A Simulation-Optimization Approach for Improved Robustness of Railway Timetables

JOHAN HÖGDAHL

Licentiate Thesis in Transport Science
School of Architecture and the Built Environment
KTH Royal Institute of Technology
Stockholm, Sweden 2019

Kungliga Tekniska högskolan
Skolan för arkitektur och samhällsbyggnad
100 44 Stockholm

TRITA-ABE-DLT-1942
ISBN: 978-91-7873-364-4

Akademisk avhandling som med tillstånd av Kungl Tekniska högskolan framlägges till offentlig granskning för avläggande av licentiatexamen i transportvetenskap onsdagen den 4 december 2019 klockan 10:15 i B26, B-huset, Kungliga Tekniska högskolan, Brinellvägen 23, Stockholm.

Tryck: Universitetsservice US AB

Abstract

The timetable is an essential part for the operations of railway traffic, and its quality is considered to have large impact on capacity utilization and reliability of the transport mode. The process of generating a timetable is most often a manual task with limited computer aid, and is known to be a complex planning problem due to inter-train dependencies.

These inter-train dependencies makes it hard to manually generate feasible timetables, and also makes it hard to improve a given timetable as new conflicts and surprising effects easily can occur.

As the demand for railway traffic is expected to continue grow, higher frequencies and more saturated timetables are required. However, in many European countries there is also an on-going public debate on the punctuality of the railway, which may worsen by increased capacity utilization. It is therefore also a need to increase the robustness of the services. This calls for increased precision of both the planning and the operation, which can be achieved with a higher degree of automation.

The research in this thesis is aimed at improving the robustness of railway timetables by combining micro-simulation with mathematical optimization, two methods that today are used frequently by practitioners and researchers but rarely in combination. In this research a sequential approach based on simulating a given timetable and re-optimizing it to reduce the weighted sum of scheduled travel time and predicted average delay is proposed. The approach has generated promising results in simulation studies, in which it has been possible to substantially improve the punctuality and reduce the average delays by only increasing the advertised travel times slightly. Further, the results have also indicated a positive socio-economic benefit. This demonstrates the methods potential usefulness and motivates further research.

Keywords: Railroad, Timetabling, Optimization, Simulation, Robustness, Punctuality, Delay prediction.

Sammanfattning

För järnvägen har tidtabellen en central roll, och dess kvalité har stor betydelse för kapacitet och tillförlitlighet. Processen att konstruera en tidtabell är ofta en uppgift som utförs manuellt med begränsat datorstöd och på grund av beroenden mellan enskilda tåg är det ofta ett tidskrävande och svårt arbete.

Dessa tågberoenden gör det svårt att manuellt konstruera konfliktfria tidtabeller samtidigt som det också är svårt att manuellt förbättra en given tidtabell, vilket beror på att de är svårt att förutsäga vad effekten av en given ändring blir.

Eftersom efterfrågan på järnväg fortsatt förväntas öka, finns det ett behov av att kunna köra fler tåg. Samtidigt pågår det redan i många europeiska länder en offentlig debatt om järnvägen punktlighet, vilken riskeras att försämras vid högre kapacitetsanvändning. Därför finns det även ett behov av att förbättra tidtabellernas robusthet, där robusthet syftar till en tidtabells möjlighet att stå emot och återhämta mindre förseningar. För att hantera denna målkonflikt kommer det behövas ökad precision vid både planering och drift, vilket kan uppnås med en högre grad av automation.

Forskningen i denna avhandling syftar till att förbättra robustheten för tågtidtabeller genom att kombinera mikro-simulering med matematisk optimering, två metoder som redan används i hög grad av både yrkesverksamma trafikplanerare och forskare men som sällan kombineras. I den här avhandlingen förslås en sekventiell metod baserad på att simulera en given tidtabell och optimera den för att minska den viktade summan av planerad restid och predikterad medelförsening. Metoden har visat på lovande resultat i simuleringsstudier, där det har varit möjligt att uppnå en väsentligt bättre punktlighet och minskad medelförsening, genom att endast förlänga de planerade restiderna marginellt. Även förbättrad samhällsekonomisk nytta har observerats av att tillämpa den föreslagna metoden. Sammantaget visar detta metodens potentiella nytta och motiverar även fortsatt forskning.

Nyckelord: Järnväg, Tidtabelläggning, Optimering, Simulering, Robusthet, Punktlighet, Förseningsprediktion.

Acknowledgements

First of all, a big thank you to my supervisors, Markus Bohlin and Oskar Fröidh, for helping me so far with my PhD-studies. You have been very supportive and engaged in my research and I have really appreciated our many discussions, short and long, on various topics. They have all mattered to me.

My research has been funded by Trafikverket, grant TRV 2016/5090, and Magnus Wahlborg has been my contact person there. I would like to thank you for providing the necessary Railsys models, timetable data, as well as for your coordination and support. Magnus Wahlborg, Åke Lundberg, Per Johansson, and Emma Solinen at Trafikverket have also participated in the reference group for my research project. Thank you all, for engaging in my research and giving valuable and relevant questions and comments to my work.

In the process of writing this thesis, Erik Jenelius has been the advance reviewer. Thank you for reviewing my thesis and giving valuable feedback. It has been very helpful. From my division I would like to thank all my colleagues, in particular those of you in the (unofficial) railway traffic group. I am happy to have you as my colleagues. I appreciate our lunches and fikas, as well as discussing simulation and python coding.

Anders Peterson and Martin Joborn: It was a pleasure to arrange RailNorrköping with you (together with Markus Bohlin, Emma Solinen, and Sara Gestrelius). Thank you for letting me be part of that.

Torbjörn Larsson and Elina Rönnberg, who supervised and examined my master's thesis: Thank you for your support during and after my master's thesis. You have a great part in that I started this career.

Finally, I would like to thank my family and friends. In particular my wife Lisa and our two children, Alva and Paul. Together, we can do anything. You are the most wonderful family and it is so much fun to be with you.

Johan Högdahl,
Stockholm, November 4, 2019

List of Publications

I. Högdahl, J., Bohlin, M., Fröidh, O., (2017). Combining Optimization and Simulation to Improve Railway Timetable Robustness. Paper presented to ICROMA RailLille2017, Lille, 4-7 April 2017.

II. Högdahl, J., Bohlin, M., Fröidh, O. (2019). A Combined Simulation-Optimization Approach for Minimizing Travel Time and Delays in Railway Timetables. *Transportation Research Part B: Methodological, 126*, 192-212.

III. Högdahl, J., (2019), Delay Prediction with Flexible Train Order in a MILP Simulation-Optimization Approach for Railway Timetabling. Paper presented to ICROMA RailNorrköping2019, Norrköping, 17-20 June 2019.

Contents

Chapter 1

Introduction

In many European countries, the demand for railway traffic has increased rapidly the last decades and is expected to continue grow. This has lead to that the capacity on many lines are insufficient and it is therefore hard to further increase the traffic without worsen the punctuality. One way to address this issue is to improve the planning of the operations by increasing the degree of automation. For example, it has been demonstrated that it is possible to schedule more trains by increasing the accuracy of railway timetables (Schlechte et al., 2011), and many approaches to improve the robustness of railway timetables have been described in the last decade (see for example Caimi et al. (2017) or Lusby et al. (2018) for recent surveys). It is therefore reasonable that improved planning can increase the capacity while not worsen, or even improving, the reliability of railway traffic, assuming that also the precision of the operation is increased.

Important quality indicators of a railway service includes departure frequency, travel time, and reliability. For a given line plan the number of departures is determined, which leaves travel time and reliability as two important properties for an attractive service. The aim with this thesis has been to develop an approach to increase robustness of a given timetable (which could be for instance the draft of a complete timetable, or a compilation of operators requests) by combining micro-simulation with mathematical optimization. In particular, the approach in this thesis has been to to minimize the weighted sum of the scheduled travel time and the predicted average delay (where a delay of a specific event is defined as the difference in the scheduled and actual time of that event), which directly relates the timetables quality to the above-mentioned important quality indicators.

