constraints on the decision variable might lead to a more efficient allocation of resources since supply can be adjusted to the present demand in a more accurate and targeted fashion than by setting the same frequency for both directions of a line. Note that the set of all symmetric solutions is a subset of the set of all asymmetrical solutions.

Finally, a further distinction regarding the formulation of the objective function is introduced by defining two different objectives: First, the minimization of total system costs as already presented in the first case study, and second, the minimization of user-related costs subject to a defined budget constraint. The latter one aims at minimizing generalized travel costs or time by changing supply in a way such that a maximum budget is not exceeded. A distinction regarding the objective function will demonstrate the capabilities of the model when applied with different practical intentions. The optimization of total costs facilitates the process of decision making and trade-off while considering the interests of both the operator and the passengers. In contrast to that, the model can also be used to find an optimal allocation of a given set of resources or budget while considering exclusively passenger benefits.

Table 4.7: Scenarios examined in the case study categorized by formulation of objective, passenger demand input and assumptions on the decision variables frequency and vehicle capacity.

vehicle fleet	homogeneous				heterogeneous	
frequency setting	symmetrical		asymmetrical		symmetrical	
Objective	Min UC	Min TC	Min UC	Min TC	Min UC	Min TC
AM peak demand	AM_UC_SYM	AM_TC_SYM	AM_UC_ASYM	AM_TC_ASYM	AM_UC_VEHCAP	AM_TC_VEHCAP
PM peak demand	PM_UC_SYM	PM_TC_SYM	PM_UC_ASYM	PM_TC_ASYM	PM_UC_VEHCAP	PM_TC_VEHCAP

Each of the 12 scenarios is formulated as an abbreviation indicating the demand period (AM or PM), the objective (UC=User cost, TC=Total cost) and constraints on the decision variables (SYM/ASYM=symmetrical/asymmetrical frequency setting; VEHCAP=consideration of multiple vehicle capacities). Those abbreviations will be used throughout the remainder of this chapter.

Table 4.8 shows the three different vehicle types considered for the 'VEHCAP'-scenarios. The normal bus is the standard city bus currently deployed in the network as depicted in Figure 4.7. According to information provided by Connexxion, two other bus types can be used for the operation of the concession as well. However, due to the marginal differences in seating and standing capacity values compared to the standard bus type which might lead to negligible small differences in the performance of potential solutions, two other types of vehicles providing that exercise more profound trade-offs regarding vehicle capacities and operational costs were selected. A minibus having about half of the seating capacity of the standard bus and fewer standing places relative to the number of seats was selected as a cheaper option with lower capacity. In contrast to that, a larger articulated bus (MAN Truck & Bus AG, 2017) was chosen as a more expensive option with a high standing capacity which is about double that of the standard bus.

Table 4.8: Vehicle-specific characteristics and operational cost components for the three different vehicle types considered.

	Minibus	Normal bus	Articulated bus
# seats	20	42	53
Total capacity [pax]	35	83	158
Length [m]	8	12	18
# doors front/rear	1/1	1/1	1/2
Boarding coeff. [s]	2.5	2	2
Alighting coeff. [s]	1.5	1	0.5
Time-base costs C_p [EUR/veh.h]	48	48	48
Fixed costs CF_c [EUR/veh.h]	4.46	4.91	6.62
Distance-based costs CD_c [EUR/veh.km]	0.37	0.58	0.93
Cost factor per veh.km β_c	0.93	1.0	1.13

Depending on the number of doors and the boarding regime, coefficients for the calculation of the passenger service time (according to equation (4.5)) have been chosen in accordance with common values used in the literature (Weidmann, 1994). It was assumed that boarding solely takes place at the front door whereas all rear doors can be used by alighting passengers only. In case of multiple rear doors (articulated bus), the marginal time per alighting passenger decreases because more passengers can simultaneously alight from the vehicle.

The operational costs consist of three components. Time-based costs CP include salaries for personnel on the bus and administration costs. This cost component is independent of the type of vehicle used. Fixed costs CF include insurance costs, vehicle taxes, supplement for carriage reserves, depreciation of investment costs etc. and are therefore dependent on the type of vehicle. Distance-based costs CD include the cost of fuel, lubricating oil, tires, spare parts etc. and are also specified per type of bus. The cost values for standard and articulated buses are based on Swedish recommendations for cost/benefit analyses (Trafikverket, 2016) (assumption 10 SEK = 1 EUR) and the values for minibuses are based on a German study on the determination of operational costs for bus services (Frank et al., 2008). Note that costs related to the fuel consumption while standing still with the engine running as originally considered in the model are very small and already included in the distance-based costs and therefore not taken into account separately. Moreover, the factor α which is used to consider indirect costs in the original model is not used either since indirect costs such as administration costs are already included in the cost factors mentioned. The total operational costs OC can eventually be computed as:

$$OC = \ CD + CP + CF = \sum_{l \in L} \sum_{c \in C} L_l \cdot f_l \cdot CD_c \cdot \delta_{l,c} + C_p \cdot \sum_{l \in L} \left\lceil \frac{T_l}{H_l} \right\rceil + \sum_{l \in L} \sum_{c \in C} \left\lceil \frac{T_l}{H_l} \right\rceil \cdot CF_c \cdot \delta_{l,c} \qquad (4.8)$$

Where L_l is the length of line l in km, f_l is the determined frequency on line l and $\delta_{l,c}$ is a binary variable indicating if bus capacity c was assigned to line l. T_l is the minimum cycle time of line l and H_l the determined headway. Note that this formulation allows considering both a line in two directions as well as each direction separately. In the former case, values for L_l and T_l are just summed over the two directions. The set of all lines and vehicle capacities is denoted as L and C, respectively.

Total costs to be borne by the passengers (or users) of the public transport service UC are computed using the value of in-vehicle time VOT and the (total) generalized travel time GTT.

$$UC = VOT \cdot GTT$$

$$= VOT \cdot \left[\theta_{w,initial} TWT_{initial} + \theta_{w,denied} TWT_{denied} + \theta_{IVT} TIVT + \theta_t TT + \ \theta_{wlk} TWLKT\right] \qquad (4.9)$$

Where $TWT_{initial}$ and TWT_{denied} are the total waiting times for initial waiting and due to denied boarding; $TIVT$ is the total nominal in-vehicle time, TT is the total number of transfers and $TWLKT$ is the total walking time. The factors θ represent the relative weights that each of the travel time components contributes to the generalized travel time, expressed as multipliers of the nominal in-vehicle time. Waiting time for the first desired boarding, either at the first stop or when interchanging, and walking time spent for access, egress and transfers are valued as $\theta_{w,initial} = \ \theta_{wlk} = 2$ (Wardman, 2004). Each transfer is valued as a fixed penalty equal to 5 minutes of in-vehicle time, thus $\theta_t = 5$ (Balcombe et al., 2004). Waiting time due to denied boarding probably imposes a higher disutility for passengers than normal waiting times since they are unpredictable and cause delays. Hence, it can be considered equal to the value of delay time for which a multiplier of 3.5 was estimated by Börjesson et al. (2012). Thus, the weight associated with waiting time due to denied boarding becomes

$\theta_{w,denied} = 2 \cdot 3.5 = 7$. Note that the values specified for the utility functions of the route choice model in BusMezzo are consistent with the weights used for evaluating user travel costs.

Crowding onboard a vehicle which imposes a disutility in form of reduced comfort to passengers is taken into account by multiplying the nominal in-vehicle time experienced by each passenger with a crowding factor that depends on the level of crowding and also considers whether the passenger is seated or standing. A meta-study by Wardman and Wehlan (2011) proposes in-vehicle time multipliers for seating and standing passengers which depend on the load factor in terms of seat occupancy levels. Figure 4.13 shows the multipliers for different load factors. For seated passengers, multipliers range from 0.95 at 50% occupancy and less to 1.71 at 200% (i.e. vehicle filled at full capacity). Standing passengers are considered separately once all seats are taken (at an occupancy level larger than 100%) with multipliers ranging from 1.78 to 2.69. Note that the in-vehicle multiplier increases non-linearly with the load factor and the relative increase is faster for standing than for seated passengers: The perceived in-vehicle time is always larger for standing passengers as comfort is generally lower when standing onboard. Note that the multiplier θ_{IVT} can be seen as an aggregated weight summarizing the total level of comfort of all passengers overall all vehicle rides during the considered period.

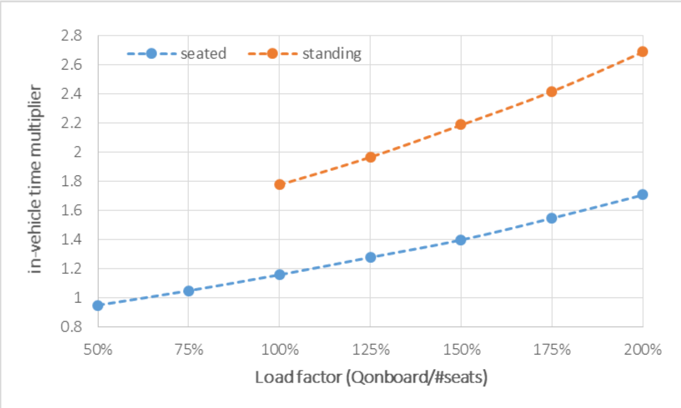


Figure 4.13: In-vehicle crowding multipliers.

The value of time VOT can be determined for different modes and travel purposes and is dependent on the location as well since economic aspects such as average income levels influence the value of time. In this case study, the value of time determined for the modes bus, tram and metro and averaged over all travel purposes in The Netherlands is used. This value currently equals 6.75 EUR per hour (Kouwenhoven et al., 2014).

For the scenarios only taking into account passenger benefits by minimizing user-related costs (UC), a budget constraint needs to be introduced. Therefore, an operational budget in terms of total vehicle kilometers provided is defined. When multiple types of vehicles are considered (UC_VEHCAP scenarios), the operational costs in terms of total vehicle kilometers need to be adjusted according to the differences in costs components per type of vehicle. That is, the amount of vehicle kilometers is computed using costs factors per vehicle type relative to the standard normal bus size. These factors are computed by relating all operational cost components to the unit of one vehicle kilometer and finally adding them up. In order to convert cost components related to the time unit vehicle hour to the distance based formulation, an average vehicle speed needs to be assumed. In accordance with literature on bus operations (Frank et al., 2008), this speed value was set to 20 km/h. For the time-based costs $C_{p,}$, for instance, this translation becomes: 48 EUR/veh.h / 20 km/h = 2.40 EUR/veh.km. The final cost values obtained per vehicle kilometer are set in relation to the standard bus. These factors β_c

are listed in Table 4.8. The operational costs per vehicle kilometer increase for instance by 13% for the articulated bus compared to the standard bus. The total operational costs in terms of vehicle kilometers equivalent per hour can be computed as:

$$TVKM = \sum_{l \in L} \sum_{c \in C} L_l \cdot f_l \cdot \beta_c \cdot \delta_{l,c} \tag{4.10}$$

A maximum amount of total vehicle kilometers is needed as a budget constraint in order to define the feasibility of potential solutions. This budget limit was set to 907.28 vehicle kilometers which corresponds to the current supply offered in the morning hour between 07:00-08:00 AM on working days (Connexxion, 2016).

The developed frequency and vehicle capacity determination model using the simulated annealing algorithm was run for all 12 scenarios depicted in Table 4.7. Due to the different degrees of freedom regarding the decision variables, the number of potential and feasible solutions is varies among scenarios. Table 4.9 shows some properties of the different TC scenarios relevant for the specification of the SA search algorithm. The number of potential solutions is highest in the ASYM scenarios since the determination of frequencies is performed for 11 individual lines and the number of combination increases exponentially with the number of lines. For the VEHCAP scenarios, the solution space is smaller since only 6 lines are considered (with symmetric frequency setting). Note that the number of feasible solutions is smaller in the UC scenarios since feasibility constraints in terms of the budget limit become binding. For the SYM scenario, the number of feasible solutions is reduced by about 96% (4.802 feasible solutions) compared to the number of all potential solutions. Note that due to the large number of potential solutions in the ASYM and VEHCAP scenarios, it is very costly in terms of computation time and internal memory to determine the exact number of feasible solutions for the UC scenarios using conventional computers. The discrete set of potential headways was chosen as: 5, 6, 7.5, 10, 12, 15 and 20 minutes.

The total runtime of the search algorithm can be approximated by the product of 4 factors: the number of days ($\#days$) needed to simulate the day-to-day adaption of passengers, the number of replications ($\#repl$) needed in order to obtain statistically sound results from the stochastic simulation, the number of iterations of the SA algorithm (i.e. the total number of solution evaluations, $iter_{SA}$) and the runtime of one simulation instance in BusMezzo (RT_{BM}).

$$RT = \#days \cdot \#repl \cdot RT_{BM} \cdot iter_{SA} \tag{4.11}$$

Table 4.9: Average total runtimes of the SA algorithm for the different TC scenarios considered.

