Delays and Timetabling for Passenger Trains

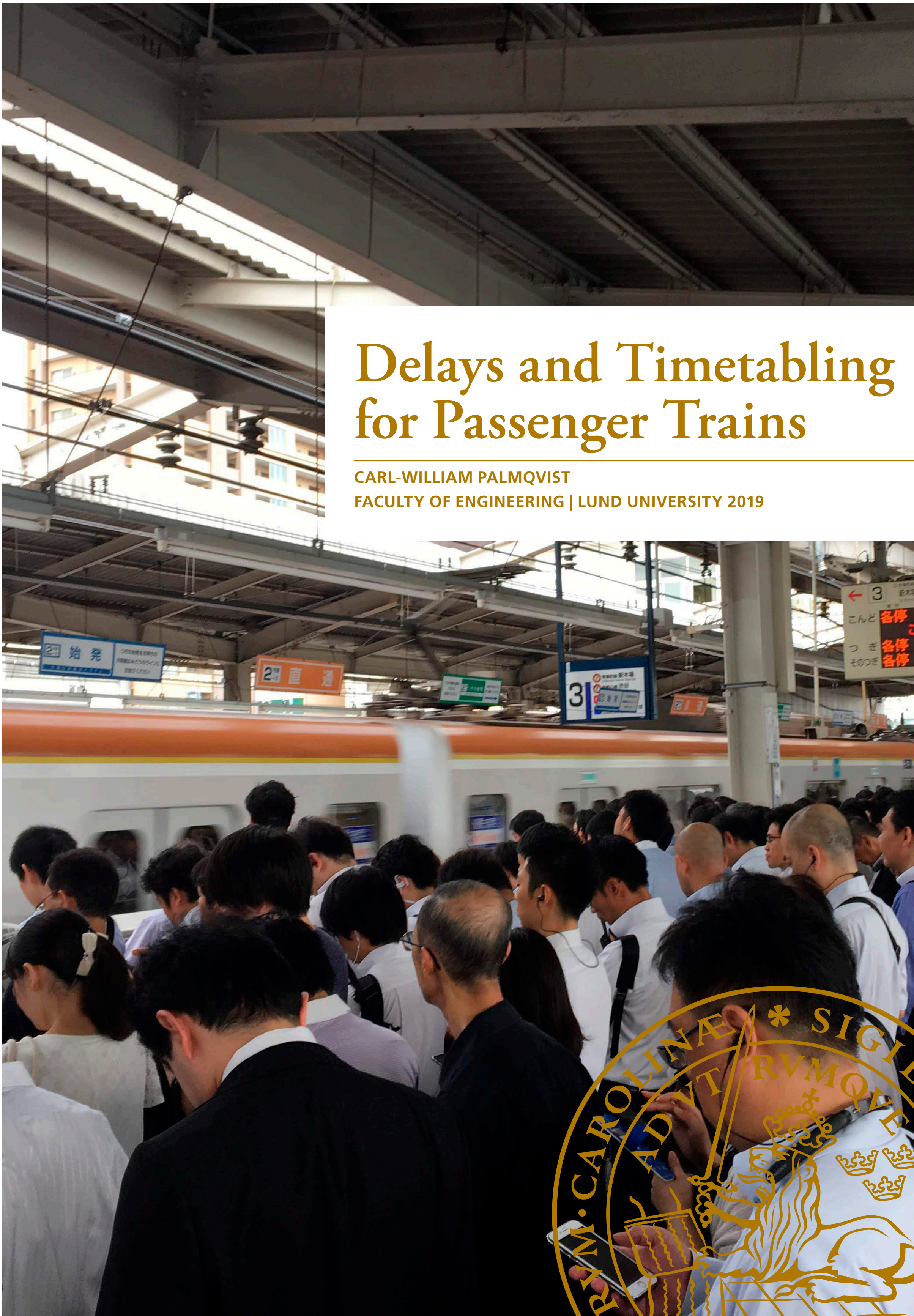
Palmqvist, Carl-William

2019

Document Version:
Publisher's PDF, also known as Version of record

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00



Delays and Timetabling for Passenger Trains

CARL-WILLIAM PALMQVIST
FACULTY OF ENGINEERING | LUND UNIVERSITY 2019

Printed by Media-Tryck, Lund 2019 NORDIC SWAN ECOLABEL 3041 0903

Travel by train has increased steadily for the last 30 years. In order to build trust in and shift even more traffic to railways, more trains must arrive on time. In practice, many train delays are caused by small disturbances at stations, which add up. One issue is that the scheduled dwell times are simply too short. Another is that punctuality falls quickly if it is either warm or cold. A third is that interactions between trains rarely go as planned. One suggestion for how to reduce delays is to paint markings that show where passengers should wait. Another is to remove switches, so that remaining ones can be better maintained. A third is to make railways more resilient to the weather variations of today, and to the climate changes of tomorrow. Cost-effective improvements can also be made with timetables. More of the planning can be automated, so that planners can focus more on setting appropriate dwell times and on improving the timetabling guidelines. This way, many more trains can be on time.

LUND UNIVERSITY

Lund University
Faculty of Engineering
Department of Technology and Society
Bulletin 317
ISBN 978-91-7895-310-3
ISSN 1653-1930
CODEN: LUTVDG/(TVTT-1059)1-105

Delays and Timetabling for Passenger Trains

Delays and Timetabling for Passenger Trains

Carl-William Palmqvist

LUND
UNIVERSITY

DOCTORAL DISSERTATION

by due permission of the Faculty of Engineering, Lund University, Sweden.
To be defended at V:C, V-House. John Ericssons väg 1, Lund.
2019-11-08 at 10:15.

Faculty opponent
Birre Nyström

Organisation
LUND UNIVERSITY

Document name DOCTORAL DISSERTATION

Date of issue 2019-11-08

Carl-William Palmqvist

Sponsoring organisation
K2 – The Swedish Knowledge Centre for Public Transport

Title and subtitle
Delays and Timetabling for Passenger Trains

Abstract

Travel by train has increased steadily in Sweden the last 30 years. The pace has been about two to three percent per year, and we now have twice as many passengers. With growing awareness of the changing climate, the pace is increasing further.

A problem that affects both passengers and businesses in Sweden is train delays. One way to describe these is as the share of trains that arrive less than six minutes delayed. About 90% of trains in Sweden meet this standard and have done so for many years. In a way this is impressive, since there are now many more trains. Unfortunately, this also means that more and more passengers are affected by delays. This leads to irritation, threatens the shift of traffic to railways, and costs a lot for society. More trains must arrive on time.

This thesis shows that delays are mostly caused by small disturbances – up to a minute or two. Over long journeys, these small disturbances accumulate and sometimes cause quite big delays. These delays mostly occur at stations, where the trains stop, but are then unable to continue on time. It is difficult to say exactly what causes these small disturbances, but the time that the trains are supposed to be at stations – the dwell times – are often too short. Another pattern is seen between delays and weather: if it is either warm or cold, delays increase rapidly. And while winter and snow return every year, they still cause major disruptions.

The thesis holds a few suggestions to reduce delays. One is platform markings that show where the trains will stop, where the doors will be, and where the passengers should wait. This is an easy and affordable way to speed up the stops, so that the trains depart on time. Another measure is to remove switches. Then there are fewer parts that can fail, and those that remain can be maintained to a higher standard. A third way is to adapt the railway, so that it better withstands the weather variations of today, and the climate changes of tomorrow. Something that has been done in other countries is to shade and air-condition electronics and signals along the railway. Then the components to not overheat, and more trains run on time.

Many things can also be done with timetables, so that more trains run on time, without a rise in costs. More of the planning can be automated. Then more time can be spent on giving trains appropriate dwell times. Infrastructure managers should also do more to evaluate and improve the rules and guidelines that govern timetabling. In this way we can improve timetables gradually from year to year, with fewer and fewer delays as a consequence. These suggestions do not solve all of the railway's issues, but they would lead to many more trains arriving on time.

Key words Railways, Trains, Delays, Punctuality, Timetable

Classification system and/or index terms (if any)

Supplementary bibliographical information

Language English

ISSN and key title 1653-1930
Bulletin – Lund University, Faculty of Engineering, Department of Technology and Society, 317

ISBN 978-91-7895-310-3 (printed)
ISBN 978-91-7895-311-0 (pdf)

Recipient's notes

Number of pages 105

Price

Security classification

Signature

Date 2019-09-30

Delays and Timetabling for Passenger Trains

Carl-William Palmqvist

Front cover photo by Carl-William Palmqvist

Back cover photo by Kennet Ruona

Faculty of Engineering
Technology and Society

ISBN 978-91-7895-310-3 (printed)
ISBN 978-91-7895-311-0 (pdf)
ISSN 1653-1930

Printed in Sweden by Media-Tryck, Lund University
Lund 2019

NORDIC SWAN ECOLABEL
Printed matter 3041 0903

Media-Tryck is an environmentally certified and ISO 14001:2015 certified provider of printed material. Read more about our environmental work at www.mediatryck.lu.se

MADE IN SWEDEN

To Irene

Table of Contents

Populärvetenskaplig sammanfattning

Resandet med tåg har ökat stadigt i Sverige de senaste 30 åren. Takten har varit cirka två-tre procent per år, och vi nu har mer än dubbelt så många resenärer. Med ökat fokus på klimatfrågan stiger resandet nu ännu snabbare.

Ett problem som drabbar både resenärer och företag i Sverige är tågförseningar. Ett sätt att beskriva dessa är den andel av alla tåg som kommer fram mer än fem minuter för sent. Räknar man så är ungefär 90 procent av alla tåg punktliga. Så har det varit i många år. Det är på ett sätt imponerande, med tanke på att det nu går många fler tåg. Tyvärr gör det också att allt fler resenärer drabbas. Detta leder till stor irritation, hotar den fortsatta överflyttningen av trafik till järnväg, och kostar mycket för samhället. Fler tåg behöver komma fram i tid.

Min forskning visar att förseningarna mest beror på små störningar – upp till någon minut. Över långa resor samlas dessa små störningar ihop och gör att den totala förseningen kan bli stor. De här störningarna sker mest på stationer, där tågen ska stanna, men där de sedan inte klarar att köra vidare i tid. Det är svårt att säga exakt vad dessa störningar beror på, men den tiden tågen ska vara på stationen – uppehållstiden – är ofta för kort. Ett annat samband syns mellan förseningar och väder: om det är varmt eller kallt ökar förseningarna snabbt. Och trots att det varje år blir vinter och snö så leder det ändå till stora besvär.

Jag ger en rad förslag för att minska förseningarna. Ett är markeringar på plattformen som visar var tågen ska stanna, var dörrarna kommer vara, och var resenärerna ska stå. Det är ett enkelt och billigt sätt att snabba på uppehållen, så att tågen kommer iväg i tid. En annan åtgärd är att ta bort växlar. Då finns det färre felkällor, och de som finns kvar kan skötas bättre. Ett tredje sätt är att anpassa järnvägen, så att den tål dagens vädervariationer och de klimatförändringar som är på väg. Något man gjort i andra länder är att skugga och ventilera elektronik och signaler längs med banan. Då blir de inte blir för varma, och tågen går oftare i tid.

Med tidtabeller kan man också göra mycket för att fler tåg ska gå i tid, utan att det kostar mera. Mer av planeringen kan göras automatiskt. Då kan mer tid läggas på att ge tågen lagom långa uppehållstider. Trafikverket bör också göra mer för att följa upp och förbättra de regler och guider som finns för att planera. På så sätt kan vi få tidtabeller som blir bättre och bättre från år till år, med mindre och mindre förseningar som följd. Dessa förslag löser inte alla järnvägens problem, men de skulle leda till att många fler tåg kommer fram i tid. Då får vi plats för ännu fler tåg och resenärer på spåren.

Acknowledgements

One often hears that writing a PhD is a lonely endeavour – but I have never for a moment felt alone during my journey. Firstly, I want to thank my two amazing supervisors: Lena Hiselius and Nils Olsson. Thank you for believing in me, for your inspiring and cheerful pragmatism, and for letting me run loose with your encouragement and light touch. I am very grateful to have shared this journey with you both, and for having had the chance to learn from you. Along the way, I have heard of difficulties and torn relationships between some supervisors and students, but with you, I have always felt encouraged and supported. Thank you! If there is one thing I want to become and accomplish in my career, it is to be a good supervisor and mentor to others, just as you have been to me.

An even greater debt is owed to my dear family. Thank you, Irene, for everything. Thank you for filling my life with both laughter and meaning. Thank you for your love and support. Thank you for sharing your life with me, and for the many adventures yet to come. You're the best, and I love you. Thank you, mom and dad, for providing a wonderful childhood, a stable and supportive home, for all your trust and support, and for watching Onega during so many of my trips. Thanks to Marie-Louise and Cecilia for being such wonderful sisters. Thanks to Oma and Grampa for inspiring me to go deeper into the world of academia and for showing me that things like this can be done.

Thanks to Kenneth Håkansson for taking me on as I wrote my master thesis. For believing in me and this research from the beginning and all the way through. Without you, your support, and your wide-reaching network, this thesis would not exist, and I would never have become a researcher at all. I am eternally grateful, and I look forward to continuing to work with you as we implement some of the recommendations. Thanks to Magnus Wahlborg for taking me under your wing and supporting our little group as we grow into this railway research business. Thank you for your vocal support, for introducing us to the community, for inviting us into projects, and for believing in us.

Thank you to all the reference group members: Martin Aronsson, Bo-Lennart Nelldal, Britt-Marie Olsson, Hans Dahlberg, Magnus Andersson, Elisabet Spross, Bengt Holmberg, Per Konrad, Kristina Abelin, Olov Lindfeldt, and Jonas Westlund. Thank you for all the long and inspiring discussions, all your questions and input, your introductions into your respective organizations. I look forward to broadening the research scope and digging into implementation together with all of you. Thank you, Anders F. Nilsson, for giving me all the data from Lupp.

Thank you to the faculty opponent and to the examination committee for taking the time to read and examine my work. I look forward to having the opportunity of discussing it with you. Thank you to Karin and Anders for reviewing an earlier version of this manuscript, for a stimulating seminar, and for providing hundreds of comments, big and small. It has undoubtedly improved the thesis and made me much less anxious in releasing this version. Thank you both.

Thank you to Emma Solinen, Anders Lindfeldt and Hans Sipilä for welcoming me as a fellow doctoral student of railways when I first began and did not know very much about the topic, or about doctoral studies. Thank you to Jennifer Warg and Ingrid Johansson for the collaboration in Plasa and the friendly discussions about both railways and studies. I am glad to have you as peers, and that we will continue working together. Thank you to Martin Joborn, Anders Petersson, Johanna Thörnquist-Krasemann and all the others who help to make KAJT function and who continue to help make Sweden world-leading in capacity-related railway research. Thank you to Sara Gestrelius and Abderrahman Ait Ali, and to Johan Högdahl and Erik Ronnle for our various discussions on research and on life as a doctoral candidate.

Thank you to all my colleagues and hosts in Japan, for giving me some of the best times of my life. To Norio Tomii, for acting as a host, mentor and a guide. I can only dream of having anywhere close to the impact that you have had on research and practice, and I hope we can write some more papers together. Thanks to Yasufumi Ochiai for making me feel at home in the lab, for teaching me about timetabling in Japan, and for taking me out to eat so many kinds of delicious food! I really appreciate it and look forward to a continued friendship for many years. Thank you to Chikara Hirai for welcoming me at RTRI so many times, for the nomikais, dinners, and inspiring management style. If I become a manager one day, I want to be a lot like you.

Thank you to Takao-san for kindly hosting me in your home and for sharing your delicious cooking with me. I hope we can cook tempura or sukiyaki again one day. Thank you to Wataru Ohnishi for showing me around, and for having one of the most delicious lunches ever, and for your deep and inspiring comments about the joys of working with and learning from your students. Thank you to Takafumi Koseki for welcoming me into your lab at the University of Tokyo, and for showing how someone so esteemed can be so modest and respectful to others. I do not think I have ever seen someone bow quite as deep as you, in all my time in Japan. I hope we can find a way to work together one day. Thanks also to all the railway companies that we visited, and to the staff, management and students at CIT. Thanks to STINT, JSPS and all the fellows of summer program 2018, for providing me the opportunity to visit Japan, and for making it so enjoyable. I will never forget it.

Thank you to Thomas Graffagnino for welcoming me to Switzerland and SBB. It was inspiring to see how you work, and to hear how you do so many of the things that we talk about doing here – big and small – and that they make a difference. Thank you to Francesco Corman for inviting me to your new lab at ETH Zürich. I had a great time with your students and colleagues and hope to maybe one day establish a similar setup in Sweden. You guys do great work.

A big set of thanks to my great colleagues at Traffic and Roads, Technology and Society and K2. An extra shout out to our awesome administrative staff: Mia, Thomas, Carin, Helena, Astrid and Maria. Without you, the entire operation would collapse. You lay an amazing puzzle. I appreciate how you are always on hand and willing to help with just about anything, and how cooperative and constructive you are with managing the finances – a great contrast to how it seems to be at many other organizations. Thank you to Anders, John and Helena. For supporting and believing in me, and for providing such healthy role models as leaders and managers. Thank you to Hanna for helping me with popularising and disseminating my work.

Thank you to Andras for your vast knowledge, unassuming manner, and sheer coolness. Thank you to Ebrahim for sharing your passion for teaching. You are one of the most inspiring teachers I have had, and I hope to one day be able to show some of that same passion and engagement in my own teaching. Thanks to Aliaksei Laureshyn, Till Koglin, Fredrik Pettersson and Fredrik Kopsch for leading the way as researchers and faculty members. It is inspiring to see you establish your own groups and careers, and I am grateful that you are so friendly and open to talking about things, rather than pulling up the ladder behind you. Thank you to Hampus and Maya for what you do in teaching, and for all the lunches and conversations we have had.

Thank you to Erik and Lia for a successful cooperation on the Thredbo-paper. Thank you, Jean, for your friendship, for serving as a role model, and for the chance to work together going forward. I hope it lasts for many years yet. Thanks to Daria, for joining the railway team, and for throwing yourself into the field with such good cheer and enthusiasm! Thank you to Emma and Joel, for your easy-going manner, and for being such good mates about sharing the office, and for not complaining about my loud keyboard. Thank you to Carmelo, Chunli and Chiara for being so approachable, friendly, and open with me and the other doctoral students. Thank you to Magnus Andersson and Christina Lindkvist at MAU for your confidence in me and for our discussions. Thank you to Sadegh. One day we will make something great of that satellite data we keep talking about.

Thank you to Malin, Jens, Alfred and Ulrik for all our conversation both in the lunchroom and at K2. You really help make it feel like a community. Thank you to Carl for our discussions about programming and for your help in the beginning. Thank you to Pajtim for showing the courage to be so outspoken, and for your big role in keeping the pavement-side of things on track at the department. Thank you to Khashayar for always being perfect – jättebra! Thanks to Azreena for giving such an inspiring defence just before my own, and for giving me hope. Thank you, Janet, for your kind questions and to Oksana, for always lighting up the room with a smile. Thank you to Ellen and Valentin, for being willing to cross the corridor and bridge the gap between our divisions. Thanks again to all of you for creating such a nice community.

Thank you to my friends, both old and new, for providing emotional support and a welcome change of scenery. Thank you to Daniel, Hampus and Jocke, for sticking with me for all these years – despite long moves and absences. Thank you to Björn, Rasmus and Eigo, for all the nice meals, games, and conversations during these past years. I treasure you all and am so grateful for your friendship.

And last of all, a very big thank you to the funders of my research. Thank you to Trafikverket for providing the bulk of the funding, for encouraging me and the research, for giving me so many connections, for opening up your doors, for nurturing our railway research here in Lund, and for continuing to fund and support our research going forwards. Without you, none of this would have happened. Thank you also to KAJT for inviting us into the capacity research community, for organizing so many great conferences and events, and for providing a great platform for research and dissemination. Thank you to K2, The Swedish Knowledge Centre for Public Transportation, for funding my small projects and for your continued faith, funding, and support. Thank you for providing such a pleasant place to work, and so many great events, connections, discussions, and much more.

Thank you once again to JSPS, the Japanese Society for the Promotion of Science, for providing the Summer Program that enabled me to visit Japan for the summer of 2018. Thank you to STINT for choosing me as one of the Swedish candidates for the programme. It has opened so many opportunities, connections, and research ideas for me. It was also one of the best times of my life. I will never forget it, and I look forward to coming back many more times. Thank you to Chiba Institute of Technology for hosting me during that first trip, and again for inviting me to return for an academic fellowship so soon. I really appreciate it, and some of the fruits of the fellowship are found in this thesis.

So, as I began to say, four pages ago, I have had plenty of good company along this journey. I am eternally grateful and could not have done it without you. Thank you, and to those many more of you whom I could not mention here. I appreciate your support, confidence, and friendship, and so much more.

Thank you all.

Introduction

This thesis analyses delays and timetabling for passenger railways. Railways are an important part of the transport system. After decades of decline, rail transportation in Sweden has been growing steadily by about 3% annually since the early 1990s (Trafikanalys 2018a), for both passenger and vehicle kilometres. The rise has been predominantly in passenger trains – mostly for local and regional journeys (Trafikanalys 2018a). In 2015 passenger trains made up 83% of all trains and contributed 10% of all motorised passenger kilometres, while freight trains were responsible for 19% of tonne kilometres (Trafikanalys 2018b). Investments in new infrastructure have not grown to the same extent (Trafikanalys 2018a), and a quarter of the metropolitan lines are now very heavily utilised, at levels associated with a high likelihood of delays, low average speeds and little time for infrastructure maintenance (Trafikverket 2017a).

One of the key factors for the attractiveness and efficiency of the railway sector is on-time performance, of which punctuality is a commonly used indicator. In Stockholm, only 56% of commuters report being satisfied with the punctuality of commuter trains, which is the lowest across all public transport options and the single most important influencing factor for their overall satisfaction with the transportation mode (Stockholms läns landsting 2017). The level of punctuality on Swedish railways has been close to 90% for the last several years, with punctuality defined as a maximum delay of five minutes at the final stop (Trafikanalys 2016). This is considered too low by the industry, which has set a goal of 95% by 2020 (Gummesson 2018). While this requires large and rapid investment, the benefits in increased attractiveness and ridership of trains would be considerable.

Even small delays of only a few minutes cause considerable inconvenience to both passengers and operators. In our data of more than 200 million observations across some 7.5 million journeys on the Swedish rail network over the years 2011-2017, delay events of three minutes or less make up 78% of delay time, while those of five minutes or less make up 85% of the delay hours. Small delays thus accumulate and are a very big part of the problem of unreliable railways. Especially so at stations, where 80% of the delay time is made up of delays up to three minutes in size. For passengers, even small delays can lead to missed connections, and to significantly larger total delays, and the punctuality for passengers is often significantly worse than for trains (Parbo, Nielsen & Prato 2016).

High quality timetable planning is one way to deal with small train delays. In most countries, timetables are revised at least once per year, and previous studies (e.g. Goverde 2005; Vromans 2005; Yuan & Hansen 2008) indicate that timetable properties can have a large impact on delays and punctuality. By scheduling the times that trains stop at stations (dwell times) and by adding margins (run time supplements) between stations, timetable planners can affect the risk of delays arising, and enable trains to make up for any delays that do occur, sometimes at the cost of longer travel times. By separating the trains from one another in time (increasing headway or buffer times) they can also reduce the risk of delays spreading from one train to another, this has a smaller effect on travel times, but consumes more capacity and reduces the number of trains that it is possible to run.

