Technical University of Denmark

Public transport optimisation emphasising passengers' travel behaviour.

Jensen, Jens Parbo; Nielsen, Otto Anker; Prato, Carlo Giacomo

Publication date:
2015

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

Public transport optimisation emphasising passengers' travel behaviour

PhD Thesis

Jens Parbo
2015

DTU Transport
Department of Transport

Public transport optimisation emphasising passengers' travel behaviour

PhD thesis

Jens Parbo

Department of Transport

Technical University of Denmark

Supervisors:

Professor Otto Anker Nielsen

Department of Transport

Technical University of Denmark

Professor Carlo Giacomo Prato

Department of Transport

Technical University of Denmark

Kgs. Lyngby 2015

Parbo, J., (2015). Public transport optimisation emphasising passengers’ travel behaviour. PhD thesis. Department of Transport, Technical University of Denmark, Kgs. Lyngby

Preface

This PhD thesis entitled ”Public transport optimisation emphasising passengers’ travel behaviour” is submitted to fulfil the requirements for obtaining a PhD degree at the Department of Transport, Technical University of Denmark. The PhD study has been supervised by Professor Otto Anker Nielsen and Professor Carlo Giacomo Prato.

The thesis consists of the following papers:

Parbo et al. (2014):	Parbo, J., Nielsen, O. A., & Prato, C. G. (2014). User perspectives in public transport timetable optimisation. *Transportation Research Part C: Emerging Technologies*, *48*, 269-284.
Parbo et al. (2015a):	Parbo, J., Nielsen, O. A., & Prato, C. G. (2015). Passenger perspectives in railway timetabling: A literature review (Accepted for publication in *Transport Reviews*)
Parbo et al. (2015b):	Parbo, J., Nielsen, O. A., & Prato, C. G. (2015). Adapting stopping patterns in complex railway networks to reduce passengers’ travel time (resubmitted after second round of review to *Transportation Research Part C: Emerging Technologies*)
Parbo et al. (2015c):	Parbo, J., Nielsen, O. A., & Prato, C. G. (2015). Improving passenger oriented line planning for high frequent railway networks (to be Submitted to the special issue: *“Integrated optimization models and algorithms in rail planning and control” of Transportation Research Part C: Emerging Technologies*)
Parbo & Lam (2015):	Parbo, J. & Lam, W.H.K. (2015). Modelling capacity degradability of a transit network (to be submitted to the *Journal of Public Transportation*)
Sels et al. (2015)	Sels, P., Meisch, K., Møller, T., Parbo, J., Dewilde, T., Cattrysse, D., & Vansteenwegen, P. Towards a better train timetable for Denmark reducing total expected passenger time. (Submitted to the journal *Public Transport*)

Jens Parbo Jensen, November, 2015

Acknowledgements

I owe a great thanks to my PhD supervisors Carlo Giacomo Prato and Otto Anker Nielsen for their guidance and supervision during the past three years. Thank you very much for giving me the freedom I needed to grow as a PhD student. Thanks to Ulla Steen Salado-Jimena for correcting my English grammar and language mistakes in the articles. Thanks to Karen Uhrup Poulsen for helping me with all the administrative tasks and especially the many travel bookings. Thanks to my dear colleagues with whom I have several great memories, both some that are related to transportation research but definitely also some that are not.

In the spring 2015, I visited Professor William Lam at the Polytechnic University of Hong Kong. Thank you for the great hospitality, your academic guidance and last but not least our informal conversations about different aspects of life and various career paths.

The Danish Strategic Research Council is acknowledged for having provided funds to make this PhD study possible under the ”Robustness in Railway OperationS”, abbreviated RobustRailS research project.

Finally, I would like to thank my parents for their constant support during the three years. I owe my sincerest thanks to my two brothers who have listened to my frustrations several times during our evening walks around the lakes in Copenhagen.

Summary

Passengers in public transport complaining about their travel experiences are not uncommon. This might seem counterintuitive since several operators worldwide are presenting better key performance indicators year by year. The present PhD study focuses on developing optimisation algorithms to enhance the operations of public transport while explicitly emphasising passengers' travel behaviour and preferences.

Similar to economic theory, interactions between supply and demand are omnipresent in the context of public transport operations. In public transport, the demand is represented by the passengers and their desire to complete particular journeys, while the supply is the transit network and its characteristics. Changing the supply (e.g. by changing line plan configuration, stopping patterns or the timetable itself), thus makes the demand adapt accordingly. Acknowledging the interaction between supply and demand is important when transit operations are planned but also when performance is evaluated. Assessing public transport performance merely by measuring vehicle punctuality would provide an unfair picture of the level of service experienced by these passengers. The unfair picture can be explained by the fact that passenger delays are often significantly larger than the vehicle delays responsible for the passengers to be late e.g. because passengers on a slightly delayed train may experience a large delay if they miss their desired connection. To overcome the discrepancy between the published performance measures and what passengers actually experience, a large academic contribution of the current PhD study is the explicit consideration of passengers' travel behaviour in optimisation studies and in the performance assessment.

Besides the explicit passenger focus in transit planning, also the applicability to real large-scale network has been a main focus of the current thesis. Consequently, heuristic (i.e. not exact) methods are developed. The PhD study contributes to the state-of-the-art by proposing

(i) A literature review outlining the discrepancy between planners, who focus on the vehicle operations and publish fixed vehicle schedules and, on the other hand, passengers, who look not only at the schedules but also at the entirety of their journey from the access to the waiting, the on-board travel, the transfers and the egress.
(ii) A metaheuristic algorithm to enhance the line plan configuration of a high frequent transit network explicitly taking into account passengers' travel behaviour.
(iii) A heuristic algorithm to optimise stopping patterns in a railway network where passengers' adapted stop-to-stop path choice is considered explicitly.
(iv) A metaheuristic algorithm minimising passengers' transfer waiting time by changing vehicle departure times from the initial stop, again passengers' route choice behaviour is considered.
(v) A methodological framework is proposed to assess the resilience of a transit network from the passengers' perspective.

Empirical evidence indicates that passengers give more importance to travel time certainty than travel time reductions as they associate an inherent disutility with travel time uncertainty. This disutility may broadly be interpreted as an anxiety cost for the need for having contingency plans in case of disruptions, and may be looked at as the motivator for delay-robust railway timetables. Interestingly, passenger oriented optimisation studies considering robustness in railway planning typically limit their emphasis on passengers to the consideration of transfer maintenance. Clearly, passengers' travel behaviour is more complex and multi-faceted, thus several other aspects should be considered as becoming more and more evident from passenger surveys identifying passengers' preferences when using transit systems. This literature review and in

particular the finding that passengers' path choice is rarely considered in the operations planning was the main motivation for the papers (ii)-(iv).