Scenario	Number of potential solutions	Average number of SA algorithm iterations	Average total runtime [min]
SYM	$1.18*10^5$	80	40
ASYM	$1.98*10^9$	500	240
VEHCAP	$8.58*10^7$	300	150

The number of replications needed in order to obtain statistically sound average results from the simulation outputs was set to 10 yielding a maximum allowable error of 1% of the average objective function value (using equation 3.1). The number of days needed in BusMezzo to simulate the day-to-day learning of passengers is dependent on a parameter which determines the degree of convergence at which the day-to-day loop is terminated. Setting this parameter to 0.2 results in an average of 10 days per simulation run. The number of iterations of the SA algorithm are finally dependent on the number of iterations per temperature step and the termination criterion which are both external parameters to be specified by the user. With respect to this, it is important to choose appropriate parameters to make the runtime of the algorithm proportional to the size of the solution space since more potential solutions require a more intensive search. The number of iterations per temperature step were chosen based on the practical experience gained in the previous application in which the number of feasible solutions was around $3.3*10^5$ and thus of a similar magnitude as in the current SYM scenarios. For the other scenarios, the parameter was increased in order to account for a larger solution space, yet still ensuring that overall running time stays within reasonable ranges of a maximum of a few hours. The termination criterion was determined by the number of possible neighbor solutions given the decision variable setting of a certain scenario, thus ensuring that the found solution is a local optimum (as described in Section 3.4.2). The initial temperature value was estimated using Equation (3.11).

4.2.4 Results

This section presents and discusses the results obtained by the model for the various scenarios. First, the found solutions in terms of supply are presented and contrasted. Then, the relative qualities and performances of the found values with respect to the objectives are presented and analyzed. Finally, resulting passenger flows through the network in different scenarios in the morning peak and the implications on capacity utilization is presented and discussed.

4.2.4.1 Decision variable values

Table 4.10 lists the solutions in terms of headways and vehicle types per line found by the SA algorithm for the 12 different scenarios including the base cases. Moreover, the total amount of vehicle kilometers is included as well which quantifies the total intensity of the service.

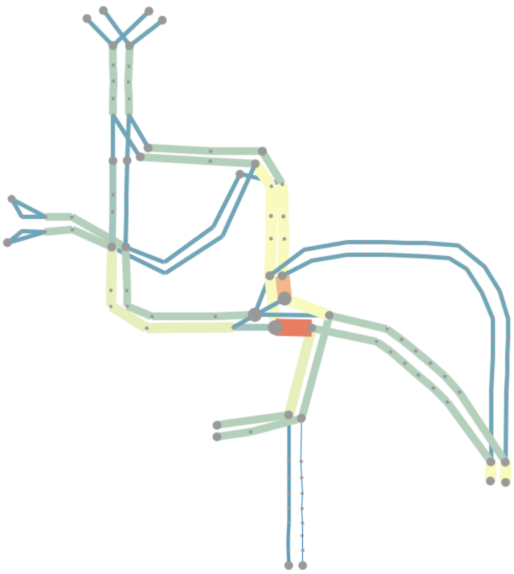
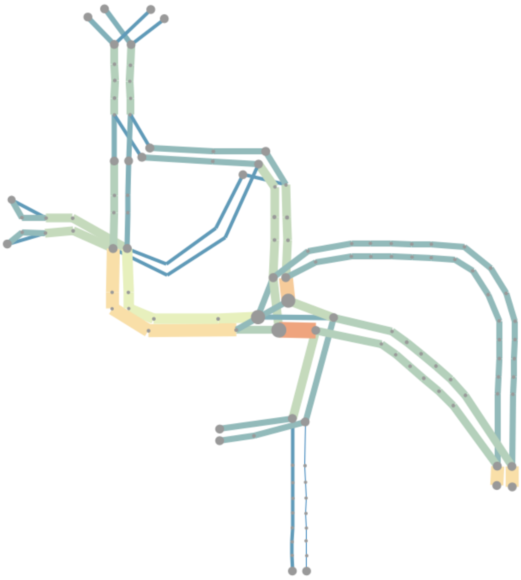
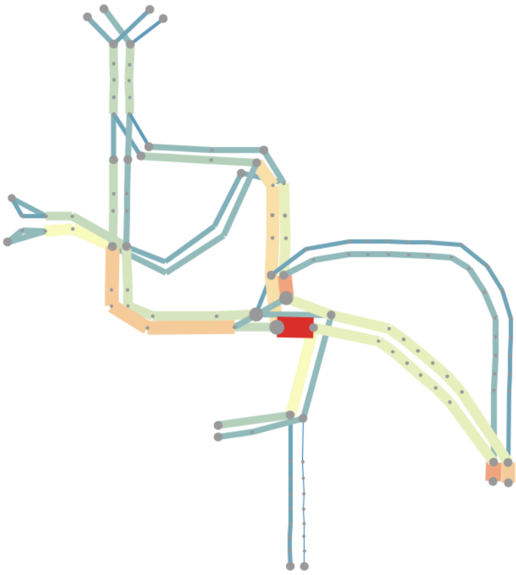
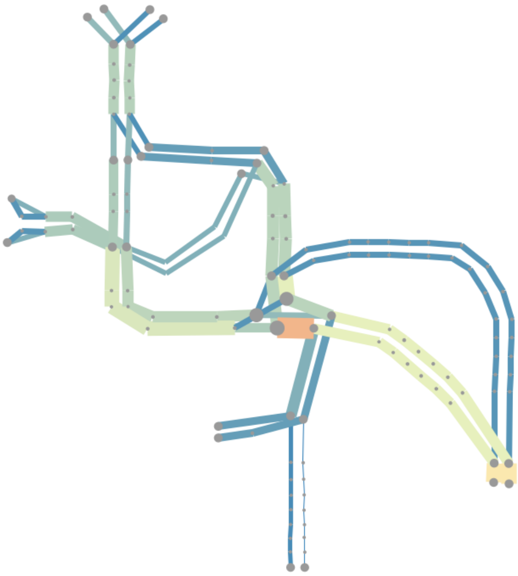
Table 4.10: Solutions found by the SA algorithm for the different scenarios in both peak hours considered.

period	Morning peak									
Scenario	AM_BASE		AM_UC_SYM	AM_UC_ASYM	AM_UC_VEHCAP		AM_TC_SYM	AM_TC_ASYM	AM_TC_VEHCAP	
Line	headway [min]	vehicle type	headway [min]	headway [min]	headway [min]	vehicle type	headway [min]	headway [min]	headway [min]	vehicle type
391a / 391b	15	normal	10	12 / 7.5	15	large	12	10 / 12	10	normal
392a / 392b	15	normal	12	10 / 12	10	mini	10	15 / 10	10	mini
394a / 394b	15	normal	15	15 / 20	10	mini	20	12 / 10	12	normal
395a1 / 395b1	15	normal	20	20 / 12	15	mini	20	20 / 15	12	mini
395a2 / 395b2	15	normal	12	20 / 15	15	normal	20	20 / 15	20	mini
398b	20	normal	12	15	15	normal	20	15	15	mini
total veh kms	753.5		903.5	905.1	920.4		779.1	898.3	979.1	
period	Evening peak									
Scenario	PM_BASE		PM_UC_SYM	PM_UC_ASYM	PM_UC_VEHCAP		PM_TC_SYM	PM_TC_ASYM	PM_TC_VEHCAP	
Line	headway [min]	vehicle type	headway [min]	headway [min]	headway [min]	vehicle type	headway [min]	headway [min]	headway [min]	vehicle type
391a / 391b	15	normal	12	10 / 12	12	large	15	15 / 15	12	normal
392a / 392b	7.5 / 15	normal	12	10 / 12	10	mini	10	7.5 / 15	10	mini
394a / 394b	15	normal	10	12 / 15	12	mini	20	15 / 15	15	mini
395a1 / 395b1	15	normal	15	15 / 15	15	mini	20	15 / 20	20	normal
395a2 / 395b2	15	normal	15	15 / 20	15	mini	20	15 / 15	20	normal
398a	20	normal	20	15	15	mini	20	20	10	mini
total veh kms	830.3		907.0	899.4	927.5		737.0	814.3	869.9	

In the UC scenarios this value is approximately equal for all scenarios being close to the budget limit. Note that the value associated with the VEHCAP scenarios is the real value in terms of vehicle kilometers provided and not adjusted by the cost factors yet. If adjusted, the equivalent value is close to the values of the other two scenarios as well. The model finds thus solutions which all (nearly) fully exploit the budget when maximizing the benefits of passengers by allocating the amount of available resources. The solutions in terms of frequency settings are significantly different between the two peak periods considered. Especially in the ASYM scenarios, differences resulting from the directions in passenger flows are clearly present and thus confirm the expectations formulated in the previous section. In general, results obtained for the UC scenarios suggest that in the evening peak supply should only be increased on lines 391, 392 and 394 whereas in the morning peak supply increments and decrements should be more balanced over all lines of the network.

In addition to the tabular description of the found solutions, Table 4.11 and Table 4.12 show a spatial visualization of the supply setting found by the model for the TC scenarios. The thickness of each link is proportional to the joint frequency (over all lines) and the color indicates the total link capacity in terms of maximum numbers of passengers transportable per hour. It is clearly observable that in the evening peak, the resulting solutions are less different from the current supply situation than in the morning peak. In the morning peak, overall network capacity is highest for the ASYM solution which significantly differs from the current supply setting. It is, moreover, remarkable that the link capacities significantly decrease for the VEHCAP scenario in both periods. The employment of minibuses seems to be a beneficial option on some lines and, although line frequencies are increased, the reduction in vehicle capacity causes overall supply capacity to decrease.

Table 4.11: Spatial visualization of the determined supply for different scenarios during the morning peak.

 AM_BASE	 AM_TC_SYM
 AM_TC_ASYM	 AM_TC_VEHCAP

Link capacity [pax/h] :

Link frequency [veh/h] :

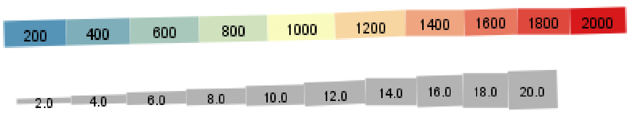
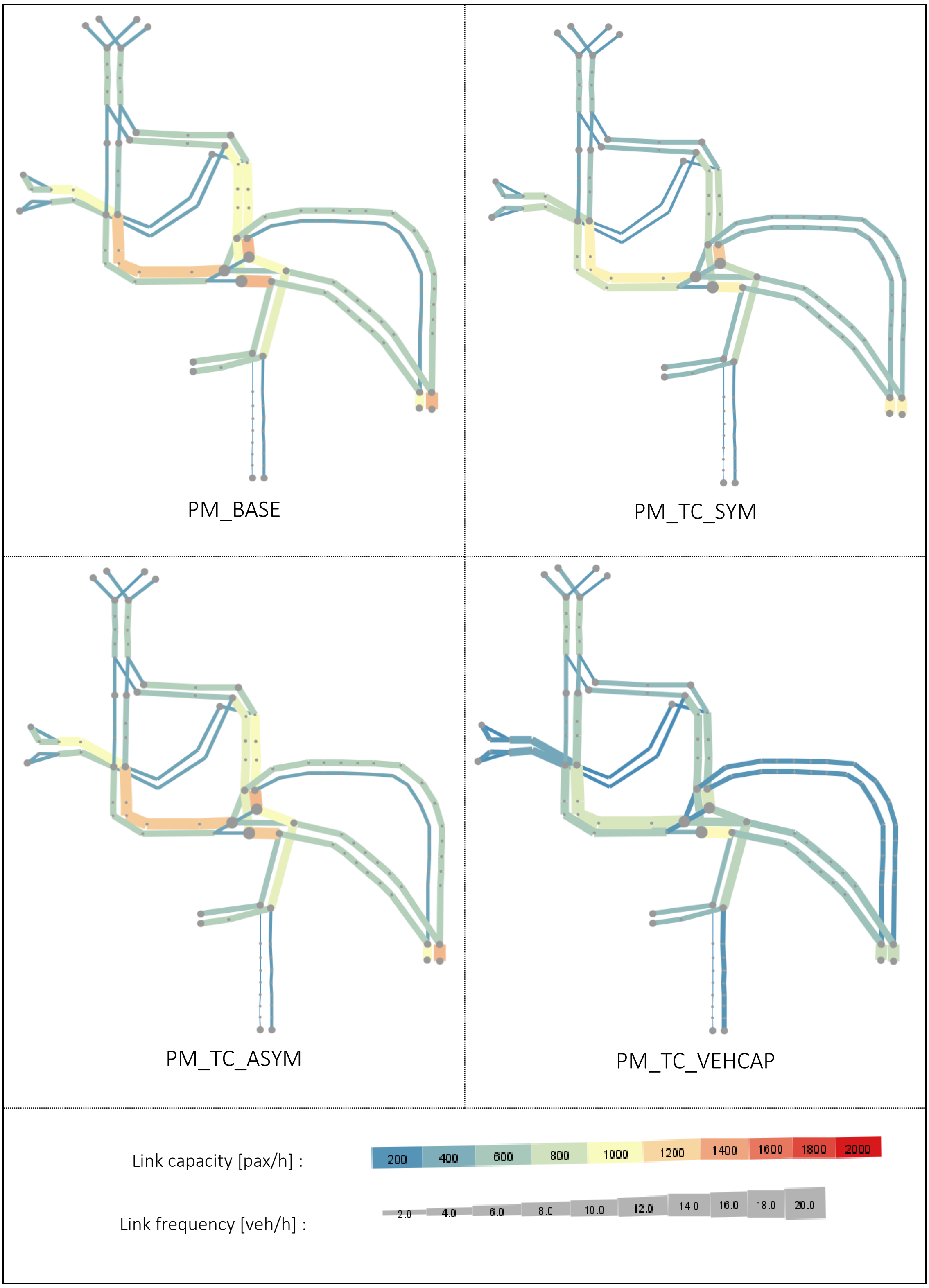


Table 4.12: Spatial visualization of the determined supply for different scenarios during the evening peak.