Structure of thesis

Following on from this short introductory chapter, the rest of the thesis is structured as follows. The second chapter further describes the *Background*, placing the thesis in a context and with overviews of earlier research on both train delays and timetable planning, and some theory on delays and punctuality. The chapter ends with two identified research gaps. Following on from the identified gaps, the third chapter describes the *Aim* of the thesis, its research questions, underlying hypotheses and delimitations. Next, the fourth chapter describes the *Method and Data* used in the thesis. Chapter Five describes the *Research Process and Results*, how the papers build on each other, an overview of the papers, and tying the findings back to the research questions. Chapter Six is a *Concluding Discussion* with recommendations and some reflections. Following a list of references is a list of the appended papers and my role in those publications, followed by the papers themselves.

Background

Timetable planning

Extensive research has been done on timetable planning in the past. In his overview of timetable research, Hansen (2009) identified some key differences between the state of the art and the state of practice: using precise and realistic estimates of dwell, run and blocking times, and reflecting the variations in behaviour of train crews. In this section a brief overview of the research on timetable planning is made. First, we consider some work with broad perspectives: (1) different ways to consider and assess timetable quality, (2) the strategic trade-off between precision and slack, and (3) some broader perspectives on timetable planning. Then we will discuss more narrow research dealing with the allocation of (4) run time margins, (5) dwell times, and (6) headways. This is followed with some more research and theory on delays and punctuality.

Quality of timetables

There are many different dimensions to the concept of timetable quality (see, for instance, Goverde & Hansen 2013 and Gestrelius, Peterson & Aronsson 2019). For passengers it is desirable to have short travel times, high reliability, high frequency in both peak and off-peak, easy and synchronised connections, and a timetable that is easy to learn (e.g. Parbo, Nielsen, & Prato 2016; Kottenhoff & Byström 2010; van Hagen & van Oort 2018). For drivers and other personnel, it is desirable to have good working hours and reliability, and to be able to start and finish the day in the same place; for operators, it is good to match supply to demand and to have a high resource utilisation (Ceder 2001). For the infrastructure manager it is also important to have time left in the timetable for maintenance (see, for instance, Lidén 2018). On a higher level, the infrastructure manager must also balance different aspects of capacity. A popular illustration of this is found in Figure 1, copied from the International Union of Railways (2004), with four dimensions: the number of trains, stability, heterogeneity and average speed.

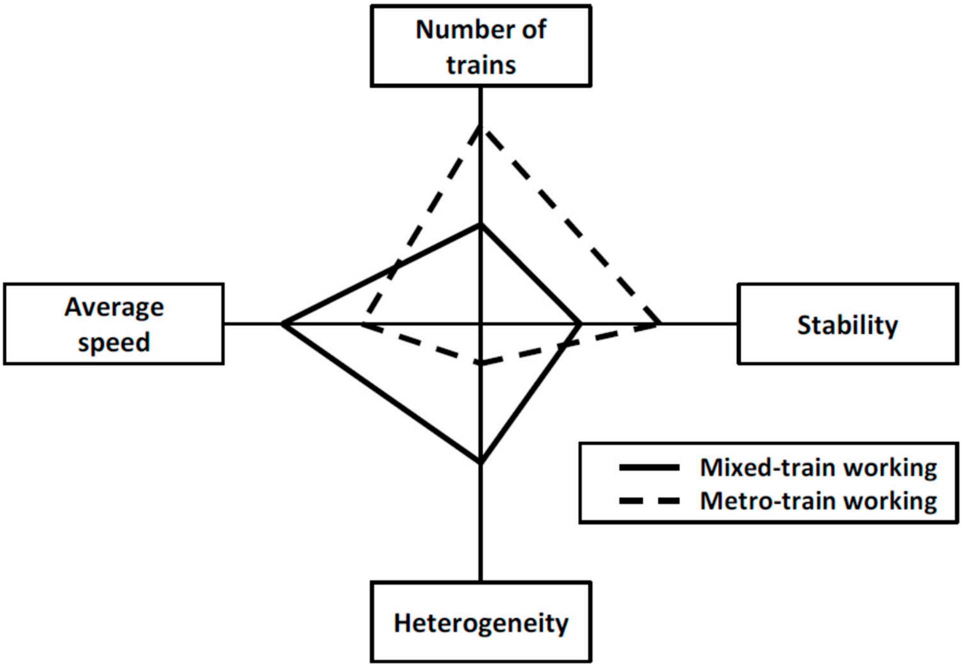


Figure 1
In timetable planning, this figure is a classic demonstration of the balance that must be struck between different dimensions of capacity: the number of trains, the stability, heterogeneity, and average speed. The figure illustrates schematically how mixed (or conventional) rail services compare to metro services, with higher speeds and more heterogeneity at the cost of fewer trains and less stability. Stability refers to reliability, robustness and related concepts, roughly understood as the absence of delays. Heterogneity refers to the variation in types of trains operated in the system. Figure from International Union of Railways (2004)

The role of planners

To an extent, timetable planners can and must balance these dimensions, but on the other hand, they are mostly bound by the requests made by the train operating companies. Each of these companies makes their own plans, making the trade-offs they prefer, and the overall timetable becomes a mix of all these different companies. The infrastructure manager can mainly influence this outcome by setting conditions in the Network Statement, and by providing rules, regulations and guidelines for how timetables are to be planned. Individual planners do not have much influence over the overall balance – their job is to follow the guidelines and to do their best to accommodate the requests of the operating companies. On the margin, however, they can and do sometimes use their discretion to balance the different parameters.

Robustness of timetables

The most relevant dimensions for this thesis are reliability and robustness – avoiding and recovering from delays – or stability, as it is referred to in Figure 1. Even these have multiple aspects, however: (1) to avoid systematic delays by having appropriate and realistic run and dwell times (Hansen 2009), (2) to make up any delays that do occur by ensuring that there are margins both at and between stations (Cerreto et al. 2016), (3) to avoid delays spreading between trains by ensuring that there are sufficient headway and buffer times (i.e. Carey 1999; Khoshniyat & Peterson 2015), (4) to maintain options for dispatchers to reschedule and reroute trains, so that they can more easily reduce the impact and spread of delays (Gestrelius et al. 2012).

Trains and passengers

Passengers often place a higher value on travel time reliability than on the travel time itself. Both delays and punctuality are also often significantly worse when considered from the perspective of the passengers rather than the trains, partly due to variations in passenger demands and loads, and partly due to missed connections. In their review of the literature on railway timetable planning with a passenger perspective, Parbo, Nielsen and Prato (2016) stated that the difference in punctuality can be as high as 10%. Rietveld (2007) wrote more on how and why supply-oriented indicators systematically give a more favourable picture than demand-oriented ones that take the passengers into account.

Passenger perspectives

Some research that explicitly takes the passengers into account has been carried out by Dollevoet (2013), in a doctoral thesis on delay management and dispatching, which examines passengers and their rerouting choices. Another doctoral thesis was presented by van der Hurk (2015), on how information on and to passengers can be used to improve rail services when there are delays. Cheng and Tsai (2014) also focused on passengers, studying which factors affect the perceived waiting time for passengers during train delays – factors that can make this time more or less tolerable. A final example discussed here is by Batley, Dargay, and Wardman (2011) using econometric models on combined demand and delay data in the UK. They find that while delays are indeed given a high value relative to travel time – and annoy passengers – they have little effect on the travel demand, in either the short or long term.

While the passenger perspective is thus important, this thesis is written from the perspective of the infrastructure manager. A precondition for any further optimisation efforts is that the timetable should routinely be executed as scheduled, or at least adhere closely to the plan. Delays ruin this by causing disruptions to passengers, train operating companies and maintenance works, and much more.

Timetable precision

Olsson et al. (2015) identified two high-level strategies in railway operations and timetable planning, that aim to achieve robustness and reliability: precision and slack. A strategy based on precision contains, as Hansen (2009) indicated, both realistic and precise estimates of run times, dwell times and headway times. In this approach, considerable effort is put into identifying and allocating these times, while little would be added in ways of margins or buffers. This requires, in turn, that considerable attention is paid during maintenance and operations, so that neither the infrastructure nor the rolling stock malfunction, and in ensuring that dispatchers, train drivers and on-board staff, as well as passengers, all behave in a disciplined and timely manner. The timetable and operations are streamlined to be efficient during normal operations, but they have less flexibility during major disruptions. The Japanese railways are prime examples of this strategy (e.g. Yabuki et al. 2018).

Timetable slack

The other high-level strategy Olsson et al. (2015) identified is slack. This approach is based on the notion that things inevitably go wrong, and that it is best to be prepared for when they do. In timetables this is achieved by ensuring that trains have large margins – with which to make up delays – and that there are sizeable buffer times scheduled between trains to limit the spread of delays between them. In operations it is achieved by ensuring that there is a reserve capacity of both rolling stock and train crews, and in infrastructure by creating many possibilities for rerouting, both at stations and across the network. The drawback of this approach is that it is costly to maintain the reserve capacity, and that travel times are extended – which is costly for both passengers and operators. Carey (1998) made a related point: as more time is allowed for an activity, there is a behavioural response to make the activity take a longer time. By extending the time to accommodate for the variation, the variation is also increased. This effect can reduce, and potentially eliminate, the increase in reliability from added slack.

Heterogeneity in timetables

Another important dimension in railway timetables and operations is heterogeneity (International Union of Railways 2004), see Figure 1. This expresses the variation in services run, often regarding the speeds or headways. A metro line where all trains run in the same way and stop at the same stations, and with equal headways, is a very homogeneous system. The Swedish conventional railway network where high speed, regional and local passenger trains all mix together with freight trains, and there are extra services during rush hours, is a very heterogeneous system.

Implications of heterogeneity

Different indicators and implications of this heterogeneity can be found in at least two other recent doctoral theses. Vromans (2005) dedicated a chapter to the issue and introduced a family of indicators centred on the Sum of Shortest Headway Reciprocals (SSHR), which has since become popular in the literature. He also listed five different ways to reduce the heterogeneity, focusing on the option to equalise the number of stops among train services. Lindfeldt (2015) instead evaluated performance using empirical data, much like we do in this thesis, using the ratio between the 95th and 10th percentiles of speeds used across the network as an indicator of heterogeneity. Both authors found, broadly speaking, that delays and heterogeneity are correlated, as delays spread more easily between trains in a heterogeneous system.

Planning and learning

While they did not write about timetable planning, Argyris and Schön (1996) presented an interesting concept. They considered learning as a process of understanding and eliminating the gap between the expected result and the actual result of an action. The gap can be eliminated by taking corrective measures within the existing values and norms – which they call single-loop learning. It can also be closed by changing the existing values and norms – double-loop learning. This perspective can be applied to timetable planning: single-loop learning means finding ways to apply the current rules, guidelines and policies, while double-loop learning implies choosing better rules, guidelines and policies. In a sense, this thesis is an attempt to achieve and support double-loop learning.

Planning in a wider context

Timetable planning fits into a broader context in several ways. Writing about public transportation more generally, Ceder (2001) described the planning process in four steps: network route design, timetable planning, vehicle planning, and crew planning. In his thesis, Watson (2008) instead considered the contrasting needs and preferences of timetable planners and their managers in an organisational context. He found that the privatisation of British Rail had a negative effect on the timetabling process, due to a poor planning and rushed implementation of the new structure. Avelino, te Brömmelstroet & Hulster (2006) compared the politics of timetable planning in the Dutch and Swiss contexts and concluded that "timetable planning is not merely an operational process to be left to engineers or economists" (pp. 20).

Capacity allocation

Timetable planning is one part of a larger process to allocate capacity on the railways. The process in Sweden is described in the Network Statement (Trafikverket 2015a). From this we have adapted Figure 2, which illustrates the process – with seven main steps. The train operating companies (1) begin by submitting requests for the timetable slots they want for the next year. The timetable planners at the Transport Administration (2) combine these requests into a draft timetable that contains all the trains for the whole year. If there are any conflicts between the requests, there is first (3) a process where the parties are encouraged to coordinate amongst themselves. If they are unable to do so, the Transport Administration (4) tries to settle the dispute in dialogue with the parties. If these attempts are also unsuccessful, the relevant parts of the network are (5) declared to be saturated, and (6) a set of prioritisation criteria is used to determine which requests have priority. The timetable is then (7) finalised and published. Not pictured is the option for dissatisfied train operating companies to appeal the decision to an administrative court. The court cannot change the timetable but may impose sanctions which lead to a change in practice for future years. This process is largely the same in most European countries, following EU regulations (e.g. DB Netz 2017, ProRail 2016, Ali & Eliasson 2019).

Prioritisation

The main differences across European countries occur when the voluntary coordination process fails, step (6) above, which happens several times every year. When requesting capacity in Sweden, each train must be classified into one of about

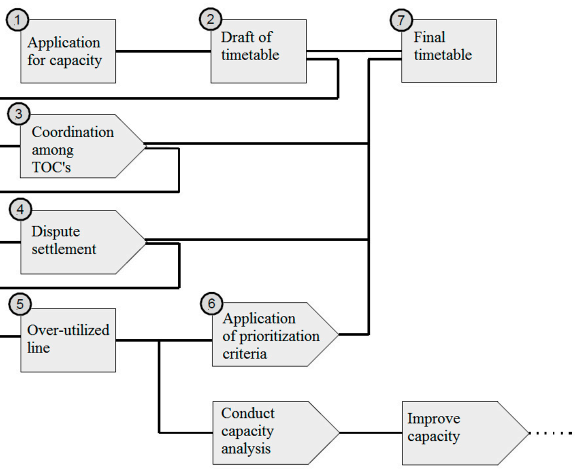


Figure 2
Capacity allocation process in Sweden. Adapted from Trafikverket (2015)

30 categories, based on the expected number and types of passengers (local or regional, business or leisure, etc.) or freight. These categories, as well as associations between trains, are given different social cost estimates, based on the methodology in Trafikverket (2018a). The solution which minimises the social welfare costs is then accepted for the final timetable. This approach has some issues, both in accurately estimating the social welfare costs, and in combining publicly subsidised commuter services with commercial long distance services (see Ali, Warg & Eliasson 2019).

Other countries have chosen alternative approaches. The United Kingdom has a more qualitative process with an overarching objective and a list of twelve evaluation criteria (Network Rail 2018). In Germany (DB Netz 2017), priority is first given to regular-interval or integrated network services, cross-border trains, and train paths for freight trains. In remaining conflicts, priority is given to the trains that would pay more in track charges. In the Netherlands (ProRail 2016), a cyclic hourly pattern is prioritised, and further conflicts are settled by increasing the track charges to the point that only one actor remains interested – a form of auction.

Ad hoc process

The process of timetable planning itself is thus mainly performed in stage (2), with final touches in (7). Once the annual timetable has been published, train operating companies can and do make requests for changes in the so-called ad hoc process. These are mostly for freight trains, where the demand for transportation often fluctuates to a greater extent (Trafikanalys 2018a). For passenger trains, many of the adjustments are in response to maintenance works on the railway that only apply for part of the year and are often announced at later stages. During a year there can be over 80,000 such changes, which are performed on a first come first served basis (Trafikverket 2019).

Dispatching

After planning, dispatching takes over as the very important activity of managing the actual operations of trains and maintenance works on a day-to-day basis, at 3.00pm every day. Dispatchers control the signals and switches, coordinate with train drivers and working crews via radio, and ensure that the operations are safe and on time. The topic of dispatching is very large, with a rich research literature of its own. See Lamorgese et al. (2018) for a recent overview. Some further, recent examples are Andreasson, Jansson and Lindblom (2018) on the complex collaboration between drivers and dispatchers, Sandblad, Andersson and Tschirner (2015) on the use of support systems, and Gholami and Törnquist Krasemann (2018) for a heuristic mathematical approach to solving rescheduling problems in real time. Finally, consider Ghaemi et al. (2018), studying the impacts of rescheduling on passenger delays. This literature is, however, largely beyond the scope of this thesis. As is that on the operation of the trains themselves.

Robust timetable planning

This section describes three commonly used ways to ensure that a timetable is robust enough to cope with delays: run time supplements (often referred to simply as margins), dwell times, and headway or buffer times. These three concepts are illustrated, together with run times, in a graphical timetable in Figure 3. The Swedish norms are summarised in Table 1.

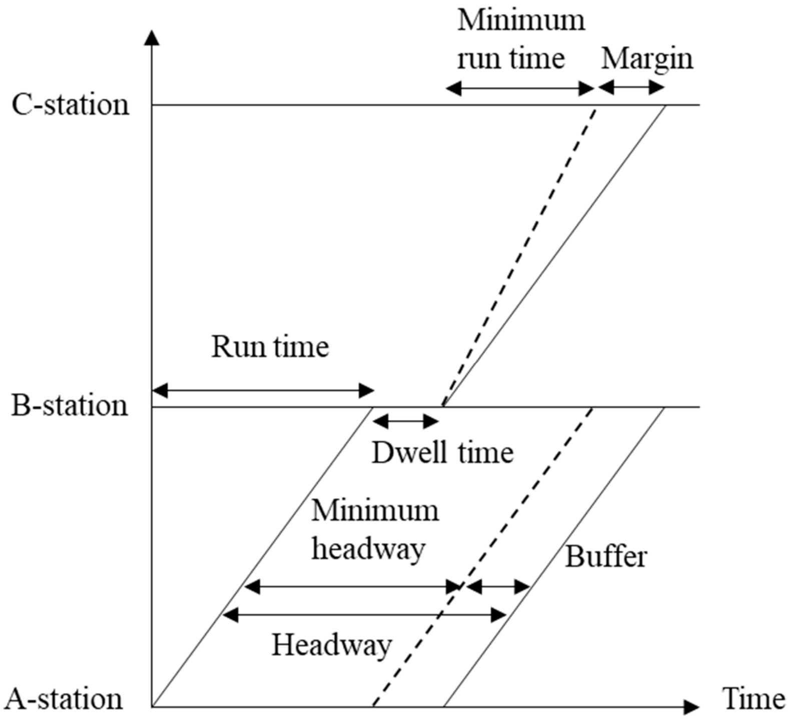


Figure 3
An illustration of run times, dwell times, headways, buffers and margins in a graphical timetable Each of the sub-sections reviews some relevant literature and presents the current norms and regulations used in Swedish timetable planning. These guidelines are summarised in Table 1.

Table 1
Timetable planning standards for passenger trains in Sweden

Robustness indicator	Norm in the Swedish regulations
Run time supplement	+ 3% across the board, included in run time calculation (Banverket 2000)
Node supplement	3 - 4 minutes run time supplements to be added per pair of nodes passed, 2 minutes if partial. 2 - 4 nodes exist per railway line (Trafikverket 2015b)
Dwell time at stations	2 minutes standard, 1 minute if the number of passengers is small (Banverket 2000)
Headway	At least 2 - 7 minutes, most commonly 3 - 5 minutes (Trafikverket 2017b)

Run time supplements

Size of margins

One of the most common and important ways to make a timetable robust as regards delays is to include margins in the form of run time supplements. These supplements are added to the calculated minimum run times (see Pachl 2002) to create flexibility which can be used either by the driver or dispatcher to reduce delays and maintain connections between trains. To an extent, these supplements are mandated by the International Union of Railways (2000): three to seven percent added to the minimum running time, plus an additional one and one and a half minutes for every 100 km. These levels are common in Europe, while six to eight percent are common in North America (Pachl 2002). In Sweden we use a base level of three percent while both the Dutch (Goverde 2005) and Swiss (Vromans 2005) automatically add margins of seven percent.

Node supplements

In Sweden (Trafikverket 2015b) and Switzerland (Vromans 2005), planners also add supplements at or between a set of strategically important locations – so-called nodes. This is one way to distribute the 90 seconds per 100 km suggested by the UIC. In Sweden, every railway line has about two to four of these nodes, and the timetable planner must allocate three or four minutes between each pair of adjacent nodes, depending on the type of train (Trafikverket 2015b). If the train only travels part of the distance between two nodes, the supplement should still be at least two minutes. In the United Kingdom, run times are based on previous performance rather than on calculations, and the supplements are thus harder to define (Rudolph 2003).

Discretion in planning

In addition to these two methods, timetable planners in Sweden use their discretion when assigning time supplements. One common practice is to add seconds so that the arrival times at stations where the train stops occur at whole minutes. For instance, if a train is meant to arrive at 12:44:27, the planner might add 33 seconds, so that the arrival instead occurs at 12:45:00. Over long journeys, these can add up. Supplements are also sometimes given for trains that are scheduled to stop at a platform which is not on the main track, because this takes slightly longer to get to, and because engineering works are being done on the track, requiring lower speeds for part of the journey (Banverket 2000). These are meant to correct cases where the run time calculation is known to be wrong, and they are not considered as margins in this thesis.

Allocation of margins

The efficiency of margins can be improved by distributing them in a good way. This has been studied by a number of authors, using different methodological approaches. For instance, Cerreto et al. (2016) presented an empirical study on the quality of run time supplement allocation in timetables, with regard to how trains make up or increase delays during journeys. To help quantify the distribution of margins along a journey, Vromans (2005) introduced the Weighted Average Distance (WAD). He also attempted to find the optimal distribution, using optimisation methods on both hypothetical and real cases, drawing the conclusion that a slight shift towards the beginning was best. Using similar methods, Vekas, van der Vlerk, and Haneveld (2012) also found that it was best for delay recovery *not* to use a uniform distribution, given some assumptions of the delay distributions.

Solinen, Nicholson, and Peterson (2017) used a combination of empirical, simulation and optimisation methods to introduce and consider a more detailed indicator, based on critical points – points in a timetable when one train enters after a preceding train, or when one train overtakes another. These points are very sensitive to delays, both because existing delays tend to increase in these cases, and because any delays here easily spread to other trains. By ensuring that there are enough margins at these points, the allocation of time supplements can be very effective.

Dwell times

One of the key issues brought up by Hansen (2009) was to use realistic and precise estimates of dwell times, and that this is often not done – either in practice or research. Peterson (2012) found evidence of this for train services in Sweden – that the dwell times were usually underestimated, without being sufficiently compensated by margins on the line. One good example of how both precise and realistic dwell times can be estimated was provided by Buchmueller, Weidmann and Nash (2008) who modelled actual dwell times in Switzerland, breaking them down into five sub-processes: (1) unlocking doors, (2) opening doors, (3) boarding and alighting, (4) closing doors, and (5) train dispatching. They then used over three million observations for calibration, to ensure that the estimates were realistic. D'Acierno et al. (2017) focused their modelling on (3) boarding and alighting, and on how this depended on the congestion and flows of passengers.

Some earlier studies

Along similar lines, there have been many publications that deal with simulations of passengers moving between the train and the platform to study various kinds of passenger management strategies (e.g. Heinz (2003) in a Swedish context, and Baee et al. (2012), Kamizuru, Noguchi & Tomii (2015), Seriani & Fernandez (2015), Zhang, Han & Li (2008) for some international examples). Focusing more on realism than on the precise breakdown of sub-processes, Pedersen, Nygreen and Lindfeldt (2018) studied how actual dwell times vary over time and running direction in Norway. Li, Daamen and Goverde (2015) did something similar with track occupancy data from short intermediate stops in the Netherlands and separating between peak and off-peak hours. Other examples are provided from Italy by Longo and Medeossi (2012), who focused on simulation, and again from the Netherlands by Kecman and Goverde (2015), who were interested in real time prediction of both dwell and run times.