In figure 1 the steps in the planning of transit operations are outlined along with the planning horizon. The arrows indicate that the outcome of a former step serves as input to a subsequent step. The current PhD study is focused around *Network Route Design* and *Timetable Development*. Although *Network Route Design* typically belongs to the strategic planning level, we approach the line planning and skip-stop planning on the tactical planning level, thereby making it possible to approximate passengers' travel choice with higher certainty. This is done by formulating bi-level optimisation problems, where the upper level solves the particular optimisation problem given passengers' route choice while the lower level derives passengers' route choice based on the updated network characteristics defined by the upper level. Due to its inherent complexity, these bi-level minimisation problems are extremely difficult to solve mathematically, since the analytical optimisation problem itself often is either non-convex non-linear or a mixed-integer linear problem, with passenger flows defined by the route choice model, where the route choice model is a non-linear non-continuous mapping of the timetable. Therefore, the bi-level optimisation problems are solved heuristically. To speed up the convergence of the bi-level algorithms, the lower level problem is incorporated in the upper level problem formulation. Integrating the upper level and lower level makes the algorithm converge faster compared to the case where the two problems are solved sequentially without taking into account interdependencies.

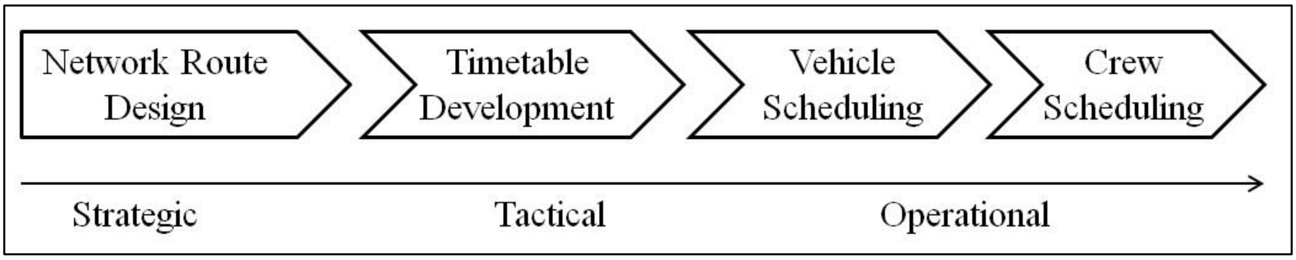


Figure 1 - Planning public transport

The PhD study develops a metaheuristic algorithm to adapt the line plan configuration in order better to match passengers' travel demand in terms of transfers as well as their waiting time experienced at boarding and transfers stations, respectively. The approach is based on swapping one part of a railway line with one part of another railway line at a station where the two lines meet. To search the solution space intelligently, a tabu search framework is applied to optimise the line plan configuration, while a passenger transit assignment model finds passengers' adapted route choices. The bi-level algorithm is validated on the suburban railway network in the Greater Copenhagen area in Denmark. Applying the improving bi-level passenger oriented line planning algorithm to this network yields a reduction of 3.83 % in railway passengers' number of transfers and 3.88 % in their waiting time.

Another part of the *Network Route Design* is determining stopping patterns on transit lines. Travel time reductions in railways are typically costly and achieved through investments in rolling stock or infrastructure. Skipping stops, on the other hand, is a cost-effective way to reduce in-vehicle travel time for on-board passengers and at the same time reduce the heterogeneity of the railway operations, which reduces the risk of knock-on delays. A passenger assignment yields passenger flows, which serve as input to the skip-stop optimisation. The updated stopping patterns and the reduced in-vehicle times then serve as input in the subsequent route choice calculation. The bi-level approach is applied to the suburban railway network in the

Greater Copenhagen area in Denmark and yields a 5.48 % reduction in in-vehicle time, while the number of transfers and the transfer waiting time increase by 1.38 % and 1.60 %, respectively.

Timetable Development is addressed by proposing a heuristic solution approach addressing the timetable optimisation from the passengers' perspective. The idea is to change bus lines' departure time from the initial station, and thereby reducing the waiting time passengers experience at any of the particular bus line's transfer stops. The offset changing heuristic is built on a Tabu Search framework, which is applied for its superiority for the particular problem type and application, but also the ability of the algorithm to escape local minima, which is important in order not to let the existing timetable affect the final outcome. In the developed passenger oriented timetable optimisation heuristic (bi-level), the lower level passenger transit assignment yields passengers' travel behaviour and thereby also their transfer choices. The solution approach is applied to the public transport network in Denmark yielding a 5.08 % reduction in transfer waiting time.

The three contributions outlined above (i.e. line plan optimisation, skip-stop optimisation and timetable optimisation) all focus on optimising operational aspects of transit services with special regards to travel demand. It is important to emphasise that the improvements obtained by applying the three different optimisation models all are achieved without any investment costs. The only costs related to the improved transit operations are administrative costs such as marketing, printing new timetables etc.

The last contribution of this PhD-study focuses on assessing the resilience of a transit network from the passengers' perspective. In this paper, a model to assess the capacity degradability of a transit network is developed. The capacity degradability of a transit network is described as the number of individual train runs that can be cancelled without violating the in-vehicle capacity constraints on the remaining trains. This is practically useful for operational and tactical planning purposes e.g. in the case where a shortage of rolling stock occurs and some trips need to be cancelled. To take the interaction between supply and demand into account explicitly and to be able to derive passengers' travel behaviour, a bi-level model is applied. In the bi-level model, the upper level determines the maximum degradable capacity by cancelling the individual train runs with the minimum maximum track segment load, while the lower level derives the individual vehicle loads by a schedule-based passenger assignment model. The explicit consideration of passengers' travel behaviour is important to ensure the validity of the results, and thus also the real-life applicability.

The capacity degradability model is tested on the transit network in the Greater Copenhagen area. Only individual train runs from the suburban railway network are subject to cancellations. From the case study it is concluded that this particular network is very resilient towards run cancellations on the suburban railway network since cancelling several transit runs only has a minor impact on passengers' travel experience. However, the inconvenience passengers may have felt due to the forced changes in path choice and in-vehicle crowding is not explicitly considered as a part of their generalised travel cost. Thereby, the actual inconvenience felt by the passengers as a result of the cancelled runs, may have been underestimated.

Summarising, the PhD study has given contributions to the state-of-the-art of several planning tasks that transit operators face by emphasising passengers' adapted travel behaviour in order for the operations to be as passenger oriented as possible. All methodologies are tested on large-scale transit networks and proved to enhance passengers' travel experience. Thus by applying these methodologies in a real-life context, passengers would have a faster and less uncertain journey from origin to destination.

Dansk resumé

Det er ikke uset at passagerer klager over rejser i den kollektive transport. Det kan dog virke paradoksalt når flere trafikoperatører verden over offentliggør stadig bedre tal for deres punktlighed og pålidelighed. Nærværende PhD studie fokuserer på udviklingen af metoder der skal optimere driften af kollektiv transport. I disse metoder vil passagerernes rejsemønstre blive betragtet eksplicit.