4.2.4.2 Objective function values

This sections elaborates on the performance of the found solutions with respect to the objective function value. The performance of the solutions found for the different scenarios are presented and compared to both the current situation as well as to the other scenarios.

Figure 4.14 shows the quality of the found solutions in terms of average generalized travel time per passenger for the morning and the evening peaks as well as the total vehicle kilometers equivalent needed to operate the found solution. The budget limit is almost reached in all solutions meaning that the full potential is exploited to maximize passenger benefits. Waiting times resulting from the determined supply significantly decrease in both peak periods whereas perceived in-vehicle times only slightly change. Note that although walking time is considered in the model its share as of the total generalized travel time is very low (about 0.2%) and thus not relevant. The number of transfers does not seem to be affected by the change in supply as this travel cost component remains at a similar level in all scenarios.

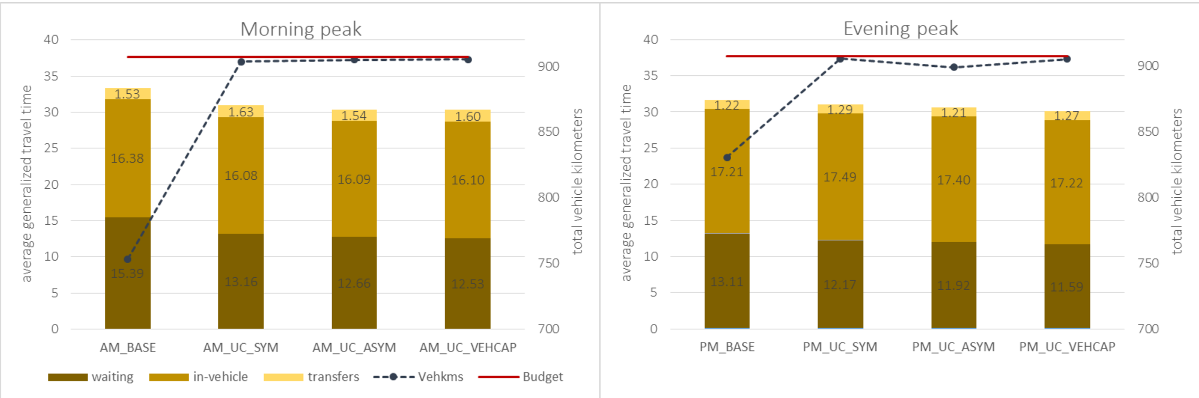


Figure 4.14: Performance of the found solutions in terms of average generalized travel time per passenger and supply in terms of total vehicle kilometers (UC scenarios).

The total costs resulting from the solutions found in the TC scenarios are depicted in Figure 4.15. Similar to the UC scenarios, transfer costs are not affected by the supply settings and costs associated with walking can be neglected. In the morning peak, the determined supplies lead to a reduction in both waiting and in-vehicle times while operational costs increase for all scenarios. In the evening, this trend is less prevalent as increases and decreases in user-related and operational costs are traded-off all the scenarios investigated.

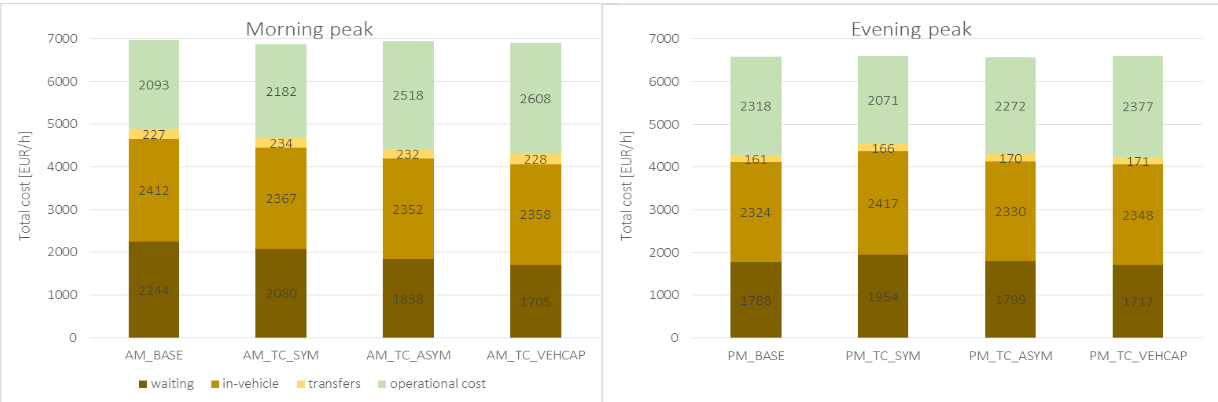


Figure 4.15: Performance of the found solutions in terms of total system costs (TC scenarios).

An overview of all solutions with respect to user-related and operational costs is shown in Figure 4.16. This visualization shows the position of each solution within a two-dimensional space. All points lying on the dashed line traversing the point marking the base scenarios have equal total costs. Moving up (to the left) on these lines implies decreasing passenger-related costs while at the same time increasing operational costs by the same amount. It also implies that all points lying to the right of the dashed line will have higher total costs while those on the left side (towards the south-west corner) will have lower total costs. It becomes evident that total system costs are currently higher in the morning than in the evening peak. This difference can mainly be attributed to the difference in total passenger-related costs as these are about 14% higher in the morning compared to the evening, although the relative difference in the average total number of passengers is only about 8%. Since the overall amount of supply in terms of total vehicle kilometers provided in the existing service provision is less in the morning than in the evening, operational costs are also higher in the latter period.

With respect to the multi-objective optimization problem of minimizing total costs, a solution is called non-dominated or Pareto optimal if neither user-related nor operational costs can be decreased without increasing the respective other component of the total costs. Hence, all solutions found for the TC scenarios in the respective periods are non-dominated since when these are compared a decrease in operational costs always implies an increase in user-related costs and vice-versa. The solution obtained for the scenario AM_UC_VEHCAP, for instance, requires more operational cost but causes lower user-related costs than the respective SYM scenario. In contrast to that, one can also observe the opposite case as the solution obtained for the scenario PM_UC_SYM is clearly dominated by the two other solutions for the VEHCAP and ASYM scenarios which result in both lower passenger-related and operational costs.

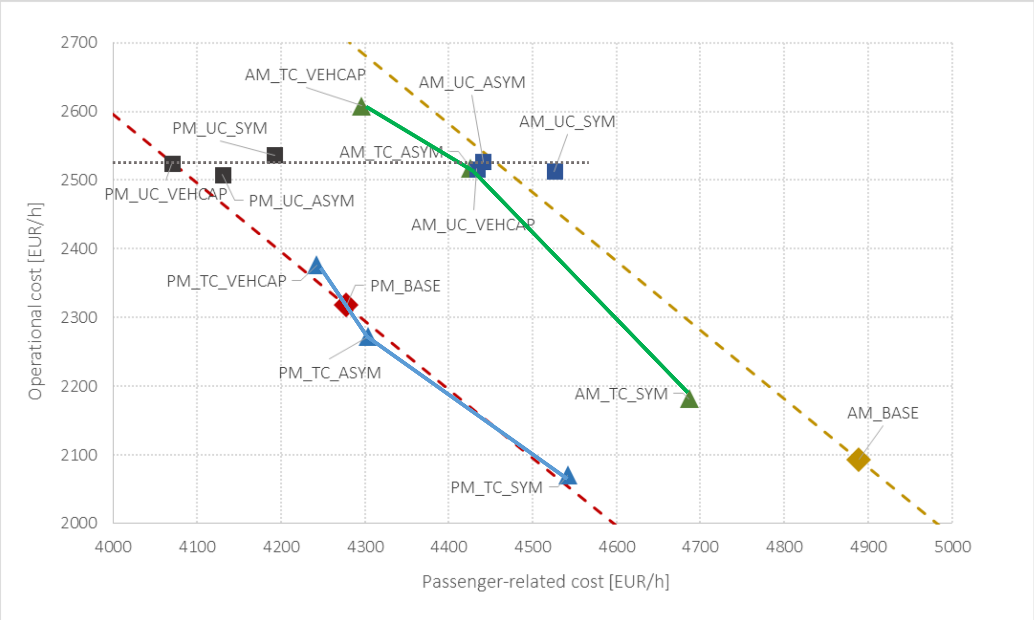


Figure 4.16: Overview of the performance of all solutions found for the different scenarios in terms of associated passenger-related and operational costs.

Overall, it can be concluded from Figure 4.16 that the relative position of the current supply provision in the evening peak in terms of user and operational costs is closer to the solutions obtained for the TC scenarios than those for the UC scenarios. This means that the current supply provided in the evening is close to system optimal conditions, yet user benefits can still be yielded when minimizing passenger-related costs in the UC scenarios. In contrast to that, solutions obtained for all scenarios in the morning peak are positioned further away towards the north-west from the current situation, thus reducing passenger-related costs while increasing operational costs. This indicates a clear potential of improvement in terms of user benefits both when optimizing for total and passenger-related costs only.

Table 4.13 shows the relative change of cost and time components compared to the base case for the respective periods and scenarios. Note that the performance indicators were determined as averages resulting from multiple simulation runs. Using a student's t-test, it can be determined whether the mean values of two performance indicators obtained from a sample of output data records produced by BusMezzo are statistically significantly different from each other. The fields marked grey in Table 4.13 are statistically insignificant (at a level of confidence of 95%), which means that the depicted mean value of these indicators is not significantly different from the value of the current situation in the respective period. Note that since operational costs are solely deterministically approximated based on the static supply setting and thus independent from stochastic outputs produced by the simulation model, the relative difference (if present) is always statistically significant.

In the UC scenarios, operational costs are increased by approximately the same share independently of the assumptions on the decision variables. The relative increase amounts to about 20% in the morning and 9% in the evening. Supply is increased up to the same level of operational costs (about 2500 EUR/h) corresponding to the budget limit set by the maximum vehicle kilometers equivalent value which can be interpreted as a vertical line the UC-OC-space (Figure 4.16). Although a similar level of operational costs is reached for all scenarios, the relative improvement in terms of passenger-related costs is significantly different among the scenarios. Both in the morning and in the evening, the VEHCAP solutions perform best with respect to user benefits. Since in this case study, the deployment of smaller types of vehicles can decrease the average operational costs per bus compared to the current situation and the other scenarios, a larger number of busses can be provided which leads to more passenger benefits (i.e. lower waiting times) compared to the other solutions at a similar level of operational expenses. Solutions obtained by the ASYM scenarios perform better than the SYM options since supply can better match the asymmetric utilization of the network during peak hours. The supply option obtained from the SYM-scenarios leads to the least user benefits. In all scenarios, most of the benefits can be attributed to savings in waiting time which amount up to 18.5% and 10.6% relative decrease in the morning and the evening peak respectively. In the morning, reductions in the perceived in-vehicle time contribute to the benefits as well by an equal amount of about 1.8% for all scenarios. In contrast to that, in-vehicle times increase by 1.7% in the evening peak for the SYM solution and do not change significantly for the other options found. When comparing the found solutions with each other, neither in the morning nor in the evening resulting in-vehicle times differ from each other. In the morning peak, the found solutions for ASYM and VEHCAP can be regarded as equivalent with respect to passenger benefits.