Swedish guidelines

The Swedish guidelines (Banverket 2000) state that dwell times for passenger trains should, in general, be scheduled to be two minutes long. Sometimes longer scheduled times are required, and other times, if the number of passengers is small and the train and station are prepared for a speedier boarding process, one minute can be used instead. If the number of passengers is very small, the guidelines state that it is possible to schedule a stop without dwell time, merely slowing the train down to a stop and then starting again immediately, but if this is done the run time on the next line section should be extended, and if passenger numbers increase, the timetable should be redrawn and longer dwell times set. In general, however, it is often difficult to know precisely how much time is required for stops at stations – and as a consequence, margins are not discussed or defined as explicitly for dwell times as they are for run times.

Headway times

The last aspect of detailed timetable planning that we will consider in this overview is headways, although they have not been studied explicitly in the thesis. Headways are the times that separate trains using the same infrastructure, and they are mostly relevant on double track lines. Sometimes there is a distinction between headway times, as the minimum time that is technically possible, and buffer times, which make up any additional buffers separating the trains from one another, see Figure 3. This is not done consistently, however, and in practice it is often difficult to separate the two.

Knock-on delays

In a paper on ways to measure timetable reliability, Carey (1999) paid special attention to so-called knock-on delays. If trains are very close to one another in time, delays to one will quickly spread to any following, connecting or meeting trains, and these delays are sometimes called knock-on (secondary) delays. By increasing the separation between trains, a buffer is created so that delays do not spread as easily, and Carey (1999) proposed that a number of indicators describing the distribution of headways in a timetable could be used to estimate the reliability of a given timetable. Yuan and Hansen (2008) found that as the scheduled buffer time between trains decreased, the knock-on delays increased exponentially.

Improving robustness

Also arguing for an increase in scheduled headway or buffer times, Nelldal, Lindfeldt and Lindfeldt (2009) performed simulation experiments on existing high speed (200 km/h) trains in Sweden. They found that the punctuality for these trains would improve by 5-10 percent if the minimum headways were increased to at least five minutes. Similarly, Dewilde et al. (2013) introduced a method to increase timetable robustness in complex stations by maximising the minimum headway time between a given set of trains – one of the heuristics proposed by Carey (1999). They found that this approach improved the robustness in the station zone of Brussels by eight percent and reduced knock-on delays in the area by half. In Sweden, minimum headways vary from two to seven minutes, depending on the location, but are most usually between three and five minutes (Trafikverket 2017b).

Train delays and punctuality

This section describes how delays can be defined, counted and deconstructed before turning to some earlier research on delays and punctuality.

Delay definitions

Delays can be considered and measured in a number of ways. The most common is arrival delay, which is defined as the difference between a train's realised and scheduled arrival time at a certain station. Closely related is the departure delay, the difference between the realised and scheduled departure times. Both arrival and departure delays can be either positive or negative, indicating early arrivals or departures, respectively. Let t^s and t^r denote the scheduled and realised times, and d the delay:

$$d = t^r - t^s \quad (1)$$

Another way to consider delays is to recognise that the delay upon arrival and departure are the result of delayed processes: either the run time between stations, or the dwell time at stations (or in some cases the depot).

The run time is the time it takes to run between two adjacent stations, A and B, and can be calculated as the difference between the arrival time (*arr*) at station B and the departure time (*dep*) at station A. This can be calculated both for the realised and scheduled times.

$$t_{run,AB} = t_{arr,B} - t_{dep,A} \quad (2)$$

The run time delay is then the difference between the realised and scheduled run times, which is equivalent to the difference between the arrival delay at B and the departure delay at A.

$$d_{run,AB} = t^r_{run,AB} - t^s_{run,AB} = d_{arr,B} - d_{dep,A} = t^r_{arr,B} - t^r_{dep,A} - t^s_{arr,B} + t^s_{dep,A} \quad (3)$$

The dwell time is the time that a train spends at a station other than the first or final station. It is the difference between the arrival and departure times.

$$t_{dwell} = t_{dep} - t_{arr} \quad (4)$$

As with run times, the dwell time can be calculated both for realised and scheduled times, and the dwell time delay is the difference between these two times.

Equivalently, it can be specified as the difference between the arrival and departure delays at the given station.

$$d_{dwell} = t^r_{dwell} - t^s_{dwell} = d_{dep} - d_{arr} = t^r_{dep} - t^r_{arr} - t^s_{dep} + t^s_{arr} \quad (5)$$

These times, delays and equations are illustrated and exemplified in Figure 4.

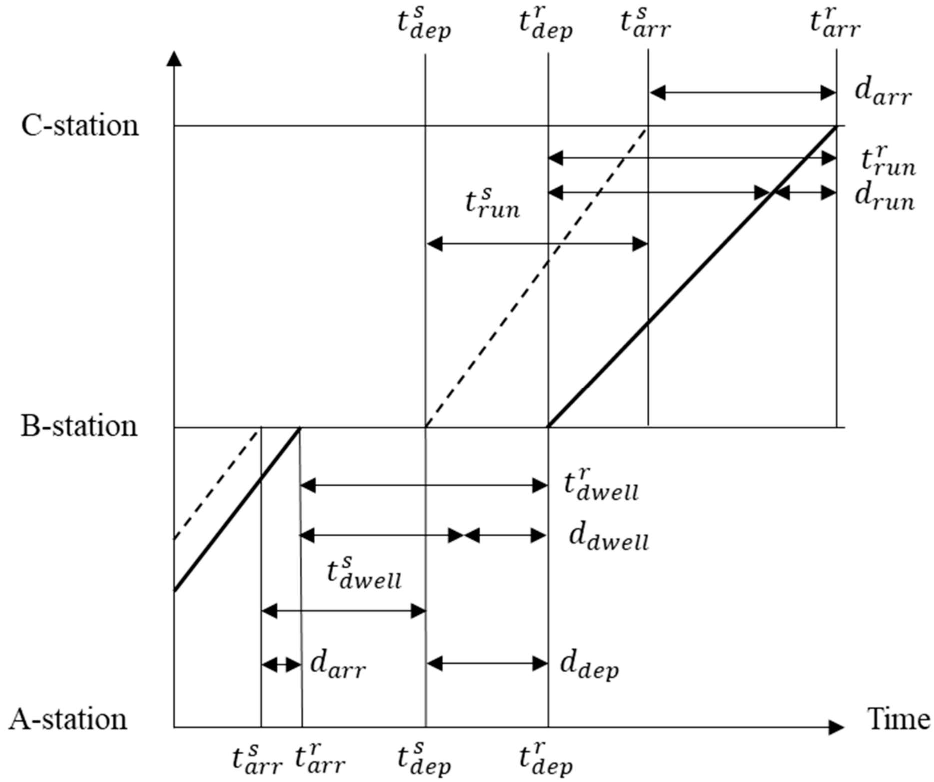


Figure 4
An illustration of arrival, departure, run and dwell times, and the associated delays. The diagonal lines represent the train paths, with the dotted line denoting the the scheduled path, and the bold line the realised path. Scheduled and realised arrival and departure times are denoted by the vertical lines, while run times, dwell times and delays are illustrated by the horisontal arrows. The scheduled run and dwell times have been pictured twice, to help illustrate the size of the run and dwell time delays.

Early arrivals

A possible objection to the definition above is that that passenger trains are often not allowed to depart early from scheduled stops. Thus, a train that arrives ahead of schedule must wait for longer, so that the following dwell time is necessarily extended, without the train necessarily departing with a delay. In some cases, this might be considered as less problematic, and not considered as a delay, but rather as some different category of time. The same could be said for run times that are extended for trains that run ahead of schedule. From the producer and timetable planning perspectives taken in this thesis, however, even these deviations from the timetable are considered problematic – because the trains are consuming capacity (occupying tracks and platforms) where and when they should not be.

Other perspectives

For transportation that does not follow timetables, delays are instead often discussed in the frame of travel time variability (i.e. Noland & Polak 2002; Rietveld, Bruinsma & Van Vuuren 2001). This is commonly the case for road-based transportation. The baseline can then be the time required in free-flowing traffic during good conditions, while factors like congestion and poor weather lead to longer travel times, which can be considered as a form of delay. When there is no timetable with which to compare, this is often the only alternative. Freight trains in the USA are traditionally run without timetables (Gorman 2009), and American conceptions of train delays are thus often quite distinct from those in Eurasian contexts. The realised run time is instead compared against the baseline of a free-flowing run time, without any stops or delays.

Earlier research on delays

Punctuality is one of the most important factors for a railway system (Gummesson 2018) and its passengers (i.e. Stockholms läns landsting, 2017) and it directly affects the competitiveness against other transport modes (Nyström 2008). This section briefly highlights some earlier research on train delays and punctuality, from a few different perspectives relevant to this thesis.

Data and delays

With increasing access to large volumes of data, researchers can study delays and their causes on a relatively detailed level. One overview of the many IT systems, sub-systems and databases in the Swedish railway industry was provided by Thaduri, Galar and Kumar (2015). They called for more researchers and practitioners to apply big data analytics to make use and sense of these large troves of data. Several different approaches have been applied successfully. In the USA, Gorman (2009) applied econometric models on freight train data with the aim of predicting delays due to congestion. Wallander and Mäkitalo (2012) used a data-mining approach to analyse train delays more generally. Another approach made by Marković et al. (2015), was to analyse the relation between train delays and various characteristics of the railway system using machine learning models.

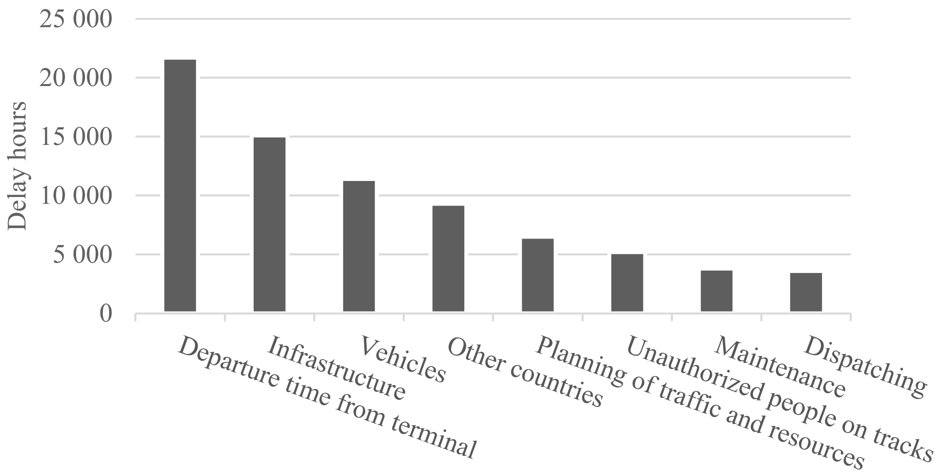


Figure 5
Delay hours by a number of causes. Adapted from (Trafikverket 2018c)

Delay cause coding

In practice, delays have often been studied using manually reported delay causes, see for instance Veiseth, Olsson and Saetermo (2007) and Trafikverket (2018b). Figure 5 illustrates data from the latter on the eight biggest causes of delays that are monitored and targeted by the main railway actors in Sweden, as they are currently categorised in the industry. However, manual attribution of delays is prone to errors and often quite inconsistent (Nyström 2008), with an estimated reliability around 80% (Nilsson et al. 2015). Importantly, only delays that are three minutes long, or more, are coded – smaller delays are not included in these statistics or datasets.

Weather and delays

Changing perspectives, the influence of weather and climate change on train delays and punctuality has received increasing attention in the literature. Some ten years ago, Koetse and Rietveld (2009) found that "studies that investigate the effects of weather or climate change on rail transport and infrastructure are scarce" (pp. 212), but they cite two studies suggesting that weather was behind about 10% of failures and accidents. Since then, however, more studies have been published.

One good example is by Xia et al. (2013), who estimated how wind, temperature and rain cause delays in the Dutch railways, mainly by damaging the infrastructure. Similar work has been done by Nagy and Csiszár (2015), who highlighted the effects of weather conditions on the punctuality of Hungarian passenger trains, and by Xu, Corman and Peng (2016) who analysed the disruptions in the Chinese high-speed railway and found that almost 90% of these were due to bad weather.

Some authors focus on more specific conditions, such as Ferranti et al. (2016), who studied how heat causes failures in the railroad infrastructure in England, particularly in the signalling systems. In a similar paper, Ferranti et al. (2018) found that a short heatwave in the UK led to more than a doubling of delays, due in a large part to speed restrictions, and they caution that such events will become more common. Another example of the opposite scenario is by Zakeri and Olsson (2017), who investigated the impact of weather on the punctuality of local trains in the Oslo area. They found strong correlations between punctuality and temperatures below minus 7°C and snowfall of at least 15 cm.

Also relevant for winter conditions Palin et al. (2016) described how seasonal forecasts of the North Atlantic Oscillation, a pressure differential related to either cold and calm or mild and stormy winters, can help predict the scope of disruptions to both railways and other transport modes in the UK, months in advance. Continuing on the use of weather forecasts, in his doctoral thesis Wang (2018), proposed methods to incorporate these into dispatching, estimating with the help of simulations that this could reduce train delays in the UK by about 20%.

Liu et al. (2018) instead wrote on the susceptibility to heavy rain and related hazards of Chinese railways. Frauenfelder et al. (2017) evaluated the vulnerability of Norwegian roads and railways to extreme weather events, also focusing on heavy rainfall and debris flows such as rock falls and avalanches, both now and in the future. Similarly, Sa'adin, Kaewunruen, and Jaroszweski (2016) evaluated the weather and climate risks facing a planned high-speed rail connection between Malaysia and Singapore, concluding that heavy precipitation – along with associated debris flows – is the biggest risk factor.

Infrastructure and delays

Another frequent and related cause of delays is infrastructure failure. Veiseth et al. (2007) linked infrastructure data with delay and punctuality data to study the infrastructure's influence on rail punctuality. They reported that some 30% of delay hours in Norway were caused by infrastructure failures and suggested that the quality of punctuality data could be improved by connecting it with infrastructure and operational databases. Stenström et al. (2015) developed a composite indicator for benchmarking and monitoring of rail infrastructure, considering four factors: failure frequency, train delays, logistic time and repair time.

Also in a Swedish context, Wiklund (2006) studied the relationship between preventative maintenance of the infrastructure and severe disruptions and delays for trains – identifying that the overhead line is the most vulnerable and critical component, but also that this is difficult to make more robust using only increased maintenance. Mattsson and Jenelius (2015) provided an overview and a discussion of the research on vulnerability and resiliency in transportation networks more generally. Ferranti et al. (2016) studied heat-related infrastructure failures in Southeast England, while Hawchar et al. (2018) discussed the vulnerability of critical infrastructure with regard to climate change, and Dobney et al. (2009) discussed the delays caused by rail buckling, to name but a few.

Trains and delays

Traffic density can also be correlated with delays, as trains interfere with one another and can cause delays to spread between trains. For instance, Olsson and Haugland (2004) found that the management of train crossings is a key success factor on single track lines. The previously mentioned study by Gorman (2009) also showed that the number of crossings, passes and overtakes consistently had high impacts on delays for freight trains in the USA, although delays are measured differently on the American freight railways. The International Union of Railways (2004) has developed a method to calculate the capacity utilisation across railway networks. This method has been adapted by the Swedish Transport Administration for use on its network (Trafikverket 2017a), and it shows that many lines are very highly utilised, with severe congestion as a consequence.

Station stops and delays

The last perspective we will cover is delays occurring at stations. For instance, Wiggenraad (2001) studied seven Dutch train stations in detail and found that the realised dwell times were longer than scheduled, that the scheduled dwell times were the same at both peak and off-peak times, and that passengers concentrated around platform access points. Yuan and Hansen (2002) also studied delays at Dutch stations and found that the mean excess dwell time was around 30 seconds. Sometimes this was due to the train having arrived early and not being permitted to depart before the schedule, but at other times it was due to a lack of discipline among train drivers and conductors.

Yet another study around Dutch stations, by Nie and Hansen (2005), found that trains there operated at lower than designed speeds, and that realised dwell times at platforms were systematically extended beyond what was scheduled, both because of other trains blocking their routes, and because of the behaviour of train personnel. Similar findings have been made elsewhere in the world. For instance, Harris, Mjøsund and Haugland (2013) studied delays at stations in the Oslo area and claimed that these delays were often small, poorly recorded, and not well understood. On the opposite side of the world, in New Zealand, Ceder and Hassold (2015) also found that one of the main delay mechanisms was increased actual dwell times, caused by heavy passenger loads that were not sufficiently compensated by increasing the scheduled dwell times.

Research gaps

In the background, we have gone through some of the work that has been done on train delays and timetable planning. This overview reveals two gaps, which this thesis aims to fill.

The first is that there are few empirical studies describing multiple causes of small delays, especially in Sweden. While there has recently been an increase in the amount of railway operations research in Sweden, much of it is focused in one way or another on simulation or optimisation. Empirical elements have been quite rare and rather limited in scope. In the international literature, the volume of empirical research is larger and increasing, however, it is typically either focused on a narrow range of factors correlated to delays, such as high temperatures, storms, or congestion. Studies that do consider multiple types of factors are usually based on manually reported causes of delay, which is problematic because of human error, and the omission of small delays. There is thus a need for more broad, empirical studies on delays that do not rely on manually reported causes.

The second gap is that there are relatively few studies that evaluate the effects of timetable planning in practice. There is certainly a sizeable literature on timetable planning, with a large segment of it building on an operations research framework with mathematical modelling and optimisation, and another big part that is based on computer simulations. Fewer authors have been concerned with evaluating the effects of the timetable planning that is done in practice. Some qualitative work has been done about the situation of planners, but quantitative and empirical work identifying the different strategies, policies and decisions made by planners is rare.

Aim

The overarching aim of the thesis is to: increase the understanding of delays that occur for passenger trains in Sweden, in order to reduce delays, primarily through improved timetable planning. This can be conceptualised as an instance of double-loop learning in timetable planning, as illustrated in Figure 9 on page 83. To break the aim down into more manageable pieces, we have created five research questions which mirror the research gaps identified in the previous section.

Research questions

RQ1. How are the delays distributed, in a broad sense of the word?

RQ2. What factors correlate with delays, and to what extent?

RQ3. How can the allocation of run time margins be improved?

RQ4. How can the allocation of dwell times be improved?

RQ5. How can the practice of timetable planning be improved?

Delimitations

Infrastructure manager's perspective

This thesis is primarily written from the point of view of an infrastructure manager responsible for coordinating and scheduling railway services. The perspectives of passengers and operators are important and have a lot to offer, but they are not the focus of this thesis. This means that we do not, for instance, delve into the issues of generalised travel time, giving delays different weights based on their lengths, or consider the extent to which passengers adapt to delays that occur, or to their justified demands for information during disruptions. Instead, the focus is on being able to schedule and manage traffic in an efficient and reliable way. The benefits for the passengers follow naturally: if the trains are on time, the passengers will be as well. Similarly, the perspective is pragmatic and empirical, rather than focusing on achieving mathematically optimal solutions. Being able to create realistic timetables that the trains can reliably stick to is a precondition for later optimisation.

Conventional passenger trains

The thesis is focused on trains with passengers, not freight. There are vast differences in both the timetabling and delays for these two types of transport, and we have chosen to focus on passenger trains. The demand for transportation of freight is much more variable than for passengers, and timetables for these trains are often created or adjusted at a much later stage. Even then, there are often large deviations between how long and heavy a train is scheduled to be, and what is actually used This makes the timetable unrealistic, and large deviations from it are frequent. This results in a split, with many freight trains having very large delays, while many others depart and arrive long before they are scheduled. The focus is on conventional passenger railways, with varying degrees of heterogeneity, not on metro, light rail or tramways.

Focus within train paths

In relation to timetable planning, the thesis is focused on dwell times and margins within each train path, not on aspects such as headways, buffer times and heterogeneity. These are also important, as they affect the spread of delays between trains, but they are beyond the scope of this thesis.

Other delimitations

The work also does not consider explicitly the influence of dispatching, maintenance, or the physical design of the railway on delays. These are all important fields, which do relate to delays, but they are also beyond the scope of this thesis. We are ambivalent as to whether timetables should be cyclic or not, although the thesis should be relevant even for such cases.

Method and Data

Quantitative, qualitative and mixed methods

This thesis includes both quantitative and qualitative methods, and this section describes some of the benefits of each approach and of combining them as we have done. The emphasis has been on quantitative methods, used in Papers 1, 2 and 5. Purely qualitative methods were used in Paper 4, while Paper 3 was written using mixed methods. This is illustrated in Table 2.

Table 2
Methodological approach across the five papers

Methodological approach	Paper 1	Paper 2	Paper 3	Paper 4	Paper 5
Quantitative	X	X	X		X
Qualitative			X	X	

Quantitative methods allow us to describe what happened. In the research field of railway timetabling and operations, quantitative methods have historically tended to fall into one of three main approaches: optimisation, simulation, and empirical. Many examples of these have been described in the chapter *Background*.

With qualitative methods such as interviews, it is possible to gain insight into the values, priorities, perspectives and perceptions of relevant people and actors. In this thesis, timetable planning plays a key part, and so interviewing timetable planners is a natural and important part of the process. The qualitative work included in this thesis draws mostly on the qualitative research interview approach of Kvale (1997).

Mixing both methods provides the best of both worlds: the quantitative data informs and reinforces the qualitative aspects, and vice versa (Johnson and Onwuegbuzie 2007). For instance, when studying timetable planning, the quantitative methods also allow us to see what decisions planners have made, which is a good complement to what is written in the guidelines or stated by the planners in interviews. Connecting this to the large sets of train movements enables us to see the consequences of their decisions, and to evaluate which of their strategies and decisions work well in practice, and which ones do not, in a form of triangulation (Fellows and Liu 2015).

Working with delays and punctuality

This section describes the different ways that delays and punctuality have been considered and measured throughout the papers included in this thesis. A brief summary is presented in Table 3, with Paper 1 considering both run and dwell times, Paper 2 considering punctuality measured at all scheduled stops, Paper 3 measuring punctuality at the final stop, and Paper 5 returning to dwell time delays. As Paper 4 used purely qualitative methods, it is not included in this section. Considering a range of different approaches in this way provides a more holistic description of the issues of delays and punctuality and helps to ensure that the findings are not sensitive to the exact specification of the specific indicator.

Table 3
Different indicators of delays and punctuality used in the papers

Indicator of delays	Paper 1	Paper 2	Paper 3	Paper 5
Run time delays	X			
Dwell time delays	X			X
Punctuality at final destination			X	
Punctuality at all scheduled stops		X		

Measuring and observing delays

The analyses in Papers 1 and 5 are based on run time delays and dwell time delays. These are the differences between the actual and scheduled process times, be they movements along line sections or station stops. Please refer to *Train delays and punctuality* for more detailed definitions and discussions on delays.

Relative and absolute effects

In Paper 1 we use the terms *average delay* and *relative risk*, with the latter indicating how much a factor increases or decreases the average delay in *relative* terms. This was done to normalise as regards the duration of activities (run or dwell time), as delays associated with bad weather, insufficient margins, or reduced traction power, for instance, might be expected to have relative rather than absolute effects. Other delays, perhaps due to a faulty signal or switch, might be better modelled as having absolute impacts, but this was left for subsequent papers. The relative risk of a delay is calculated as the average delay conditional on some value of an explanatory value, divided by the average delay across all station stops or line sections. This makes it easier to discern the effect of factors (e.g. the level of run time supplements, the scheduled dwell time, or the amount of precipitation) where the values or differences are small in absolute terms. The analysis of this paper also mentions *delay contribution.* Here we multiplied the average delay by the number of observations and compared this to the total amount of delays.

Issues with manual coding

We do not use manually reported causes of delays in any of the papers, for a number of reasons described in the section *Train delays*. Most of the delays considered in the papers are also small, up to one or two minutes, whereas the causes are only reported for bigger delays of at least three minutes in size. In the data from the Swedish Transport Administration, 55% of all delay time is too small to be categorised, and in Paper 5 this figure was even higher.