1.1 Background

The research in this thesis belong to the field of railway timetabling, which deals with modelling and solving train timetabling problems. The train timetabling problem is defined as the problem to decide the optimal arrival and departure times, at each station, for a set of trains while satisfying operational constraints for railway traffic. The problem is commonly divided into a periodic and a non-periodic problem. In periodic timetabling the problem is to find the timetable for one period with length T (e.g. 15 minutes or one hour), which then repeats during for example an entire day so that trains within the same line operate once every period T. This has computational benefits as it reduces the problem size (which allows application on larger and more complex networks), and is convenient for passengers as it makes the timetable easy to remember. In the non-periodic timetabling problem no periodic dependencies between the trains exists and they can therefore be scheduled freely. This has the benefit that it may lead to a better optimal value compared with the periodic timetabling problem (Caimi et al., 2017). For this reason it may also consume less capacity than the periodic timetable and is therefore useful in timetabling of long distance railway corridors with heterogeneous and dense traffic, where periodic timetables may be infeasible.

The timetabling problem is a well-studied problem and several models have been proposed to deal with the periodic and the non-periodic problem (Szpigel (1973), Serafini and Ukovich (1989), Jovanovic and Harker (1991), Brännlund et al. (1998), and Caprara et al. (2002)). Recent surveys of railway timetabling models have been conducted by for example Harrod (2012), Cacchiani and Toth (2012), and Caimi et al. (2017).

An important concept in the field of railway timetabling is robustness. It stems from the observation that optimal timetables (for instance in terms of short travel times or high capacity utilization) often are sensitive to disturbances, even small ones. To be useful in practice it is therefore necessary that they are constructed to be robust. The concept lack a single commonly accepted definition, however it is often agreed that robustness is obtained by allocating sufficiently large margins in the timetables, both to allow delay recovery and to reduce delay propagation between trains. Recent surveys on robustness have been conducted by for instance Caimi et al. (2017) and Lusby et al. (2018).

A classical approach to achieve robustness is to apply robust optimization. However, in railway timetabling it has been ruled out as a suitable approach as it would lead to unacceptable poor solutions (Caimi et al., 2017; Fischetti and Monaci, 2009; Liebchen et al., 2009). This is due to that a strict robust solution (e.g. as in Soyster (1973)) must be feasible for all possible combinations of the uncertain parameters which together with the combinatorial nature of timetabling would lead to very low efficiency. For this reason, most of the approaches in the literature have been developed specifically for robust railway timetabling.

To overcome the above-mentioned issue of over-conservatism with classical robust optimization, Liebchen et al. (2009) and Fischetti et al. (2009) have developed the concept of recoverable robustness and light robustness, respectively. The concept of recoverable robustness, proposed by Liebchen et al. (2009), considers a solution to be robust if a set of delay scenarios can be recovered in finite time using a recovery algorithm, and thus share similarities with stochastic programming. In light robustness, proposed by Fischetti et al. (2009), the formulation of the problem resemble classical robust optimization, but the deterioration of the optimal solution to the nominal problem[1] (which has to be computed beforehand) is bounded and violation of the robustness constraints is allowed but penalized. In Fischetti et al. (2009), light robustness was also evaluated in a computational experiment and they found that it could achieve similar results as stochastic models but in shorter time. More recently, Cacchiani et al. (2012) proposed a Lagrangian heuristic approach, in which they iteratively solved a relaxation of the timetabling problem in combination with a scheme to dynamically updating parameters to control the robustness. In computational experiments they found that the results from their approach was similar to the results of light robustness, but in much shorter time.

Kroon et al. (2008) applied stochastic programming to allocate time supplements and margins such that the expected delay is minimized. The approach was evaluated using both computational experiments and a field experiment, which showed that average delays can be significantly reduced with only minor modifications of the initial timetable. Sels et al. (2016) proposed to increase robustness by minimizing the total expected passenger travel time and derived the expected value of the total passenger delay in a closed form, assuming that primary delays[2] are distributed exponentially. One advantage with this approach was that it were deterministic, which avoids the intractability-issue with stochastic programming (which also has been ruled out as a suitable approach as the problem tend to become too hard to solve for real-world instances (Fischetti and Monaci, 2009; Liebchen et al., 2009)), another benefit was also that upper bounds on travel times could be avoided, which solved an issue with infeasibility for earlier attempts using the periodic timetabling model PESP for robust timetabling.

Approaches based on optimizing a robustness metric have also been proposed by e.g. Andersson et al. (2015) and Khoshniyat and Peterson (2017). Andersson et al. (2015), proposed the robustness measure robustness in critical points, in which some events in the timetable are identified as time-sensitive critical points (which for a double track line is all overtaking events and all occasions where a train passes a station from which another train is scheduled to start its journey immediately after the passing train). These events are considered important as small delays have the potential to cause unrecoverable delays if they are not planned to be robust. Khoshniyat and Peterson (2017) propose a strategy called “travel time dependent scheduled minimum headway”, in which the scheduled minimum headway between trains increase with the traveled distance of each train. This is motivated from the observation that the average delay of trains increases with the traveled distance, and the strategy is therefore a mean to compensate against this loss of accuracy.

So far, the above-mentioned approaches has all been based on single-objective

[1] This is the non-robust problem.

[2] Primary delays are delays caused by some event, independent of the timetable. Examples of primary delay causes can be increased running times due to weather conditions or a delayed departure time due to some train error. The condition that primary delays are independent of the timetable rule out delays caused by other delayed trains, earlier delays that have not been recovered, and delays caused by insufficient running or dwell times from being primary delays.

optimization formulations. However, in railway it exists several different stakeholders, who each have objectives and aims that may be in conflict with each other, and it may therefore be insufficient to only optimize a single objective. For this reason, a compromise has to be found and in the following articles multi-objective optimization formulations have been proposed.

Schöbel and Kratz (2009) and Schlechte and Borndörfer (2010) considered the robustness problem as a trade off between the conflicting goals efficiency and robustness, and proposed bi-objective optimization problems to address this. In Schöbel and Kratz (2009), robustness measures for three delay management strategies was proposed and conditions for the Pareto-optimal solutions was derived, while Schlechte and Borndörfer (2010) instead proposed a robustness function based on buffer time allocation and solved a scalarized formulation of the optimization problem to find the Pareto-optimal solutions.

Chow and Pavlides (2018) and Yan et al. (2019) have both described multi-objective optimization problems. In Chow and Pavlides (2018), a customer's perspective is adopted and they include journey time, waiting time, crowdedness, and punctuality as objectives, in which the punctuality for each candidate timetable is estimated using Monte Carlo simulation and new candidate timetables are generated using a genetic algorithm in combination with Dijkstra's algorithm. In Yan et al. (2019), periodic timetabling of railway corridors is considered and they propose journey time, regularity deviation, the number of overtakings, and vulnerability as objectives, where vulnerability is based on penalizing too small and too large headways.

A common characteristic for most timetabling models is that they model the timetabling problem macroscopically, meaning that they do not include for example switches and signals in the problem and that stations are modelled as points (or black-boxes). By doing this, the problem becomes easier to solve but may lead to issues with microscopic infeasibility. This may sound like a theoretical problem but is in fact a real and well-known issue, which for instance can lead to that it is impossible to route trains through stations according to the announced timetable. To address this "micro-macro" issue, a few approaches has been found in the literature.