Table 4.13: Relative change of cost and time components compared to the base case for both peak periods considered.

demand	Morning peak						Evening peak					
objective	User cost minimization			Total cost minimization			User cost minimization			Total cost minimization		
scenario	SYM	ASYM	VEHCAP	SYM	ASYM	VEHCAP	SYM	ASYM	VEHCAP	SYM	ASYM	VEHCAP
Total cost	0.9%	-0.1%	-0.4%	-1.6%	-0.5%	-1.1%	2.2%	0.8%	0.1%	0.3%	-0.3%	0.4%
User cost	-7.3%	-9.0%	-9.2%	-4.1%	-9.5%	-12.1%	-1.8%	-3.2%	-4.6%	6.2%	0.6%	-0.8%
Operat. Cost	20.1%	20.7%	20.2%	4.2%	20.3%	24.6%	9.4%	8.2%	8.9%	-10.7%	-2.0%	2.5%
waiting time	-14.5%	-17.7%	-18.5%	-7.3%	-18.1%	-24.0%	-7.2%	-9.1%	-10.6%	9.3%	0.6%	-4.0%
in-vehicle time	-1.9%	-1.8%	-1.7%	-1.9%	-2.5%	-2.3%	1.7%	1.1%	-0.3%	4.0%	0.2%	1.1%

The solutions obtained for the TC scenarios significantly differ among the two periods regarding their performances relative to the respective base cases. In the morning peak, passenger-related costs can be reduced up to about 12% whereas in the evening peak no passenger benefits can be yielded. As with the UC scenarios, most of the savings in user costs can be attributed to reductions in waiting times, yet in-vehicle times can also be slightly reduced by up to 2.5%. In contrast to that, the solution found by the SYM scenario in the evening peak decreases passenger benefits while operational costs are decreased. A consistent feature that can be observed in all solutions and both periods is that user-related costs are always lowest for the VEHCAP scenario and highest for the SYM solution. As a consequence, the inverse relation holds for operational costs which increase with increasing passenger benefits.

Total costs can only be significantly reduced in the morning peak in the scenarios SYM and VEHCAP. In the evening, none of the solutions found can significantly reduce total costs. Figure 4.17 shows the total user benefits in terms of monetary savings per hour in relation to the additional operational expenses needed to generate these benefits in the morning peak. As already mentioned, these cost values are highest for the VEHCAP scenario. The benefit-cost-ratio is slightly higher for the VEHCAP than for the ASYM scenario which means that the marginal decrease in user costs per unit of additional operational expenses is higher when determining both vehicle capacities and frequencies simultaneously than just determining frequencies asymmetrically per line using the same type of vehicle. Interestingly, the benefit-cost-ratio is highest for the SYM scenario which suggests that this solution is most efficient in terms of marginal cost savings per additional operational expense. Note that in the evening peak, no significant user benefits can be generated for the ASYM and VEHCAP scenarios and the operational cost savings that result from the solution obtained from the SYM scenario do not exceed the extra costs to be borne by the passengers. Hence, none of the solutions found by the model for the TC scenarios in the evening peak can beneficially improve the current situation. Yet, solutions turned out to be beneficial during this period as well if only passenger costs are minimized subject to an available budget.

All in all, it can be stated that the obtained solutions and results of the model are in line with the expectations formulated in Section 4.2.3. It could be shown that using small vehicles can decrease the average operational costs per deployed vehicle which allows to increase overall network capacity and thus reduce passenger-related costs. Moreover, it proved clearly evident the relaxation of constraints related to the decision variable frequency in terms of independence on the line's direction leads to a more efficient allocation of resources. That is, lower user costs are generated at the same amount of operational expenses than if frequencies were set equally in both directions of a line. The fact that the potential of improvement of the current supply provisions is greater in the morning than in the evening peak is also in line with previous expectations. Notwithstanding, it is remarkable that the model is able to find solutions at similar total cost levels compared to the current supply provision in the evening. This emphasizes the quality of the current service provision from a total cost point of view and also proves that the model can produce reasonable and realistic results.

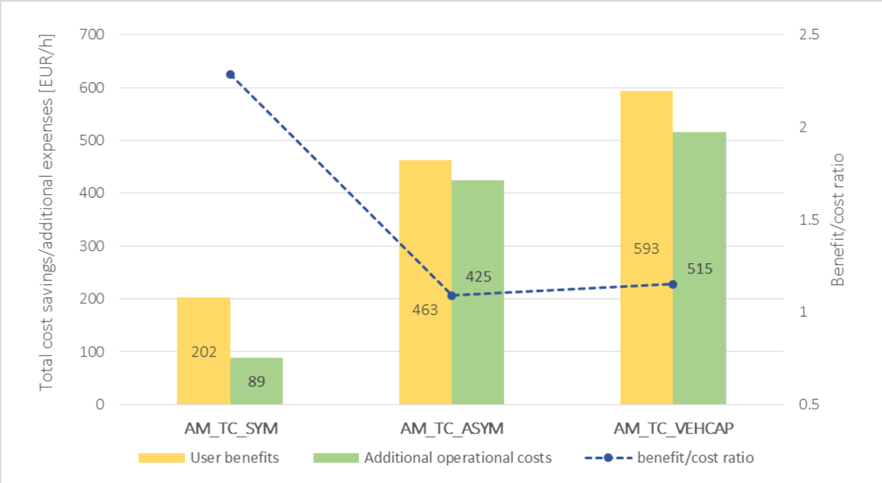


Figure 4.17: Total user cost savings vs. additional operational expenses found for the TC scenarios in the morning peak.

4.2.4.3 Passenger flows: morning peak

This section analyzes the different results obtained for the scenarios minimizing total costs in the morning peak in terms of passenger flows through the network and supply utilization on individual lines. The latter attribute is investigated using the average seat occupancy levels as an indicator.

Figure 4.18 shows a visualization of the average total passenger flows on the network during the morning peak hour for the three scenarios and the current situation. The thickness of a link is proportional to the magnitude of flow between two successive stops. Note that the flows are the results of a dynamic passenger assignment averaged over multiple simulation runs. Different colors are indicating the average level of seat occupancy which is computed by the ratio of the load to the total hourly seat capacity of a line segment which is the product of the number of seats of the employed vehicle and the line's frequency. It can be clearly observed that overall flow patterns do not change significantly among the scenarios and thus passengers' route choices do not change in consequence of the found supply settings. However, since supply is different in each scenario the resulting levels of vehicle utilization do significantly differ. The supply setting resulting from the SYM and ASYM scenarios significantly reduces the average occupancy of vehicles since total supply is increased compared to the base case without changing the vehicles' capacities. The VEHCAP supply setting, however, significantly increases vehicle utilization on some line segments since link capacities are reduced due to the deployment of smaller vehicles. These observation are in line with and directly resulting from the network capacities presented earlier in Table 4.11.

A more comprehensive visualization of the results concerning passenger flows is given in Table 4.14 which shows capacity utilization for each scenario at the individual line level. Overall, these results indicate that occupancy levels are decreasing on all lines in the ASYM and on some lines in the SYM and VEHCAP scenarios. This is in line with the fact that the relative decrease of the total perceived in-vehicle time is also the largest among all scenarios. On line 391, current seat occupancy levels of around 80% on a large part of the line in both directions can be significantly reduced by the supply setting found for all three scenarios. Line 392 shows current seat utilizations of nearly 100% on a small part in one direction which can be significantly reduced by the SYM and ASYM scenarios. Supply proposed by the VEHCAP scenario, however, is lower in terms of overall seat capacity than the current situation. Therefore, occupancy levels increase up to 130% on a small part. A similar result can be observed on line 395 and 398 on which also smaller buses are used in the VEHCAP solution. Results of the SYM scenario will increase seat occupancy on some lines (394, 395 and 398) as well since supply in terms of frequency is reduced from 4 to 3 vehicles per hour on these lines.

It is moreover remarkable that, although crowding levels will increase on some lines for the SYM and VEHCAP solutions, overall in-vehicle times do still decrease compared to the current situation. Note that nominal in-vehicle times do not change significantly in any of the new supply settings found. Hence, all savings regarding in-vehicle times can be attributed to reductions in crowding levels. Since crowding multipliers start to increase the nominal in-vehicle times at occupancy levels of about 70% and higher (see Figure 4.13), only changes in vehicle occupancies on lines 391 and 392 contribute to the savings. Especially the benefits on line 391 seem to be decisive since crowding levels are significantly reduced for all scenarios and will affect a relatively large number of passengers.

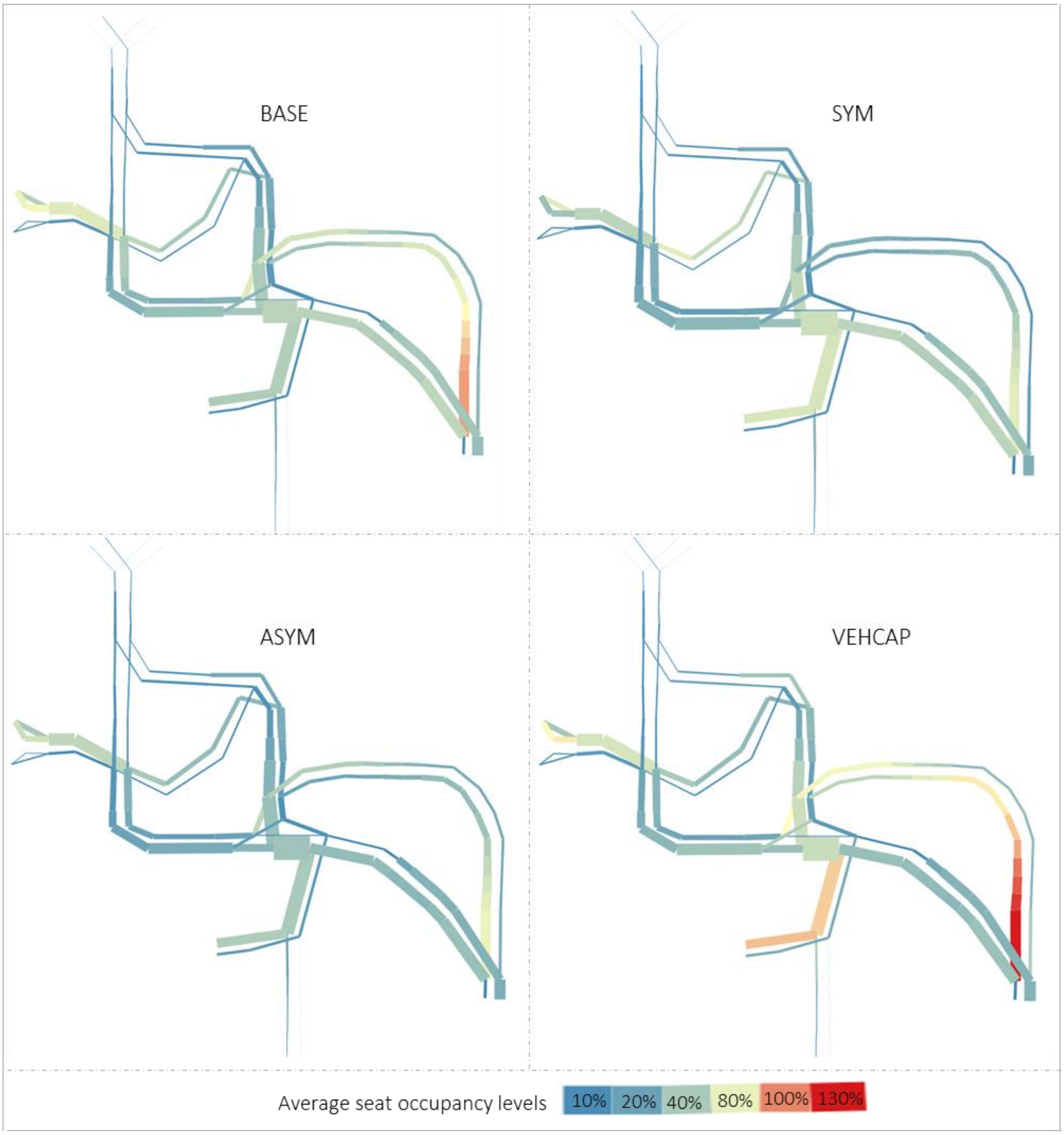


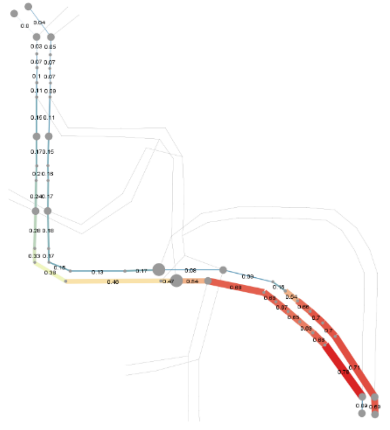
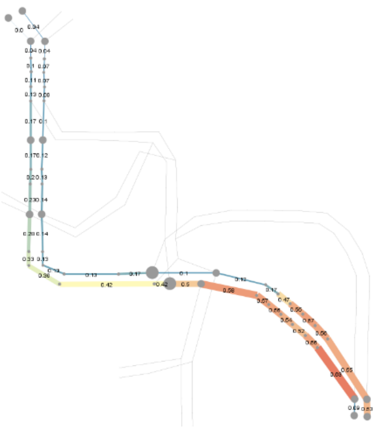
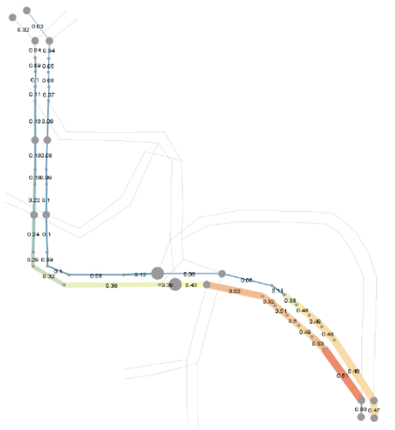
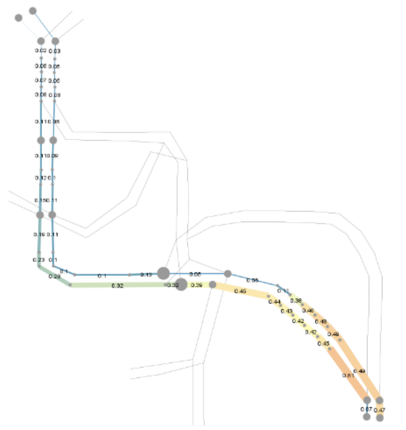
Figure 4.18: Total passenger flows on the network during the morning peak hour for different supply settings.