Whether to normalise or not

To ease comparisons between the differing lengths of line sections, between the speeds of trains, and so on, in Paper 1 we chose to use ratios instead of absolute time units. Throughout we chose to use the scheduled duration as the denominator. If the scheduled duration is 120 seconds and the total margin for that activity is 12 seconds, we registered that as a margin of 10%. In the train movement data, times are only given in whole minutes, which we simply converted to 60 seconds – the margins in the timetable were, however, presented in seconds. For durations, of run times or dwell times, we normalised with the average duration for the respective activity. In the sampled data those averages were approximately 100 seconds for station stops and 197 seconds for run times on line sections. A scheduled stop of two minutes was thus translated to a duration of 1.20 while a scheduled movement along a line section of two minutes would translate to 0.61. These ratios were then rounded to limit the number of distinct values, and to ensure that there were enough observations in each bin for the average delays to be reasonably stable. In Paper 5, which focused on dwell time delays, there was much less variation in the scheduled dwell times, so this normalisation was not required.

Varying punctuality definitions

The need for aggregation

The data volumes across networks are often very large, which makes computations across individual delay observations quite demanding and difficult to perform. Considering individual delays also makes it difficult to follow trains during their journeys, and to gain a holistic picture of what happens during these journeys. For these reasons, it can be useful to study punctuality. This is an aggregate indicator, useful for providing an overview of delays across entire journeys, many trains, and entire networks. It is easier to work with than delays: the number of observations is much lower. It is also a well-known metric to both passengers and practitioners. However, it can be defined in different ways.

Delay thresholds

In Sweden the convention is to use a threshold of five minutes, so that trains with a delay of up to and including five minutes is considered punctual, but those with delays of six or more minutes are considered unpunctual (Trafikanalys 2019). In other countries, the threshold can be on different levels, e.g. one or three minutes (SBB 2018).

Measurement locations

Where punctuality is measured can also vary. The traditional method in Sweden has been to use the final stop, which we did in Paper 3. This was considered quite relevant for long-distance trains mostly serving the end-markets, but less so for regional and especially local trains, which can have many important stops and passengers who only travel part of the journey, and where the most important stations are often in the middle of the journey, although this varies from region to region. To address this, organisations like Trafikverket, Skånetrafiken and SJ have shifted to measuring punctuality at a wider range of stations. The most inclusive version of this is to include all stops, which we did in Paper 2, while more restrictive versions only include major hubs.

Mathematical notation

In mathematical terms, the punctuality of train i's arrival at station j is a function of its arrival delay $d_{i,j}$ and the punctuality threshold k such that:

$$p_{i,j,k} = f(d_{i,j,k}, t) = \begin{cases} 1, d_{i,j,k} \leq t \\ 0, d_{i,j,k} > t \end{cases} \quad (6)$$

Punctuality $P_{j,k}$ at station j in the time period k is then calculated as the weighted average of the punctuality $p_{i,j,k}$ of n arrivals of train i during period k (which can be any given day, week, month, or year), and the weights w_i (usually set to 1).

$$P_{j,k} = \frac{\sum_{i=0}^{n}(p_{i,j,k} * w_i)}{\sum_{i=0}^{n} w_i} \quad (7)$$

The inclusivity of the punctuality indicator depends primarily on the setting of j: this can be a specific station, the final stop for each train i (Paper 3), all stations where the train i has scheduled stops (Paper 2), or all points where arrivals are registered.

Channel precision

This last option, to include stations or control points where the trains do not stop at all, is less relevant to passengers, because they are not affected by this. From a traffic-management perspective, however, there might be some benefits in measuring how well the trains stick to their paths even barring any stops, but then even early arrivals should be identified. In Sweden such an indicator is sometimes used and discussed, called *kanalprecision*, or channel precision. The exact definition is that the train is allowed to be at most three minutes late, and at most two minutes early, to be considered as being inside the channel. The trains can then be measured continually, as often as the signal system allows, and an aggregate indicator can be calculated for a whole train journey, a part of the network, or a period of time.

Weighted punctuality

Of course, it is also possible to consider different trains, so that i only covers passenger trains and not freight trains, or only commuter trains, or trains from a specific company. The weights are rarely used, so that by default w_i equals one. If passenger data is available, however, it is possible to set the weights based on the number of passengers getting on or off, so that trains and stations with more passengers are given higher weight.

Treatment of cancellations

In both Papers 2 and 3, we considered cancellations as unpunctual. The alternative with cancelled trains is to exclude them entirely from the calculation and to instead present them separately in an indicator known as regularity (considering only the proportion of trains that complete their entire journey, without taking delays into account). In the data used for this thesis, about one percent of trains are cancelled, and excluding them from the punctuality figure would thus give an apparent improvement in punctuality of about one percentage point. By counting cancelled trains as unpunctual, the impact of any adverse weather conditions, for instance, should be captured in the analysis. Overall, however, this has made little difference to the results.

Summary

Thus, while the exact definitions can and do vary depending on who is counting and why, punctuality is thus a convenient and commonly used way to aggregate delay data to present a much simpler and more holistic figure, which can be easily discussed and evaluated over time.

Many other indicators exist – such as the maximum deviation, the time to recover, and the deviation area (Nicholson et al. 2015) – and can be found throughout the literature, but punctuality is certainly one of the most common.

Quantitative data used

This section describes the five sets and sources of quantitative data which were combined and used in Papers 1, 2, 3 and 5. Table 3 summarises which datasets were used in each paper.

Table 3
Overview of quantitative data used in the papers

Quantitative data	Paper 1	Paper 2	Paper 3	Paper 5
Train movements	X	X	X	X
Timetables	X	X	X	X
Weather	X	X		
Infrastructure		X		
Passengers				X
Capacity utilisation	X			

Train movement data

Scope of data

Papers 1-3 contain Swedish data for the year of 2015. As a pilot study, the first paper only used data on one regional line, with about 363,000 train movements. The later Papers 2 and 3 instead consider all – approximately 32.4 million – movements in the national network. Paper 5 draws partly from the same data source, for the period of 2011-2017, but only covering the commuter trains in and around Stockholm, approximately 16.6 million movements. The paper also includes corresponding data from a commuter railway company in Tokyo from 2013-2018, with about 63.7 million train movements. These figures are summarised in Table 4.

Table 4
Overview of train movement data used for the papers

Paper	Underlying train movements	Years	Comment
1	363k	2015	Regional line in Sweden (Skånebanan)
2	32.4M	2015	Sweden, nationwide
3	32.4M	2015	Sweden, nationwide
5	16.6M, 63.7M	2011-2017, 2013-2018	Stockholm, Tokyo

Structure of data

In Sweden, this data is structured so that one row covers the departure from one station (A) and the arrival at the next station (B), with both scheduled and realised times for each. Both Norwegian and Japanese train movement data are instead structured so that one row covers both the arrival and departure at one station (A). Yet another structure is found in Danish data: one row there covers only the arrival, departure, *or* passing time at a given station (A). There are thus several different ways to specify this kind of data, although they are functionally equivalent: from all of them it is possible to restructure the data and to calculate the arrival and departure delays. It is also possible to calculate the scheduled and actual durations of both dwell times at stations and the runtimes between them. The stations are not necessarily stations where the train stops – in Swedish the term is instead *driftplatser* – but also include old stations that are no longer serviced, technical stations where trains can meet, and such. The distance between these stations varies across the country. Good maps of the lines and stations can be found at Trafikverket (2018d).

Precision of data

The core of the data in this thesis is made up of train movements registered by the signalling system. The system includes track circuits that detect the presence of trains and sends a message when a train enters or exits the circuits. These messages are very precise, of the order of milliseconds, and include both the scheduled and observed times. However, the signals and track circuits that send timestamps are usually located at the edges of the station areas, rather than at the middle of the platform, where the train stops. To adjust for this, an automatic adjustment is made within the signalling system, to account for the time that it takes for a train to move from the signal to the middle of the station, or vice versa, usually of the order of 10-20 seconds. In Sweden, this adjustment does not differentiate between trains with different characteristics, and the size of the adjustment is not routinely recalibrated.

Truncation of data

This causes some imprecision in the timing of train movements. For historical reasons, and partly to cover for this imprecision, the data layer Lupp (Trafikverket, n.d., 2018b) that receives the messages from the signals (sent via UTIN) truncates the data to the minute level. Thus, while the messages have a precision of milliseconds, the data stored only contain minutes. This complicates the processing and interpretation of the data. For one, it is difficult to study activities which take a short amount of time, as well as small delays. One must also be very cautious when studying small samples or individual observations and be aware that a deviation of ±1 minute can be caused by fluctuations of ±1 second, and that errors of ±30 seconds are to be expected.

Usefulness of data

With large enough samples and systematic effects, however, even small variations can be captured using these data. Consider that something systematically causes a delay of ten seconds, then about one sixth of the trains will register as having been delayed by one minute, while five sixths will appear to not be delayed at all. Looking at the individual trains, this would be misleading, but averaging across the sample, the average will be a delay of ten seconds.

Interpretation of data

In summary, this imprecision and truncation leads to some difficulties in interpreting the train movement data. The data are thus not suitable for precise studies of individual trains with small numbers of observations. However, the errors are reduced when considering the duration of activities – be they run or dwell times – rather than arrivals or departures, and they tend to cancel our over-large numbers of observations. And when considering large samples, even small effects on the level of seconds can, if they are systematic, be captured.

Operational variables

The first three papers also contain what we call operational variables. These include the distance travelled, the number of movements for each train set, the number of days a certain train service is run, the number of trains running, and train movements, a certain day, the number of interactions between trains at and between stations, the number of trains arriving at a certain station within a given hour, and other estimates of capacity utilisation at stations. One of these variables is the number of interactions between trains, defined as instances when more than one train is present at one station, or moving along the same line section, in the same direction, at the same time. These operational variables have all been derived from the train movement data described above and provide additional information.

Dwell times and passenger counts

Alternative datasets were used to conduct more detailed studies of dwell times in Paper 5, from commuter trains in Stockholm and Tokyo. Both these alternative datasets enabled more precise dwell time studies, with a connection to the passenger flows and without the issues of truncation or adjustment to timing points. Such data are quite rare, and often considered a trade secret. Cross-referencing against train movement data improves the reliability of the observations. By further limiting Swedish data to the same time periods as in the Japanese, the analysis was performed over very comparable sets of high-quality data.

Swedish data

About an eighth of the trains in Stockholm are equipped with an automatic passenger counting system, with sensors in the doors to detect the boarding and alighting of passengers at stations. This system also logs arrival, departure and dwell times with a precision of seconds. The data used in the paper span 5.5 million observations over the years 2013-2017, where one observation represents one door at one station stop. While this does not cover all doors of all trains at all stations, over time it gives a good picture of the passenger flows and dwell times. The sensors need to be calibrated from time to time, and they are not very precise for large flows of passengers – these large flows are relatively rare, however.

Japanese data

For the trains in Tokyo, the paper utilised a dataset of manual dwell time and passenger count observations. At least twice a year the railway company manually counts passengers on trains at stations, to estimate the level of on-train congestion. This is mandated by the Japanese government, as a way to monitor the peak congestion of trains – a matter of considerable public interest. As such, the observations are centred on the biggest stations, during the busiest times, where the levels of congestion are the highest. From about 6:30am to 9:00am, staff at a number of stations (eight of which are done regularly, the others sporadically) observe all trains that stop or pass by, making note of exact arrival and departure times, the train number and the number of cars, as well as the estimated congestion rate on the train. Approximately 50 trains are observed in this manner, per station, and over the years 2013-2018, more than 4,000 observations were made.

Connecting the Swedish data

On a practical level, there is no explicit way to connect the Swedish data with the train movements: there is no train identification number in the detailed observations. As the timestamps are based on the doors opening and closing, rather than entering or exiting the station area, they are also different between the two datasets. Instead, we created an algorithm which matched the two datasets together, based on the origins and destinations, locations and timestamps of the observed trains. Because these trains run with high frequencies, we had to be conservative in the matching process, and overlook trains which deviated from the timetable by more than two minutes. In this way, we could be confident that the datasets were matched together correctly. The cost was a reduction in the number of observations included, but the resulting number was still in the thousands, high enough for the needs of the analysis.

Connecting the Japanese data

As the measurements in Tokyo are made manually, some errors can occur. Fortunately, these data did allow for an explicit connection to the train movement data. By cross-referencing these two datasets, it was possible to identify and exclude suspicious observations, where either the signalling system or the human observer could be at fault, again to ensure that only reliable observations were considered in the analysis.

Timetable data

The main purpose in using these detailed timetables was to calculate how much of a margin has been added, and where it was placed. The size of the margins can be expressed in two ways: as a percentage of the scheduled run time without margins, and as seconds per kilometre, the first one being more common.

Scope of the data

Timetables for trains in Sweden are created and stored by the Swedish Transport Administration in the tool TrainPlan. We have received data dumps from this system by the Swedish Transport Administration spanning 2011-2018, and in Papers 2 and 3 we use data from 2015, covering almost 46,000 distinct timetable versions and over 1.1 million train journeys totalling about 32.4 million train movements. To give an overview: 83% of these were for passenger trains, 14% freight trains, and 3% service trains. Forty-three percent of journeys were longer than 100 km, 31% shorter than 50, and 25% between 50 and 100. These data include roughly 80,000 changes that have been made to the timetables during the ad hoc-process (Trafikverket 2019).

Structure of the data

Each timetable variant comes with a calendar reference, a long string which specifies on which days during the timetable year it applies. These long strings can be translated into dates, which can then be matched to the train movement data.

In addition to the calendars, information is given on the level of trains, train arcs, and link usage. Link usage roughly corresponds to train movements, although the infrastructure is slightly more detailed. Train arcs are series of links which have some common properties, such as top speed, train length, and train type. For instance, a train might change configuration during its journey, and the displayed train number identifier might also change during the journey.

The infrastructure model also deviates somewhat from that in the train movement database. Many stations exist in both datasets, but the timetable data have some additional detail, with more timing points. These must be aggregated, so that all margins are counted.

Contents of the data

The time stamps in the timetable are specified in seconds, both for arrival, departure and dwell times. Time supplements are listed explicitly, broken down across a few categories: mainly for performance, adjustments, maintenance works and phasing against other trains. Time supplements for maintenance works were not considered as margins, as they compensate for longer run times, but the three other types have been included and aggregated as margins. Performance and adjustment supplements are the most common by far, phasing supplements are uncommon.

Unfortunately, there is no explicit difference made between dwell times that are required for a stop, and those that are intended to act as margins. Similarly, while it is possible to calculate the headways between trains using the timetables, there is no explicit way of separating between what is the technical minimum, and what is the additional buffer time.

Processing the data

To measure the distribution of margins within a timetable, we use the indicator of Weighted Average Distance (WAD) described in Vromans (2005). This is used to describe how the various time supplements in a timetable are balanced, being more towards the beginning or end of the journey, or in between. It is expressed as a value between 0 and 1, with lower values expressing a shift towards the beginning and higher values a shift towards the end of the journey.

In many timetables, there are instances of negative margins: cases where the scheduled time has been manually set to be shorter than the technical minimum, the effects of this have been studied in Papers 2 and 3. Finally, variables such as the travel time without margins, the average speed of the trains, and the average distance between stops, can be derived from the timetable data and add further information.

Weather data

Papers 1 and 2 used weather data from the Swedish Meteorological and Hydrological Agency (SMHI). From their website we were able to download all historical observations of snow depth, temperature, wind strength, precipitation in Sweden, which we then used to estimate the weather conditions in which the trains operated.

Scope of the data

There are more weather stations in the southern and more populated parts of the country, so that the distance between train stations and meteorological stations differs from place to place. The average and maximum distances for each type of observation are summarised in Table 5. While the stations are by no means perfectly matched in geography, the weather is often quite similar on the scale of 10-20 km, which is the typical range. As with the train movement data, however, the data are better suited for analyses across large samples, rather than individual observations, where the imprecisions in the data can cause problems.

Table 5
Distance between train station and meteorological station across the country

Distance to station	Precipitation	Snow depth	Temperature	Wind speed
Average (km)	10	19	16	21
Maximum (km)	44	97	50	91

Processing the data

The weather data in these papers were linked to the punctuality of trains, which is an aggregate indicator for trains that often travel long distances, through varying weather conditions. There are several ways in which to convert these different values to one single variable.

With wind, we were interested in the highest speeds and chose to take the maximum. With temperature, we tried both the average, the minimum and the maximum. We ended up choosing the minimum temperature for cold weather, and the maximum temperature for hot weather, arguing that we are most interested in the extremes, but the choice made little difference in the analysis. For precipitation and snow depth, we considered the average, maximum and sum of the measured variables. In the end, we found that the sum best explains the effect of precipitation on punctuality, while for snow depth, we found the average to work very well. These steps were not necessary in Paper 1, where we studied delays and could simply use the closest observations in both time and space, without aggregation.

Temperature is measured about 18.7 times per day and station, on average. The average of these was taken to get a daily temperature value for each station. Wind strength was measured at fewer stations, with an average of 23 observations per day. To convert to a daily wind value for each station, we took the maximum value for each day, because we are mainly interested in stronger winds. Snow depth and precipitation is measured daily, but with data missing on average 9% and 0.5 % of the days, respectively.

Connecting the data

To match the datasets, we first had to link the train stations to the nearest meteorological stations, for each of the four types of data, transforming the GPS coordinates of the meteorological data to the SWEREF99 coordinates used for the railway stations. The matching was done separately for each weather variable, because not all meteorological stations observe the same variables. As some stations lack observations on some days, the algorithm was set to match the two station sets for each day, to ensure that an observation could always be given.

Infrastructure data

From the Swedish rail asset management database, BIS, we have information on the type and location of eight categories of infrastructure elements. These are signals, switches, bridges, tunnels, level crossings, cuttings, embankments and fences, all in all 82,700 elements. As a pilot study in investigating this data, all but 1,500 of the elements could be matched to the railway stations and links found in the train movement data. While there are some issues with the reliability and updating of the data, knowing the number of switches or signals, for instance, that trains pass by can help explain part of their delays, and to explain part of the phenomenon that trains travelling longer distances often tend to have more delays. And while the exact locations of the elements may not be easy to map to the train movement data, the data can be used to describe the varying complexity of infrastructure across different parts of the network. In the future, information might be added to the infrastructure data based on the age or condition of the elements, and this could then be used to further help explain delays that occur.

Analytical methods used

This section describes how the data have been pre-processed, and the different analytical methods we have used. An overview of the different categories and variables is presented in Table 6, along with their units, and in which papers they have been included.

Table 6
Overview of influencing, or explanatory, variables studied in the papers, and their units.

Category	Variable	Unit	Papers
Weather	Precipitation	Millimetres	1, 2
	Snow depth	Centimetres	1, 2
	Temperature	Degrees Celsius	1, 2
	Wind speed	Metres per second	2
Timetable	Average speed	Kilometres per hour	2
	Distribution of margins	Percentage	2, 3
	Dwell time	Percentage, seconds	1, 5
	Negative margins	Binary	2, 3
	Number of stops	Count	2
	Run time	Percentage	1
	Size of margins	Percentage	1, 2, 3
	Supplements following stops	Seconds	3
	Travel time	Hours	2
Operational	Capacity utilisation	Percentage	1
	Number of days operated	Count	2
	Distance travelled	Kilometres	2
	Earlier delay	Seconds	1, 5
	Line interactions	Count	2, 3
	Month	Count	1
	Movements per day	Count	2
	Movements per vehicle	Count	2
	Passenger count	Count	5
	Station interactions	Count	2, 3
	Trains per station and hour	Count	2
	Weekday	Count	1
Infrastructure	Bridges	Count	2
	Cuttings	Count	2
	Embankments	Count	2
	Fences	Count	2
	Level crossings	Count	2
	Signals	Count	2
	Switches	Count	2
	Tunnels	Count	2

The basic approach

Throughout Papers 1-3 & 5, the basic approach has been to identify and quantify the link, or correlation, between various delay-variables and other explanatory variables. An important distinction here relates to the terms of correlation, covariation and causation. In these papers, we do not put forward causal mechanisms, or attempt to describe the course of events leading up to a particular delay for a given train. It is not feasible to do this with such large numbers of observations. It is difficult, even when considering individual cases, to disentangle various causes and events from one another. A faulty switch might lead to delays, but the fault might be because of poor maintenance, because of bad weather, because of a flaw in the design or manufacturing process, because of a train passing through it at too high speeds, or with wheel defects, or because of a combination of any or all of these and more.

Causation and correlation

Rather than trying to sort out what, precisely, was the cause of events in cases like these, we have stuck to the level of correlation and covariation. While we cannot say what, precisely, caused the delay described above, we might, by using many observations, conclude that delays are more likely when trains pass by many switches, when the weather is bad, when the maintenance of either the infrastructure or the vehicles is not up to par, and when the drivers do not respect the speed limits, for instance. The variables we consider are thus in many cases proxy-variables for what is happening on the ground. While this is not perfect, and it can sometimes be difficult to disentangle the effects of some variables from one another – such as the effect of travelling long distances from the effect of passing many signals – following this approach can point us in the direction of problematic areas, and sometimes suggest mechanisms that lead to delays and possible measures that can be taken to reduce them. A schematic view of this be seen in Figure 6.

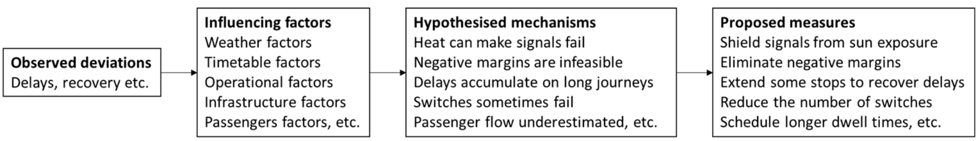


Figure 6

By linking observed deviations, such as delays and recovery, to data covering weather, timetable planning, operational factors, infrastructure and passengers (and more), we can identify some patterns. From these patterns we can hypothesise about the underlying mechanisms, which lead to the delays. These hypotheses can be investigated and tested more thoroughly, and if they are found to be believable, we can then propose measures which address the mechanisms and should lead to a reduction in delays.

Pre-processing the data

Data quality can be an important issue for quantitative methods – especially with empirical data. Missing, faulty or outlying observations might influence the results so that they are not representative. There are several approaches to deal with this, and which have been used in the making of this thesis.

Large datasets

The quantitative work in this thesis has covered large samples of trains and train movements. The analysis in Paper 1 spanned 561,000 observations of run and dwell times, Paper 2 analysed the punctuality of over 883,000 trains, Paper 3 was more restrictive with 470,000 trains. Paper 5 had the smallest samples of the quantitative papers, with the analysis covering about 6,000 dwell times. Especially for Papers 1-3, this suggests that the effects of the odd error or outlier would be to a large extent diluted.

Aggregation

With such large numbers of observations, it can often be useful to aggregate observations. The underlying dataset used for Papers 2 and 3, for instance, covered about 32.4 million train movements – aggregated to about 1.1 million train journeys, which were then filtered out further in pre-processing, to the 883,000 and 470,000 observations mentioned in the previous paragraph. Aggregating from observing individual run and dwell time delays to the punctuality of train journeys meant that the variation was reduced, and the impact of faulty, missing or outlying observations reduced: as the punctuality for a given train at a given station is binary, either 0 or 1, even an outlying observation with a very large deviation from the timetable is reduced to the same scale as all the others.