Schlechte et al. (2011) and Bešinović et al. (2016) have proposed approaches based on transforming the microscopic network into a macroscopic problem, Bešinović et al. (2016) additionally proposed an iterative method to compute a timetable which is microscopically feasible, stable and robust. In the evaluation of Schlechte et al. (2011) they showed that it was possible to generate a microscopically feasible timetable (evaluated with OpenTrack[3]) using their approach and that capacity can be increased by increasing the precision in the time discretization of the timetable. In Bešinović et al. (2016), the approach was evaluated in a computational experiment and it was demonstrated that their approach was able to find a microscopically feasible and stable timetable that addionally satisfied the UIC infrastructure occupation rate norm. de Fabris et al. (2014) proposed a heuristic approach to construct

[3]URL: http://www.opentrack.ch/

timetables using a mesoscopic model of the infrastructure, in which the network is divided into line and station tracks which are connected with switch areas (which describe how the line and station tracks are connected). Computational experiments demonstrated that their approach could generate timetables that can be accepted by the timetable planners in a few minutes. For timetabling of large and complex station areas, Burggraeve and Vansteenwegen (2017b) proposed a two-step approach in which a routing plan for each train is computed on the microscopic level and a timetable (which satisfies the routing plan) is computed on the macroscopic level. Using this approach they where able to improve the robustness with 11 % compared with a timetable from the Belgian railway infrastructure manager Infrabel. Finally, Lamorgese et al. (2017) extended the regular train timetabling problem with a station track model and proposed a decomposition approach (which extended an approach previously developed for dispatching) to efficiently find an exact solution. In a computational experiment based on data from the Norwegian railway they demonstrated that the approach can find a feasible timetable in short time for small real world instances.

Most of the approaches described above have one thing in common: they rely on optimization - either for constructing a timetable from scratch or improving a given one. In other areas, optimization is often found to be used in combination with simulation. However in the railway research community, this is an approach that seems to have received less attention, even though it potentially could help to improve the robustness and mitigate for instance the micro-macro problem. In the literature, examples of such methods are found in Bešinović et al. (2016), Lee et al. (2017), Burggraeve and Vansteenwegen (2017a), and Chow and Pavlides (2018).

In the micro-macro approach described by Bešinović et al. (2016), macroscopic simulation is integrated in a heuristic algorithm for constructing a robust macroscopic timetable (in which simulation is part of evaluating the objective function value for each generated timetable). Another heuristic approach has been described by Lee et al. (2017) and Burggraeve and Vansteenwegen (2017a), who also propose iterative methods. A common characteristic in their methods is that a sequence of candidate timetables are computed by iteratively simulating the most recent candidate timetable and solving an optimization problem (in each iteration a set of parameters in the optimization problem, which govern the allocation of margins in the timetable, is updated based on the simulation result). Finally, Chow and Pavlides (2018), propose another heuristic approach based on combining a genetic algorithm with Dijkstra's algorithm to iteratively construct a timetable that optimizes a multi-objective optimization problem (in which punctuality is one of the objective functions, which is evaluated using simulation).

1.1.1 Relation to Previous Work

The approach in this thesis has been to combine micro-simulation with macroscopic timetable optimization to improve the robustness of a given timetable by minimizing the weighted sum of scheduled travel time and predicted average delay.

A closely related objective has been proposed by Sels et al. (2016), who proposed to minimize the total expected passenger travel time, which is the sum of the scheduled passenger travel time and the expected delay. However, the weighted sum of scheduled travel time and predicted average delay, which is the objective in this thesis, has a different interpretation and does not result in that the predicted average travel time necessarily is minimized. Further, the approach in this thesis differs from the approach in Sels et al. (2016) as optimization is here combined with simulation.

Another example in which delays are minimized is found in Kroon et al. (2008). However, they do not include the scheduled travel times in the objective function, and therefore have to formulate upper bounds on the feasible running times. Their approach, stochastic programming, can also be considered to be a combined simulation-optimization approach, in which the second-stage problem is a simulation of the timetable obtained as the solution to the first-stage problem. However, stochastic programming is a different approach, in which the timetable and the expected delays in it is jointly optimized, while the approach in this thesis is based on formulating a model to predict the effects of modifications in the initial timetable.

As described in the end of the previous section, several simulation-optimization approaches for railway timetabling has already been proposed. A distinct difference between those approaches and the approach in this thesis is that those are iterative approaches, in which a new timetable is generated and then simulated in each iteration. The result from the simulation is then either used to update parameters in the optimization problem or added to the objective function value. In this thesis the approach has instead been to simulate a given timetable and based on the simulation result form a delay prediction model which is then included explicitly in the objective function of the optimization problem and minimized directly.

1.2 Thesis Organization

This thesis is organized in the following five chapters:

- Chapter 1: Introduction to the thesis.
- Chapter 2: Statement of the research goals.
- Chapter 3: Description of the research methodology.
- Chapter 4: An overview of the included papers.
- Chapter 5: Conclusions and future work.

Chapter 2

Research Objectives

Planners in most railway companies rely on simulation to evaluate their timetables. Often the simulation tools are micro-simulators, which includes the position and function of switches, signals, and track geometry, aswell as dynamic models of the trains. The main advantage of this, highly detailed, representation is that running times and minimum headway times can be computed accurately (as the computations are based on physical models and blocking time theory), which makes it possible to accurately detect conflicts.

In the research community, most approaches for generating railway timetables are based on solving mixed integer linear programming problems using a macroscopic representation of the network. However, if such macroscopic timetables are imported to a micro-simulation tool this may lead to problems with microscopic infeasability, caused by microscopic conflicts between trains or unrealizible running times, which can be observed with simulation as it is a cause of delays.

One approach that has been proposed in the literature (Schlechte et al., 2011) is to, in advance of the timetable optimization procedure, transform the microscopic network into a macroscopic representation to find a timetable that is microscopically feasible. However, compared to a traditional macroscopic formulation this can make the optimization problem harder to solve, which may be problematic for large-scale timetabling. As the regular macroscopic timetabling approach yield a more aggregated formulation, which potentially is easier to solve (although it is still NP-hard), it would be of interest to investigate approaches in which micro-simulation data are used as input to reduce delays, and thus also to mitigate the consequences of the micro-macro problem.

The purpose of this thesis has therefore been to develop models and methods to combine micro-simulation with macroscopic timetable optimization to improve the robustness of a given timetable. By the term robustness we, in this thesis, refer to the ability of a timetable to avoid large delays due to minor disturbances, and as a measure of robustness the total average delay (which is measured as average arrival delay at the end station for all trains) is used.

Considering the above, this has been formulated as the following two research goals:

RG1: Develop a method that combine micro-simulation and mathematical optimization for robust railway timetabling.

RG2: Integrate delay prediction in non-periodic timetable optimization.

Research Goal 1 (RG1) refers to the overall aim of developing methods to combine micro-simulation with macroscopic timetable optimization to improve the robustness, which is in line with the overall aim of the research project that has funded this thesis. Research Goal 2 (RG2) correspond to the definition of robustness and distinguish the method in this thesis from related approaches (as was discussed in Section 1.1.1).

2.1 Limitations

The research has focused on prioritized lines in the Swedish railway network. Two of the most important lines in Sweden are the Southern and the Western Main Line, which connects Stockholm with Malmö and Göteborg, respectively. These two lines can be considered as corridors, with a highly heterogeneous traffic mix, with sections having at least double tracks. For this reason, the research in this thesis has been limited to non-periodic timetabling of double track railway lines with mixed passenger and freight traffic.

As the research has focused on corridors, trains running solely on a branch line (i.e. trains that are entering and leaving the main line in the same station) have been excluded. The effect of this has been expected to be small as traffic on a main line have higher priority than traffic on a branch line. As branch line trains have been excluded, connections between trains (e.g. for passenger transfers) have also been neglected.

Field experiments has not been conducted, instead the evaluation has been limited to simulation experiments. This means that the conclusions inferred in this thesis are in, a strict sense, only valid within the simulation environment. However, for the following reasons we still judge that results from simulation can be viewed as an indication of the real system's response: firstly, simulation is an established method in the railway community; secondly, the simulation tool that has been used in this research (RailSys[1]) is based on research; and thirdly, the simulation tool is a standard tool for capacity analysis in Sweden and has been used for more than 10 years by the capacity planners at the Swedish Transport Administration for capacity analysis and timetable evaluation.