Table 4.14: Average total loads and seat occupancy levels for all TC scenarios and lines in the morning peak hour.

Line 391

Average seat occupancy levels [Passenger load per hour/(#seats * line frequency)]

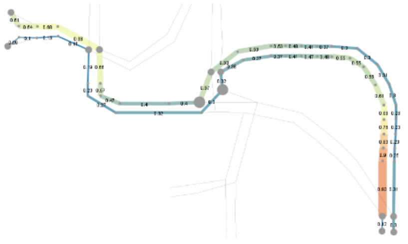
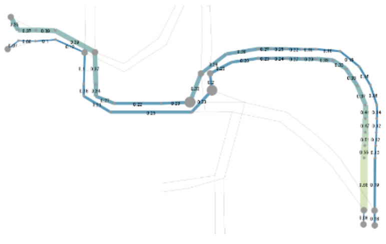
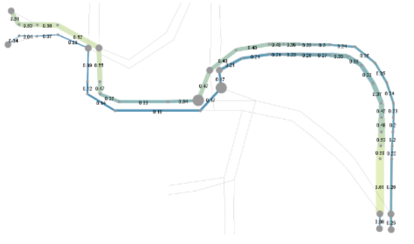
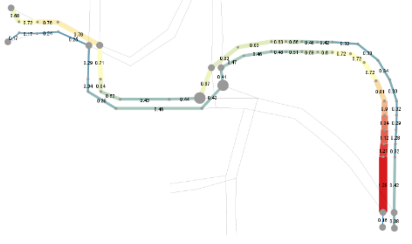
0.1 | 0.15 | 0.2 | 0.4 | 0.6 | 0.8

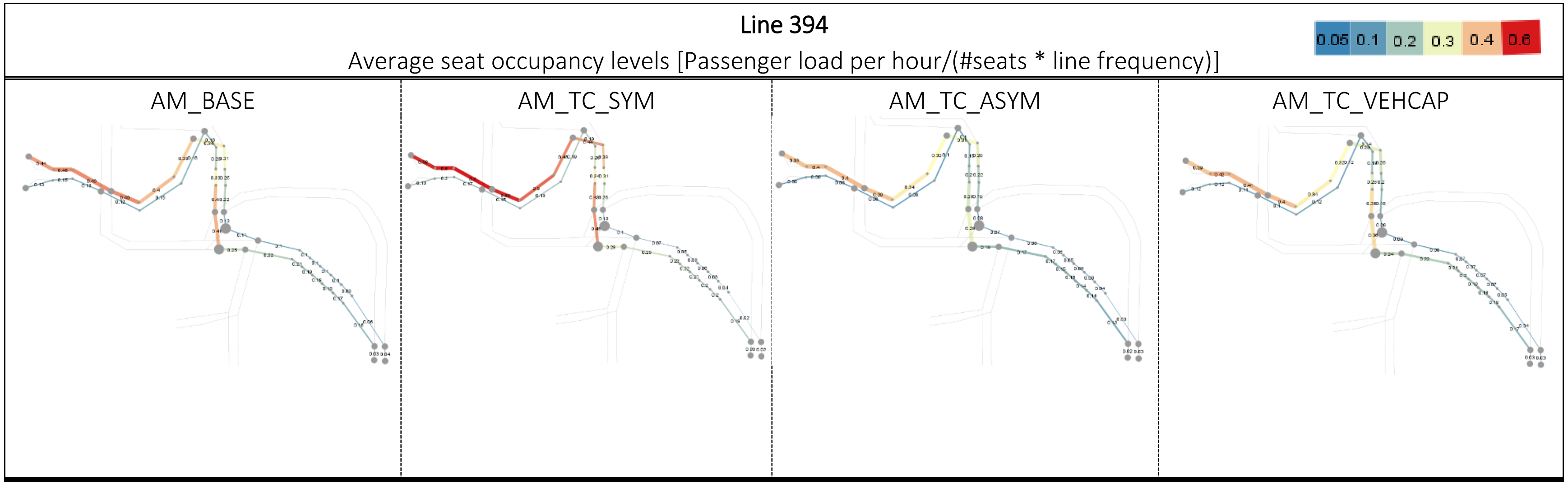
AM_BASE	AM_TC_SYM	AM_TC_ASYM	AM_TC_VEHCAP
			

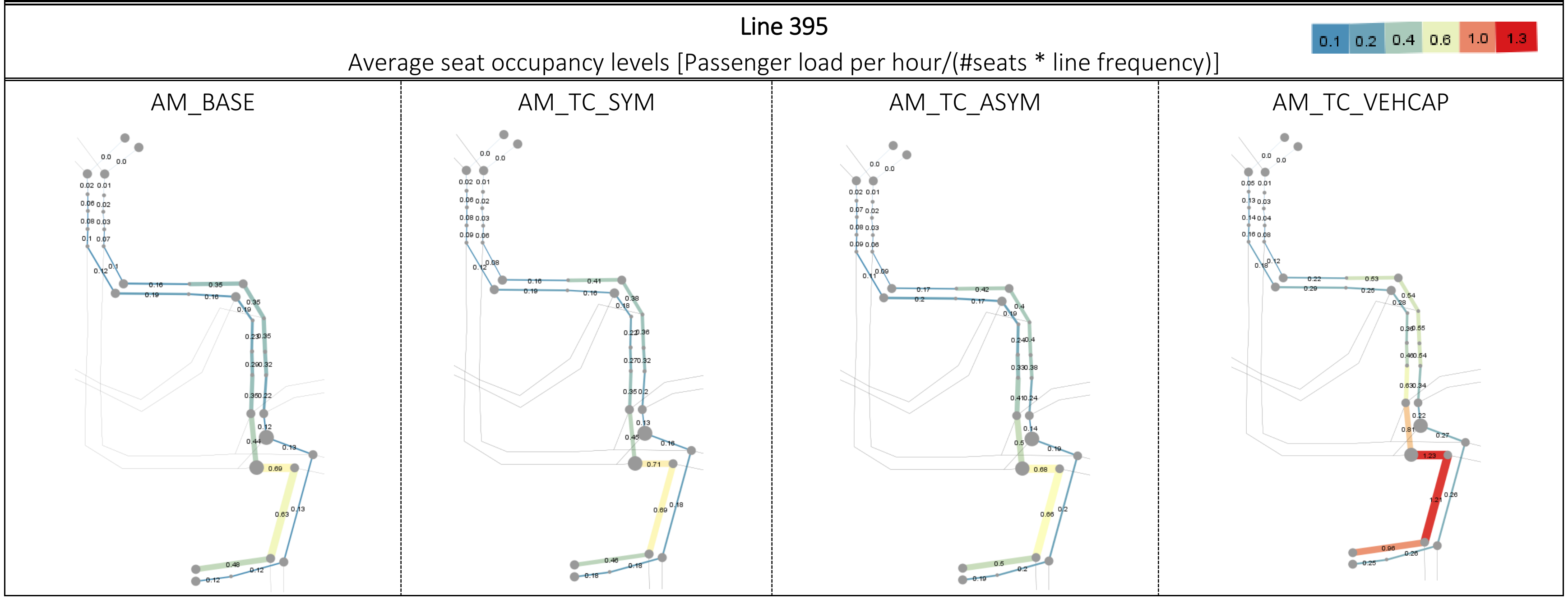
Line 392

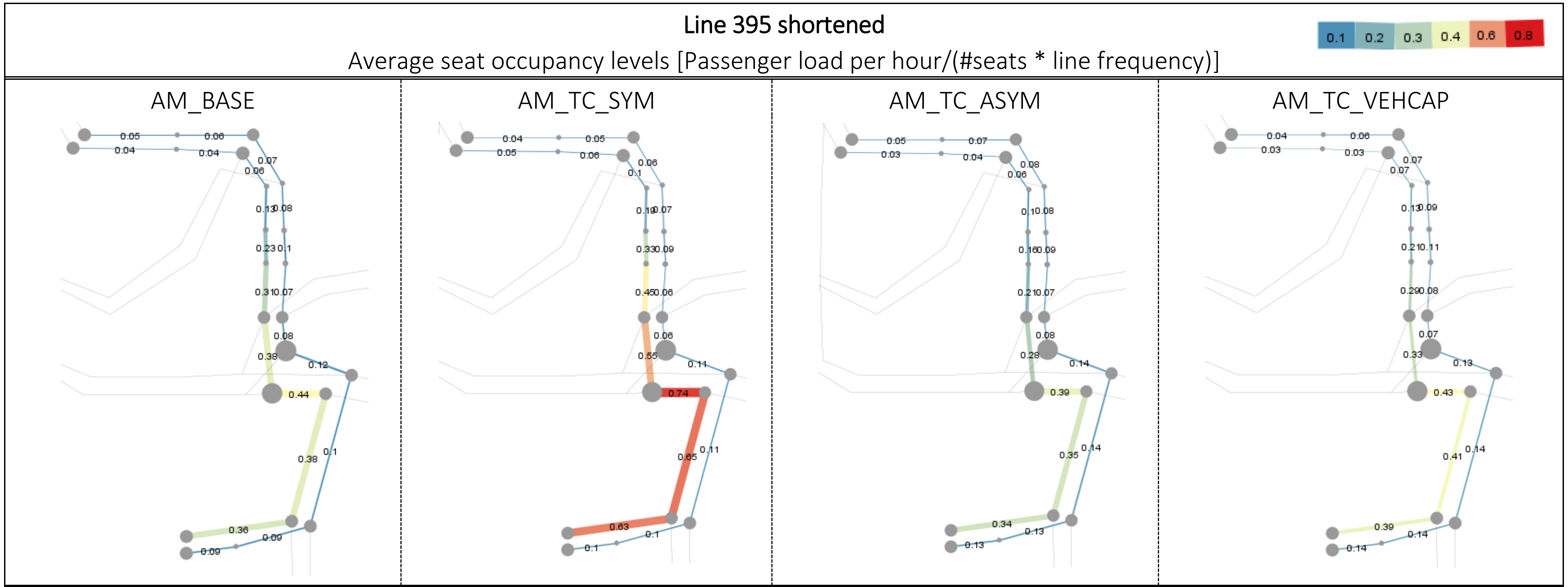
Average seat occupancy levels [Passenger load per hour/(#seats * line frequency)]