Interpolation

In Paper 2 we were interested in calculating punctuality across all intermediate stops, not only the last one. This introduced some difficulty, as there are some observations that are missing from the data. These are points where the train was not cancelled but there was no record of the train arriving or departing, while there were records of it arriving at surrounding stations. Some of these missing observations are seemingly random, others are more frequent at stations like Arlanda Södra, which are only a few hundred metres from another station, Arlanda Norra. Here we used a process of iteratively interpolating the missing observations from surrounding and existing ones. The details are described in depth in the paper, but we were able to go from 7.5% to 0.1% missing observations. We did not perform this process for the other papers because: (1) when we study the delays explicitly as in Paper 1 and 5, we want to focus on real observations and (2) when we only consider punctuality at the final stop, it is not possible to interpolate the relevant missing observations or to be sure that the train did not stop and turn around at an earlier stop. While the interpolation may introduce some small inaccuracies, these are less important when the data are transformed into an aggregate indicator like punctuality.

Normalisation

In Paper 1 we remained on the level of run and dwell times. As it was still relevant to compare across differing lengths of line sections and speeds of trains, we normalised all durations, delays and margins – to use ratios instead of absolute time units, with the scheduled duration as the denominator. This transformation of the variables made it easier to compare the figures with each other, and to identify those that were outside the normal variation. In Paper 5, where we also remained on the level of dwell times, the variation in scheduled times was much narrower to begin with, and this step of normalisation was not required.

Unfiltered delays

In neither of these papers, dealing explicitly with delays, did we distinguish between primary or secondary delays, or filter the delays based on size (although for technical reasons, larger delays could not be considered for the commuter trains in Stockholm). While filtering based on size can make sense in some circumstances – larger delays causing disproportionate displeasure among passengers (Börjesson and Eliasson 2011), and smaller delays being easier to deal with using timetables – the scope of this thesis is broader, and seeks to give a better understanding of delays in general. Since small delays are so much more common, however, filtering out the larger delays (with a threshold of say 15 or 20 minutes) would have relatively little impact on the averages that we have considered. The standard deviations would be more affected by such a filter, but we have not focused on these in this thesis, and in any case, the effect would be to make the indicators less representative.

Cross-referencing

Yet another approach is to use alternative data-sources to cross-reference and validate observations. This was used for both cities studied in Paper 5 and is described in the section *Dwell times and passenger counts*. In the case of Tokyo in particular we were able to use the automatically registered data to identify errors in the manual observations, and vice versa. In Stockholm, the restrictive set of assumptions used to combine the two datasets also limited the variation in the data, filtering out any large deviations between the two. This approach is not always feasible, and it can be quite restrictive, but it can be a good way to ensure that any remaining observations are reliable.

Visual analysis and Welch's t-test

In all the quantitative papers we used visual analysis of the data, both in tables and in plots, to identify the normal range as well as any outlying or abnormal observations. This was primarily relevant for the variables used to explain the variation in delays or punctuality in Papers 1-3. With such large numbers of observations, the visualisation had to be performed after a process of further aggregation – it is not feasible to plot hundreds of thousands of observations. In Papers 1-3 we performed this aggregation in essentially the same way, grouping observations by the value of the studied variables.

Variation across papers

In Paper 3 we varied this approach slightly, to see if the results were sensitive to the exact grouping. Instead of grouping together trains that exceed a certain threshold, Paper 3 rounded the variables and considered the punctuality for trains within these groups. The groups could not overlap, which they did in Paper 2. How the results vary in detail across the two approaches can be seen in the section *Overview of papers and findings* and in the appended papers, but overall the differences were minor.

Welch's t-test

In Papers 1-3 we used a statistical test known as Welch's t-test, partially to aid in the visual analysis – ensuring that we only plotted values where the difference in punctuality was statistically significant. This is a variation on the commonly known Student's t-test, which allows for comparisons between samples with unequal size and variance (Welch 1947). If $\bar{X}$ is the sample mean, N the sample size and s^2 the sample variance, the t-statistic t and the degrees of freedom v are defined as:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}, v \approx \frac{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}\right)^2}{\frac{s_1^4}{N_1^2 v_1} + \frac{s_2^4}{N_2^2 v_2}}, v_i = N_i - 1 \quad (8)$$

These statistics were then compared to a two-tailed t-distribution, to see if the difference in sample means was statistically significant or not. In Paper 1 this test was used twice per studied variable: once to see if it had an influence on run time delays, and once for dwell time delays. In Papers 2 and 3 it was used to ensure that we did not include values which covered so few observations that we could not be sure of their effects. A threshold of 0.01 was set, but in many cases the p-values were much lower, frequently approaching 0. In other cases, however, the differences were not significant. Examples for this are when wind speeds were studied in Paper 2: only once the wind speed reached 5 m/s did the punctuality begin to be affected, speeds of up to 4 m/s had no statistically significant impact.

Curve fitting

For each plot in Papers 2 and 3, we were interested in describing the shape of the relationship and thus tried to fit various functions to the scatter diagrams – choosing that which provided the best fit. As this was done on aggregated data, the plotted observations each had different weights – sometimes differing by multiples of several hundred. In Paper 2 we relied on the t-tests and did not use these weights when fitting the trendlines, rather assigning equal weight to every (aggregated) observation and fitting according to the Ordinary (unweighted) Least Squares. In Paper 3 we instead used Weighted Least Squares, with the number of observations per group used as weights. Overall, this choice made little difference to the results.

Ordinary Least Squares Regression

Papers 1, 2 and 5 also used ordinary least squares regression to varying extents. In Papers 1 and 3 this was mostly to see which factors had statistically significant impacts, and which did not. In Papers 2 and 5, regression analysis was also used to see to which extent the explanatory variables could explain the variation in punctuality or dwell time delays, respectively. One estimate of this is the coefficient of determination, the R^2 or adjusted R^2, of the regression model. Paper 5 included a separate set of regressions for this express purpose, including both squared variables and interaction terms, intended to catch non-linear and interaction effects in the data. Such a model is, however, too complicated use in practice. A second set of models was estimated with only linear effects, and no interactions, to provide at least a rough model for how to adapt the scheduled dwell times to the flow of passengers at a given station. In this case, the coefficient estimates were more important. Thus, the regressions were used to (1) check the statistical significance of any effects, (2) check the predictive power of the data, and (3) provide useful coefficient estimates.

Predictive power

The predictive power of these regression models has varied substantially – from about 4% in Paper 2, to 40% in Paper 5. There are several reasons for this difference. For one, the scope is more well-defined, with only commuter trains in either Stockholm or Tokyo, in the morning rush hours of weekdays in April or November – compared to all passenger trains in the Sweden for one full year. The variation was thus much smaller to begin with. Second, is that punctuality, studied in Paper 2, is an aggregate indicator which is less granular and detailed than the dwell time delays studied in Paper 5. The cause of events is much more complex for the former. Third, the models in Paper 2 did not include squared or interacting variables – in Paper 5 this inclusion approximately doubled the coefficient of determination. This simply suggests that the problem studied in Paper 2 is more complex, and less well-suited for predictions – not that the regressions there are any worse or less valid.

Interviews with timetable planners

To produce the material for Paper 4 (and parts of Paper 3), we carried out semi-structured interviews with timetable planners working at the Swedish Transportation Administration's office in Malmö. Each interview was approximately an hour long, recorded, and transcribed in full, which resulted in written material of around 50 pages. The results were analysed by manual categorisation and concentration of meaning.

The interviewees

The Swedish Transport Administration employs about 20 long-term timetable planners, who work chiefly in the annual timetabling process. In addition to these, there are short-term timetable planners who work in the ad hoc-process. The Swedish railway is divided into eight regions, and the southernmost region is planned from the office in Malmö by four timetable planners, all of whom we interviewed. Two of the planners were men and two women. All of them have worked in the industry for many years, at least since 2003 and going back as far as 1985, and with timetables for nine or more years. The region they plan for is a sort of microcosm of the railway network in Sweden, and it contains a very diverse mix of railway lines, train traffic and capacity utilisation.

Motivation

We used a qualitative method because this allowed us to effectively study the values and priorities of those involved. In this choice of method, we thus applied a qualitative approach on a topic that is typically studied using quantitative methods. We prepared an interview guide based on four areas which were identified before the interviews: (1) guidelines and support, (2) rules of thumb, (3) feedback loops, (4) trade-offs, with a handful of guiding questions in each area.

Processing

The process of analysing the transcribed material was carried out in sequence. The first step was to sort the different interviewer-interviewee exchanges by area, rather than chronologically – what is often called categorisation. The second step was to concentrate the meaning of the answers by cutting superfluous words and sometimes reformulating entire paragraphs into a few sentences. This process reduced the volume of text from 24,500 words to 4,500 and enabled a better overview of what was said. After the interview answers were concentrated, they were sorted into 16 new sub-areas, and the contents then summarised further, reducing the volume from 4,500 to 500 words. This made obtaining an overview of the contents manageable.

Alternative reading

The analysis in Paper 4 is based on an alternate reading of the interview responses: instances where the planners described feedback from dispatchers about errors in the timetable. Several of their statements could explain why timetabling errors sometimes happen; these were condensed into a list of eleven reasons. At this point, we looked for different ways to group and categorise the answers, looking for themes on a higher analytical level, and came up with the three following categories, which we use in the analysis: (1) “missing tools and support”, (2) “role conflict”, and (3) “single-loop learning” (see Argyris & Schön, 1996). These three categories are used to explain and discuss the reasons behind the errors more deeply.

Research Process and Results

This section briefly describes each of the five papers: how they are structured, how they lead into one another, and how they fit together. The connections are illustrated in Figure 7.

The papers follow the basic format of *describing* the situation, *analysing* it, and presenting recommendations for how to *improve* the situation. *Describing* the situation entails collecting, processing and summarising the data. In Papers 1 and 5 this is about the distribution of delays, in Papers 2 and 3 about the level of punctuality and how this varies over time and for different types of trains, and in Paper 4 about describing the situation of the timetable planners and the different errors that they sometimes make. *Analysing* the situation consists of linking different types of data together, running regressions, performing t-tests and creating plots, in the quantitative papers. In Paper 4 we analysed the transcripts more closely to find the reasons that errors are sometimes made, and to categorise them into themes.

Finally, based on the analysis, all the included papers propose practical measures that can be taken to *improve* the situation and reduce delays. Most of these measures relate to timetable planning, as such changes are quite inexpensive and quick to carry out, while, as we show in the papers, they would have real effects on delays and punctuality. Others focus on physical changes to the infrastructure, or on changing some operational aspects. These recommendations have been collected in a separate section.

How the papers build on each other

This section briefly describes how the pilot study in Paper 1 set the stage for the following papers, and how these flowed from and built upon one another. This is illustrated schematically in Figure 7.

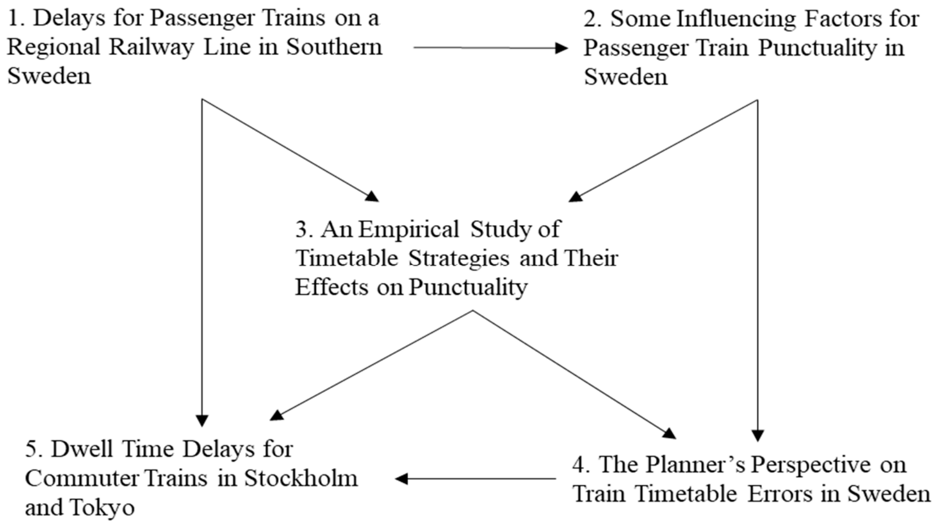


Figure 7
A schematic illustration of how the five papers are connected. Paper 1 is the pilot study, and both its findings and basic methods lead directly to Papers 2 and 3. The finding that most delays occur at stations, rather than between them, is also the basis of Paper 5. Paper 2 extends the approach from Paper 1 to more data and more factors, but still shows a big role for timetable planning, which inspires Papers 3 and 4. Paper 3 uses the same approach as Papers 1 and 2, but focuses on timetable planning. It includes some interviews with timetable planners, the analyses of which are expanded greatly upon in Paper 4. These interviews also shed light upon the scheduling of dwell times, which combined with the findings in Paper 1, lead well into Paper 5, which is focused on dwell times and dwell time delays.

The pilot study

Paper 1 is based on a pilot study, intended to test the basic approach and data to be used in the research, and to help direct the topic of the research. We limited the geography to *Skånebanan*, a regional railway line in Southern Sweden, and used one full year of observations, 2015.

The basic approach

The basic approach was to link train movement data with other relevant datasets that might help both explain the delays that occur and provide ways to reduce them. In this paper we also had access to timetable data showing the level and distribution of margins, to weather data including temperature, wind, precipitation and snow depth, and to data on theoretically estimated capacity utilisation.

Run and dwell time delays

One of the innovations in this first paper was to consider run and dwell time delays separately, to identify more clearly when and where the delays occurred. This led to the result that the risk of being delayed was a great deal higher at stations, where as many as 40% of dwell times took at least one minute longer than scheduled. This was one of the key drivers behind later performing the study described in Paper 5, focusing more on one aspect of what happens at stations, namely the exchange of passengers.

Other findings

The other main findings, to be followed up in later papers, were that the details of a timetable were strongly associated with the risk of delays, that weather had important but somewhat complicated impacts, and that theoretically derived indicators of capacity utilisation were not so strongly linked to delays. These findings, and the success of the basic method, led on to Papers 2 and 3.

The subsequent papers

Extending and broadening

Paper 2 is basically an extension of Paper 1, both in that it considers the national railway network instead of a single line, and in that the number of influencing factors considered is greatly increased. Beyond some refinements and new variables from the previous datasets, a dataset of infrastructure components such as switches, signals, tunnels, bridges and more was added. The idea is that breakdowns in infrastructure sometime cause delays, and that trains passing by many components are exposed to a higher risk of such delays. To deal with the vastly increased number of observations, we shifted from studying individual delays to the punctuality of trains. The findings basically validated those from Paper 1, and as it is difficult to do much about the weather, and adaptations in the infrastructure can be quite expensive and take a long time to perform, while timetables are redone every year at very little cost, we chose to look more closely at the role of timetable planning in Papers 3 and 4.

Focusing on the timetables

In Paper 3 we used the same train movement and timetable data as in Paper 2. However, the aim was much more focused on describing the process and strategies in timetable planning - particularly around the issue of allocating various kinds of run time margins - and evaluating the effects of these strategies. Based on literature and interviews with planners, we identified a few different ways to think about when to allocate margins and how to distribute them. We then used the actual timetable data to see to what extent they were used, and empirical data on train movements to identify the real effects on punctuality.

Focusing on the planners

Paper 4 is focused on the timetabling process, extending the analysis of the interviews with planners, and their situation, from Paper 3. As previous papers had illustrated that even seemingly small decisions in timetables impact punctuality, it seemed prudent to interview the planners who make the decisions. The interviews were centred on four topics: feedback loops, support and guidelines, trade-offs, and rules of thumb. In the analysis we then returned to the transcribed material with the aim of finding out what errors are made in timetable planning, and what the consequences and causes of these errors are. The results align well with the finding in Paper 1, that most delays happen at stations. This combination of results leads into Paper 5, where we focused explicitly on dwell time delays.

Focusing on the dwell times

In Paper 5 we looked closely at dwell times and dwell time delays, following up on findings from Papers 1 and 4. The former identifying that most delays occur at stations, rather than between them; the latter identifying both that dwell times are often scheduled unrealistically, and that these outcomes are not evaluated by planners. It is also a complement to the focus on run time margins found in Paper 3. The paper is also the result of gaining access to new datasets, from commuter trains in both Stockholm and Tokyo, containing observations of dwell times and passenger counts, along with the train movements. These data enabled much more detailed study of the delays that happen at stations, and into one of the key causes of such delays – the exchange of passengers.

Overview of papers and findings

This section briefly summarises the aims, analyses and findings of the five included papers. This is summarised in Table 7.

Table 7
Overview of papers with aim, dataset, analysis and delay indicator used

Paper	Aim	Dataset	Analysis	Delay indicator
1	Improve knowledge on delays and the methods to study them	One regional line, 1 year	Regressions + t-tests + plots	Run & dwell time delay
2	Identify and quantify the impact of several weather, timetable, operational and infrastructure variables on punctuality	Sweden, 1 year	Regressions + t-tests + plots	Punctuality at all stops
3	Study the punctuality effects of strategies for allocating margins	Sweden, 1 year	Regressions + t-tests + plots	Punctuality at final stop
4	Describe the situation for timetable planners in Sweden. Identify common errors in timetables, as well as the reasons behind them.	Four planners	Qualitative based on interviews	Errors in timetable planning
5	To study dwell time delays and to see how much they can be explained using passenger data.	Stockholm + Tokyo, several years	Regressions	Dwell time delay

Paper 1. The pilot study

The main aim of this paper is to learn more about how delays are distributed and about how they are associated with various weather, timetable and operational variables. As a pilot study intended to test the approach and methods for future papers, the analysis is based on one year (2015) of data from one regional single-track line in Sweden.

Method

In order to analyse the link between the studied variables and the delays we make use of three basic steps. The first is to determine if the studied variables had a statistically significant impact on the delays, using Welch’s t-test. For example, we looked at temperatures below zero and temperatures above zero and used Welch’s t-test to test whether the average delays for these two samples are significantly different from each other. The second step is to perform linear regressions for delays at station stops and line sections respectively, in order to give an overview of the trend of each studied variable. The third step of the analysis includes a set of plots, each emphasising different aspects of the data. These plots together give a more nuanced picture than either the t-tests or the regressions, and visual inspection can reveal if the studied variables vary along with delays in any recognisable pattern.

Findings

We find that the delays mostly occur at stations – both the frequency and severity are much higher there than on the line sections between stations, where the average delay is very close to zero. At stations, the scheduled dwell time explains more of the delays than any other factor we consider in the paper. The average delay is significant if the scheduled dwell time is lower than 160 seconds, while the greatest delay reduction occurs at dwell times of 210 seconds, beyond that the average delays increase. On the line sections, the single most important factor is the level of margins: most of these delays are caused when trains have no margins, although negative margins do occur and contribute to a small degree.

We find that there are clearly diminishing marginal returns, and that the most effective level for delay reduction is around 10%. Both temperature and snow had small but statistically significant impacts on delays, and similarly there were small but statistically significant differences in average delays across weekdays, which can be explained by the variation in the number of trains run. Finally, small arrival delays to stations tended to speed up the stops, while early or on-time arrivals are on average delayed by around 30 seconds when departing the station. Larger arrival delays are associated with even larger dwell time delays.

Paper 2. Four types of factors

The purpose of this paper was to identify and quantify the link between several weather, timetable, operational and infrastructure variables and the punctuality of passenger trains in Sweden. We use data on all passenger trains in Sweden during the year of 2015. The method is similar to Paper 1, a major difference being that we study punctuality (across all scheduled stops and with a delay threshold of five minutes), rather than run and dwell time deviations. This greatly reduces the number of observations and the computational load compared to studying delays directly, and facilitates the larger study area, while giving a more holistic picture than only considering punctuality at the final destination.

Weather factors

The results indicate that punctuality is clearly correlated with weather. The relationship with temperature is exponential, for both low and high temperatures. Compared to the average, punctuality drops by 50% at -30 °C, and by 26% at 27 °C. Strong winds also lower punctuality: when they exceed 23 m/s, it is about 9% lower than average, following a power curve. Precipitation lowers punctuality moderately, in a linear fashion, as it is accumulated throughout the journey. And while less than 6% of the observations in our dataset have average snow depths larger than 1 cm, the effect is quite large: at an average of 5 cm the drop in punctuality is about 17.5%-points, following a logarithmic curve.

Timetable factors

The results regarding timetable variables suggest that margins are correlated to punctuality up to a point of around 12 s/km, or 25-30% of the minimum run time. Then it is around 2% higher than average, at even higher levels the correlation turns negative. Similarly, the weighted average distance (WAD) of margins is correlated to punctuality up to a point of about 0.60, where it is about 1% higher than average, before the correlation switches sign. When they exist, negative margins are linked to, on average, 2.8% lower punctuality.

Operational factors

We also consider a range of operational factors. The single best indicator for punctuality is the distance travelled by a train, with about 3% per 100 km – the correlation coefficient of -0.20 is the highest in our findings. Highly correlated to this is the duration in time, where every hour is linked to a decrease in punctuality of about 1.6%. Still, increasing average speeds of trains is linked with decreasing punctuality, with the airport trains being an exception. This suggests that the ability to accommodate heterogeneous traffic, with widely varying speeds on the same line, has been overestimated.

The average distance between stops also appears to be linked to lower punctuality in a mostly linear fashion, by about 1.3% for every 10 km. The link between punctuality and the number of trains is best described using a quadratic function, but the effect is relatively small – the highest punctuality drop we see is 1.2%.

At stations, the number of trains that arrive per hour is instead linked to punctuality in a linear and positive manner: at volumes of at least 20 trains per station and hour, the punctuality is about 2.5% higher than average – though this may be explained by a high proportion of short-distance commuter trains, with higher punctuality. We also see that interactions between trains – instances where trains are at the same place at the same time – are associated with lower punctuality: by about 1% and 2.2% for interactions at and between stations, respectively.

Infrastructure factors

Turning finally towards infrastructure, plotting the number of switches, tunnels and fences against punctuality, we find that they fit best to quadratic functions. The mechanisms behind these negative synergies are not clear, and should be studied further, but we can perhaps speculate that it has to do with difficulties of maintenance and complex environments with more trains, people and animals in the vicinity that can cause problems. Bridges, signals, level crossings, and cuttings show linear relationships to punctuality. Signals have the largest effects in terms of magnitude, being associated with punctuality drops of around 22% at the most, although this is highly correlated to the distance travelled.

Paper 3. Margins and interactions

The purpose of this paper was to study the effects on punctuality of some strategies for allocation of margins in timetables for passenger trains by analysing empirical data. The analysis focuses on aspects that relate to the planning of a timetable, i.e. the size and distribution of margins, and on margins "within" train paths rather than those "between" them. As such, this study does not address headway or buffer times.

The method and data are similar to Papers 1 and 2, with all passenger trains in Sweden during 2015. As we are interested in how margins are distributed along the journey and timetable, it is important to consider an aggregated indicator like punctuality, rather than looking at each delay directly. Here we use the conventional Swedish definition of punctuality: considering the final destination and a delay threshold of five minutes. This enables comparisons against Paper 2, and for us to see the difference between using different punctuality definitions. Another difference is that we use weighted least squares and the bi-square method to help produce more robust estimates.

Interactions between trains

The results indicate a negative association between interactions and punctuality, by about 1.2 and 3.9% each at and between stations, respectively. There is a wide variation in the number of interactions at stations: the mode is 1, the mean 2.89 and the 95th percentile 8. Interactions between stations are uncommon, occurring only for 2-3% of trains.