[1]URL: https://www.rmcon.de

Chapter 3

Research Methodology

This chapter describes the research methodology of this thesis, with focus on the proposed approach for improving the robustness of a given timetable. The chapter is divided in two sections, in which the first section describe the proposed approach and the second section describe the evaluation of the approach.

3.1 The Simulation-Optimization Approach

The proposed approach is based on combining micro-simulation and macroscopic timetable optimization to improve the robustness of a given timetable. It features two main parts, simulation and optimization, and can in summary be described as follows: (1) Begin with an initial timetable, which could for instance be a draft timetable or a compilation of operators requests; (2) simulate the initial timetable and collect delay data; and (3) use the delay data as input parameters in the timetable optimization problem and minimize its objective, which has been the weighted sum of the scheduled travel time and predicted average delay. The resulting timetable is called the rescheduled timetable.

The approach is illustrated in Figure 3.1, which also points out how the approach has differed between the included papers. Initially, in Paper I, the approach was proposed to be iterative. However, as it was unlikely that the approach would converge it was in Paper II instead reformulated into a sequential three-step approach, which included a new calibration step (compared with the initial approach). The calibration step featured an optimization problem to optimize the parameters of the delay prediction model and was proposed to be solved using a random search method. In the most recent paper, Paper III, the calibration step was not included and the parameters in the delay prediction model (which differed from the model in Papers I and II) were instead chosen manually. However, the calibration step could (with only minor modifications) be included in the approach also with the new delay prediction model.

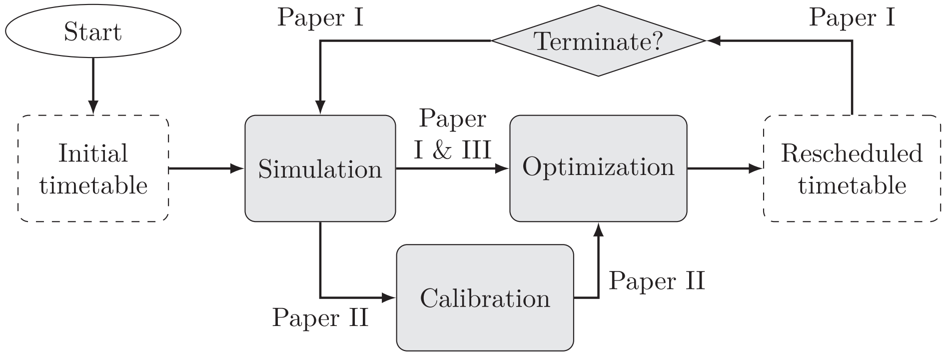

Figure 3.1: The simulation-optimization approach.

3.1.1 Simulation

Simulation has had two roles in this thesis: first, it's part of the proposed method in which it has been used to generate delay data for the models used when rescheduling the initial timetable; and second, the proposed method has been evaluated in simulation experiments (which is described in Section 3.2). All simulations have been carried out using Railsys (Version 9.8.25, Rail Management Consultants GmbH[1], 2014), which is a commercial software for microscopic simulation of railway systems.

The infrastructure in Railsys is represented by a graph, in which the nodes represents physical and virtual objects (such as signalling equipment, switches, crossings, station borders, stopping and timing locations, etc.) and the links represents physical tracks. Further, the tracks are divided into block sections which enables to simulate the signalling system. A graphical representation of the infrastructure model is included in Figure 3.2.

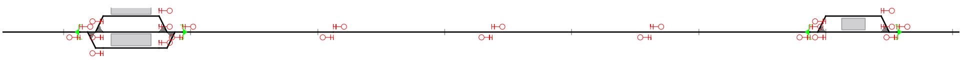

Figure 3.2: Example of a single track line connecting two stations (exported from Railsys). The black lines shows the physical track, the red markers shows signals and the green symbols shows station borders.

The timetable in Railsys contains the planned train formation and the scheduled arrival time, departure time, route, and stopping or timing location at each station for all trains. By combining the timetable data with train dynamics and the infrastructure, running times are calculated. The running time calculations makes it possible to detect infeasible travel times and conflicts between trains, where a

[1] URL: https://www.rmcon.de

conflict is detected if two trains are scheduled to use the same block section at the same time. An example of the graphical timetable is included in Figure 3.3.

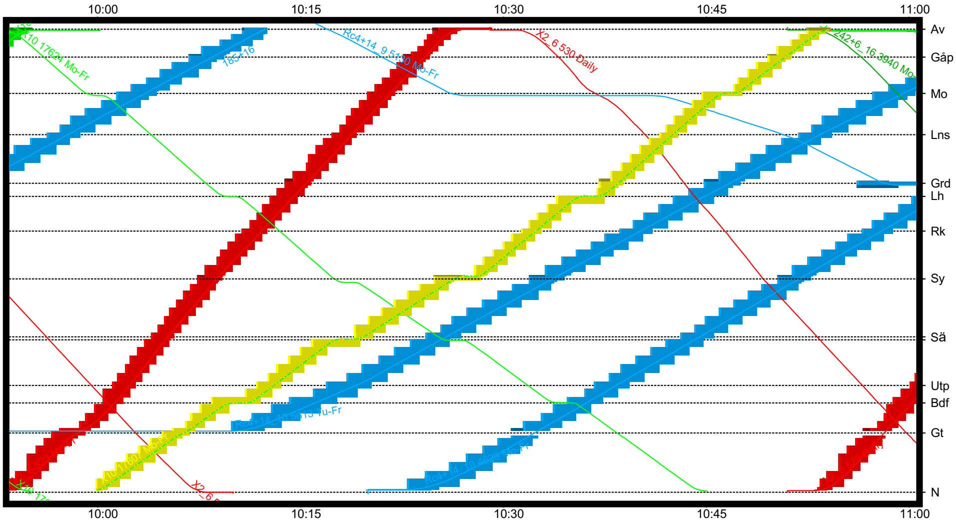

Figure 3.3: Example of a graphical timetable (exported from Railsys). The graphical timetable shows each trains position (the y-axis) as a function of time (the x-axis). The thin lines shows the the scheduled slot for each train, and the thick stairway-shaped lines shows the scheduled block occupation for each train. Dashed horizontal lines shows the stations.

When simulating a timetable, a set of randomly generated perturbed timetables are given as input to the simulator and the output is a set of realizations of the traffic (in which each realization correspond to one perturbed timetable), which is illustrated in Figure 3.4.

The perturbed timetables represent a set of traffic days and each perturbed timetable contains a declaration of all primary delays (i. e. delays that are independent of the timetable). The primary delays are generated randomly and are divided into entry delays (which are applied when a train enters the system), dwell time delays (which are applied when trains stop at stations), and running time extensions (which are applied when trains are travelling between stations).

The realizations contains the actual arrival and departure times at each station for all trains, which is the result by applying the primary delays of the perturbed timetable, using the scheduled time supplements to recover delays, and having a dispatching algorithm to solve all conflicts that (as a consequence of the delays) may occur.

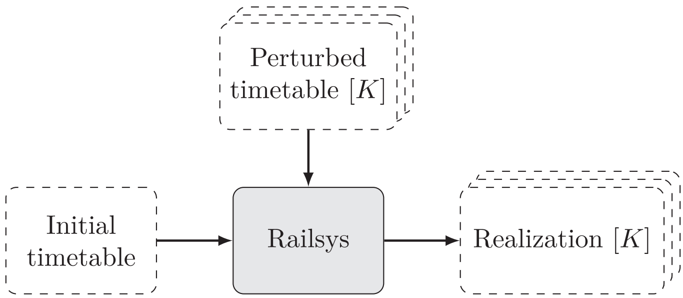

Figure 3.4: Simulation input and output.