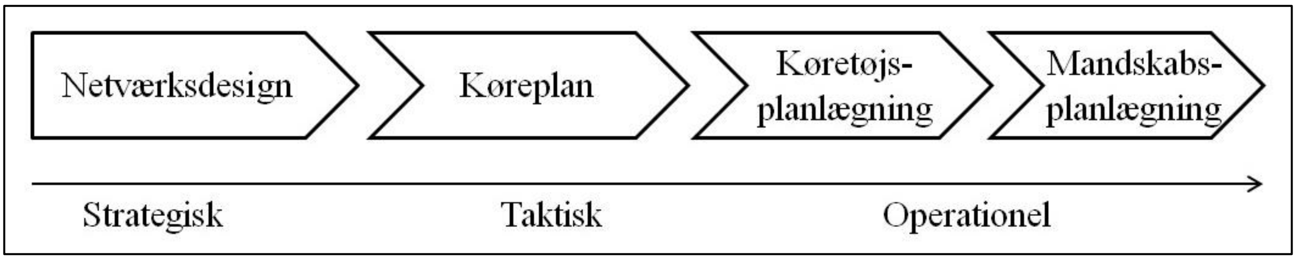
Ligesom indenfor økonomisk teori eksisterer der en klar sammenhæng mellem udbud og efterspørgsel når kollektiv transport betragtes. I kollektiv transport dækker efterspørgslen over passagererne og deres vilje til at gennemføre bestemte rejser. Udbuddet dækker og selve det kollektive transport netværk og dets karakteristika. Ændres udbuddet, eksempelvis ved at ændre på linjeplanerne, standsningsmønstrene eller køreplanen, tilpasser efterspørgslen sig tilsvarende. Dette forhold er væsentligt at tage højde for når planlægningen af kollektiv trafik foretages eller når kvaliteten af det reelt udførte evalueres. Vurderes den kollektive trafik kun på om de kollektive transportmidler ankommer rettidigt vil der således ikke kunne sættes lighedstegn til passagerernes rejseoplevelser. Den manglende sammenhæng bunder i at passagerernes forsinkelser ofte er større end det enkelte transportmiddels forsinkelse. Dette ses eksempelvis hvis passagerer kommer for sent til et skift som følge af en lille togforsinkelse. I det tilfælde vil køretøjet kun opleve en ubetydelig forsinkelse mens passageren er tvunget til at vente på det næstkommende tog. I nærværende PhD studie er det store akademiske bidrag udviklingen af optimeringsmodeller, der eksplicit tager højde for passagerernes tilpassede rutevalg som følge af ændringer i den kollektive transport.

Foruden det eksplicitte passagerfokus, så har udviklingen af heuristiske modeller, der kan anvendes på virkelige kollektive transport netværk også udgjort en stor del af nærværende PhD studie. Et PhD studie, der bidrager til den nyeste forskning indenfor kollektiv transport planlægning gennem det følgende

- (i) En gennemgang af eksisterende litteratur, der anskueliggør forskellen mellem operatørernes planlægning, der ofte har et ensidet fokus på de enkelte transportmidler, og så på den anden side passagererne, der betragter rejsen fra A til B som en helhed.
- (ii) En algoritme til at optimere linjeplanerne i et højfrekvent urbant kollektivt transport netværk.
- (iii) En algoritme til optimering af standsningsmønstre i et jernbanenetværk.
- (iv) En algoritme til at minimere passagerernes ventetid når der skiftes mellem to kollektive transportmidler. Algoritmen tilpasser afgangsminuttallet for de enkelte køretøjer.
- (v) En detaljeret metode udviklet til evaluere hvor meget kapaciteten af den kollektive kan blive forringet før antallet af passagerer i det enkelte køretøj overstiger dets egentlige kapacitet.

Forskning i passagerers præferencer viser at rejsetidsvariationer opleves som en dobbelt så stor gene som rejsetiden i sig selv. Den oplevede gene udspringer af det irritationsmoment der opstår når rejsetiden er uvis og behovet for alternative ruter må granskes. På baggrund af dette er det vigtigt, at operatørerne forsøger at gøre køreplanerne robuste overfor forsinkelser. Eksisterende studier i opbygningen af robuste køreplaner i kollektiv transport har ofte kun fokus på at sikre gennemførelsen af de planlagte skift. Som fremhævet før betragter passagererne den samlede rejse, hvorfor det så vidt som muligt er nødvendigt at medtage alle relevante aspekter når driften planlægges. Litteraturgennemgangen og særligt den opdagelse, at passagerernes samlede rejse sjældent betragtes når optimeringsalgoritmer udvikles til brug i kollektiv transport udgjorde den primære motivation for studierne (ii)-(iv).

Af figur 1 ses de forskellige trin der udgør rygraden af den kollektive transport planlægning. Pilene indikerer den sekventielle proces, hvor resultatet af forrige trin udgør beslutningsgrundlaget for et senere trin. Nærværende PhD studie fokuserer på netværksdesign og køreplansdesign. Om end netværksdesign traditionelt er et strategisk planlægningsproblem, så betragtes problemet på det taktiske planlægningsniveau i dette studie. Den kortere planlægningshorisont betyder, at der kan laves nogle antagelser omkring passagerernes rutevalg som normalt vil være forbundet med større usikkerhed ved en længere tidshorisont. Konkret betragtes passagerernes tilpassede rutevalg ved at formulere optimeringsproblemerne i to niveauer, hvor det øvre niveau forsøger at optimere driften af den kollektive ved at indføre bestemte ændringer. I det nedre niveau beregnes det hvorledes passagererne forventes at tilpasse deres rutevalg til de indførte ændringer. Dette problem er yderst kompliceret at løse analytisk da optimeringsproblemet (øvre niveau) ofte er enten ikke-konveks, ikke-lineær eller formuleret som en kombineret heltals problem, mens rutevalget (nedre niveau) beregnes som en ikke-kontinuert og ikke-lineær afbildning af køreplanen. Derfor er der, i nærværende studie, udviklet forskellige heuristiske algoritmer, der løser de to niveauer sekventielt. For at reducere regnetiden forsøges det, at forbinde problemerne på de to niveauer således at optimeringsproblemet adresserer rutevalget. Optimeringen (øvre niveau) tager således hensyn til hvorledes passagererne forventes at tilpasse deres rutevalg. Dermed konvergerer de sekventielle algoritmer hurtigere.



Figur 1 - Planlægning af kollektiv transport

Dette PhD studie udvikler en algoritme der kan optimere linjeplanerne i et kollektivt transportnetværk således at behovet for at skifte og dermed også den ventetid skift er forbundet med minimeres. Algoritmen baserer sig på såkaldte ”swaps”, hvor enkelte dele af krydsende linjer ombyttes. For at gennemsøge løsningsrummet effektivt opbygges en såkaldt ”Tabu Search” algoritme, der skal optimere linjeplanerne mens en rutevalgsmodel beregner hvorledes passagererne tilpasser deres rejsemønstre til de indførte ændringer. Algoritmen er afprøvet på S-togsnetværket i København. Algoritmen reducerede antallet af skift med 3,83 % og den samlede ventetid med 3,88 %.

Foruden linjeplaner, er standsningsmønstre også en del af netværksdesign. Rejsetidsreduktioner er ofte resultatet af store investeringer i infrastruktur eller rullende materiel. Ændring af standsningsmønstre har dog også potentiallet til at reducere rejsetiden mellem udvalgte stationer. Udover at være en billig måde til at opnå rejsetidsbesparelser, kan ændringer af standsningsmønstre også reducere heterogeniteten af jernbanedriften, hvilket har potentialet til at reducere risikoen for at forsinkelser spreder sig til andre tog. I dette studie udvikledes en algoritme til at optimere standsningsmønstre med det formål at reducere passagerernes samlede rejsetid og på samme tid reducere heterogeniteten af jernbanedriften. På baggrund af en rutevalgsberegning kunne tilpasningen af passagerernes rejsemønstre fastlægges når standsningsmønstrene var blevet ændret. Algoritmen er afprøvet på S-togsnetværket i København. Algoritmen formåede at reducere køretiden med 5,48 % samtidig med at antallet af skift steg med 1,38 % og den samlede ventetid steg med 1,60 %.