Size of margins

The results on margins indicate a positive correlation, but the slope is shallow at about 1/8: to achieve an improvement in punctuality by five percentage points, margins of the order of 40% of the minimum run time must be added. However, we estimate that when negative margins exist in a train's timetable, its punctuality is on average 4.1% lower.

Distribution of margins

The distribution of margins is also of interest. We found that the average WAD of margins was 0.56, and that the slope as regards punctuality is about 0.11. Supplements are also frequent directly following scheduled stops. They cluster around the "round" numbers of 30, 60, 90 and 120 seconds, although not having any supplements was the most common. We found that larger supplements, of at least 60 seconds, were associated with higher punctuality, while it was worse to have small supplements than to have none. These findings are statistically significant. We hypothesise that there is a tension between the added time supplements and the drivers' behaviour, such that drivers sometimes overcompensate when given small supplements, believing their effects to be larger than they really are.

Paper 4. Problems in planning practice

The paper is based on interviews with experienced timetable planners in southern Sweden and gives a description of their current situation. It identifies common errors in timetables, which influence the punctuality, as well as the reasons behind them.

Responsibilities

The planners stated that they have a large individual responsibility in learning what is necessary and in performing quality control, and that there is not enough time for either. They described receiving frequent comments from dispatchers, centred on four areas: (a) crossing train paths at stations, (b) wrong track allocation of trains at stations, especially long trains, (c) insufficient dwell and meet times at stations and (d) insufficient headways leading to delays spreading. Yet, there is no established system or routine to keep track of or utilise comments from dispatchers. The planners also feel that important conditions change from year to year, making it difficult to draw direct comparisons and learn over time.

Difficulties

The planning method has been largely the same for the last 20 years or so, but the work is becoming more difficult due to the increasing number of trains and engineering works. Problems increasingly occur at the stations, where there is insufficient capacity. This is exacerbated by ad hoc planners bending the rules to fit in more trains. Trainplan is the main tool used at the Swedish Transport Administration and it does not handle track allocation, conflict management, or provide topographical information. The more experienced planners work based more on discussions with the train operating companies than on a strict application of the guidelines. The guidelines were mentioned by all the interviewees, but they are interpreted liberally and were not described as helpful.

Margins

All four planners use different methods to assign margins. Another common practice is to adjust arrival and departure times at stations to occur at whole minutes: usually adding, but sometimes subtracting, seconds up to the whole minute. Negative margins are often used for local trains on single-tracks at the request of train operating companies. The explanations for this vary from person to person, but they say that: “it has always been like this”. Only the most experienced planner adds a minute after a scheduled stop, although others were aware of the practice.

Dwell times

The planner only deviates from requests when necessary, and this is done in dialogue with the applicant. The latter sets the dwell times. The planners say that the standard is two minutes, but they give the impression that shorter times dominate. Local trains are often given the same arrival and departure times, with no scheduled dwell time, to avoid waiting unnecessarily in case the train is delayed or the number of passengers is small. Dwell times in excess of two minutes are primarily for connections and phasing reasons, and in some cases during winter breaks.

Paper 5. Passengers and dwell times

The aim of this paper is to study the dwell time delays that occur for commuter trains in Stockholm and Tokyo, and to see how many of them can be explained using passenger data. Different timetabling policies are discussed, with the intent of improving timetable planning and punctuality, and to reduce delays in both cities.

Method

In the analysis we used passenger counts to explain the variation in dwell time delays. We also had information of the arrival delay to the station. The effects are studied using regression analysis, with two purposes: (1) to estimate the degree to which we can explain the variation in the dwell time delays, using passenger data, and (2) to estimate the impact of passenger volumes on the dwell time delays, in a comprehensible way. The first point relates to the coefficient of determination, the adjusted R^2, while the second has to do with coefficient estimates. To fulfil the first purpose, it is relevant to include both squared variables, as well as variable interactions, as these may well, in fact, have impacts on the delays and add to the share which we can explain. The second purpose, however, is better served by keeping the models simple and straightforward, sacrificing explanatory power for interpretability.

Delay distribution

In Stockholm and Tokyo, minor delays of at most five minutes make up 96 and 97% of delay hours respectively. In both cities, the delays mostly occur at stations in the form of dwell time delays. In Stockholm, 91% of the total delay time is generated at stations, while the corresponding figure in Tokyo is 88%. Thus, small dwell time delays make up a clear majority of all delays.

Scheduled dwell times

The railway companies in the two cities have different policies on scheduling dwell times. The default in Stockholm is to use 42 seconds regardless of day, time or station, with a few exceptions. In Tokyo, the dwell times are adjusted to a much greater extent. The range is quite similar to that found in Stockholm, but the variation is greater: the 5th, 25th, 50th, 75th and 95th percentile values are 40, 45, 50, 60 and 115 seconds, respectively. Almost across the board, the dwell times in Tokyo are also longer than the 42 seconds used in Stockholm, and adjustments are made in five-second intervals, across different train services, stations, and hours.

Dwell time delays

Dwell time delays are similar across the two combined datasets. In Stockholm the median delay is 6 seconds, compared to 5 in Tokyo, and the 95th percentile values are 34 and 30 seconds respectively. About the same proportion of trains make up time during the stops, in the two cities, but in Tokyo the ones that do make up slightly more time, with the fifth percentile being 21 seconds early there, compared to 8 seconds in Stockholm. The range of dwell time delays, from the 5th to the 95th percentiles, are quite small and comparable: 42 seconds in Stockholm and 51 seconds in Tokyo. As these are commuter trains, however, stops are frequent, and the seconds add up.

Explaining the variation

In both cases, about 40% of the variation is explained using the full models, with squares and interactions, and about half as much by the simple linear models. In Tokyo, we find that a rise in the congestion rate of 10% corresponds to an increased dwell time delay of about one second. For arrival delay and scheduled dwell time, we find small negative effects – this implies that trains that arrive late have slightly shorter dwell times than otherwise, and that the influence of the personnel and driver are somewhat successful in reducing delays. Longer scheduled dwell times see slightly less delays, indicating that they include some margins, not just the minimum required time for the increased congestion rates. In Stockholm, the estimates are about 0.4 seconds per person and car, in either direction. We also see a slight effect from congestion, as the number of passing travellers further increases the delays.

Results

In this section we link the results from each of the papers to the overarching aim and research questions of the thesis, summarise them, and discuss our findings in relation to the research literature. In this way we work towards the overarching aim: to better understand the smaller delays that occur for passenger trains in Sweden, in order to improve timetable planning so that the delays are reduced. Schematically, the process of moving from observed deviations through influencing factors on to proposed measures can be illustrated as in Figure 8.

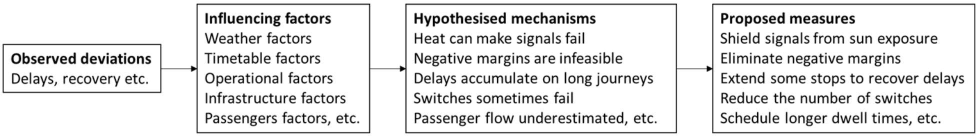


Figure 8
By linking observed deviations, such as delays and recovery, to data covering weather, timetable planning, operational factors, infrastructure and passengers (and more), we can identify some patterns. From these patterns we can hypothesise about the underlying mechanisms, which lead to the delays. These hypotheses can be investigated and tested more thoroughly, and if they are found to be believable, we can then propose measures which address the mechanisms and should lead to a reduction in delays.

Linking the papers to the research questions

A brief summary of how the papers link to the research questions is shown in Table 8. The links are then discussed in turn, before we turn to the results.

Table 8
Connection between papers and research questions

Research Question	Paper 1	Paper 2	Paper 3	Paper 4	Paper 5
RQ1. How are the delays distributed?	X				X
RQ2. What factors affect delays, and to what extent?	X	X			X
RQ3. How can the allocation of run time margins be improved?	X	X	X		
RQ4. How can the allocation of dwell times be improved?	X		X		X
RQ5. How can the practice of timetable planning be improved?				X	X

RQ1. How are the delays distributed, in a broad sense of the word?

To understand and analyse the delays, we first need to know how they are distributed. The distribution can be considered along many dimensions, such as: among trains, in time, in space, and by size. The question is primarily addressed in Papers 1 and 5, which both consider delays explicitly, without using the aggregated indicator of punctuality. Paper 1 makes the distinction between dwell and run time delays – happening *at* or *between* stations. It also describes how delays vary with weekdays and the months of the year, while Paper 5 describes the distribution by size, with over 95% of all delay time being made up of delays smaller than six minutes for commuter trains.

RQ2. What factors correlate with delays, and to what extent?

Beyond understanding their distribution, we are interested in quantifying the link between delays and certain influencing factors, ranging from weather factors, timetable factors, operational factors, infrastructure factors and passenger factors. Papers 1 and 2 both consider the weather variables temperature, precipitation and snow depth, along with various timetable and operational factors, such as various indicators of capacity utilisation. Paper 2 also considers wind speed, more operational factors, and the complexity of the infrastructure while using a much larger dataset. Paper 5 considers the influence of passengers, using more detailed data.

RQ3. How can the allocation of run time margins be improved?

One common way of reducing delays is to allocate run time margins in timetables, so that the train can run faster than scheduled and catch up to the timetable. How this allocation can be improved is addressed in Papers 1, 2 and 3. Paper 1 considers the size of margin, section by section. Papers 2 and 3 instead consider the whole journey, with the total size of margins, the distribution of these margins across the journey, the occurrence of negative margins, and, in Paper 3, margins directly following station stops.

RQ4. How can the allocation of dwell times be improved?

Another way to make a timetable robust as regards delays is to schedule dwell times that are longer than necessary, so that any delays are reduced once the train reaches the station. These dwell times are considered explicitly in Papers 1 and 5, both of which present recommendations. Paper 1 considers the length of dwell times, and Paper 5 focuses on shorter stops, particularly for commuter trains. Paper 3 considers run time supplements directly following scheduled stops, intended to make up for delays that occur when dwell times are insufficient.

RQ5. How can the practice of timetable planning be improved?

The process of timetable planning itself can both introduce delays and help reduce them. Paper 4 discusses at length how aspects of the timetabling process can be improved in this regard, with recommendations about both tools and support, routines for systematically evaluating outcomes, and clarifying the role of planners at different organisations. Paper 5 also touches upon the timetable planning process, focusing on the allocation and evaluation of dwell times.

RQ1. How are the delays distributed?

At stations, not between them

In Paper 1 we found that the average delay for a train on a line section between two stations is very close to zero, while at stations almost half of all stops take longer than scheduled, and the delays that do occur also tend to be greater at station stops than on line sections. This pattern was found again in Paper 5, where more than 90% of all delay time occurred at stations.

Most are small

In Paper 5 we found that in both Stockholm and Tokyo, minor delays of at most five minutes make up more than 95% of delay hours for commuter trains. For passenger trains in Sweden overall, the corresponding figure is about 85%.

In time

There are small variations between months, with slightly fewer delays during spring and autumn. We also saw small but statistically significant differences in average delays across weekdays – although these were almost entirely explained by how many trains run each day. Traffic volume does not appear to be a major concern however: variations due to a higher number of trains per day are small, and in Paper 2 we found that punctuality is marginally higher at more busy stations and times.

Long distances

Another way to consider how delays are distributed, is that they are larger and more frequent for trains travelling long distances. In Paper 2, the single best indicator for punctuality we find is the distance travelled by a train, with about -3% per 100 km. Closely related to this is the travel time of the journey, where the slope is about -1.6% of punctuality per hour. These are well in line with estimates in the literature showing that distance covered was statistically significant in determining punctuality.

RQ2. What factors correlate with delays, and to what extent?

Weather factors

Both high and low temperatures are associated with large problems in operations, and we have identified exponential relationships between how more extreme temperatures lead to increasing drops in punctuality. In the extremes, with temperatures of positive or negative 30˚C, punctuality can drop by 50 percentage points or more, with an almost total collapse of operations. In the face of increasing temperatures and more frequent heat waves, this suggests that more ought to be done to increase the railway systems' resilience to high temperatures.

Snow also has a clear effect on punctuality. Even with as little as five centimetres of snow, average punctuality was 17.5% lower than normal. These effects are substantially larger than we have found in the literature, as is the case for cold temperatures. The relation appears to be logarithmic, however, so while increasing amounts of snow does lead to more problems, the rate is diminishing. This may suggest an increased preparedness and ability to deal with snow in the regions where large snow depths are often found.

When it comes to wind, delays appear to increase with increasing wind speeds: at ten metres per second punctuality is 2% lower than normal, and when we approach storm-level wind speeds of 23 metres per second, the punctuality drop is 9%. Again, this is a larger effect than we have found in comparable studies from other countries.

Finally, we saw that the more precipitation a train is exposed to, the lower the punctuality. The figures are more modest, however, and the variation is only of the order of a couple of percentage points.

In summary, weather can cause considerable delays, and the problems often begin even during seemingly normal conditions.

Timetable factors

These are discussed under RQ.3-5.

Operational factors

Punctuality is negatively correlated with the length of the journey, on average dropping by about 3% per 100 km, with the length between stops, and with the maximum speed of the trains. These patterns are likely due to problems stemming from heterogeneous speed profiles – long-distance and high-speed trains traversing multiple shorter distance and lower speed commuter train systems. This hypothesis is supported by repeated findings that interactions between trains, especially between stations, are associated with lower punctuality (by about 1% and 2-4% each, at and between stations, respectively). This issue appears to be more severe than the volume of traffic, per se, which only has a small impact.

Infrastructure factors

The overall picture is that a simple infrastructure with less components performs better. We found a quadratic, negative relationship between punctuality and the number of switches. This suggests a potential of gaining disproportionately large punctuality benefits by limiting their number, particularly in large stations, where the numbers are high and even small gains in punctuality are very valuable. With signals, the relationship is linear, and more clearly related to the distance traversed.

Passenger factors

The data on passengers we used in Paper 5 could explain approximately 40% of the variation in dwell time delays in rush hours for commuter trains. While by no means perfect, this is high compared to other studies attempting to explain delays or punctuality with empirical data. It is also a reasonable number, considering the range of other factors which affect delays.

Comparing Stockholm to Tokyo, we find that the latter has about 60% less delays per degree of passenger congestion – they are much more efficient at reliably boarding and alighting. About half of this effect is due to Tokyo's trains having twice as many doors – the other half can be explained by a combination of more appropriate scheduled dwell times, markings on the platforms, platform screen doors, more staff, and more discipline among both staff and passengers.

RQ3. How can the allocation of run time margins be improved?

Size of margins

We find that levels of around 10% are the most efficient and effective, this is where the big gains are made, and that the benefits diminish at higher levels. This means that run time margins for Swedish trains are, in general, sufficiently large, and that they could in many cases be reduced somewhat, without major consequences for the punctuality of operations. Conversely, it would be quite costly to make major improvements in punctuality using only increased margins: to improve punctuality by five percentage points, journey times need to be extended by approximately 40%. This is not acceptable in most cases. One possible way of explaining the diminishing returns of margins is a behavioural response of the driver and other personnel.

No negative margins

It is also important to avoid allocating any negative margins. We have seen that in 2015 about 40% of passenger trains had negative margins on at least one section, and that this was associated with a punctuality drop of 3-4%, depending on the definition of punctuality. In a time when the punctuality of the railways is about 5% lower than the stated goal of the industry, this is a very big and clear effect. Arguments about how the negative margins are compensated by positive ones on the next sections thus do not seem to hold in practice.

Distribution of margins

How the margins are distributed can make a big difference. The important thing is that there are margins everywhere – not only for each train, but on every part of the journey. Otherwise the risk of delays increases dramatically. Beyond this, we have found that punctuality is highest when the centre of gravity of margins is about 0.6-

0.7. That is, close to the middle, but shifted a little towards the end. While there are large variations between trains, the average in Sweden is about 0.56, and could thus be shifted slightly to the second half of the journey.

Margins after stops

Margins can also be placed directly after station stops – as stops are known to generate delays, which can then be recovered directly by a strategically placed time supplement. It is important that these margins are large enough – at least one minute per stop. If they are smaller than that, delays instead tend to increase.

One possible explanation for this is a behavioural response: by allowing more time exiting stations, the train crew takes more time in executing that activity. In this case, they overcompensate. It is possible that this is made worse by the truncation of seconds in some sub-systems in the Swedish railway, so that the driver believes that the supplements given are larger than they really are.

These problems disappear if the dwell time is long enough to begin with, without compensation from run time margins, and in these cases, punctuality is on average about three percentage points higher. The results are slightly better still if the compensation is large enough, and with a supplement of at least one minute after the stop, the punctuality is yet another percentage point higher.

RQ4. How can the allocation of dwell times be improved?

Reallocating time

In Paper 1 we drew attention to the importance of dwell times. We saw a clear relation between the scheduled dwell time and the risk of delays for passenger trains on the studied regional line, Skånebanan. In the paper we presented an example of how dwell times could be re-allocated to reduce the delays. By setting 80% of dwell times to a relatively short 50 seconds, and 20% to 210 seconds, which was the level most effective at reducing delays, we estimated that delays on the line could be reduced by about 80%, without any increase in travel times.

The importance of passengers

Towards the end of the research we returned to the topic of dwell times. Paper 5 in particular deals to a large extent with how they can be adapted to the number of passengers, with case studies from both Stockholm and Tokyo. In the first case, dwell times for commuter trains are generally not adapted to passenger volumes, while such adjustments are very common and important in Tokyo. This helps to keep their punctuality higher than in Stockholm, despite having many more trains, and many more passengers.

Often overlooked

We have also found that dwell times are often overlooked in Swedish timetable planning. This is seen in the interviews in Papers 3 and 4, where timetable planners describe that they often receive feedback from dispatchers about how dwell times are insufficient, while there is no discussion or thought about how to adjust and improve these dwell times.

This is also seen in Paper 5, where the difference in approach and object of focus between planners in Sweden and Japan becomes evident. While planners in Japan work in a very conscious and dedicated way to fine tune dwell times, planners in Sweden have long used the same template, with only a few exceptions. Both Papers 1 and 5 describe how most delays in Sweden arise at stations, indicating that the dwell times are not as robust, and do not contain the same level of margins, as the sections between stations.

Old strategy

We suggest that timetable planners in Sweden focus more on allocating appropriate dwell times. For too long, dwell times have been treated like some sort of residual, been scheduled with a rough template that has been applied too liberally, or knowingly been given too little time. There has been a strategy that delays can arise at stations, to be recovered on the line, where there are margins.

While this strategy could work, in principle, we have not found evidence that it is successful. While there are often large margins on the line sections between stations, they do not make up for the delays that occur at stations. Instead, it leads to systematic delays at stations, which introduces instability in the timetable, disrupts the scheduled interactions between trains, and largely leaves the operations to dispatchers rather than planners. Regardless of whether this strategy has been successful in the past, it does not work with today’s complex and dense traffic.

New strategy

The baseline must be to allocate realistic dwell times, to allocate the time that it takes to complete a stop. Not the least amount of time that is possible, or a time that is sufficient for half of all trains, but the time that it should take for the given train at the respective station and time. This implies a move from rough rules of thumb, and that dwell times may well vary from hour to hour, station to station, and potentially from train to train. It may well be permissible to schedule longer dwell times than necessary, and to use them as a type of margin to absorb delays, but not to deliberately create delays to artificially speed up station stops.

Of course, it will not be possible to entirely avoid dwell time delays, just as it is unrealistic to allocate run times that are never exceeded. But the current state is shifted heavily towards too short dwell times, and to dwell time delays being much more common than run time delays.

RQ5. How can the practice of timetable planning be improved?

Improved planning software

For Papers 3 and 4 we interviewed timetable planners and found several weaknesses in the process for timetable planning. Based on the interviews, we found that the main planning software, Trainplan, supports neither conflict detection nor track allocation at stations. Both of these lead to recurring mistakes in planning, to delays, and to problems for dispatchers. The software was introduced around the new millennium, and has thus been used for almost two decades, without fixes to these issues.

While a new tool (TPOS) is currently under development, which is intended to address these issues and modernise the planning process, the project is delayed, and it is unclear whether all the features will be implemented. This is reminiscent of the introduction of Trainplan, and it is thus important to emphasise these issues and to ensure that they are addressed by the new tools, and that these are implemented.

Improved documentation

There are also deficiencies in the documentation, in a broader perspective. Timetable planners describe how important information about which interlocking systems are in use at various stations, how they work and how they must be handled from a planning perspective, is primarily documented in binders which the planners themselves must navigate, interpret and memorise. This is difficult and leads to mistakes being made. The guidelines used for planning have also remained in use for many years without major revisions, despite not giving much support in the issues which planners describe struggling with, and they are interpreted very liberally.

More systematic evaluation

Connected to the lacking documentation, we see in Papers 3 and 4 that there is no systematic evaluation of timetable planning. It is up to each planner to find out what works well, and what should be improved. There are no tools to help with this, no detailed statistics, and far too little time. They also describe how important preconditions in the requests for train paths change from year to year, so that the problems and complaints that have arisen throughout one year cannot so easily be applied to the next year. This makes it very difficult to achieve a systematic evaluation, and evolution, of the timetable quality.

Clarified roles and responsibilities

We see in Papers 3 and 4 that there is a lack of clarity and a role conflict in the role of timetable planner at the infrastructure manager – the Swedish Transport Administration.

On the one hand, they are responsible for making sure that the timetable comes together, is feasible, and maintains a high quality. On the other hand, they are in close contact with timetable planners at the train operating companies. These companies run the trains, are closest to the customers, and ought to know best how they want to their timetables to be planned. Sometimes this leads to conflicts and situations where the train operating company would rather that the timetable planners at the infrastructure manager be less stringent with the guidelines and what would make for a robust timetable, to be able to run more trains.

At the infrastructure manager the planners also appear to interpret the timetable requests as well motivated and thought through. They strive to make as few deviations from these as possible. The picture at the train operating companies can be quite different: that they do not have the same resources for timetable planning, that they usually stick to what they did last year, or to commercial demands, trusting that the infrastructure manager will ensure that the timetables maintain a high level of quality. Of course, this situation can lead to a gap, and to the quality suffering, as both parties assume that the other one has the responsibility.

Improved feedback loops

The results presented in Paper 4 suggest that both researchers and practitioners should focus more on identifying and improving the relevant feedback-loops, to achieve a higher level of learning among those involved.

Single-loop learning is both a technical and organisational issue. Since the tools are lacking, planners are hard-pressed just to finish their work. There is simply not sufficient time to perform quality control. Because there is no systematic review of the quality and outcome, there is no way to begin to improve the rules and guidelines, or to create a better timetable. As the tools do not provide enough assistance, the focus is, and must be, on creating a timetable before creating a better timetable.

Creating a better timetable is what we imagine the timetable planners of the future will be tasked with doing, when more of the work has been automated and the software tools provide far more assistance. Rather than trying to manually execute all the details, they will choose which heuristics, goal-functions and constraints to apply in different scenarios, to achieve the best overall results. An illustration of this wider feedback-loop, or double-loop learning, is found in Figure 9.