3.1.2 Optimization

Optimization has in this thesis been used to generate the rescheduled timetable, which is generated by solving a train timetabling problem (which is the problem of finding the optimal timetable for a set of trains, T, that operate on a network with a set of stations, S, such that the resulting timetable satisfies the operational constraints for railway traffic).

In a compact form, that highlights the relation between the scheduled travel time and predicted average delay, this optimization problem has been formulated as follows:

$$\begin{aligned} \min \; & f(t) = \alpha F(t) + (1-\alpha)G(t) \\ \text{st.} \; & t \in X, \end{aligned} \tag{3.1}$$

where t is a timetable, $F(t)$ is the total scheduled travel time, $G(t)$ is the total predicted average delay, α is a weighting parameter, and X is the set of all feasible timetables.

In the literature it exists several approaches to model the train timetabling problem, and in this research a non-periodic event-based model has been used. This has been suitable as the initial timetable has been a non-periodic timetable and as one aim has been to allocate time supplements and buffer times, which require flexibility in the scheduled running and dwell times. Table 3.1 summarizes a collection of aspects, that points out differences and commonalities, for the optimization problem that has been used in the included papers. For details regarding the formulation of the optimization problem, the reader is referred to the included papers of the thesis.

3.1.2.1 The Predicted Average Delay

The simulation and optimization step in the proposed approach is connected via the predicted average delay, $G(t)$, in Problem (3.1). For this reason, and since it has been a vital part of the research, the two delay prediction models proposed in

	Problem type	Delay prediction model	Operational constraints	DoF	Restrictions	Scope
Paper I & II	LP	Model 1 (see Section 3.1.2.1)	Feasible running times, arr./dep. headway, (overtakings only at stations)	Time of events, running times, dwell times	Fix departure time from the first station, fix sequence of the trains at each station	Double track line, one direction
Paper III	MILP	Model 2 (see Section 3.1.2.1)	Feasible running times, arr./dep. headway, overtakings only at stations	Time of events, running times, dwell times, sequence of the trains, and stop pattern for freight trains.	Bounded event times (within $\pm D$ min of the initial event time, where $D \geq 0$ is a constant)	Double track line, both directions

Table 3.1: A comparison of the optimization models used in the included papers. For Papers I and II, "overtakings only at station" is placed inside parenthesises as it was modelled but not relevant as the sequence of the trains were fixed at every station. DoF=degrees of freedom (which should be interpreted as what can be modified in the initial timetable), LP=linear programming, MILP=mixed integer linear programming.

Papers I and II and in Paper III will here be described in detail. The main difference between these two models has been that in Papers I and II the formulation of the delay prediction model limited the set of feasible timetables, such that the sequence of the trains (at every station) in the rescheduled timetable must be equal to the sequence of the trains in the initial timetable.

For the purpose of describing the delay prediction models, the following notation is now introduced. Let T denote the set of all trains, S denote the set of all stations, and let E_a^h denote the set of all arrival events for train $h \in T$. Let the timetable t be a vector, where t_i^h denote the i:th event of train h, and let τ denote the initial timetable. Finally, let y be the vector of all realizations (i.e. the simulation output), and let $y_{h,i}^r$ denote the actual time of the i:th event for train h in realization r, where $r = 1, \ldots, K$ and K is the number of realizations of the initial timetable.

Paper I and II: Delay prediction model 1

In Papers I and II, the delay prediction model is based on computing the empirical probabilities of each observed delay for all arrival events, $p^h_{i,\omega}$, which is done as follows:

$$p^h_{i,\omega} = \frac{1}{K}\sum_{r=1}^{K} \mathbb{1}_\omega \left(y^h_{i,r} - \tau^h_i\right), \qquad \forall \omega \in \Omega^h_i,\ \forall (h,i) \in E^h_a,\ \forall h \in T, \tag{3.2}$$

where Ω^h_i is the set of all observed delays for the event $(h,\ i) \in E^h_a$, and $\mathbb{1}_\omega(z)$ is the indicator function, which is 1 if and only if $z = \omega$.

The model is based on the following assumptions: (1) the departure time from the first station is fixed; (2) train drivers aim to drive as fast as possible, (so that the expected arrival time to a given station does not depend on the scheduled arrival time), which means that the average delay at a given station can be reduced by extending the scheduled arrival time; and (3) if the headway to the preceding train is reduced then the average delay increases, given that the two trains have a sufficiently small headway. Based on these assumptions, the following delay prediction model for changes in the initial timetable was proposed:

$$G(t) = \sum_{h \in T} \sum_{i \in E^h_a} \lambda^h_i g^h_i(t), \tag{3.3}$$

$$g^h_i(t) = \sum_{\omega \in \Omega^h_i} p^h_{i,\omega} \max\left\{0, \left(\tau^h_i + \omega\right) - t^h_i + \beta\phi\left(t^h_i\right)\right\}, \tag{3.4}$$

$$\phi\left(t^h_i\right) = \begin{cases} t^k_j - \tau^k_j, & \text{if } t^h_i - t^k_j < b, \\ 0, & \text{else,} \end{cases} \tag{3.5}$$

where Eq. (3.3) defines the total predicted average delay, $G(t)$, as the weighted sum of the predicted average delay, $g^h_i(t)$, for each arrival event with the weighting factor λ^h_i (which can be e.g. the number of disembarking passengers). Eq. (3.4) computes the predicted average delay for the i:th event of train h, which is computed by multiplying the empirical probability that a delay ω occurs with the delay that is assumed to be observed if the arrival time were t^h_i instead of τ^h_i. Finally, Eq. (3.5) defines the the train dependency penalty function, $\phi\left(t^h_i\right)$, which penalizes the predicted average delay of event $(h,\ i)$ if the margin to the preceding event $(k,\ j)$ is reduced (given that the rescheduled headway between the two events is less than the threshold b) .

Paper III: Delay prediction model 2

In Paper III, the realizations of the initial timetable is used to compute the average delay of each event. The delay prediction model is then based on computing the difference to the initial timetable, with the following assumptions: (1) It is assumed that any event will occur as early as possible, which means that a train is assumed to run at the highest possible speed and that the driver will depart as soon as it is allowed (i.e. either at the scheduled departure time, or when the driver receives a green signal); and (2) it is assumed that the delay will depend on either the self-induced delay or the knock-on delay from the preceding train, where the self-induced delay is the delay of an event that does not directly depend on another train. Based on this, the delay prediction model was formulated as follows:

$$G(t) = \sum_{\forall (h,i) \in E} \lambda_i^h \hat{\Delta}_i^h, \tag{3.6}$$

$$\hat{\Delta}_i^h = \max \left\{ 0,\ \hat{\delta}_i^h,\ \hat{\kappa}_i^h \right\}, \tag{3.7}$$

$$\hat{\delta}_i^h = \hat{\Delta}_{i-1}^h + \left(\bar{\delta}_i^h - \bar{\delta}_{i-1}^h \right) + \beta f_i^h \left(\sigma_{i-1,i}^h - s_{i-1,i}^h \right), \tag{3.8}$$

$$\hat{\kappa}_i^h = \max_{(k,\ j)} \{ \hat{t}_j^k + h_{min}^s - M x_{ij}^{hk} \} - t_i^h, \tag{3.9}$$

$$\hat{t}_i^h = t_i^h + \hat{\Delta}_i^h, \tag{3.10}$$

$$f_i^h(y) = \begin{cases} y, & -M < y \leq 0, \\ 0, & 0 < y \leq a_i^h, \\ y - a_i^h, & a_i^h < y \leq \sigma_{i-1,i}^h, \end{cases} \tag{3.11}$$

where Eq. (3.6) gives the total predicted average delay of all timing events E (i.e. all events where punctuality is important), computed as the weighted sum of the predicted average delay for each event. Eq. (3.7) gives the predicted average delay of event $(h,\ i)$, which is the non-negative maximum value of the predicted self-induced delay, $\hat{\delta}_i^h$, and the predicted knock-on delay, $\hat{\kappa}_i^h$. Eq. (3.8) gives the predicted self-induced delay for event $(h,\ i)$, where $\bar{\delta}_i^h$ and $\bar{\delta}_{i-1}^h$ is the average delays for the events $(h,\ i)$ and $(h,\ i-1)$ using the initial timetable, and $s_{i-1,i}^h$ and $\sigma_{i-1,i}^h$ is the time supplements between events $i-1$ and i for train h in the rescheduled and the initial timetable, respectively. Eq. (3.9) gives the predicted knock-on delay for event $(h,\ i)$, which is the difference between the sum of the predicted time for the preceding event $\hat{t}_j^k$ and a safety margin h_{min}^s for the station s, and the scheduled time for the event, t_i^h. Eq. (3.10) defines the predicted time of event $(h,\ i)$ as the sum of the scheduled time and the predicted delay. Finally, Eq. (3.11) defines a piecewise linear function, which allows reduction of time supplements up to the value of a_i^h without increasing the predicted average delay, where M is a large constant and $\sigma_{i-1,i}^h$ is the scheduled time supplement for the initial timetable.