Planlægningen af køreplanen adresseres ved at udvikle en algoritme, der har til formål at ændre på bussernes afgangsminuttal således at passagererne oplever en reduceret ventetid når de skal skifte fra et kollektivt transportmiddel til et andet. Algoritmen baserer sig igen på en ”Tabu Search” struktur, der har vist sig at være yderst velegnet til denne type køreplansoptimering. Derudover er strukturen valgt da den udfører en intelligent gennemløbning af løsningsrummet, hvor det undgås at ende i et lokalt minimum, hvilket er vigtigt for kvaliteten af det endelige resultat. Rutevalgsberegninger anvendes til at evaluere den samlede ændring i passagerernes skiftetid som følge af ændringerne i afgangsminuttal. Algoritmen er afprøvet på det kollektive transport netværk i hele Danmark samt det i Storkøbenhavn. Køreplansoptimeringen medførte en reduktion i passagerernes skiftetid på 5,08 %.

De tre algoritmer, som er gennemgået herover (optimering af linjeplaner, standsningsmønstre og afgangsminuttal), fokuserer alle på optimering af forskellige dele af driften i den kollektive transport. Fælles for de tre algoritmer er, at passagerernes tilpassede rutevalg betragtes eksplicit. De forbedringer som er noteret i afsnittene herover kan alle opnås stort set uden investeringer. De eneste investeringer, der kunne være, vil være dem relateret til markedsføring af den nye køreplan, standsningsmønstre eller linjeplaner.

Sidste bidrag fra nærværende PhD studie fokuserer på at vurdere hvor stor en kapacitetsforringelse et kollektivt transportnetværk kan påføres før passagererne nægtes adgang til de kollektive transportmidler fordi der ikke er plads til flere passagerer. Kapacitetsforringelser forstås her som en aflysning af et tog, en bus eller et andet af de kollektive køretøjer. En sådan analyse er brugbar på det operationelle planlægningsniveau eksempelvis i tilfælde af mangel på rullende materiel. For at gøre analysen så realistisk som muligt, er det vigtigt at betragte effekten af kapacitetsforringelser på passagerernes rejsemønstre eksplicit. Igen udvikles en algoritme i to niveauer. Det øvre niveau forringer kapaciteten ved at aflyse udvalgte afgange, mens det nedre niveau beregner hvorledes passagererne tilpasser deres rejsemønstre til den forringede situation. Denne analyse er foretaget på det kollektive transport netværk i Storkøbenhavn, hvor kun S-togene kunne aflyses. Fra denne test kan det konkluders at det specifikke transportnetværk er yderst robust overfor togaflysninger. Dette skyldes især det meget tætte busnetværk, som generelt tilbyder attraktive rejsemuligheder. Det bør dog pointeres at indeværende analyse ikke eksplicit har medregnet den gene passagererne oplever i overfyldte toge og busser i de generaliserede rejseomkostninger, hvorfor resultatet sandsynligvis ligger til den optimistiske side.

PhD studiet har således givet bidrag til den nyeste forskning indenfor passagerorienteret planlægning af kollektiv transport. For at understøtte de samfundsmæssige bidrag er alle metoder og algoritmer udviklet så de kan anvendes på virkelige transportnetværk. Alle metoder er testet på virkelige kollektive transportnetværk og resultaterne er lovende. For at opsummere kan resultaterne af nærværende PhD-studie være med til at sikre at passagererne i den kollektive trafik i fremtiden kommer hurtigere frem til destinationen samtidig med at rejsen er mere komfortabel og er forbundet med en mindre risiko for forsinkelser.

Contents

1. Introduction

Mitigating road congestion is becoming a large challenge these years. One of the biggest concerns is how to move car users into the transit system. An increase in public transport's market share can only be obtained when passengers find the service sufficiently competitive. There are several ways to enhance the quality of a transit system, some more costly than others. Expectably, there would be a large correlation between the money invested and the quality of the transit service provided. However, simply planning transit operations efficiently has an appreciable impact on the service provided to the passengers.

The present study focuses on optimisation of different transit operations that can be obtained at a negligible cost. Initially, attention is given to the existing literature, where the discrepancy between passengers' perception of transit operations and the performance measurements published by the operating companies is addressed. To accommodate this discrepancy, optimisation models handling various transit planning tasks are developed, all taking into account passengers' complete journey from origin to destination. The study contributes to the state-of-the-art by developing bi-level formulations for the line planning problem, the skip-stop problem and the timetabling problem explicitly accounting for the dependency between transit network changes and passengers' travel behaviour adaptations. Finally, a method used to assess the degradable capacity of a transit network is also developed and tested on a real large-scale transit network.

1.1 Background

The quality of the public transportation network is known to have a large influence on potential customers and in particular how inclined they are to choose it (Ceder, 2007). Adopting microeconomic terms, passengers' desire to travel would be the demand while the transit network and its particular characteristics would be the supply. Supply, in the transit context, could be e.g. the line configuration, the frequency, the stopping patterns, the structure of the timetable, the number of available seats, the in-vehicle comfortability etc. All these characteristics would, if changed, have an impact on the travel demand. Demand reflects e.g. mode choice, route choice, departure time choice and boarding stop choice (Nuzzolo et al., 2012).

Enhancing the attractiveness of transit networks is on the agenda in several cities around the globe. The purpose is ideally to move more people from the cars into the transit vehicles, thereby solving two large problems, namely, reducing emissions from the cars and relieving road congestion. There exists several ways to enhance the attractiveness of a transit network. Building new metro lines, enhancing the frequency by purchasing new rolling stock or expanding existing railway lines from single to double track are among the more costly examples. On the other hand, there exist methods to enhance the operations simply by reconfiguring the current operational plan. Changes as e.g. optimising line plan configuration, making the line frequencies and stopping patterns demand responsive or minimising passengers' transfer time by adapting the timetable are all examples requiring barely any investments.

According to Parbo et al. (2015a), there is a discrepancy between how transit operations are typically planned with a vehicle oriented focus and how passengers perceive the performance when considering their entire journey from origin to destination. Even in the case where transit services are running according to the planned schedule, it is not given that e.g. the connectivity (including transfers) between different zones in the network is planned with the passengers in focus. To put passengers in focus when planning transit operations, it is important to know the zone-to-zone demand as well as passengers' path choice behaviour. Only in the case where planners take travel demand explicitly into account in their planning, it is possible to utilise the resources available effectively. However, planners and passengers may have contradicting desires, hence the optimal timetable for the operator could be far from optimal for the passengers (Medeossi et al., 2009; Schöbel & Kratz, 2009).