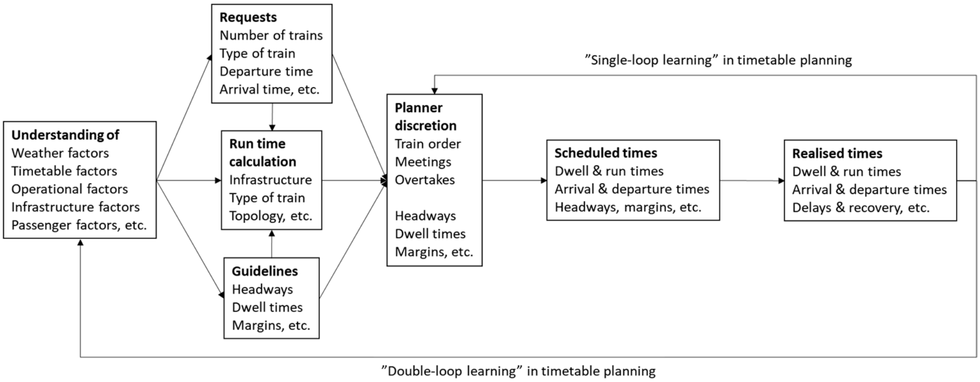


Figure 9
A schematic illustration of the timetable planning process with single- and double-loop learning. In the case of single-loop learning, the timetable planner uses comparison between scheduled and realised time to inform their own discretion. In double-loop learning, the comparison is used to inform the understanding of various factors relating to weather, timetables, the infrastructure, passengers, and operational concerns, and this improved understanding is used to improve the requests, run time calculation, and timetable planning guidelines. The outcome is then less dependent on the discretion of individual planners, and the planners can focus more on improving the run time calculations and guidelines, as well as potentially helping the train operating companies make more informed requests, than on remembering countless details, tweaks and exceptions to the guidelines.

International relevance

This study focuses on Sweden. In the literature, we have seen large similarities with the United Kingdom, which uses the same tools, and has a similarly deregulated market. The new tools that are currently being implemented in Sweden have recently been implemented, by the same supplier, in both Norway and Denmark. Experts from the infrastructure manager and largest train operator in the Netherlands (Olink and Scheepmaker 2017) describe similar issues with errors causing delays and infeasibility there.

We believe that the planning process is largely similar across most European countries, although the level of deregulation and competition between train operating companies for capacity may vary, as will the tools and contexts.

Concluding Discussion

Recommendations

This section describes our recommendations about what to do, in practice, based on the results and conclusions. Of course, there is some overlap between this and the conclusions themselves. The recommendations are placed in four clusters: allocating margins, scheduling dwell times, timetabling process, and physical changes. All of them are intended to reduce delays and improve punctuality, while being highly cost-efficient. In practice, these can be considered for implementation by the Swedish Transport Administration in their new planning tool TPOS, and by other infrastructure managers and railway operators.

Size and allocation of margins

Margins – or run time supplements – clearly play an important part in timetables. However, the process for allocating them currently takes too long, and is too arbitrary, and it both could and should be automated to a greater extent, so that planners can spend more time and energy on more difficult and important issues.

New baseline

When timetables are created, one of the first steps is to calculate run times. Currently, these calculations assume that trains run 3% slower than possible, essentially adding an automatic level of margins of 3%, in excess of the levels we have discussed in the thesis.

We recommend that the procedure is changed, to instead automatically add margins of 10% to all run times. This is a good base level and compromise, higher than what is common internationally, but lower than what is currently added manually, and it is very simple to do. This automatic addition should replace the current practice of node supplements, which is confusing, time-consuming and highly arbitrary. Ideally, the margins should be made explicit and displayed so that both timetable planners and train drivers can easily see them.

Automatic distribution

Taking this a step further, the computer could be instructed to add margins in such a way that the weighted average distance comes in at around 0.67, while keeping the margins as evenly distributed as possible. This is a slightly more complicated adjustment, but it can be performed automatically either ahead of time, or on the fly, and it should improve punctuality by about one half of a percentage point.

In parts of the network that are not so highly congested, it might also be worth adding a time supplement of 60 seconds directly following a scheduled stop, to further absorb delays that often happen there. While this slows down the flow of trains leaving the station and reduces the capacity, making it infeasible in some cases, in many cases this is not a problem, and the punctuality of the affected trains would likely improve by about one percentage point.

No manual rounding

Planners should be instructed not to concern themselves with adding or subtracting seconds so that the trains arrive to and depart from stations at whole minutes. These adjustments take a lot of time to do while they neither improve the quality of the timetable, nor are necessary for the technical systems, passengers, or train drivers. In the rare cases that the systems do require times on whole minutes, these adjustments can and should be made automatically by the computer. Timetable planners should not spend any time on this.

No negative margins

Negative margins should neither be tolerated nor possible to add in the timetabling software. With an increase in the automatic, base-line margins, it is conceivable that manual subtractions of a few seconds would still leave a positive level of margins, but it should never be possible to reduce margins to below a level of about 5%. If the trains do not fit in the timetable, they should not be run.

Role for manual additions

Manually adding to the margins should still be allowed, but with the purpose of coordinating train paths, rather than absorbing delays. Another permissible case is when maintenance works are being carried out and margins are needed to compensate for reduced speeds, although properly recalculating the run times and rescheduling meetings and overtakes is much better than simply adding margins, this may not be realistic in the short term.

Summary

If all these recommendations were implemented, the size of margins would decrease somewhat for most trains. On average, the difference may be of the order of about five percentage points of the run time. Because there are diminishing marginal returns, this reduction will not have a significant impact on punctuality, likely of the order of five tenths of a percentage point, and it will be more than made up for by the improved efficiency of the allocation, and the reduction in mistakes. The time is better spent on increasing dwell times, which are systematically too short.

Scheduling dwell times

Bigger role in planning

Instead of focusing so much on margins, which can easily be automated, this thesis suggests that more time and attention should be spent on dwell times. This includes the efforts of the planners who schedule the times, the dispatchers and train crews that are engaged in operations, and for managers, analysts, planners and others who evaluate the outcomes. The ambition should be to use realistic dwell times, so that the stops routinely take the amount of time that they are supposed to. Again, it is illustrative to compare against run times, as in Paper 1: the probability of delay is three times higher for dwell times. If dwell times were scheduled realistically, instead of optimistically, the frequency and magnitude of these delays would decline rapidly, and a large portion of all delays would be eliminated. While the scheduled travel times would increase somewhat, which could be compensated by reduced margins between stations, adjusting for systematic delays does not increase the total realised travel times.

Higher precision in planning

Our recommendation is to vary the dwell times by train type and station, in intervals of 5-10 seconds, depending on what is most practical from a technical standpoint. In principle, the times should also be varied over time, so that they are shorter in off-peak hours, at weekends and other times when the flows of passengers are lower. Practically speaking, however, this step is difficult to implement right away, and there are benefits to having the same timetable all day and every day, even though there are some losses in efficiency. In this case, it is much better to plan for the high flows during peak hours, when dwell times are longer and the risk of delays is higher, and to accept slightly longer times in the off-peak periods, than to be efficient while off-peak and constantly struggle with delays when the flows of trains and passengers are the highest. The arrival and departure times should be displayed to the drivers and dispatchers, but not necessarily to the passengers, with a precision of seconds.

Periodic delay recovery

To further improve punctuality, a good strategy can be to make about one in five stops significantly longer than the others, of the order of 3-4 minutes. This has a stabilising effect on the traffic and helps to return trains to their scheduled paths. It is more relevant to use for trains that travel long distances, and where the stops are less frequent. The punctuality is much lower for these trains, and the extensions to travel time are not so big across the whole journey. For local trains that make many stops, the increases to travel time would be considerable, while the punctuality is already quite good.

Systematic evaluation and adjustment

Finally, it is important to evaluate the outcomes of the dwell times, and to adjust them going forward, at least once per year. These evaluations should be done by station and train type, and the lessons are easy to apply in practice. In order to do this in a rational way, the infrastructure manager should systematically collect data on actual dwell times – on the level of seconds – from on-board systems, or at least require that the operators report the distributions of dwell times by train type and station, as a part of the information required to allocate capacity. These data are much more direct and useful for timetable planning than aggregate indicators such as punctuality and are in fact a precondition for proper scheduling.

The practice of timetabling

Clarify roles and responsibilities

The Swedish Transport Administration should assume clear responsibility in ensuring that timetables maintain a high quality. Currently, there is some uncertainty if they should do so, or if that should be the domain of the train operating companies. As this requires some highly specialised expertise and resources, which not all companies possess, the baseline must however be that the infrastructure manager performs this role until the companies clearly express the interest and display the ability to do so. The infrastructure manager is also the only actor that can properly coordinate both traffic and maintenance projects from a multitude of companies. If some companies later decide that timetable planning is a core function for them, and invest in this capability, this responsibility can perhaps be shifted. As it stands, however, the only feasible solution is that the infrastructure manager assumes the responsibility.

Improve the tools

To enable the timetable planners to fulfil this responsibility, better tools are required. These tools should, among other things, be able to: (1) automate the allocation of margins, (2) solve the issues of rounding to whole minutes, (3) eliminate the practice of negative margins, (4) detect conflicting train movements, (5) display the topography, (6) suggest reasonable dwell times, (7) support track allocation at stations and (8) keep track of feedback from dispatchers. In order to make the feedback loops shorter and address small mistakes, it should also be possible to make minor changes to timetables during the year, without having to create entirely new train identification numbers. Either on a continual basis, or on one or two occasions during the year, before the next annual timetable takes effect. While each change might only affect one or two trains, and only marginally affect punctuality, it promotes a culture of constant, gradual improvements and fine-tuning, which is very important in the long term.

Systematic evaluation and improvement

Equally important is to make the evaluation of previous timetables and outcomes a routine part of the process and job description of timetable planners, and to allocate enough time for this. Many of the other recommendations we make regarding timetable planning free up time, which can then be used for these purposes. In line with shifting more attention towards evaluation, one key step is to improve the documentation of policies, strategies, guidelines and analyses, so that these things are clearly written down. Another part of this is to organise joint conferences and seminars where timetable planners can meet across regions, organisations, and sometimes countries, to share experiences, discuss policies, communicate with one another, and learn from research.

Seconds, not minutes

Finally, some of the problems we have identified in timetables and timetable planning occur because planning, operation and evaluation is done on the level of minutes, rather than seconds. That dwell times are considered as residuals is often due to planners wanting to have departure times at whole minutes. The same holds for many both positive and negative margins, caused by a belief that some scheduled times need to occur at whole minutes. The first leads to waste and the other causes delays, and in either case the timetable planners spend time and mental capacity on matters that do not create value, rather than spending time on creating a high-quality timetable. Using minutes rather than seconds also makes it more difficult for train drivers and dispatchers to operate in a good way, because they receive feedback about being ahead of or behind schedule only when the deviation has reached one or more minutes.

Some physical changes

Markings on platforms

Finally, we can recommend some physical changes in the infrastructure that would likely reduce delays. The first is to make physical markings on platforms where the trains and doors will stop. If train drivers stick consistently to these, it will become clear to the waiting passengers where they should stand. This makes the process of boarding and alighting faster and more reliable, both because it distributes passengers more evenly along the platform, and because those waiting to board are not standing in the way of those getting off the train. For trains with seat reservations, the benefits are even greater, as passengers who know where to board will do so more quickly, without having to move along the platform once the train has arrived. As small delays at stations are a very big part of the problem in Swedish railways, addressing the root cause in this way is an important step towards increasing overall punctuality.

Reduce the number of switches

Another measure that can be taken, is to gradually begin removing non-critical switches that are rarely used. They are statistically disproportionately connected to decreased punctuality, introduce more complexity, risk breaking down, and need to be maintained. This maintenance diverts resources from more useful infrastructure, which in turn breaks down more frequently than would otherwise be the case. By getting rid of these rarely used switches, there will be fewer elements that can break down, and more time and money left to maintain those that remain, in a virtuous circle, with less delays and higher punctuality as a result.

Improve the resiliency

Finally, a programme should be initiated to improve the resiliency of infrastructure with regard to weather and climate, particularly temperature and snow. The Swedish railways are currently very vulnerable to adverse weather, which will only become more frequent and extreme as a result of climate change. Types of measures include shielding signals and other electronics from direct sunlight, which can lead to overheating even during moderate air temperatures, limiting the number of switches so that snow and ice cause less problems and are easier to deal with, and making sure that the drainage is sufficient, to avoid flooding.

Some Reflections

Further research

More on dwell times

One of the contributions of this thesis is to show the importance of dwell times – a large part of all delays occur at station stops. Meanwhile, one of the key differences when interviewing timetable planners in Sweden and Japan is precisely the awareness of dwell times. Put together, this suggests that dwell times should receive considerably more attention in research as well as practice. This should cover the scheduling of appropriate dwell times, better understanding the different mechanisms that lead to delays at stations, coming up with measures to address these mechanisms and delays, and testing the effectiveness of these methods both in simulations and in practice.

Interactions and dispatching decisions

Interactions between trains is one of the factors most clearly linked to decreased punctuality in this thesis. Many of these interactions are due to some kind of dispatching decisions – holding a train back, shifting a train meeting from one station to another, etc. These interactions and decisions can be identified and evaluated using historical data such as that used in this thesis. This can even open a window into understanding dispatching strategies used in practice – an understanding which could improve the quality of dispatching and the punctuality of operations. The same methods can also inform simulation models about how dispatching is performed in practice, and thus help to further calibrate and validate dispatching modules.

Climate adaptation pathways

We have shown that the Swedish railways are not robust when it comes to variations in weather and are particularly vulnerable to high temperatures and extreme weather conditions. With climate change, these conditions will become increasingly prevalent, and the railways must adapt. There is a large body of research around this subject in other sectors, partially under the label of *adaptation pathways*, and there is a lot more to be done to this end in railway research. Understanding more precisely which components break down, how these breakdowns can be avoided, where the problems are the largest and where they are expected to increase the most, when measures can and must be taken, and how to avoid investing in the wrong things, are just some of the questions that can and must be addressed.

Reflections on methodology

The basic approach

The large amount of data used in this thesis provided a good opportunity for quantitative methods. With such large datasets they are able to detect and quantify even subtle influences on delays and punctuality and – with the help of visualisation – to describe the approximate shapes of these effects. Another strength of a quantitative and empirical approach is that it is possible to detect and evaluate both interactions of trains, and the various choices and strategies used by timetable planners. There are some difficulties, however. Visualising data, for instance, becomes difficult – on a practical level – when the number of observations range from hundreds of thousands to tens of millions.

In some of the papers where we studied a wide range of factors, it can also be difficult to disentangle the effects of one influence from another. While methods like regression help with this – and are to some extent made for this – the results can be difficult to visualise and interpret when the number of factors is large. It can also be quite difficult to establish precise mechanisms of causation or chains of events that are easy to explain to practitioners, using large amounts of data. It is one thing to establish that there is a statistically significant link between high temperatures and decreased punctuality, for instance, and quite another to explain why a certain train that ran on a given day was delayed to the extent that it was.

Truncation

One issue with train movement data in Sweden is that it is truncated to the level of minutes, such that the seconds are lost. This, in combination with the fact that the data are based on the occupation of signal blocks rather than the arrival to and departure from the platform, causes imprecision in the data. When run and dwell times are calculated from this data, however, these errors will tend to cancel out over large enough samples. One should take care when considering individual observations, however, as these imprecisions could mean that random variations of a few seconds give observations that vary by up to two minutes in length. These errors are not systematic, however, and with large enough samples, the estimates should not be biased to any real extent. The issues around truncation also become much less relevant when dealing with punctuality, than with individual observations.

Aggregation and punctuality

An additional benefit of working with punctuality rather than series of individual run and dwell time observations is that it dramatically reduces the number of observations – often by a factor of one hundred or more. This means that any calculations are much easier to perform. As the computational power increases, however, this aggregation into punctuality becomes less relevant, and it is preferable to work with more direct observations of delays. Partly because the increased number of observations makes it easier to determine whether effects are statistically significant or not, and partly because delays are a more precise and direct indicator of what is going on. This can result in a greater explanatory power of the estimated models.

The interviews

Interviewing planners was very rewarding. They were very frank and forthcoming, and provided good leads and new information that would not be possible to acquire in other ways. Transcribing the material was very time consuming, but being able to cross-examine the transcribed material, and return to it with new sets of questions, was very interesting and fruitful. A potential drawback with the approach is that it is based on the subjective and not necessarily accurate descriptions of the interviewees. Being able to check these against the actual data is very valuable. For instance, according to the planners themselves, the practice of giving trains negative margins on some line sections should not be very problematic at all, since they are compensated by positive margins on the next sections – but in practice, the data show that the compensation is not successful, and that the delays remain.

The interviewees

The question of who is interviewed is also very important, and interviewing planners at train operating companies, or dispatchers, would certainly result in a different description of the issues at hand. Naturally, the subjectivity of the interviewer and interpreter of the transcribed material are also involved, and to some extent this is unavoidable. This is the case with quantitative work as well, however, and in both cases, there is often quite a long process going from raw data to a finished analysis and a set of conclusions. On a personal level, I found that the qualitative approach gave quite a lot of interesting insights, and I believe that there is a lot more such work to be done in the railway sector.

Alternative methods

Moving on to alternative methods, simulation and optimisation are other commonly used approaches for research on timetable planning. They rely, however, on good delay data and descriptions of planning strategies. Knowledge and data on the sources or distributions of delays cannot be generated with these approaches themselves. Instead, it is useful to have a dialogue between the different approaches, where empirical work – such as that found in this thesis – can inform the more theoretical work on simulation and optimisation, and vice versa.

Delay cause data

On the empirical side of research, the main alternative is to base the research on so-called delay cause data. The benefit, compared to what has been presented in this thesis, is that the explicit causes of the delays are known – and coded in the data. There is thus less reliance on statistical relations, and on collecting and combining multiple datasets and sources. One drawback is that the delay cause data omits small delays – the exact threshold varies by country, but in Sweden delays smaller than three minutes in size are not coded. As has been shown in this thesis, small delays make up over half of all delay time – it is not prudent to simply omit them.

Other drawbacks are that the data on delay causes is already well known and utilised in the industry, and that there are well documented errors (of the order of 20%) in the coding of delays, as well as more philosophical criticism of how feasible it really is to determine one "cause" for a specific delay. A more fruitful approach might be to combine the two approaches – using alternative datasets and statistical tools to validate the delay cause data, and vice versa. Once this has been done, it might be possible use machine learning approaches to find patterns and classify even the small delays.

Qualitative methods

Another approach would be to perform more qualitative work. Timetable planning, and the railway sector in general, has rich processes and traditions to study using qualitative approaches with interviews, document studies, anthropological methods, and others. There are very important interfaces to study, between timetable planning and dispatching, drivers, maintenance, and more. These are quite difficult to approach from a quantitative perspective, and qualitative methods are thus the only real alternative. A risk with relying too heavily on interviews and such, however, is that the research and findings might be limited to what practitioners already know and consider to be problems. It is also difficult to quantify the scope of problems, without using quantitative methods.

Contributions of the thesis

The contribution to research

The contribution of this thesis to research has mainly been to present new empirical studies that: (1) describe and quantify how a range of factors are associated with delays, (2) evaluate the effects of timetable planning practice on train delays and punctuality, and (3) the overarching method and conceptual figures used to gather, connect, process and analyse vast amounts of data from different sources. The thesis is written in a time when big data is discussed and used seemingly everywhere, and my work has been a step in bringing this approach to railways and timetable planning.

The contribution to practitioners

To practitioners, the most significant contribution of the work has been to demonstrate the central importance of small dwell time delays – smaller than three minutes. In Sweden, these are often overlooked, partly because of a conscious but misguided strategy, partly due to lacking tools and routines. We have shown that, in practice, most delays are small and occur at stations. If punctuality is to improve noticeably, these delays must be addressed. The other contribution to practitioners is to show that a great deal can be done with timetable planning. Partly by reducing mistakes, both conscious and inadvertent, that systematically cause delays. Partly by further improving the capacity of timetables to absorb delays. We have shown that the distribution of margins is often more important than their overall size, and that the returns from increasing them quickly diminish, suggesting a possibility of reducing margins and travel times without the risk of delays increasing. The thesis also contains a few other suggested measures around stations that would help to reduce delays in a very cost-effective manner.

The contribution to the public

The main contribution to society at large is the increased knowledge about and attention to train delays – a problem which affects the daily lives of millions of people. If the recommendations in this thesis are implemented, the problem of delays should diminish significantly. Trains would not systematically be delayed at stations, and neither the summers nor the winters would cause nearly as much disruptions. The timetable would continue to get better year by year, without necessarily extending travel times, and punctuality would climb steadily. Perhaps not quite to the level achieved in countries like Japan, but the industry target of 95% is certainly within reach. This would be a great service to the public. It would raise the level of trust and confidence in the railway and serve to further increase the number and share of journeys made by train. It would also do a lot to reduce the annoyance and frustration that many people who already travel by train.

References

Ali, Abderrahman Ait, and Jonas Eliasson. 2019. "Railway Capacity Allocation : A Survey of Market Organizations , Allocation Processes and Track Access Charges." VTI Working paper 2019:1. Linköping: VTI.

Ali, Abderrahman Ait, Jennifer Warg, and Jonas Eliasson. 2019. "Pricing Commercial Train Path Requests Based on Societal Costs." VTI Working paper 2019:2. Linköping: VTI.

Andreasson, Rebecca, Anders A. Jansson, and Jessica Lindblom. 2018. "The Coordination between Train Traffic Controllers and Train Drivers: A Distributed Cognition Perspective on Railway." *Cognition, Technology and Work*. https://doi.org/https://doi.org/10.1007/s10111-018-0513-z.

Argyris, Chris, and Donald Schön. 1996. *Organizational Learning II: Theory Method and Practice*. Reading, Mass.: Addison-Wesley.

Avelino, Flor, M.C.G. te Brömmelstroet, and G Hulster. 2006. "The Politics of Timetable Planning: Comparing the Dutch to the Swiss." In *Proceedings of the Colloquium Vervoersplanologisch Speurwerk 2006*.

Baee, Sonia, Farshad Eshghi, S. Mehdi Hashemi, and Rayehe Moienfar. 2012. "Passenger Boarding/Alighting Management in Urban Rail Transportation." In *Proceedings of the 2012 Joint Rail Conference*. https://doi.org/10.1115/jrc2012-74102.

Banverket. 2000. "Riktlinjer För Tidtabellskonstruktion För Tåg På Statens Spåranläggningar." Internal document. Borlänge: Banverket.

Batley, Richard, Joyce Dargay, and Mark Wardman. 2011. "The Impact of Lateness and Reliability on Passenger Rail Demand." *Transportation Research Part E: Logistics and Transportation Review* 47 (1): 61–72. https://doi.org/10.1016/j.tre.2010.07.004.

Börjesson, Maria, and Jonas Eliasson. 2011. "On the Use of 'Average Delay' as a Measure of Train Reliability." *Transportation Research Part A: Policy and Practice* 45 (3): 171–84. https://doi.org/10.1016/j.tra.2010.12.002.

Buchmueller, S., U. Weidmann, and A. Nash. 2008. "Development of a Dwell Time Calculation Model for Timetable Planning." *WIT Transactions on the Built Environment* 103: 525–34. https://doi.org/10.2495/CR080511.

Carey, Malachy. 1998. "Optimizing Scheduled Times, Allowing for Behavioural Response." *Transportation Research Part B: Methodological* 32B (5): 329–42. https://doi.org/10.1016/S0191-2615(97)00039-8.

———. 1999. "Ex Ante Heuristic Measures of Schedule Reliability." *Transportation Research Part B: Methodological* 33 (7): 473–94. https://doi.org/10.1016/S0191-2615(99)00002-8.

Ceder, Avishai. 2001. “Public Transport Scheduling.” In *Handbook of Transport Systems and Traffic Control*, edited by K.J. Button and D.A. Hensher, 1st ed., 539–58. Oxford: Elsevier.

Ceder, Avishai, and Stephan Hassold. 2015. “Applied Analysis for Improving Rail-Network Operations.” *Journal of Rail Transport Planning and Management* 5 (2): 50–63. https://doi.org/10.1016/j.jrtpm.2015.06.001.