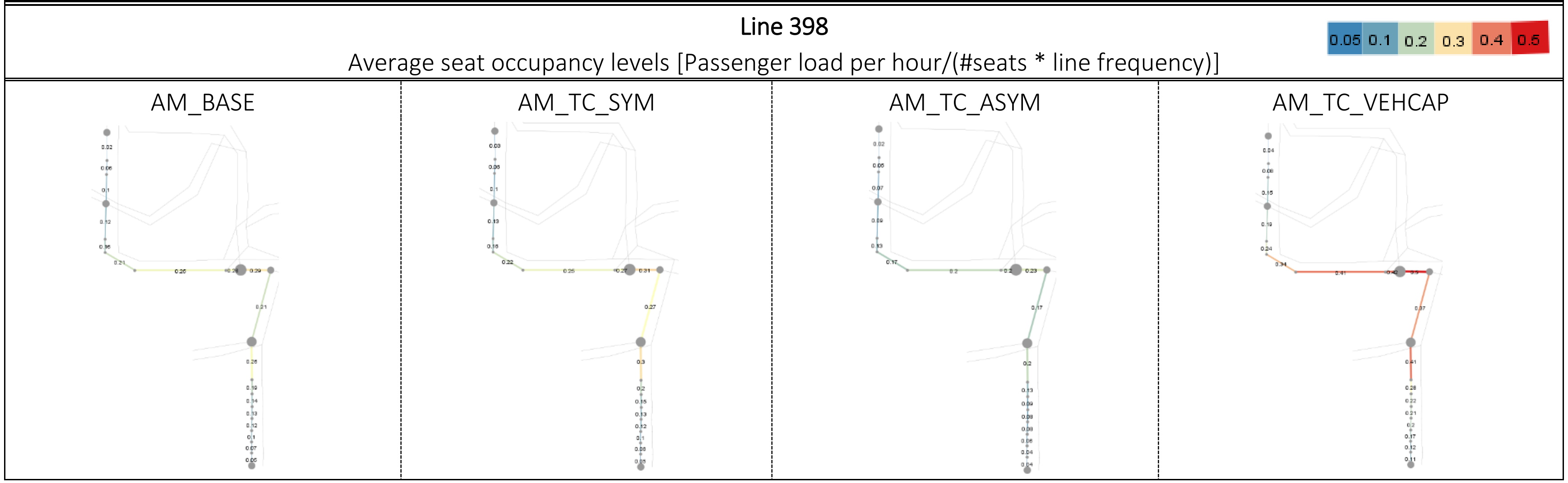
0.1 | 0.2 | 0.4 | 0.8 | 1.0 | 1.3

AM_BASE	AM_TC_SYM	AM_TC_ASYM	AM_TC_VEHCAP
			


Line 394
Average seat occupancy levels [Passenger load per hour/(#seats * line frequency)]
0.05
0.1
0.2
0.3
0.4
0.6
AM_BASE
AM_TC_SYM
AM_TC_ASYM
AM_TC_VEHCAP


Line 395
Average seat occupancy levels [Passenger load per hour/(#seats * line frequency)]
0.1
0.2
0.4
0.6
1.0
1.3
AM_BASE
AM_TC_SYM
AM_TC_ASYM
AM_TC_VEHCAP


Line 395 shortened
Average seat occupancy levels [Passenger load per hour/(#seats * line frequency)]
0.1
0.2
0.3
0.4
0.6
0.8
AM_BASE
AM_TC_SYM
AM_TC_ASYM
AM_TC_VEHCAP


Line 398
Average seat occupancy levels [Passenger load per hour/(#seats * line frequency)]
0.05
0.1
0.2
0.3
0.4
0.5
AM_BASE
AM_TC_SYM
AM_TC_ASYM
AM_TC_VEHCAP

5 CONCLUSIONS

This study presented the formulation and application of a model which simultaneously determines line frequencies and vehicle capacities in a public transport network by minimizing total costs in terms of operational costs and passenger-related generalized travel costs. The latter ones are computed by a dynamic public transport assignment model which simulates the individual movements of passengers and vehicles in the network using an agent-based approach. Since this model is able to simulate the dynamic interaction of demand and supply, effects related to crowding and service reliability can be fully captured by the model. A search algorithm based on the method of simulated annealing was developed to iteratively select potential feasible solutions based on their performance regarding the objective function and finally finds a well-performing supply setting.

This chapter presents the conclusions of the present study and is structured as follows: Section 5.1 comments on the main scientific contributions of this research, Section 5.2 presents the main findings and gives answers to the research questions which lead to the practical implications and recommendations discussed in Section 5.3. The chapter closes with an examination of the limitations and shortcomings of the developed model and further proposes some aspects for future research projects and model enhancements in Section 5.4.

5.1 Scientific contribution

The main scientific contribution of this study is the use of a dynamic assignment model for the tactical planning purposes of frequency and vehicle capacity determination. Recent existing studies within this domain of research use conventional static assignment models which determine passenger flows based on average demand and supply conditions. Hence, these models are not able to capture the dynamic interaction between demand and supply which may strongly influence overall system performance.

The dynamic and stochastic simulation model used in this study for the assessment of the performance of a supply setting can fully capture the three congestion effects in public transport networks, those are (Cats et al., 2016):

1) Deteriorating comfort onboard a crowded vehicle

2) Denied boarding in case of insufficient vehicle capacity

3) Service headway fluctuations resulting from riding and dwell time variations

The first effect is included in the assessment of passenger benefits as the perceived in-vehicle time computed by vehicle-load-dependent crowding multipliers. The second effect imposes an additional disutility to passengers in form of an extra waiting time which is taken into account separately in the generalized cost function. The third effect results indirectly from the dynamic interaction between demand and supply since dwell times at stops are dependent on boarding and alighting passenger flows as well as vehicle occupancy levels and vehicle riding times between stops are modelled stochastically. The mutual relation between passenger flows, dwell times and headways between successive vehicles results in a positive feedback loop that magnifies variations in headways and can thus cause delays and a reduction in service reliability. In order to capture the implications of all these effects on passengers' route choices, an iterative network loading is performed in the simulation that accounts for the day-to-day learning of passengers who may adjust their routes based on experienced service reliability attributes in terms of waiting times and on-board crowding levels. Using this feature results in network-wide steady-state conditions which can be regarded as an equivalent to the user equilibrium computed in conventional static assignment models.

A second important contribution of the present study is the simultaneous determination of line frequencies and vehicle capacities with one supply optimization model. This contrasts with most of the reviewed models for tactical public transport supply optimization which determine either one of the two decision variables and consider the other one as an exogenous input (vehicle capacity) or a direct result of the decision variable value (frequency resulting from determined vehicle capacity). Only two recent studies in the domain of bus (Dell'Olio et al., 2012) and rail transport (Canca et al., 2016) performed a simultaneous approach (see Section 2.2.2.2). Unlike these two studies, the formulation of the present model allows to simultaneously determine both decision variables, or consider either of them as an exogenous input. This feature makes the model even more suitable for practical applications.

Finally, it is worth mentioning that the metaheuristic search method of simulated annealing applied in this study has insofar not been used by any of the reviewed models on public transport supply optimization. Although SA algorithms were applied in models dealing with the strategic design of public transport networks (Zhao & Zeng, 2008; Fan & Machemehl, 2006), the use of other metaheuristics such as Genetic Algorithms (GAs) or Tabu Search seems to be more common for the tactical decision level. Hence, the present study offers an alternative method to those commonly used in this domain. An advantage of SA is the simple and easily understandable structure of the algorithm itself which makes it convenient to implement and adapt to any kind of problem at hand.

5.2 Main findings

This section concludes on the main findings gained during the development and application of the proposed model to the case studies. It provides answers to the research questions stated in Section 1.2.

The main objective of this study was the formulation of a model as a tactical decision tool for frequency and vehicle capacity determination which accounts for the dynamic behavior of demand and supply components in public transport networks and considers the interests of both passengers and the operator. The latter aspect was taken into account by formulating the objective function as the total costs of the system in terms of user and operator costs. Since these two components are inherently in conflict, a trade-off is required for which well-balanced solutions can be found using a multi-objective optimization approach. As parts of objective function are computed based on stochastic simulation outputs no closed analytical formulation describing the mathematical relation between the decision variables and the objective function value is available and the topology of the solution space is also unknown. Hence, dedicated standard methods for finding the optimum of a given objective function such as gradient-based optimization methods are not suitable for this specific problem. Finding a solution by merely evaluating all possible combinations of decision variable values is not feasible in many real-sized instances of the problem since the number of solutions increases exponentially with the size of the problem and thus increases computational time. Therefore, an intelligent search algorithm is needed that systematically explores the solution space by solely using the objective function value as an input. Simulated annealing was chosen as a suitable metaheuristic search algorithm because of its applicability to diverse optimization problems, the absence of requirements on the type of problems and the topology of the solution space as well as the ability to avoid getting trapped in local optima and thereby increasing the probability of finding solutions within globally optimal regions. A discrete formulation of the problem's decision variables as a binary matrix enables a convenient and easily traceable generation of sets of neighboring solutions. Since the feasibility of the solutions does not depend on simulation outputs, infeasible options can be excluded from the evaluation upfront and thus reduce the size of the solution space as well as computation time.

The second objective of this thesis was to test and show the practical applicability and benefits of the proposed model. Therefore, the first case study on a hypothetical public transport network aimed at exploring the influence of different input parameters of the developed model on the objective function value and decision variables of obtained solutions. Hence, this numerical experiment aimed at gaining insight into the sensitivity of model outputs towards certain input parameters.

In a first test, different initial solutions were provided to both the simulated annealing (SA) and a simple local search (LS) algorithm in order to test the effect of different starting solutions as stated in the research sub-question in Section 1.2. Solutions found by SA in terms of both decision variables and objective function values are similar and thus independent from the initial solution whereas solutions found by LS indicate a clear dependence on the initial solution and perform worse than the solutions obtained by SA. These results indicate the presence of multiple locally optimal solutions in the specific problem and thus confirm and justify the suitability and advantages of the application of the SA methodology.

In order to test the effect of different model parameters on the final solution obtained, two parameters were selected that relate to the strategy of the search method and the model used for evaluation of potential solutions. Selecting these two parameters allows to investigate the sensitivity of those two different sub-models as presented in Figure 3.1.

A test on a SA algorithm parameter setting influencing overall runtime showed that a longer execution time of the algorithm, meaning a more intensive search, can lead improve the quality of the obtained solutions. Hence, investing more resources in terms of computation time can pay off in yielding a lower objective function value. Nevertheless, marginal improvements in the objective function value are limited once a certain number of iterations is exceeded.

The results of another third test indicate that increasing the relative weighting of waiting costs in the objective function can lead to an overall increase of supply and thus operational costs while user costs remain relatively stable. Hence, when the marginal increase in operational costs is lower than the increase in user costs resulting from a higher weighting of waiting costs, it is beneficial to increase supply such that additional waiting costs are compensated and the average waiting time per passenger decreases. This result clearly shows how the relative importance attached to a certain aspect in the objective function can significantly affect the final solution found by the model. Hence, special attention needs to be paid to the weighing factors in the objective function which reflect the (potentially conflicting) interests of the different stakeholders involved.

In a second step, the model was applied to a real case study in order to demonstrate its practical applicability to real-sized problems and identify potential benefits of supply optimization. The model was run under two different demand levels (morning and evening peak) in order to test the robustness/sensitivity of the model against varying demand conditions. Multiple scenarios regarding decision variable settings and objective function formulations were executed and the results were compared to the current reference situation in order to reveal the potentials of improvement. The determination of supply for different objectives allowed to investigate the effect of different interests of involved stakeholders on the solutions obtained and to contrast the results against each other.

Overall results indicate that the potential for improving total costs and passenger-related costs is larger in the morning than in the evening peak. While passenger benefits can be attained in both periods, total costs can only be significantly reduced in the morning peak. Moreover, results obtained for the two different demand settings are clearly different which indicates a significant influence of both the structure and overall demand level on the final model outputs. Hence, the model is sensitive against varying demand conditions.

New supply settings were found when minimizing passenger costs subject to a budget constraint resulting with significant reductions in generalized travel times (user costs) of up to 9.2% and 4.6% in morning and evening peaks, respectively. While most of the saving are attributed to waiting time reductions and operational costs have increased up to a similar level in all scenarios. Absolute annual travel cost savings when determining line frequencies symmetrically, asymmetrically and both vehicle capacities and frequencies simultaneously amount to approximately 92.000, 114.000, and 116.000 EUR in the morning peak as well as 21.000, 39.000, and 56.000 EUR in the evening peak periods, respectively. These results are consistent with the results found by Dell'Olio et al. (2012) who showed that using a mixed vehicle fleet in terms of bus capacities may lead to lower costs (i.e. higher benefits) than if a homogenous fleet was used. Moreover, it becomes evident that an asymmetric frequency setting allows for a more effective allocation of available resources in case of a directed demand profile.

The minimization of total costs results in user cost reductions of up to 12.1% compared to the current situation in the morning peak. Absolute annual benefits amount to approximately 52.000, 120.000 and 154.000 EUR for the different scenarios SYM, ASYM and VEHCAP respectively. It is remarkable that additional operational costs required to increase passenger benefits are always lower than the benefits attained, meaning that an additional investment in supply is socially viable. In contrast to that, the solution found for the TC_SYM scenario in the evening peak allows to decrease operational costs without changing total costs by 10.7% which corresponds to an absolute annual value of about 64.000 EUR. These results suggest that a potential reduction of current supply in the evening peak may be reasonable when trading off user against operational costs, while in the morning peak an increase of the current supply provision would be favorable. Overall results obtained when determining both frequencies and vehicle types suggest slightly higher line frequencies and a larger number of smaller vehicles in the morning compared to the evening (during which overall passenger demand is slightly lower). These findings match with the observations by Walters (1980) who established that for increasing passenger volumes, vehicles sizes and corresponding headways decrease.