Another issue related to the gap between the operator's intentions and what passengers experience is the way transit performance is typically measured. One of the generic performance measures for transit services is the punctuality, often also referred to as reliability. The punctuality reflects the number of trains that are on time (or within a small threshold from this) according to the planned operations. Although being deployed by most transit companies, some studies have found a systematic discrepancy between the vehicle punctuality and, on the other hand, how large delays passengers experience (Vansteenwegen & Oudheusden, 2007; Nielsen et al., 2008). Empirical evidence shows that passenger on time performance is up to 10 percentage points below train punctuality during peak hours, with the reasons being cancelled trains, missed transfers and/or route choice adaptations (Nielsen et al., 2008).

1.2 Aim and main contributions

The present study aims at developing passenger oriented optimisation models for various transit operations. The contribution of the present study is the explicit focus on passengers' travel behaviour. Taking passengers' travel behaviour from origin to destination into account enhances the real-life applicability of the methods. Only considering single vehicle trips, could lead to suboptimal planning of the transit operations. Another large contribution of this PhD study is on the application side. All models are applied to real large-scale networks, which require heuristic solution algorithms due to the computational complexity of the optimisation problems.

In particular, the study aims at:

1) Visit the previous literature on transit planning and passenger oriented transit planning to clarify the gap between passengers' perception of railway operations and the way transit operating companies measure the transit performance. Addressing this gap is vital when focus latter turns to the development of optimisation methods addressing and trying to eliminate/narrow this gap.
2) Improve the line plan configuration of an existing transit system with the aim to reduce the number of passenger transfers but without increasing the vehicle operation costs related to the number of kilometres traversed.
3) Adapt stopping patterns with the aim to reduce passengers' in-vehicle travel time and avoid that extra transfers, transfer waiting time and reduced vehicle availability outweighs the reduction in in-vehicle time. Also, the aim will be to reduce the heterogeneity of the railway operations.
4) Change bus departure times from the initial station with the aim to minimise passengers' waiting time at transfer stations. Additionally, to produce a timetable from scratch with the aim to reduce passengers' average journey time.
5) Cancel individual transit vehicles to explore the degradable capacity of a transit network without increasing the number of rejected passengers.

1.2.1 Literature review

Optimising transit operations with an explicit passenger focus requires, on the one hand, deep knowledge about how transit operations are planned, performed and later measured, and on the other hand, how passengers perceive transit performance and how this perception influences their travel behaviour.

Parbo et al. (2015a) review the literature and start by looking at the parameters that railway optimisation/planning studies are focused on and the key performance indicators typically deployed in railway planning. When looking at railway planning, a discrepancy exists between planners, who focus on the train operations and publish fixed railway schedules, and passengers, who look not only at the schedules but also at the entirety of their trip from the access to the waiting, the on-board travel and the egress.

Exploring this discrepancy is essential as assessing railway performance merely by measuring train punctuality yields an unfair picture of the service provided to the passengers. Firstly, passengers' delays are often significantly larger than the train delays responsible for the passengers to be late. Secondly, train punctuality is often related to too tight schedules that in turn might translate into knock-on delays, thereby also increasing risk of missing transfer connections. A key aspect is the robustness of railway timetables. Empirical evidence indicates that passengers give more importance to travel time certainty than travel time reductions as they associate an inherent disutility with travel time uncertainty. This disutility may be broadly interpreted as an anxiety cost for the need for having contingency plans in case of disruptions, and may be looked at as the motivator for the need for delay robust railway timetables. Interestingly, passenger oriented optimisation studies considering robustness in railway timetabling typically limit their emphasis on passengers to transfer maintenance. Passengers' travel behaviour is far more complex and multifaceted, thus several other aspects should be considered, as becoming more and more evident from passenger surveys.

1.2.2 Line Plan configuration

Passengers are usually reluctant to include transfers in their journey (Nielsen & Frederiksen, 2006). One of the planning tasks that have the biggest influence on the number of transfers between different origin and destination pairs (O/D-pairs) is the line plan configuration. The line planning problem is well known from the literature (e.g. Lee & Vuchic, 2005; Zhao et al., 2005; Schmidt & Schöbel, 2010; Wang et al., 2011; Schöbel, 2012). Nearly all studies either simplify or neglect passengers' path choice behaviour, which could lead to suboptimal situations due to the simplified assumptions about passengers' travel behaviour.

Parbo et al. (2015c) optimise the line configuration of a railway network so that passengers are accommodated in a way that minimises their number of transfers as well as their waiting time experienced at boarding and transfer stations, respectively. An improving algorithm is developed to solve the line planning problem with explicit consideration of passengers' travel behaviour. The solution algorithm is based on swapping existing railway lines.

The contribution of this model is an explicit consideration of passengers' route choice behaviour combined with a line planning model applicable at the tactical planning level. The shorter planning period (line planning is typically addressed at the strategic planning level) enables the model to be more demand responsive, since short term demand can be predicted with larger certainty than long term demand. Additionally, based on the shorter planning period, the existing timetable structure is maintained in order also to be able to assess the impact on passengers' waiting time at the boarding stops and transfer stops. Adopting the existing timetable structure, it is assumed that O/D-travel demand, departure time choice abd boarding stop choice are fixed, which means that passengers' route choice and mode choice are adaptable when the new line plan is implemented.

1.2.3 Skip-stop optimisation

In order for public transport to form an attractive alternative to cars, the service needs to be competitive in terms of travel time and reliability (Ceder, 2007). One way to reduce the travel time between different station pairs is by skipping stops in between. The skip-stop optimisation problem is a well-known transit planning problem; one of the first papers considering the planning of stopping patterns is by Kikuchi & Vuchic (1982). No studies so far have considered a model explicitly taking into account how passengers adapt their route choice when changing stopping patterns. Failing to do so could lead to unreliable results since the results obtained when passengers' route choice are assumed to be unaffected by the changes may deviate significantly from what will be observed in reality.

Parbo et al. (2015b) deal with skip-stop optimisation for railway lines with the aim to reduce passengers' travel time and at the same time reduce the heterogeneity of the railway operations in all corridors of a railway network. For each stop skipped, affected passengers are assumed to adapt their travel behaviour accordingly. The challenge is thus to form stopping patterns intelligently, such that the reduction in passengers' in-vehicle travel time is not outweighed by the extra transfers and extra boarding waiting time. Additionally, including heterogeneity in the objective function creates a timetable which reduces the risk of minor delays propagating onto subsequently running trains.

1.2.4 Timetable optimisation

Having defined the line plan configuration and frequencies as well as the stopping patterns for a transit network, the next task is to create a timetable. The primary objective from the passengers' perspective would be to create an integrated timetable making the transfers as smooth as possible. This problem is known as the synchronisation problem, where the objective is to coordinate the transfers at the major hubs of the transit network. The problem has been studied by e.g. Ceder (2007), Zhigang et al. (2007) Wong et al. (2008) and Ibarra & Rios-Solis (2012).

The timetable is typically renewed once every year. The fairly short planning period allows an accurate estimation of passenger demand, both in terms of station-to-station demand but also passengers' path choice behaviour. Although, it is known that the timetable structure has a significant impact on passengers' route choice behaviour, so far no studies have managed to incorporate this explicitly into their optimisation model.