Cerreto, Fabrizio, Otto Anker Nielsen, Steven Harrod, and Bo Friis Nielsen. 2016. “Causal Analysis of Railway Running Delays.” In *11th World Congress on Railway Research (WCRR 2016)*.

Cheng, Yung-hsiang, and Yu-chun Tsai. 2014. “Train Delay and Perceived-Wait Time : Passengers ’ Perspective.” *Transport Reviews* 34 (6): 710–29. https://doi.org/10.1080/01441647.2014.975169.

D’Acierno, Luca, Marilisa Botte, Antonio Placido, Chiara Caropreso, and Bruno Montella. 2017. “Methodology for Determining Dwell Times Consistent with Passenger Flows in the Case of Metro Services.” *Urban Rail Transit* 3 (2): 73–89. https://doi.org/10.1007/s40864-017-0062-4.

DB Netz. 2017. “Network Statement 2018.” Frankfurt am Main: DB Netz.

Dewilde, Thijs, Peter Sels, Dirk Cattrysse, and Pieter Vansteenwegen. 2013. “Robust Railway Station Planning: An Interaction between Routing, Timetabling and Platforming.” *Journal of Rail Transport Planning and Management* 3 (3): 68–77. https://doi.org/10.1016/j.jrtpm.2013.11.002.

Dobney, K., C. J. Baker, A. D. Quinn, and L. Chapman. 2009. “Quantifying the Effects of High Summer Temperatures Due to Climate Change on Buckling and Rail Related Delays in South-East United Kingdom.” *Meteorological Applications* 16: 245–51. https://doi.org/10.1002/met.

Dollevoet, Twan. 2013. “Delay Management and Dispatching in Railways.” Doctoral thesis. Rotterdam: Erasmus University Rotterdam.

Fellows, R., and A. Liu. 2015. *Research Methods for Construction*. 4th ed. Hoboken, NJ: Wiley-Blackwell.

Ferranti, Emma, Lee Chapman, Susan Lee, David Jaroszweski, Caroline Lowe, Steve McCulloch, and Andrew Quinn. 2018. “The Hottest July Day on the Railway Network: Insights and Thoughts for the Future.” *Meteorological Applications* 25 (2): 195–208. https://doi.org/10.1002/met.1681.

Ferranti, Emma, Lee Chapman, Caroline Lowe, Steve McCulloch, David Jaroszweski, and Andrew Quinn. 2016. “Heat-Related Failures on Southeast England’s Railway Network: Insights and Implications for Heat Risk Management.” *Weather, Climate, and Society* 8 (2): 177–91. https://doi.org/10.1175/WCAS-D-15-0068.1.

Frauenfelder, Regula, Anders Solheim, Ketil Isaksen, Bård Romstad, Anita V. Dyrrdal, Kristine H. H. Ekseth, Alf Harbitz, et al. 2017. “Impacts of Extreme Weather Events on Transport Infrastructure in Norway.” *Natural Hazards and Earth System Sciences Discussions*, no. December: 1–24. https://doi.org/10.5194/nhess-2017-437.

Gestrelius, Sara, Martin Aronsson, Martin Forsgren, and Hans Dahlberg. 2012. “On the Delivery Robustness of Train Timetables with Respect to Production Replanning Possibilities.” In *2nd International Conference on Road and Rail Infrastructure*.

Gestrelius, Sara, Anders Peterson, and Martin Aronsson. 2019. “Timetable Quality from the Perspective of an Infrastructure Manager in a Deregulated Market: A Case Study of Sweden.” In *Proceedings of the 8th International Conference on Railway Operations Modelling and Analysis (ICROMA) - RailNorrköping*, 505–24.

Ghaemi, Nadjla, Aurelius A. Zilko, Fei Yan, Oded Cats, Dorota Kurowicka, and Rob M.P. Goverde. 2018. “Impact of Railway Disruption Predictions and Rescheduling on Passenger Delays.” *Journal of Rail Transport Planning and Management* 8 (2): 103–22. https://doi.org/10.1016/j.jrtpm.2018.02.002.

Gholami, Omid, and Johanna Törnquist Krasemann. 2018. “A Heuristic Approach to Solving the Train Traffic Re-Scheduling Problem in Real Time.” *Algorithms* 11 (4): 55. https://doi.org/10.3390/a11040055.

Gorman, Michael F. 2009. “Statistical Estimation of Railroad Congestion Delay.” *Transportation Research Part E: Logistics and Transportation Review* 45 (3): 446–56. https://doi.org/10.1016/j.tre.2008.08.004.

Goverde, and Ingo Arne Hansen. 2013. “Performance Indicators for Railway Timetables.” In *IEEE ICIRT 2013 - Proceedings: IEEE International Conference on Intelligent Rail Transportation*. https://doi.org/10.1109/ICIRT.2013.6696312.

Goverde, Robert Michael Petrus. 2005. “Punctuality of Railway Operations and Timetable Stability Analysis.” Doctoral thesis. Delft: Delft University of Technology.

Gummesson, Mats. 2018. “Tillsammans För Tåg i Tid Resultatrapport 2018.” Report. Stockholm: Järnvägsbranschens Samverkansforum.

Hagen, Mark van, and Niels van Oort. 2018. “Improving Railway Passengers Experience : Two Perspectives.” Presented at *Conference for Advanced Systems in Public Transport - CASPT 2018*.

Hansen, I. A. 2009. “Railway Network Timetabling and Dynamic Traffic Management.” *International Journal of Civil Engineering* 8 (1): 19–32.

Harris, Nigel G., Christian S. Mjøsund, and Hans Haugland. 2013. “Improving Railway Performance in Norway.” *Journal of Rail Transport Planning and Management* 3 (4): 172–80. https://doi.org/10.1016/j.jrtpm.2014.02.002.

Hawchar, Lara, Paraic C. Ryan, Owen Naughton, and Paul Nolan. 2018. “High-Level Climate Change Vulnerability Assessment of Irish Critical Infrastructure.” In *Civil Engineering Research in Ireland Conference*.

Heinz, Wiktoria. 2003. “Passenger Service Times on Trains.” Licentiate thesis. Stockholm: Royal Institute of Technology.

Hurk, Evelien van der. 2015. “Passengers, Information, and Disruptions.” Doctoral thesis. Rotterdam: Erasmus Rotterdam University.

International Union of Railways. 2000. “UIC Code 451-1 Timetable Recovery Margins to Guarantee Timekeeping.” Paris: Union International de Chemin de Fer.

———. 2004. “UIC CODE 406 Capacity 1st Edition.” Paris: Union International de Chemin de Fer.

Johnson, R. Burke, and Anthony J. Onwuegbuzie. 2007. “Mixed Methods Research: A Research Paradigm Whose Time Has Come.” *Educational Researcher* 33 (7): 14–26. https://doi.org/10.3102/0013189x033007014.

Kamizuru, Tomomi, Tomoyuki Noguchi, and Norio Tomii. 2015. "Dwell Time Estimation by Passenger Flow Simulation on a Station Platform Based on a Multi-Agent Model." In *RailTokyo2015*, 1–15.

Kecman, Pavle, and Rob M. P. Goverde. 2015. "Online Prediction of Train Event Times." *IEEE Transactions on Intelligent Transportation Systems* 16 (1): 465–74.

Khoshniyat, Fahimeh, and Anders Peterson. 2015. "Robustness Improvements in a Train Timetable with Travel Time Dependent Minimum Headways." Presented at *RailTokyo2015*.

Koetse, Mark J., and Piet Rietveld. 2009. "The Impact of Climate Change and Weather on Transport: An Overview of Empirical Findings." *Transportation Research Part D: Transport and Environment* 14 (3): 205–21. https://doi.org/10.1016/j.trd.2008.12.004.

Kottenhoff, Karl, and Camilla Byström. 2010. "När Resenärerna Själva Får Välja." Report. Stockholm: The Royal Institute of Technology.

Kvale, Steinar. 1997. *Den Kvalitativa Forskningsintervjun*. Lund: Studentlitteratur.

Lamorgese, Leonardo, Carlo Mannino, Dario Pacciarelli, and Johanna Törnquist Krasemann. 2018. "Train Dispatching." In *Handbook of Optimization in the Railway Industry*, edited by R. Borndörfer, 268:265–83. Springer. https://doi.org/10.1007/978-3-319-72153-8_12.

Li, Dewei, Winnie Daamen, and Rob M.P. Goverde. 2015. "Estimation of Train Dwell Time at Short Stops Based on Track Occupation Event Data." In *6th International Conference on Railway Operations Modelling and Analysis*, 1–20.

Lidén, Tomas. 2018. *Concurrent Planning of Railway Maintenance Windows and Train Services*. *Linköping Studies in Science and Technology. Dissertations NV - 1957*. https://doi.org/10.3384/diss.diva-152491.

Lindfeldt, Anders. 2015. "Railway Capacity Analysis: Methods for Simulation and Evaluuation of Timetables, Delays and Infrastructure." Stockholm: The Royal Institute of Technology.

Liu, Kai, Ming Wang, Yinxue Cao, Weihua Zhu, and Guiling Yang. 2018. "Susceptibility of Existing and Planned Chinese Railway System Subjected to Rainfall-Induced Multi-Hazards." *Transportation Research Part A: Policy and Practice* 117 (December): 214–26. https://doi.org/10.1016/j.tra.2018.08.030.

Longo, G., and G. Medeossi. 2012. "Enhancing Timetable Planning with Stochastic Dwell Time Modelling." *WIT Transactions on the Built Environment* 127: 461–71. https://doi.org/10.2495/CR120391.

Marković, Nikola, Sanjin Milinković, Konstantin S. Tikhonov, and Paul Schonfeld. 2015. "Analyzing Passenger Train Arrival Delays with Support Vector Regression." *Transportation Research Part C: Emerging Technologies* 56: 251–62. https://doi.org/10.1016/j.trc.2015.04.004.

Mattsson, Lars Göran, and Erik Jenelius. 2015. "Vulnerability and Resilience of Transport Systems - A Discussion of Recent Research." *Transportation Research Part A: Policy and Practice* 81: 16–34. https://doi.org/10.1016/j.tra.2015.06.002.

Nagy, Enikő, and Csaba Csiszár. 2015. "Analysis of Delay Causes in Railway Passenger Transportation." *Periodica Polytechnica Transportation Engineering*. https://doi.org/10.3311/pptr.7539.

Nelldal, Bo-Lennart, Olov Lindfeldt, and Anders Lindfeldt. 2009. “Kapacitetsanalys Av Järnvägsnätet i Sverige Delrapport 1 Hur Många Tåg Kan Man Köra? En Analys Av Teoretisk Och Praktisk Kapacitet.” Report. Stockholm: The Royal Institute of Technology.

Network Rail. 2018. “Network Code.” London: Network Rail.

Nicholson, Gemma L., David Kirkwood, Clive Roberts, and Felix Schmid. 2015. “Benchmarking and Evaluation of Railway Operations Performance.” *Journal of Rail Transport Planning and Management* 5 (4): 274–93. https://doi.org/10.1016/j.jrtpm.2015.11.004.

Nie, Lei, and Ingo Arne Hansen. 2005. “System Analysis of Train Operations and Track Occupancy at Railway Stations.” *European Journal of Transport and Infrastructure Research* 5 (1): 31–54.

Nilsson, Jan-Eric, Roger Pyddoke, and Inge Vierth. 2015. “Regress - En God Idé i Järnvägssektorn?” Report. Linköping: VTI.

Noland, Polak J.W. 2002. “Travel Time Variability: A Review of Theoretical and Empirical Issues.” *Transport Reviews* 22 (1): 39–54. https://doi.org/10.1080/01441640010022456.

Nyström, Birre. 2008. “Aspects of Improving Punctuality.” Doctoral thesis. Luleå: Luleå University of Technology.

Olink, H., and G. Scheepmaker. 2017. “RailSys Use by RailwayLab.” Oral presentation at the Royal Institute of Technology.

Olsson, N., A. H. Halse, P. M. Hegglund, M. Killi, A. Landmark, and A. Seim. 2015. *Punktlighet i Jernbanen – Hvert Sekund Teller*. Oslo: SINTEF Akademisk Forlag.

Olsson, Nils O.E., and Hans Haugland. 2004. “Influencing Factors on Train Punctuality - Results from Some Norwegian Studies.” *Transport Policy* 11 (4): 387–97. https://doi.org/10.1016/j.tranpol.2004.07.001.

Pachl, J. 2002. *Railway Operation and Control*. Mountlake Terrace: VTD Rail Publishing.

Palin, Erika J., Adam A. Scaife, Emily Wallace, Edward C.D. Pope, Alberto Arribas, and Anca Brookshaw. 2016. “Skillful Seasonal Forecasts of Winter Disruption to the U.K. Transport System.” *Journal of Applied Meteorology and Climatology* 55 (2): 325–44. https://doi.org/10.1175/JAMC-D-15-0102.1.

Parbo, Jens, Otto Anker Nielsen, and Carlo Giacomo Prato. 2016. “Passenger Perspectives in Railway Timetabling: A Literature Review.” *Transport Reviews* 36 (4): 500–526. https://doi.org/10.1080/01441647.2015.1113574.

Pedersen, Timothy, Thomas Nygreen, and Anders Lindfeldt. 2018. “Analysis of Temporal Factors Influencing Minimum Dwell Time Distributions.” *WIT Transactions on the Built Environment* 181: 447–58. https://doi.org/10.2495/CR180401.

Peterson, A. 2012. “Towards a Robust Traffic Timetable for the Swedish Southern Mainline.” In *WIT Transactions on the Built Environment*, 127:473–84. https://doi.org/10.2495/CR120401.

ProRail. 2016. “Network Statement 2017.” Utrecht: ProRail.

Qi, Zhang, Han Baoming, and Li Dewei. 2008. "Modeling and Simulation of Passenger Alighting and Boarding Movement in Beijing Metro Stations." *Transportation Research Part C: Emerging Technologies* 16 (5): 635–49. https://doi.org/10.1016/j.trc.2007.12.001.

Rietveld, P., F. R. Bruinsma, and D. J. Van Vuuren. 2001. "Coping with Unreliability in Public Transport Chains: A Case Study for Netherlands." *Transportation Research Part A: Policy and Practice* 35 (6): 539–59. https://doi.org/10.1016/S0965-8564(00)00006-9.

Rietveld, Piet. 2007. "Six Reasons Why Supply-Oriented Indicators Systematically Overestimate Service Quality in Public Transport." *Transport Reviews* 25 (3): 319–28. https://doi.org/10.1080/0144164042000335814.

Rudolph, Raphaela. 2003. "Allowances and Margins in Railway Scheduling." In *Proceedings of the World Congress on Railway Research*.

Sa'adin, Binti Sazrul Leena, Sakdirat Kaewunruen, and David Jaroszweski. 2016. "Risks of Climate Change with Respect to the Singapore-Malaysia High Speed Rail System." *Climate* 4 (4): 65. https://doi.org/10.3390/cli4040065.

Sandblad, Bengt, Arne W. Andersson, and Simon Tschirner. 2015. "Information Systems for Cooperation in Operational Train Traffic Control." In *Procedia Manufacturing*, 3:2882–88. The Authors. https://doi.org/10.1016/j.promfg.2015.07.793.

SBB. 2018. "SBB Facts and Figures 2018." Bern: SBB-CFF-FFS. https://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx.

Seriani, Sebastian, and Rodrigo Fernandez. 2015. "Pedestrian Traffic Management of Boarding and Alighting in Metro Stations." *Transportation Research Part C: Emerging Technologies* 53: 76–92. https://doi.org/10.1016/j.trc.2015.02.003.

Solinen, Emma, Gemma Nicholson, and Anders Peterson. 2017. "A Microscopic Evaluation of Railway Timetable Robustness and Critical Points." *Journal of Rail Transport Planning and Management* 7 (4): 207–23. https://doi.org/10.1016/j.jrtpm.2017.08.005.

Stenström, Christer, Aditya Parida, Jan Lundberg, and Uday Kumar. 2015. "Development of an Integrity Index for Benchmarking and Monitoring Rail Infrastructure: Application of Composite Indicators." *International Journal of Rail Transportation* 3 (2): 61–80. https://doi.org/10.1080/23248378.2015.1015220.

Stockholms läns landsting. 2017. "Trafiknämnden/Trafikförvaltningen Årsrapport 2017." Annual report. Stockholm: Stockholms läns landsting.

Thaduri, Adithya, Diego Galar, and Uday Kumar. 2015. "Railway Assets: A Potential Domain for Big Data Analytics." *Procedia Computer Science* 53 (1): 457–67. https://doi.org/10.1016/j.procs.2015.07.323.

Trafikanalys. 2016. "Punktlighet På Järnväg 2015." Report. Stockholm: Trafikanalys.

———. 2018a. "Bantrafik 2017." Report. Stockholm: Trafikanalys.

———. 2018b. "Transportarbete 2000–2017." Report. Stockholm: Trafikanalys.

———. 2019. "Punktlighet På Järnväg 2018." Report. Stockholm: Trafikanalys.

Trafikverket. n.d. “Lupp v2.3. UH/Järnvägssystem. Systemöversikt.” Internal document. Borlänge: Trafikverket.

———. 2015a. “Network Statement 2017.” Borlänge: Trafikverket.

———. 2015b. “Noder i Järnvägssystemet.” Internal document. Borlänge: Trafikverket.

———. 2017a. “Järnvägens Kapacitetsutnyttjande 2016.” Report. Borlänge: Trafikverket.

———. 2017b. “Riktlinjer Täthet Mellan Tåg.” Internal document. Borlänge: Trafikverket.

———. 2018a. “Analysmetod Och Samhällsekonomiska Kalkylvärden För Transportsektorn: ASEK 6.1.” Report. Borlänge: Trafikverket.

———. 2018b. “Lupp Uppföljningssystem.” System Och Verktyg - Förvaltning Och Underhåll. 2018. https://www.trafikverket.se/tjanster/system-och-verktyg/forvaltning-och-underhall/Lupp-uppfoljningssystem/.

———. 2018c. “Tillsammans För Tåg i Tid Resultatrapport 2017.” Report. Borlänge: Trafikverket.

———. 2018d. “Trafikledning.” För Dig i Branschen - Järnväg. 2018. https://www.trafikverket.se/for-dig-i-branschen/jarnvag/Trafikledning/.

———. 2019. “Trafikverkets Årsredovisning 2018.” Annual report. Borlänge: Trafikverket.

Veiseth, M., N. Olsson, and I.A.F. Saetermo. 2007. “Infrastructure’s Influence on Rail Punctuality.” *WIT Transactions on the Built Environment* 96 (August 2007): 481–90. https://doi.org/10.2495/UT070451.

Vekas, Peter, M.H. van der Vlerk, and Willem K. Klein Haneveld. 2012. “Optimizing Existing Railway Timetables by Means of Stochastic Programming,” In: *Stochastic Programming E-print Series*. 8.

Vromans, Michiel. 2005. “Reliability of Railway Systems.” Doctoral thesis. Rotterdam: Erasmus University Rotterdam.

Wallander, Jouni, and Miika Mäkitalo. 2012. “Data Mining in Rail Transport Delay Chain Analysis.” *International Journal of Shipping and Transport Logistics*. https://doi.org/10.1504/ijstl.2012.047492.

Wang, Ying. 2018. “Incorporating Weather Impact in Railway Traffic Control.” Doctoral thesis. Leeds: University of Leeds.

Watson, Robert. 2008. “Train Planning in a Fragmented Railway: A British Perspective.” Doctoral thesis. Loughborough: Loughborough University.

Welch, B. L. 1947. “The Generalization of Student’s’ Problem When Several Different Population Variances Are Involved.” *Biometrika* 34 (2): 28–35.

Wiggenraad, P. B. L. 2001. “Alighting and Boarding Times of Passengers at Dutch Railway Stations.” Report. Delft: Delft University of Technology.

Wiklund, Mats. 2006. “Förebyggande Underhållsåtgärders Effekt På Järnvägstransportsystemets Sårbarhet Försök Med Delfimetoden.” VTI Notat 23-2006. Report. Linköping: VTI.

Xia, Yuanni, Jos N. Van Ommeren, Piet Rietveld, and Willem Verhagen. 2013. “Railway Infrastructure Disturbances and Train Operator Performance: The Role of Weather.” *Transportation Research Part D: Transport and Environment* 18 (1): 97–102. https://doi.org/10.1016/j.trd.2012.09.008.

Xu, Peijuan, Francesco Corman, and Qiyuan Peng. 2016. “Analyzing Railway Disruptions and Their Impact on Delayed Traffic in Chinese High-Speed Railway.” In *IFAC-PapersOnLine*, 49:84–89. Elsevier B.V. https://doi.org/10.1016/j.ifacol.2016.07.015.

Yabuki, H., Y. Takatori, M. Koresawa, and N. Tomii. 2018. “Increasing Robustness of Timetables by Deliberate Operation of Trains to Shorten Headways.” *International Journal of Transport Development and Integration* 1 (3): 432–41. https://doi.org/10.2495/tdi-v1-n3-432-441.

Yuan, Jianxin, and Ingo Arne Hansen. 2002. “Punctuality of Train Traffic in Dutch Railway Stations” *Proceedings of the International Conference on Traffic and Transportation Studies (ICTTS) 2002*: 522–29. https://doi.org/10.1061/40630(255)73.

Yuan, Jianxin, and Ingo Arne Hansen. 2008. “Closed Form Expressions of Optimal Buffer Times between Scheduled Trains at Railway Bottlenecks.” In *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 675–80. IEEE. https://doi.org/10.1109/ITSC.2008.4732539.

Zakeri, Ghazal, and Nils O.E. Olsson. 2017. “Investigation of Punctuality of Local Trains - The Case of Oslo Area.” *Transportation Research Procedia* 27 (2214): 373–79. https://doi.org/10.1016/j.trpro.2017.12.080.

List of Included Papers

Paper 1. Palmqvist, C.-W., Olsson, N. O. E. & Hiselius, L. (2017). Delays for passenger trains on a regional railway line in southern Sweden. International Journal of Transport Development and Integration, 1(3), 421–431.

Paper 2. Palmqvist, C.-W., Olsson, N. O. E. & Hiselius, L. (2017). Some influencing factors for passenger train punctuality in Sweden. International Journal of Prognostics and Health Management.

Paper 3. Palmqvist, C.-W., Olsson, N. O. E. & Hiselius, L. (2017). An Empirical Study of Timetable Strategies and Their Effects on Punctuality. Proceedings of the 7th International Conference on Railway Operations Modelling and Analysis – RailLille 2017 (peer-reviewed).

Paper 4. Palmqvist, C.-W., Olsson, N. O. E. & Winslott Hiselius, L. (2018). The Planners' Perspective on Train Timetable Errors in Sweden. Journal of Advanced Transportation, 2018, 1–17.

Paper 5. Palmqvist, C.-W., Tomii, N. & Ochiai, Y. (2019). Dwell Time Delays for Commuter Trains in Stockholm and Tokyo. In Press. 8th International Conference on Railway Operations Modelling and Analysis – RailNorrköping 2019 (peer-reviewed).

Contribution to papers

For Papers 1-4 I was the main author, responsible for the research questions, collecting and processing both quantitative and qualitative data, as well as most references. I designed and performed the analyses, in dialogue with the co-authors, drew conclusions and wrote the papers. The co-authors helped in finding supplementary literature, in discussing the analyses, and in structuring and editing the papers. For Paper 5 I was also the main author, responsible for the research question, processing and analysing the data, as well as writing the paper. The co-authors provided access to, and explained, the Japanese data. They also helped to find literature on Japanese railways, explained the policies and practices of Japanese railway companies, and assisted in fact-checking the manuscript.