3.2 Evaluation

As stated previously, simulation experiments with Railsys have been conducted to evaluate the approach. Each experiment has been setup such that a common base project was created, originating from the Swedish Transport Administration's (Trafikverket) Railsys-model of the Swedish railway network. From the base project a partial network, covering a certain line segment, was generated, and all trains (except trains on branch lines) that operates on any part of the partial network during a specific time period were selected (further details are given in the included papers). The initial timetable was then obtained by solving all conflicts between the selected trains, and the rescheduled timetable by applying the approach developed in this thesis on the initial timetable. Finally, both timetables were simulated for 200 cycles using the same (partially synthetic) distributions for the random primary delays. For an overview illustration of the simulation experiments, see Figure 3.5.

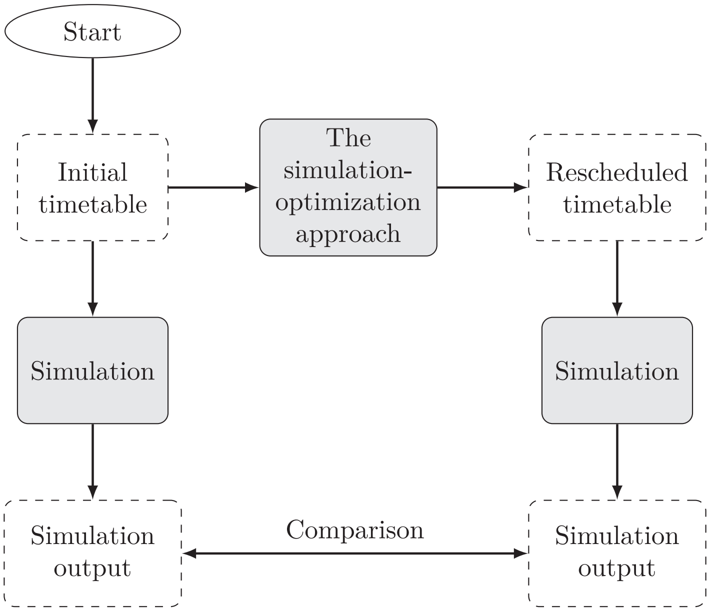

Figure 3.5: Overview of simulation experiment

From the raw simulation output, arrival and departure times at each station for all trains were gathered. Simulation cycles with deadlocks were removed from the data set. Data analysis was conducted using Matlab, Python and Microsoft Excel.

Chapter 4

Overview of the Included Papers

This chapter contains an overview of the research in this thesis. The first section lists the papers and includes a reference to each paper, its abstract, and a declaration of contribution. Then, in the following section a mapping between the included papers and the research goals is provided. This is followed by a section that highlights how the method of this thesis has evolved in the three papers. Finally, in the last section of this chapter the scientific contribution for each paper is summarized, including references to where the stated contributions can be found in the papers.

4.1 Included Papers

This section summarize the included papers and includes a declaration of contribution to each paper.

4.1.1 Paper I

Högdahl, J., Bohlin, M., Fröidh, O., (2017). Combining Optimization and Simulation to Improve Railway Timetable Robustness. Paper presented to ICROMA RailLille2017, Lille, 4-7 April 2017.

Abstract

The Train Timetabling Problem (TTP) is the problem of finding the timetable that utilizes the infrastructure as efficient as possible, while satisfying market demands and operational constraints. As reliability is important to passengers it is important that timetables are robust. In this paper we propose a method that combines optimization and simulation to find the timetable that minimizes the travel times and maximizes the expected punctuality. The core method consists of iteratively re-optimizing a bi-objective mixed integer sequencing timetable model, where both planned travel time and simulated delays are taken into account. Each generated timetable is validated and re-evaluated using the micro-simulation tool RailSys.

The advantage of the method is that it captures both the uncertainty of a timetable at the planning stage and the validity of the generated timetable. The method is evaluated on a unidirectional track section of the Western Main Line in Sweden and shows promising results for future research.

Declaration of Contribution:

The idea of this paper originated from Markus Bohlin. In cooperation we all formulated the delay prediction model. The optimization model was inspired on a formulation proposed in Fischetti et al. (2009). I implemented the optimization model and conducted the simulation experiments (using Railsys models provided by the Swedish Transport Authority (Trafikverket)). The approach, results, and conclusions were continuously discussed among the authors. I, as main author, contributed to all sections. The co-authors Markus Bohlin and Oskar Fröidh contributed mainly to the introductory section.

4.1.2 Paper II

Högdahl, J., Bohlin, M., Fröidh, O. (2019). A Combined Simulation-Optimization Approach for Minimizing Travel Time and Delays in Railway Timetables. *Transportation Research Part B: Methodological, 126*, 192-212.

Abstract

Minimal travel time and maximal reliability are two of the most important properties of a railway transportation service. This paper considers the problem of finding a timetable for a given set of departures that minimizes the weighted sum of scheduled travel time and expected delay, thereby capturing these two important socio-economic properties of a timetable. To accurately represent the complex secondary delays in operational railway traffic, an approach combining microscopic simulation and macroscopic timetable optimization is proposed. To predict the expected delay in the macroscopic timetable, a surrogate function is formulated, as well as a subproblem to calibrate the parameters in the model. In a set of computational experiments, the approach increased the socio-economic benefit by 2-5% and improved the punctuality by 8-25%.

Declaration of Contribution:

This paper was an extension of Paper I. I formulated a stochastic calibration problem and proposed a solution approach based on a basic random search method described in Spall (2003). The approach, results, and conclusions were continuously discussed among the authors. I conducted the experiments and analyzed the results. As the main author, I contributed by writing most of the text in all sections. The co-authors, Markus Bohlin and Oskar Fröidh, contributed mainly to the introductory section.

4.1.3 Paper III

Högdahl, J., (2019), Delay Prediction with Flexible Train Order in a MILP Simulation-Optimization Approach for Railway Timetabling. Paper presented to ICROMA RailNorrköping2019, Norrköping, 17-20 June 2019.

Abstract

This paper considers the problem of minimizing travel times and maximizing travel time reliability, which are important socio-economic properties of a railway transport service, for a given set of departures on a double-track line. In this paper travel time reliability is measured as the average delay, and a delay prediction model for MILP timetable optimization is presented. The average delay prediction model takes into consideration time supplements, buffer times and propagation of delays in the railway network and is not restricted to a fixed order of the trains. Validation of the average delay prediction model, and an evaluation of the approach with combined simulation-optimization for improving railway timetables, are conducted by a simulation study on a part of the Swedish Southern Main Line. Results from the simulation study show that the average delays are reduced by up to approximately 40% and that the punctuality is improved by up to approximately 8%.