Parbo et al. (2014) develop a model that explicitly takes into account how passengers adapt their route choice behaviour when the timetable is changed. The model is an improving model, which means that an existing timetable is needed as input. The idea is to impose changes in vehicle departure time from the initial stop, which impacts passengers' transfer time on all stops along the line. The aim is to minimise passengers' transfer waiting time when transferring either to or from a bus.

Towards a Better Train Timetable for Denmark Reducing Total Expected Passenger Time

Another way of creating passenger oriented timetables rather than improving existing ones is by creating a timetable from scratch. Thereby, avoiding that existing timetable structures bias the final result.

Sels et al. (2015) apply a model based on the periodic event scheduling problem (known from e.g. Liebchen 2007). The model is able to produce a timetable for all passenger trains from scratch. The objective is to minimise total expected passenger journey time including the probability of missing a transfer. The contribution is the addition of a particular cycle constraint set that reduces computation times. It is demonstrated that the innovation result in a method that quickly generates cyclic timetables for a railway network spanning an entire country and that these timetables also reduce passengers' expected travel time.

1.2.5 Capacity degradability

All the before mentioned optimisation models are developed under the assumption that every track segment is functioning and that all vehicles are functioning. In reality, there might sometimes be a lack of rolling stock or there might be speed reductions on certain track segments. In such disrupted cases, contingency plans are used to ensure an efficient operation.

Parbo & Lam, (2015) develop a model to assess the capacity degradability of a transit network, which is practically useful for operational and tactical planning purposes. The aim of the capacity degradability model is to cancel all train runs possible without violating vehicle capacity constraints, thereby exploring how much the capacity can be degraded without deteriorating passengers' experienced level of service below a certain threshold. The result of the model can thereby be treated as a guideline on which transit services to cancel in the case not all vehicles (or track segments) are fully functioning.

1.3 Structure and reading guide

The remainder of the present thesis is structured as follows. Section 2 gives general insights in transit planning. Focus is first on the different planning tasks ranging from the strategical to the operational. After that, attention is given to passenger route choice models. Finally, it is outlined how bi-level programming can be used when the interaction between transit operations and passengers' travel behaviour needs to be considered in the optimisation. Section 3 presents the conclusions of the present study and outlines directions for future work. In the appendices 1-6, the papers produced as a part of the PhD study can be found. It is suggested that these are read after visiting sections 1 and 2, but before visiting section 3, where the conclusions and the future outlook are presented. Finally, appendices 7 and 8 provide additional insights on the case networks used for testing the algorithms and a description of how to alter the specific transit network databases through C#, respectively. Appendices 7 and 8 are especially meant for researchers who, in the future, are interested in extending the models and methodologies developed in this PhD study.

2. Transit planning

The main contributions of this PhD study are found in the appendices 1-6. However, due to the limitations of most journals, some of the fundamental theory in transit planning is omitted from these papers, thus making it hard for interested readers without particular knowledge on transit planning to grasp all of it. The purpose with this section is to introduce the basic transit planning theory. First, focus is on how transit planning is performed, while later, attention is given to the different ways of representing passengers' travel behaviour. Finally, a description of a methodological framework on how to plan transit operations while explicitly taking into account passengers' travel behaviour is provided.

2.1 Supply: Different aspects of transit planning

From the operator's perspective, four different steps related to the planning of public transport are generally present. In figure 1, these planning steps are outlined on a horizontal axis describing the planning horizon of each step. *Network Route Design* is focused on the route configuration and which stops each route serve. *Timetable Development* plans frequencies, headways and departure/arrival times for all lines operated. *Vehicle Scheduling* has the aim to minimise the fleet size. Finally, *Crew Scheduling* defines the crew schedules and duty rosters based on the trips defined in the previous planning step.

Although, the figure shows each step as a one way arrow, where output from a former step serves as input to a latter step, there might be some feedback from a latter step to a former step. Therefore, the ideal case would be to plan all four steps simultaneously. However, due to the complexity of the problems, it is not possible to combine all these planning tasks in a single optimisation problem.

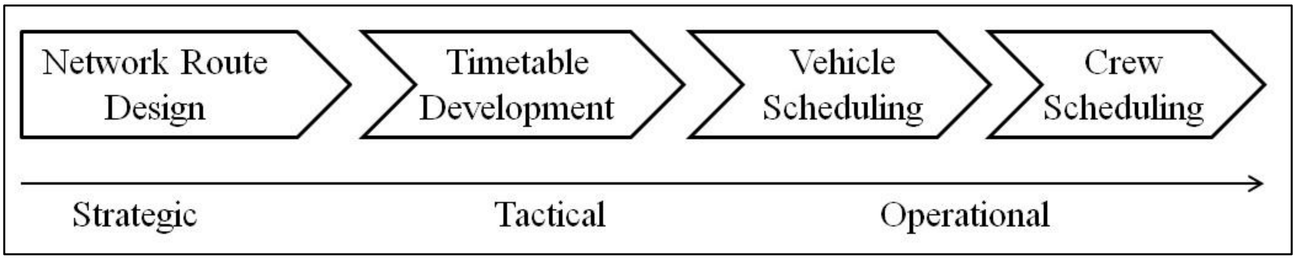


Figure 2 - Transit planning activities

In the present study, the focus is on the passenger related aspects of the planning i.e. the *Network Route Design* and the *Timetable Development*. Only these two planning steps directly affect the service provided to the passengers. Passengers are neither aware of the specific driver nor the vehicle rotation and how many vehicles are operated to complete all the planned trips. However, passengers are highly affected by the route network design, the stopping patterns, the frequency, the transfer synchronisation and the timetable.

2.1.1 Passenger related performance measures

In the ideal case, the timetable is completed just as planned. However, in real life disturbances occur, thus disrupting the planned schedule. To reflect the transit service actually provided to the passengers, various performance measurements can be applied. Among the most general ones are reliability, regularity and the number of cancelled transit vehicles.

Reliability is generally used as a measure of schedule adherence, i.e. the number of vehicles (often in percentage) arriving on time or within a certain threshold from the published timetable. Schedule adherence is said to be one of the predominant performance measures in public transport together with door-to-door travel time (Vromans et al., 2006).

Regularity is the ability to keep an equal distance (in both time and space) between subsequently running vehicles. Especially, for bus services the lack of regularity is well known and even has its own research field, bus bunching. The last of the three passenger related transit performance measurements is the number of cancelled transit vehicles. The performance measure is typically measured as the percentage cancelled vehicles out of the planned vehicles.

It is vital that the transit companies provide both a regular and punctual service in order to stay attractive; an unreliable service can act as a deterrent both to existing as well as potential passengers (Ceder, 2007). Disruptions can occur, but while these are small, the timetable should be sufficiently robust to absorb these, and hence adhere to the planned schedule at a satisfactory level.

2.2 Demand

Assignment models within the field of public transport are either frequency-based or schedule-based. The frequency-based approach considers data in a more aggregate way, where only average values related to lines are represented. The scheduled-based approach is more disaggregate, which makes it possible to assess the attributes for every single run.

Transportation flows or network flows are a kind of equilibrium between supply and demand. Demand in this context is the O-D matrix i.e. passengers' desire to perform certain trips. The supply is the public transportation network and its characteristics, e.g. timetable, line frequency, line plan etc. Then, based on the supply and the demand, it is possible by use of a transit assignment model (supply-demand interaction model) to calculate the network flows (passengers' travel patterns).