All in all, it can be said that the objectives of this study have been successfully met and it could be shown that the proposed model can be applied to problems of real scale and yield practical benefits. Depending on the objective, the degrees of freedom of the decision variables and the level of demand, the model yields different solutions that partly differ from the current situation and can lead to significant cost savings.

5.3 Practical implications and recommendations

This section elaborates on the practical implications resulting from the findings gained in this research. Furthermore, practical recommendations to potential users of the model such as public transport authorities and operators are provided.

5.3.1 Implications of the gained findings

Due to the flexible and universal formulation of the model and the inclusion of dynamic effects, some generic implications and recommendations for practical applications of the model as a decision tool for frequency and/or vehicle capacity determination can be given.

Thanks to its ability to fully consider the dynamic interaction of demand and supply settings, the model is able to capture the complex implications of a potential supply provision on overall transport system performance already at the early planning stage of tactical decision making. Hence, operational issues such as service reliability are already explicitly taken into account during this planning phase and may thus lead to smoother daily operations and consequently fewer real-time measures are required in order to improve the level of service than if conventional tactical decision tools or mere reactive adjustments of supply provision were used. The application of the model for the tactical revision of supply provision is therefore expected to be particularly beneficial in public transport networks that are highly-utilized and potentially crowded given the current level of supply, independent of the type of mode.

The detailed output of the simulation model enables the formulation of different objective functions and constraints allowing the application to be used for different planning proposes. There are various stakeholders involved having different and potentially conflicting viewpoints and interests related to the public transport service. The public authority seeks for an overall attractive and socially acceptable transport service that maximizes social welfare in terms of minimum total costs. Depending on the type of concession awarded, the operator may want to improve the provided service in order to attract more passengers and thus create more revenue or reduce his operational costs by cutting inefficiently utilized supply. The first objective can be met by minimizing passengers’ costs subject to a budget constraint which corresponds to a re-allocation of available resources in order to increase effectiveness (as demonstrated in the real case study). The latter objective can be met by minimizing operational costs subject to a defined level of service which may for instance relate to vehicle occupancy levels. As explained in the previous section, the formulation of the objective function in terms of the relative weighting of individual components may strongly influence the final results. Hence, it is advised to choose these values with care to ensure that they reflect the perspective of the stakeholder under consideration.

The application of the model to a real case study demonstrated its practical applicability and usefulness. Overall results indicate that the potential of improvement of the current supply provision is largest in the morning peak between 08:00 and 09:00 o’clock. During this period, significant travel cost savings can be generated by a change in supply resulting from both total and user cost minimizations. Thus, it is advised to the incumbent operator Connexxion to increase supply during this period.

In the evening peak, the decision on a change in service provision is dependent on the objective considered. From a social point of view (total costs) a change of the current provision is not necessary since no significant travel cost savings can be generated. This result confirms the quality and optimality of the current situation. Although one of the found solutions would allow to decrease operational costs while not changing total costs, it would need to be investigated whether this reduction in supply would indeed generate benefits for the operator given the reduction in demand and thus revenues which is usually triggered by a decrease in supply.

When determining supply from the passengers' point of view, significant travel cost savings can be generated by increasing supply up to the maximum operational cost level which is currently provided in the morning peak (between 07:00 and 08:00 o'clock). In case of a concession issued as net-cost contract which gives an incentive to operators to increase ridership and revenue by improving the level of service, this option could be beneficial for the operator as well. However, it would need to be checked whether the additional fare revenues by induced demand can generate sufficient profit given the additional operational costs and the available amount of subsidy.

The separate determination of line frequencies per direction yielded lower user costs compared to a conventional symmetric supply setting at the same level of operational expenses in both demand periods considered. This result clearly highlights the advantages of asymmetric service provision during periods of directed passenger demand which is currently present in the regarded network. However, the associated benefits may be accompanied by additional operational costs resulting from a larger fleet size. This can be a consequence of a more complex vehicle scheduling involving longer layover times due to the asymmetric frequency settings. Depending on the difference in overall frequencies between incoming and outgoing lines at a certain terminal, often applied strategies may also include deadheading or short-turning trips. The realization of a certain frequency setting is therefore constrained by the deadhead travel times for vehicles as well as the locations of vehicle depots within the network. Nevertheless, the application showed that the use of asymmetric frequency settings can lead to a more effective satisfaction of the present demand and also confirms the suitability of the current asymmetric supply setting in the evening peak.

A simultaneous determination of vehicle capacities and line frequencies indicates the benefits of deploying different vehicle sizes per line. Found solutions suggest that smaller busses are an attractive alternative to the currently used type of bus both in the morning and evening peaks. Although occupancy levels will increase on some line segments, overall user benefits prevail and cost savings are even slightly higher than in the found asymmetric frequency settings. So, if the occasionally raised vehicle loads are justifiable given potential level of service requirements, a deployment of a mixed vehicle fleet on the regarded network is clearly advisable. Yet, having available a mixed vehicle fleet might also cause disadvantages to particular small operators with a small fleet who are less flexible in responding to changes in demand or bidding for a different concession. The determination of the size and composition of an operator's vehicle fleet is usually regarded as a strategic long-term decision requiring a significant amount of capital investment. Therefore, this decision often imposes a hard constraint to problems situated further down the decision cycle of public transport service planning such as the tactical frequency determination. A solution to this could be the flexible provision of vehicles by third parties such as leasing companies.

5.3.2 Generic recommendations for practical users

Finally, it is worth mentioning some recommendations for the practical application of the model to real problems. These practical advices are based on the experiences gained during the development and application of the model.

Since an exact determination of the SA algorithm parameters influencing runtime by an analytical expression is not possible, values have to be estimated based on the size of the problem and prior experience in using the model. It is therefore recommended to run the model multiple times using different numbers of iterations per temperature step to check the qualitative effect of prolonged computation times on the final solution found. The advantage of this procedure is also that multiple non-dominating solutions obtained can be weighed against each other.

As the runtime of the simulation model contributes the most to the overall execution time of the algorithm, it is worth knowing the effect of certain model properties on the simulation time. Due to the agent-based approach of BusMezzo, especially the total number of simulated passengers significantly affects the runtime. Hence, in cases with large number of passengers (multiples of 10.000), one simulation run can take several seconds using conventional computers. It might therefore be interesting, particularly in case of large-scale networks, to use high-performance computers or cloud computing techniques to reduce overall algorithm execution time when evaluating a larger number of solutions and scenarios.

Since the path set needed for the dynamic route choice model is only dependent on the static network properties such as line routes (i.e. independent from supply provision), it is always recommended to generate it once for a specific network only prior to the execution of the algorithm and provide it as an input to BusMezzo. This approach can save a lot of computation time especially in large networks having many stops.

Finally, it is important to note that the simulation set-up in terms of the provision of warm-up and cool-down phases of supply generation is a crucial factor ensuring that full supply is present in all areas of the network once the first passengers are generated and that all passengers can eventually reach their destination within the simulation period. This becomes particularly relevant when evaluating low supply solutions. To this end, the timespan of supply simulation before and after demand generation should be chosen sufficiently high considering the size of the network. Otherwise, the simulation model might produce errors leading to biased outputs.

5.4 Limitations and future research

The present study has fulfilled its objective of developing a public transport supply determination model which considers network-wide dynamic interactions of passengers and vehicles and proved its practical applicability as well as potential to yield benefits. Notwithstanding, there are some crucial aspects that are not or only inadequately considered by the model. This section presents these limitations and discusses the resulting consequences as well as proposes ideas for future enhancements to the model in order to overcome the limitations and expand the functionalities. Moreover, potential future directions of research in the field of public transport service planning that may build up on the gained findings are outlined.

For the sake of simplicity, vehicle scheduling was not considered in the present model. That is, a vehicle is assigned to one trip on a specific line only, which implies that potential delays being present at the destination terminal of a line cannot affect the punctuality of the following departure from the terminal. Hence, effects related to the propagation of delays and degraded service reliability among multiple lines and line directions were not properly accounted for. Another consequence of not including vehicle scheduling considerations in the present model is the potentially incorrect determination of operational costs resulting from a certain supply setting. These costs are computed based on an estimated fleet size which is given by the ratio of cycle time to headway of a line. In fact, the exact cycle time and fleet size is a result of vehicle scheduling which might lead to significant differences in operational costs in reality. Especially in cases of asymmetric frequency settings per line direction as elaborated on in Section 4.2, constraints imposed by vehicle scheduling considerations may severely affect the required fleet size and thus operational costs. In order to improve these limitations, future research should focus on the integration of a vehicle scheduling model into the present framework. This would allow to further increase the practical utility of the developed tool.

Another aspect that the model disregards is the implication of supply changes on the overall passenger demand. As overall travel demand between a specific OD pair usually splits up to multiple available modes according to the relative utility associated with each mode, an increase in the utility of one mode usually leads to a shift of demand between competing modes, or more generally to induced/latent demand if a certain mode gets more attractive. Since a change of supply in terms of line frequencies and or vehicle capacities will lead to changed generalized travel costs for certain OD pairs, the number of passengers travelling between this OD pair will also change and cause itself demand-dependent changes to the generalized travel costs such as altered crowding levels. In order to account for these implications, an additional iterative procedure needs to be implemented into the modelling framework that adjusts the OD matrix based on the relative changes of generalized travel times per OD pair using a demand elasticity function. Note that this feature will increase the overall complexity of the model and might lead to instability issues related to the convergence of the algorithm since demand is introduced as an additional dynamic factor that affects system performance. Incorporating this feedback loop will allow to investigate crucial topics such as forecasting ridership growth and resulting additional revenues or supply determination given the objective of operator profit maximization or subsidy minimization.

Further areas of research within the domain of public transport service planning in which the model could be applied beyond the tactical level are the strategic network design and the management of supply provision during special events. In the former case, the model could be used to identify attractive lines given a set of potential routes by applying the frequency determination procedure for all potential lines without taking into account a lower bound for the frequency setting. Hence, lines resulting in zero supply (or very low frequencies) can be excluded from the final network layout. Running the model on a modified network or demand configurations in case of special circumstances such as construction works or big events can create valuable outputs which can be used as a tactical basis for service plans aimed at mitigating the negative consequences of exceptional situations in the network. For instance, the model can determine the supply of the entire network in case there is a high trip attractor/generator at a specific location or the routes of certain lines have been changed due to exceptional circumstances. Thus, a basis for a predefined strategy can be found and its robustness (for instance regarding demand fluctuations and uncertainty) can be evaluated. Canca et al. (2016) concluded that their tactical frequency and capacity determination model for railway services could also deal with exceptional situations such as track failures by removing the affected link.

Given the recent attention of research and technological advancements within the field of automated traffic and transportation, it is worth investigating the capabilities of the developed modelling framework as a decision tool for the design of automated public transport services. Winter et al. (2016) have already demonstrated the usefulness of event-based simulation techniques for the design of automated demand-responsive point-to-point services. Since on-board staff costs will disappear in the case of fully automated vehicles, the operational cost differences between different types of vehicles will get more pronounced and thus enable a greater potential for vehicle capacity optimization. Moreover, operational constraints imposed by vehicle scheduling might become more relaxed since automated vehicles are more flexible in terms of intermediate vehicle depots and empty deadheading trips. As driving staff is no longer required, the parking and movement of vehicles that are not in service would be possible at any location in the network for any period of time without increased operational costs related to personnel. From a passenger's perspective, it should be examined whether and to which extent the automatization of public transport services leads to a change in user perceptions and thus costs as compared to conventional public transport services. Increased levels of comfort and safety and different perceptions of crowding may imply a significant change in generalized travel costs.

Another future study could conduct a comparative analysis between the developed model and conventional supply optimization models using static assignment approaces or no passenger route choice implications at all (i.e. a mere adjustment of supply given observed demand conditions). In this way, the added value of the present model could be validated and quantified. In terms of practical considerations, issues relating to runtime, data requirements and model calibration should be contrasted against the quality of the results obtained by the different modelling approaches in order to identify practical advantages and disadvantages. Moreover, the robustness of the proposed model and other approaches against uniform changes in demand should be examined and compared as well, since this aspect was not thoroughly analyzed in this study.

5.5 Personal reflection

This final section provides a personal reflection on the thesis work and process as well as the lessons learned during the project. It is therefore written from a personal and subjective point of view.