Declaration of Contribution:

The basic idea of this paper originated from discussions with my supervisors during our joint work with Papers I and II. However, in this paper I formulated and implemented the model, conducted the experiments and wrote the paper by my self.

4.2 Mapping of the Papers to the Research Goals

Table 4.1 shows how the papers relate to the research goals.

	RG1	**RG2**
Paper I	✓	✓
Paper II	✓	✓
Paper III		✓

Table 4.1: Mapping of the papers to the research goals.

4.3 Relation Between the Included Papers

Table 4.2 shows the relation between the included papers. In particular, it highlights how the method and the delay prediction model have evolved between the papers, as well as which quality measures that have been included in the evaluation.

	Approach	**Delay prediction model**	**Train sequence restrictions**	**Evaluation**
Paper I	Iterative	Model 1 (see Section 3.1.2.1)	Fixed during optimization	Punctuality Pareto Front (average delay and travel times)
Paper II	Sequential	Model 1 (see Section 3.1.2.1)	Fixed during optimization	Punctuality
Paper III	Sequential	Model 2 (see Section 3.1.2.1)	Flexible during optimization	Punctuality, Average delays, Travel times, Running time supplements

Table 4.2: Evolution of the method during Papers I to III. The restriction that the train sequence is "Fixed during optimization" means that the sequence of the trains at all stations must be equal to the sequence of the trains in the initial timetable.

4.4 Summary of the Scientific Contribution

This section summarize the scientific contribution for each paper.

4.4.1 Paper I

In this paper, an iterative method to combine micro-simulation and optimization to improve the robustness of a given timetable by minimizing the weighted sum of scheduled travel time and the predicted average delay was proposed (Section 4).

A delay prediction model based on scheduled travel time and headway to the previous train was also proposed (Section 4.1). The delay prediction model was formulated such that the sequence of the trains (at every station) had to be equal to the sequence of the trains in the initial timetable.

The method was partially evaluated (the iterative part of it was not included) in a simulation study and it was found that the punctuality increased when the valuation of reducing delay time compared with reducing the scheduled travel time was increased (Section 5.2).

4.4.2 Paper II

In this paper a sequential three-step approach combining micro-simulation and optimization to improve the robustness of a given timetable was proposed (Section 2.3). An important contribution to the method was the formulation of a calibration problem (Section 2.2) to set the parameters of the delay prediction model (which was the same as in Paper I).

It was demonstrated (Section 3.2.3) that by including the calibration of the delay prediction model, the solution to the optimization problem was near-optimal with respect to the weighted sum of scheduled travel time and observed average delay (in which the observed average delay is the value obtained by simulating the resulting timetable), when compared with timetables that was generated by solving the optimization problem with random parameter values.

The performance of the method was evaluated in a simulation study using a set of delay scenarios (Section 3.2.4). It was observed that the punctuality was increased by 8-25% compared with the initial timetable (p. 208). The effect on the scheduled travel times was not evaluated explicitly, instead a socio-economic measure was used to indicate if the overall effect of extending the travel times was positive, which it seemed to be as the socio-economic cost of the traffic was observed to decrease by 0-5% depending on delay scenario and valuation of delay time (p. 208). The evaluation also indicated a positive relation between low punctuality of the initial timetable and larger potential benefits of the method (p. 207).

4.4.3 Paper III

In this paper a new model for predicting the total average delays was proposed and linearized (Sections 3.1.2 and 3.1.3). It differed from the model proposed in Paper I as it did not require that the sequence of the trains in the rescheduled timetable and the initial timetable has to be equal.

The approach was evaluated in a simulation experiment on a part of the Swedish Southern Main Line (Section 5.2). In the experiment the effect of either rescheduling all train, all freight trains, or all passenger trains was investigated and it was observed that for the rescheduled timetables the average travel time differed by -1% to +5%, the total average delays was reduced by 8-40%, and the punctuality was increased by 2-8% compared with the initial timetable (these figures are computed from Table 4). In addition, it was also observed that the socio-economic cost[1] was reduced by 2-5% (also computed from Table 4). The largest effects was observed when either the passenger trains or all trains were rescheduled.

[1]Denoted TPI (timetable performance index) in Table 4, which is computed as a weighted sum of scheduled travel time and average delay.

Chapter 5

Conclusions and Future Work

The aim with this thesis has been to develop an approach to improve the robustness of a given timetable. Three papers have been included in the thesis and they all contribute to the research goals. In this chapter the results of all papers are summarized and discussed.

5.1 Conclusions

The most interesting results from the simulation experiments has been the following:

- It was observed that the scheduled travel times were increased by integrating a prediction of the average delay in the optimization problem. In the experiments of Paper III, the difference compared to the initial timetable was in the range of -1 to +5%.
- By integrating the predicted average delay in the timetabling model it was possible to reduce the total average delay. In the experiments of Paper III, it was observed that the average delay was reduced by approximately either 8% or 40%, where an approximate 40% reduction was observed when either all trains or the passenger trains were rescheduled.
- As the average delays were reduced, the punctuality was improved. In the experiments of Papers II and III, improved punctuality up to 25% and 8% (respectively) was observed.
- By applying the approach, the socio-economic cost of railway traffic was reduced. In Papers II and III, the socio-economic cost was reduced by up 5%.

The above-mentioned results are interesting as it indicate that the proposed approach can lead to a substantially more reliable railway system, by only increasing the advertised travel times slightly. It has also been shown that the socio-economic cost is reduced by applying this approach. As punctuality and delays are among the most important factors that influence the attractiveness of railway, it seems reasonable that the approach would also have positive effect on the demand, and therefore that the overall socio-economic benefit of this approach also would be positive.

However, it must be pointed out that the above figures are averages for all trains, and that the effects differ between individual trains. Several assumptions have also been made that influence the results, and the most important ones have been: (1) the capacity is sufficient to handle increased travel and buffer times, and (2) the distributions of the primary delays can be estimated with sufficient precision.

Another reason to be cautious when interpreting the results from this thesis is that they are observations from simulation experiments. The main objection against simulation as a method is that it can be questioned whether or not the conclusions from a simulation experiment can be generalized to the corresponding real-world system. When it comes to micro-simulation of railway traffic, the distribution of the primary delays and the dispatching algorithm are two important aspects that potentially can lead to false inferences. In this thesis, these aspects have been considered as uncertain parameters and to mitigate the risk that they are chosen badly they have been treated as is described in the two following paragraphs.

The distribution of the primary delays have in this research been synthetic (meaning that they they do not come from real observations), which have the potential to cause unrealistic results. In order to mitigate this issue, the experiments in Paper II were conducted over a set of different delay scenarios. The delay scenarios were composed of entry delays, dwell time delays, and running time delays and the difference between the delay scenarios was in the magnitude of the running time delays (which varied in the span of approximately 0-20% of the minimum running time). The main difference in outcome between the experiments with different delay scenarios was that the observed effects increased with larger running time delays, however the inferred conclusions remained for each delay scenario (which lead to a conclusion in it self). As the main conclusions of Paper II did not depend on the delay scenario, only one set of primary delays was used in Paper III. This set of primary delays was chosen such that the initial timetable obtained a punctuality comparable to the punctuality in Sweden.

When it comes to the influence of the dispatching algorithm two different dispatching strategies have been evaluated. In Papers I and II, all dispatching actions were disabled which lead to that the traffic was operated according to a first-in-first-served principle. In Paper III, the dispatching algorithm was configured such that the trains were assigned movement authorities according to train priority and where delayed trains were less prioritized than trains on-time (which follows the governing principle for dispatching in Sweden).

In combination with that all observations have been made based on comparing differences between similar simulation scenarios, it is therefore anticipated that the results would hold if the method were applied in reality for traffic with similar conditions (i.e. Swedish double track lines), although the actual values of the observed effects might differ compared with the results obtained in this thesis.