2.2.1 Frequency-based Approach

The frequency-based approach is also referred to as the line-based approach, while the schedule-based approach may be referred to as the run-based approach. The frequency-based approach is simpler in its network representation than the schedule-based approach. This is due to the fact that instead of the exact timetables for each run, only line frequencies and average travel times are taken into account. Practically speaking, the frequency-based model does not include the time dimension which is a major part of the schedule-based model. The frequency-based approach can thus advantageously be applied in a system, where frequencies are not changed throughout the day and transit services are not notably affected by congestion (resulting in varying travel time, and thus also varying headways) and travel demand is close to uniform. Figure 2 exhibits an example of a frequency-based network representation. In this figure, *Or* and *De* are the origin and destination, respectively. Walking links (dashed lines) are used to connect the origin and destination to the stations, A+B and F+G, respectively. Walking links are also used to connect the stations with the transit services (access and egress). Stations are represented by black squares. The solid lines represent transit lines, above which, the line number and the frequency are outlined. Between station pairs where more than one line is operating, lines are typically "merged". Consequently, passengers perceive the lines as one transit line with a resulting higher frequency. The "merge" implies some uncertainty in the distinction between the actual loads on the different parallel operating lines, known as the common lines problem.

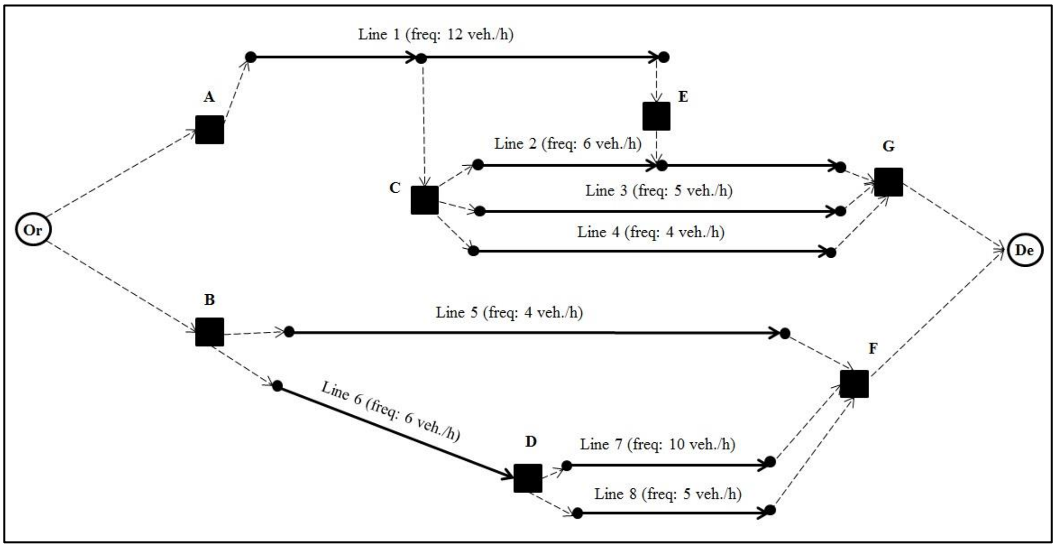


Figure 3 - Network representation (frequency-based transit assignment model)

On the positive side, the frequency-based approach allows a shorter calculation time, which is important especially for large-scale networks. On the negative side, the frequency-based approach only derives average values for each line, which thus may lead to a wrong picture of the service level since e.g. variations in vehicle occupancy and travel times are neglected.

When using a frequency-based approach to model public transport, the path choice approach is often based on a mixed pre-trip/en-route (indifferent) choice behaviour (Spiess & Florian, 1989; Schmöcker et al., 2011). This choice behaviour implies that instead of a single path, a set of attractive paths (called hyper path) is determined before (pre-trip) boarding any public transit services. This means that the passenger boards the first arriving vehicle on one of the lines in the hyper path (Lam & Bell, 2002). An intelligent en-route choice behaviour cannot be observed when using the frequency-based approach since the approach is line-based, which entails that it is not possible to distinguish between the different runs of a line. The frequency-based approach is thus most suitable in urban areas with a dense and high frequency transit network, or when examining future scenarios, where timetables are unknown (Friedrich & Wekeck, 2004).

The deterministic utility function (i.e. travellers' perceived travel impedance) associated with a hyper path j assuming the described indifferent en-route choice behaviour can be expressed as an average generalised cost C_j, which is equal to the average generalised cost C_k on each path k multiplied by the probability of choosing path k among the paths in hyper path j, $q_{k,j}$ (Lam & Bell, 2002).

$$C_j = \sum_{k \in j} q_{k,j} C_k$$

The public transit path choice models are based on random utility theory, where every user is assumed to be a rational decision-maker trying to maximise personal utility relative to the available choices (Cascetta, 2009). Considering the hyper paths belonging to the set I_{od}, this is the set of all hyper paths connecting the origin and destination. Each hyper path in the set has a perceived (i.e. an error term ε_j is added) utility U_j based on the attributes of the hyper path.

$$U_j = C_j + \varepsilon_j , \forall\, j \in I_{od}$$

C_j represent the deterministic disutility cost (or travel impedance) of all relevant attributes related to a certain transit trip. The relevant attributes are presented in the different papers (appendices 1-6) together with the estimated attribute values.

2.2.2 Schedule-based Approach

The schedule-based models are developed more recently than the frequency-based ones and they can be considered as an extension of the frequency-based ones. In the schedule-based models individual runs are represented individually. Therefore, schedule-based models are superior when demand is unevenly distributed or if headways are irregular. When passengers have knowledge of the network and they are provided reliable real-time information on the state of the system, their route choice tend to be run-based rather than line-based. Therefore, the common lines problem described earlier rarely applies when real-time information is provided to the travellers (Schmöcker & Bell, 2009). The schedule-based approach to public transit planning has been applied by e.g. Nielsen & Frederiksen (2006) and Nuzzolo et al. (2001).

To get a better understanding of the schedule-based model, figure 3 is created. The main differences from figure 2 are the individual representation of every single run belonging to a line (solid lines) and that parallel runs are not "merged" together. Figure 3 shows the paths that a traveller may consider between origin and destination. τ_{Dt} is the departure time from the origin. Comparing this network representation, with the frequency-based network from figure 2, calculating accurate performance measures of every single run is now possible. The more detailed network representation entails a higher calculation time though.

When considering the schedule-based assignment approach, the time at which travellers desire to start or end their trips, known as target times, plays a key role. The following target times are the most commonly used: desired departure time (DDT), actual vehicle departure time (VDT), desired arrival time (DAT) and actual vehicle arrival time (VAT). Because of the discrete nature of transit networks, some researchers introduce a disutility component usually referred to as late/early schedule penalty. This penalty reflects the difference between DDT and VDT or DAT and VAT, respectively. Only in low frequency environments desired and actual departure/arrival times are considered to be different (Nuzzolo & Crisalli, 2004).