The origin of this thesis can be ascribed to an internship I performed as a part of my studies at the consultancy company Goudappel Coffeng. During this internship, I got acquainted with the simulation tool BusMezzo and deepened my passion for transport modelling and especially public transport. After the internship, I started, together with Oded Cats, thinking about potential topics. Since I really liked working with BusMezzo and also appreciated the good cooperation and support of my internship supervisor Oded, he encouraged me to incorporate this tool in my Master thesis and offered me his support during this project.

During the initial phase, it was difficult for me to define the right topic since there were many possibilities available and I had to read a lot of literature. Thanks to the support of Oded and intensive reflection about my personal interests, I could define a topic that teased me and offers the right portion of challenge and scientific demand to be considered as a Master thesis. After the first kick-off meeting with the committee, my vision about the topic was even clearer since they provided me with helpful feedback and guidance to get on the right track from the beginning of the work. Particularly Henk and Bruno, who are both experienced in mathematical optimization models and techniques, provided me with some helpful advice regarding the implementation and application of certain methods. For instance, they convinced me not to use the complex population-based method of genetic algorithms which I originally intended to apply. In retrospect, I think that this decision was crucial for facilitating the development and implementation of my model since I decided for a less complex single-search method (SA) which is easier to understand and implement than population-based methods (at least from my perspective as a student having little experience with such methods).

For the implementation of my proposed model, BusMezzo needs to be executed by an automated procedure and a transfer of input and output data between the simulation model and the procedure needs to be implemented. This was a challenging task that took me a lot of hours of work until everything went smoothly. During the programming work, I really experienced what it is like to iteratively learn by doing, which can be quite frustrating sometimes. Next time, I would first construct the entire framework of the entire model including all details and then start implementing it step by step instead of immediately diving into the matter and working on small modules without seeing the big picture. This would surely help to get work better structured and avoid repetitive and redundant steps. Nevertheless, I can definitely say that by doing this work I have evolved from a beginner to an advanced user when it comes to programming in Matlab and this makes me proud.

When designing the scenarios of the case studies, it was hard for me to decide on which things to focus since there a so many aspects you could theoretically examine. In retrospect, I would have also included a scenario investigating the model's robustness against uniform changes in demand, not only structural differences as shown by AM and PM demand distributions. Investigating this aspect could certainly reveal further valuable insights.

Overall, I dare to say that I managed the graduation project quite well and I am proud of my accomplishments. I think that by working on this project, I could particularly further improve my programming and abstraction skills and learned what it means to comply with a scientifically sound way of working. But I also have to say that without the help and support of the graduation committee and particularly my daily supervisor Oded, I would not have been able to achieve all this.

REFERENCES

Balcombe, R., Mackett, R., Paulley, N., Preston, J., Shires, J., Titheridge, H., . . . White, P. (2004). *The Demand for Public Transport: A practical Guide (Report No. TRL593).* London, UK: Transportation Research Laboratory.

Börjesson, M., Eliasson, J., & Franklin, J. P. (2012). Valuations of travel time variability in scheduling versus mean-variance models. *Transportation Research Part B, 46*(7), 855-873.

Burghout, W. (2004). *Hybrid microscopic-mesoscopic traffic simulation.* Stockholm, Sweden: Royal Institute of Technology.

Canca, D., Barrena, E., De-Los-Santos, A., & Andrade-Pineda, J. L. (2016). Setting lines frequency and capacity in dense railway rapid transit networks with simultaneous passenger assignment. *Transportation Research Part B, 93*, 251-267.

Cats, O. (2011). *Dynamic modelling of transit operations and passenger decisions.* Stockholm, Sweden: Royal Institute of Technology.

Cats, O. (2017). Determinants of Bus Riding Time Deviations: On the Mutual relation between Driving Patterns and Transit Performance. *Working paper*.

Cats, O., & Gkioulou, Z. (2014). Modeling the impacts of public transport reliability and travel information on passengers' waiting-time uncertainty. *EURO Journal on Transportation and Logistics*, 1-24.

Cats, O., Burghout, W., Toledo, T., & Koutsopoulos, H. N. (2010). Mesoscopic Modeling of Bus Public Transportation. *Transportation Research Record 2188*, 9-18.

Cats, O., West, J., & Eliasson, J. (2016). A dynamic stochastic model for evaluating congestion and crowding effects in transit systems. *Transportation Research Part B, 89*, 43-57.

Ceder, A. (1984). Bus frequency determination using passenger count data. *Transportation Research Part A, 18A*(5/6), 439-453.

Ceder, A. (2007). *Public transit planning and operation : theory, modelling and practice.* London: Elsevier.

Ceder, A., & Wilson, N. H. (1986). Bus Network Design. *Transportation Research Part B, 20B*(4), 331-344.

Cerny, V. (1985). Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm. *Journal of Optimization Theory and Applications, 45*(1), 41-51.

Chang, S. K., & Schonfeld, P. M. (1991). Optimization models for comparing conventional and subscription bus feeder services. *Transportation Science, 25*(4), 281-298.

Chien, S. I.-J. (2005). Optimization Of Headway, Vehicle Size and Route Choice for Minimum Cost Feeder Service. *Transportation Planning and Technology, 28*(5), 359-380.

Chien, S., Spasovic, L. N., Elefsiniotis, S. S., & Chhonkar, R. S. (2001). Evaluation of feeder bus systems with probabilistic time-varying demands and non-additive value of time. *Transportation Research Record, 1760*, 47-55.

Connexxion. (2016). *Dienstregeling 2017, Regio Zaanstreek.* Retrieved March 23, 2017, from https://www.connexxion.nl/data/upload/Zaanstreek2017_v3.pdf

Constantin, I., & Florian, M. (1995). Optimizing Frequencies in a Transit Network: a Nonlinear Bi-level Programming Approach. *International Transactions in Operational Research, 2*(2), 149-164.

dell'Olio, L., Ibeas, A., & Ruisánchez, F. (2012). Optimizing bus-size and headway in transit networks. *Transportation, 39*(2), 449-464.

Du, K., & Swamy, M. N. (2016). *Search and Optimization by Metaheuristics.* Springer International Publishing AG Switzerland.

Eglese, R. W. (1990). Simulated Annealing: A tool for Operational Research. *European Journal of Operational Research, 46*, 271-281.

Fan, W., & Machemehl, R. B. (2006). Using a Simulated Annealing Algorithm to Solve the Transit Route Network Design Problem. *Journal of Transportation Engineering, 132*(2), 122-132.

Frank, P., Friedrich, M., & Schlaich , J. (2008). Betriebskosten von Busverkehren schnell und genau ermitteln. *Der Nahverkehr, 11*.

Furth, P. G., & Wilson, N. H. (1981). Setting Frequencies on Bus Routes: Theory and Practice. *Transportation Reserach Record*(818), 1-7.

Glover, F., & Greenberg, H. (1989). New approaches for heuristic search: A bilateral linkage with artificial intelligence. *European Journal of Operational Research, 39*, 119-130.

Gronau, R. (2000). Optimum diversity in the public transport market. *Journal of Transport Economics and Policy, 34*, 21-42.

Gwilliam, K. M., Nash, C. A., & Mackie, P. J. (1985). Deregulating the bus industry in Britain - (B) The Case Against. *Transport Reviews, 5*(2), 105-132.

Han, A. F., & Wilson, N. H. (1982). The allocation of buses in heavily utilized networks with overlapping routes. *Transportation Research Part B, 16B*(3), 221-232.

Herman, M. (n.d., 03 15). *Simulated Annealing & the Metropolis Algorithm: A Parameter Search Method for Models of Arbitrary Complexity.* Retrieved March 15, 2017, from http://kestrel.nmt.edu/~mherman/publications/metropolis_final.pdf

Hooke, R., & Jeeves, T. A. (1961). Direct search solution of numerical and statistical problems. *Journal of the Association for Computing Machinery*, 212-229.

Huang, Z., Ren, G., & Liu, H. (2013). Optimizing Bus Frequencies under Uncertain Demand: Case Study of the Transit Network in a Developing City. *Hindawi: Mathematical Problems in Engineering*.

Ibarra-Rojas, O., Delgado, F., Giesen, R., & Muñoz, J. (2015). Planning, operation, and control of bus transport systems: A literature review. *Transportation Research Part B*, 38-75.

Jansson, J. O. (1980). A simple bus line model for optimization of service frequency and bus size. *Journal of Transport Economics and Policy, 14*(1), 53-80.

Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by Simulated Annealing. *Science, 220*(4598), 671-680.

Kouwenhoven, M., de Jong, G., Koster, P., van den Berg, V., Verhoef, E., Bates, J., & Warffemius, P. (2014). New values of time and reliability in passenger transport in The Netherlands. *Research in Transportation Economics*, 37-49.

Laarhoven, P. v., & Aarts, E. H. (1987). *Simulated Annealing: theory and Applications.* Dordrecht: Reidel.

Lee, K. K., Kuo, S. H., & Schonfeld, P. M. (1995). Optimal mixed bus fleet for urban operations. *Transportation Research Record, 1503*, 39-48.

MAN Truck & Bus AG. (2017). *MAN Lions City.* Retrieved May 5, 2017, from http://www.bus.man.eu/man/media/en/content_medien/doc/business_website_bus_master_1/Lions_City.pdf

Martinez, H., Mauttone, A., & Urquhart, M. E. (2014). Frequency optimization in public transportation systems: Formulation and metaheuristic approach. *European Journal of Operational Research, 236*, 27-36.

Mathforum. (2017). *The mathforum*. Retrieved June 6, 2017, from http://mathforum.org/kb/message.jspa?messageID=19446

Mazloumi, G., Currie, G., & Rose, G. (2010). Using GPS data to gain insight into public transport travel time variability. *Journal of Transportation Engineering, 136*, 623-631.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., & Teller, A. H. (1953). Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics, 21*, 1087-1092.

Newell, G. (1971). Dispatching Policies for a Transportation Route. *Transporatation Science, 5*(1), 91-105.

Oldfield, R. H., & Bly, P. H. (1988). An Analytical Investigation of Optimal Bus Size. *Transporation Research Part B, 22B*(5), 319-337.

Ruisanchez, F., dell'Olio, L., & Ibeas, A. (2012). Design of a tabu search algorithm for assigning optimal bus sizes and frequencies in urban transport services. *Journal of Advanced Transportation, 46*, 366-377.

Salzborn, F. J. (1972). Optimum Bus Scheduling. *Transportation Science, 6*(2), 137-148.

Spiess, H., & Florian, M. (1989). Optimal Strategies: A new assignment model for transit networks. *Transportation Research Part B, 23B*(2), 83-102.

Still, C., & Westerlund, T. (2001). Extended cutting plane algorithm. *Encyclopedia of Optimization, 2*, 53-61.

Tisato, P. (2000). A comparison of optimisation formulations in public transport subsidy. *International Journal of Transport Economics, 27*, 199-228.

Trafikverket. (2016). *Analysmetod och samhällsekonomiska kalkylvärden för transportsektorn: ASEK 6.0.* Borlänge: version 2016-04-01.

Verbas, I., Mahmassani, H., Frei, C., & Chan , R. (2015). Stretching resources: sensitivity of optimal bus frequency allocation to stop-level demand elasticities. *Public Transport, 7*, 1-20.

Walters, A. (1982). Externalities in Urban Buses. *Journal of Urban Economics, 11*, 60-72.

Wardman, M., & Wehlan, G. (2011). Twenty years of rail crowding valuation studies: evidence from lessons from British experience. *Transport Reviews, 31*(3), 379-398.

Wardmann, M. (2004). Piblic transport values of time. *Transport Policy, 11*(4), 363-377.

Weidmann, U. (1994). *Der Fahrgastwechsel im öffentlichen Personenverkehr.* Zürich: IVT.

Winter, K., Cats, O., Homem de Almeida Correia, G., & van Arem, B. (2016). Designing an Automated Demand-Responsive Transport System. *Journal of the Transportation Research Board, 2542*, 75-83.

Yoo, G.-S., Kim, D.-K., & Chon, K. S. (2010). Frequency Design in Urban Transit Networks with Variable Demand: Model and Algorithm. *KSCE Journal of Civil Engineering, 14*(3), 403-411.

Yu, B., Yang, Z., & Yao, J. (2010). Genetic Algorithm for Bus Frequency Optimization. *Journal of Transportation Engineering, 136*(6), 576-583.

Yu, B., Zhongzhen, Y., Xueshan, S., Baozhen, Y., Qingcheng, Z., & Erik, J. (2011). Parallel genetic algorithm in bus route headway optimization. *Applied Soft Computing, 11*, 5081-5091.

Zhao, F., & Zheng, X. (2006). Optimization of transit network layout and headway with a combined genetic algorithm and simulated annealing method. *Engineering Optimization, 38*(6), 701-722.