Finally, to summarize the discussion the experiments have shown interesting, promising results. However, the generalization of the conclusions from the simulation experiments to a real-world railway system is not trivial. Therefore, the conclusion of this thesis is as follows: In simulation the approach (1) improve the reliability of railway traffic, by only increasing the scheduled travel times slightly, and (2) improve the socio-economic benefit of railway traffic.

However, as of the limitations with simulation, the findings so far does not provide sufficiently strong evidence to prove that the approach would be effective for real-world applications in practice. The results so far should therefore be considered as a strong motivation for further research, which ultimately could either strengthen the confidence in the approach, or rule out its practical usefulness.

5.2 Future Work

As discussed in the previous section, the approach described in this thesis have achieved promising results in simulation experiments. However, to yield the full potential in practice it is important to have good knowledge of delay distributions. In this thesis it has been assumed that we have perfect knowledge of this, and therefore the same delay distributions have been used when simulating the initial timetable (to generate delay data for the optimization problem) as when the rescheduled timetable was validated. This is clearly an idealization that cannot be satisfied in reality and for this reason it would be interesting to evaluate the methods sensitiveness (or robustness) to errors in the delay distributions, which would constitute more realistic circumstances.

As pointed out in the previous section, strict capacity restrictions have not been considered. It would therefore also be interesting to evaluate the performance of the method when capacity utilization is restricted.

An observation that was made in Paper III is that the arrival times can have a relatively large variance. This could possibly be due to assumptions in the simulation software or it can be an effect of the method. Irrespective of cause, it is an important topic to address in future research as it is unlikely that the operations would be perceived as reliable if arrival times varies largely, even though punctuality would be high. From another perspective, it would also be of importance as reduced variability could allow less margins, which would increase capacity.

The focus in this research has mainly been on the allocation of time supplements, however a topic that would be interesting to explore further could be to apply a similar approach but instead to put focus on the relation between allocated buffer times and delays.

Bibliography

Andersson, E. V., Peterson, A., and Törnquist Krasemann, J. (2015). Reduced railway traffic delays using a MILP approach to increase robustness in critical points. *Journal of Rail Transport Planning & Management*, 5(3):110–127.

Bešinović, N., Goverde, R. M., Quaglietta, E., and Roberti, R. (2016). An integrated micro–macro approach to robust railway timetabling. *Transportation Research Part B: Methodological*, 87:14–32.

Brännlund, U., Lindberg, P. O., Nõu, A., and Nilsson, J.-E. (1998). Railway timetabling using lagrangian relaxation. *Transportation Science*, 32(4):358–369.

Burggraeve, S. and Vansteenwegen, P. (2017a). Optimization of supplements and buffer times in passenger robust timetabling. *Journal of Rail Transport Planning & Management*, 7(3):171–186.

Burggraeve, S. and Vansteenwegen, P. (2017b). Robust routing and timetabling in complex railway stations. *Transportation Research Part B: Methodological*, 110:228–244.

Cacchiani, V., Caprara, A., and Fischetti, M. (2012). A lagrangian heuristic for robustness, with an application to train timetabling. *Transportation Science*, 46(1):124–133.

Cacchiani, V. and Toth, P. (2012). Nominal and robust train timetabling problems. *European Journal of Operational Research*, 219(3):727–737.

Caimi, G., Kroon, L., and Liebchen, C. (2017). Models for railway timetable optimization: Applicability and applications in practice. *Journal of Rail Transport Planning & Management*, 6(4):285–312.

Caprara, A., Fischetti, M., and Toth, P. (2002). Modeling and solving the train timetabling problem. *Operations Research*, 50(5):851–861.

Chow, A. H. and Pavlides, A. (2018). Cost functions and multi-objective timetabling of mixed train services. *Transportation Research Part A: Policy and Practice*, 113:335–356.

de Fabris, S., Longo, G., Medeossi, G., and Pesenti, R. (2014). Automatic generation of railway timetables based on a mesoscopic infrastructure model. *Journal of Rail Transport Planning and Management*, 4(1-2):2–13.

Fischetti, M. and Monaci, M. (2009). Light robustness. In Ahuja, R. K., Möhring, R. H., and Zaroliagis, C. D., editors, *Robust and Online Large-Scale Optimization. Lecture Notes in Computer Science*, volume 5868. Springer, Berlin, Heidelberg.

Fischetti, M., Salvagnin, D., and Zanette, A. (2009). Fast approaches to improve the robustness of a railway timetable. *Transportation Science*, 43(3):321–335.

Harrod, S. S. (2012). A tutorial on fundamental model structures for railway timetable optimization. *Surveys in Operations Research and Management Science*, 17(2):85–96.

Jovanovic, D. and Harker, P. T. (1991). Tactical scheduling of rail operations. The SCAN I system. *Transportation Science*, 25(1):46–64.

Khoshniyat, F. and Peterson, A. (2017). Improving train service reliability by applying an effective timetable robustness strategy. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*, 21(6):525–543.

Kroon, L., Maróti, G., Helmrich, M. R., Vromans, M., and Dekker, R. (2008). Stochastic improvement of cyclic railway timetables. *Transportation Research Part B: Methodological*, 42(6):553–570.

Lamorgese, L., Mannino, C., and Natvig, E. (2017). An exact micro–macro approach to cyclic and non-cyclic train timetabling. *Omega*, 72:59–70.

Lee, Y., Lu, L.-S., Wu, M.-L., and Lin, D.-Y. (2017). Balance of efficiency and robustness in passenger railway timetables. *Transportation Research Part B: Methodological*, 97:142–156.

Liebchen, C., Lübbecke, M., Möhring, R., and Stiller, S. (2009). The concept of recoverable robustness, linear programming recovery, and railway applications. In Ahuja, R. K., Möhring, R. H., and Zaroliagis, C. D., editors, *Robust and Online Large-Scale Optimization. Lecture Notes in Computer Science*, volume 5868. Springer, Berlin, Heidelberg.

Lusby, R. M., Larsen, J., and Bull, S. (2018). A survey on robustness in railway planning. *European Journal of Operational Research*, 266(1):1–15.

Schlechte, T. and Borndörfer, R. (2010). Balancing efficiency and robustness – A bi-criteria optimization approach to railway track allocation. In Ehrgott, M., Naujoks, B., Stewart, T. J., and Wallenius, J., editors, *Multiple Criteria Decision Making for Sustainable Energy and Transportation Systems. Lecture Notes in Economics and Mathematical Systems*, volume 634, Berlin, Heidelberg. Springer.

Schlechte, T., Borndörfer, R., Erol, B., Graffagnino, T., and Swarat, E. (2011). Micro–macro transformation of railway networks. *Journal of Rail Transport Planning & Management*, 1(1):38–48.

Schöbel, A. and Kratz, A. (2009). A bicriteria approach for robust timetabling. In Ahuja, R. K., Möhring, R. H., and Zaroliagis, C. D., editors, *Robust and Online Large-Scale Optimization. Lecture Notes in Computer Science*, volume 5868. Springer, Berlin, Heidelberg.

Sels, P., Dewilde, T., Cattrysse, D., and Vansteenwegen, P. (2016). Reducing the passenger travel time in practice by the automated construction of a robust railway timetable. *Transportation Research Part B: Methodological*, 84:124–156.

Serafini, P. and Ukovich, W. (1989). A mathematical model for periodic scheduling problems. *SIAM Journal on Discrete Mathematics*, 2(4):550–581.

Soyster, A. L. (1973). Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations Research*, 21(5):1154–1157.

Spall, J. C. (2003). *Introduction to stochastic search and optimization: Estimation, simulation, and control*. John Wiley & Sons, Inc., Hoboken, New Jersey, USA.

Szpigel, B. (1973). Optimal train scheduling on a single track railway. *Operations Research*, 72:343–352.

Yan, F., Bešinović, N., and Goverde, R. M. (2019). Multi-objective periodic railway timetabling on dense heterogeneous railway corridors. *Transportation Research Part B: Methodological*, 125:52–75.