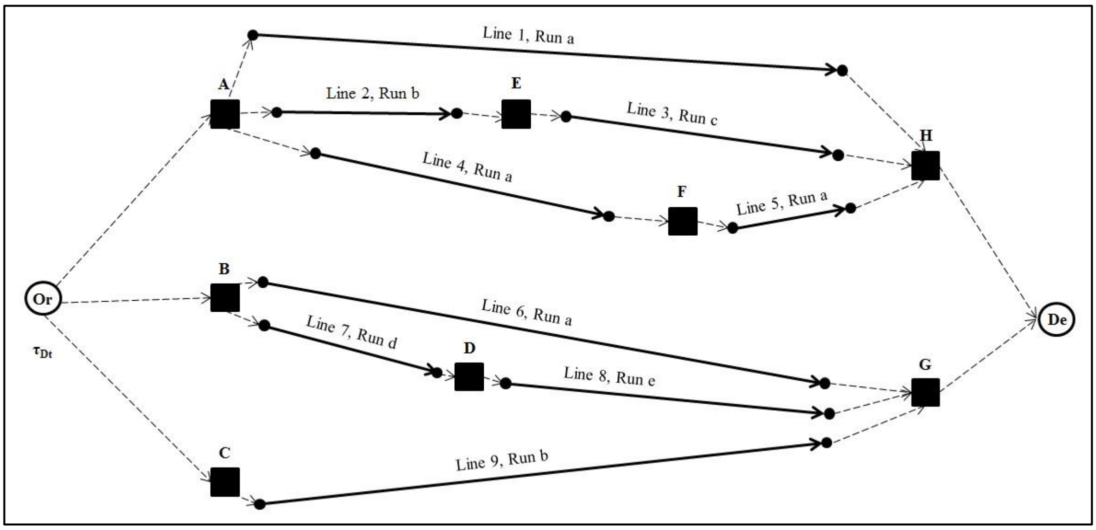


Figure 4 - Network representation (schedule-based transit assignment model)

To get a better understanding of the run choice in public transport and in particular within schedule-based models the following utility function is considered.

$$U_r = \sum_j \beta_j X_{jr} + \varepsilon_r$$

U_r is the perceived utility of a given run *r*, while β_j is the weight of attribute j, X_{jr} is the value of attribute *j* for run *r* and ε_r is the stochastic error component for run *r*. In transit contexts, the attributes typically considered are waiting time, in-vehicle time, transfer time, number of transfers, in-vehicle comfort and ticket fare.

2.3 Bi-level programming

Transport networks behave in a similar manner as economic networks/markets. The equilibrium of these systems can be found as a balance between supply and demand. Modifications of the transit supply are e.g. improving the infrastructure, building new roads, extending existing roads etc. These modifications share the characteristic that they are all expensive to realise. A change in supply that is achievable nearly for free is changing the timetable settings in the transit network. No matter how the supply is changed, it is probable that the modifications influence travel time, transfer time, in-vehicle congestion etc. These factors all have an impact on passengers' route choice and should therefore be considered, when optimising timetables. In figure 4, it is exemplified how a timetable change (indicated as a change in supply from S1 to S2) can have a positive impact on the travel time, which at the same time, with the existing demand curve, will increase the patronage.

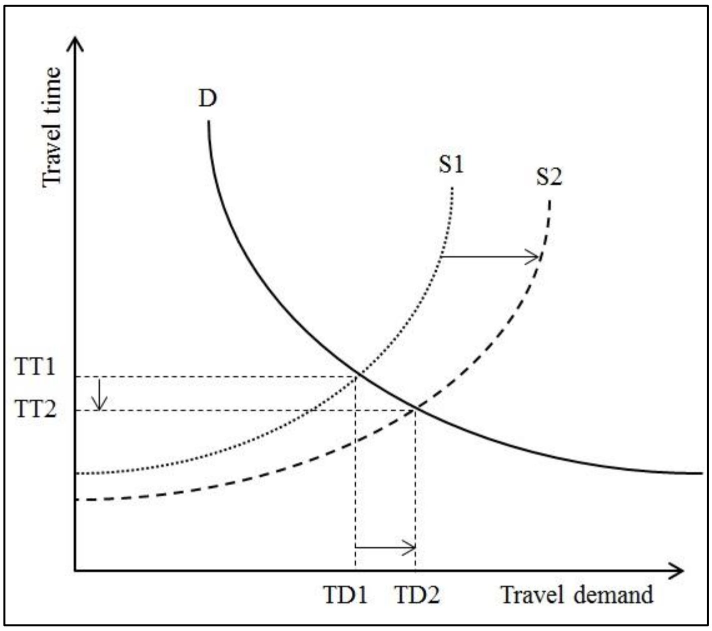


Figure 5 - Supply and demand

Existing literature within the field of transit optimisation rarely capture changes in demand explicitly. They assume that changing the supply do not affect passengers' route choice. One paradigm that allows taking into account the interaction between supply and demand is bi-level programming. Bi-level programming problems are nested optimisation problems characterised by the decision maker (e.g. the planner) at one level influencing the behaviour of a decision maker (e.g. the passenger) at another level. Therefore, an important feature of the bi-level programming problem is that the objective functions of each unit may be partially determined by variables controlled by other units operating at other levels (Oduguwa & Roy, 2002). In the transit context, the lower level may be a passenger route choice model, while the upper level could be an optimisation model with the objective to improve a specific part of the transit operations. In that case, would the changes in supply determined by the upper level serve as input to the passenger route choice model, since passengers' travel behaviour is affected by the changed transit operations. Similarly, would passenger demand affect the way the transit operations could be optimised.

3. Conclusions and future research

The PhD study presents theoretical and practically applied contributions to the state-of-the-art within optimisation of passenger oriented transit planning. The study starts by conducting a large literature review. From this review, the gap between general vehicle oriented transit planning and passengers' more elaborate perceptions of transit operations becomes evident. Based on the gap, a range of bi-level optimisation models are developed. The bi-level approach allows an explicit consideration of passengers' travel behaviour. The tangible outcomes of this study are transit optimisation models that are all tested on real large-scale transit networks with promising results. These models will thus be able to make the transit operations more attractive for the passengers at a very low cost, which mean they are more likely to be applied by operators compared to the changes that require large investment costs. The contribution of this study can thus be seen as a step towards increasing public transport's market share of the overall transport work.

The contributions are (i) a review of existing literature within transit timetabling and passengers' perception of transit operations, (ii) three different tools with an explicit passenger focus for optimising transit operations, and (iii) a model for an assessment of the degradable capacity of a transit network.

3.1 Literature review

The study is the first to provide an extensive review of the existing literature that has aimed at enhancing operational characteristics related to the robustness of railway timetables and the literature that has focused on passengers' perception of railway performance, respectively.

The operational characteristics considered most often in railway timetabling when the aim is to enhance the robustness, are time supplements and buffer times. All reviewed studies agreed on lowering capacity utilisation to reduce the risk of delay propagation. However, the recommendations vary based on the network layout, how performance is measured and on other network specific characteristics.

Passengers' travel behaviour (e.g. mode choice, route choice, departure time choice) depends on their perception of operational attributes as e.g. in-vehicle time, transfer time, waiting time, access/egress time, crowding level and delays. Using train punctuality as performance measurement thus turns out to be inadequate when trying to reflect the level of service passengers experience since train punctuality does not say anything about passengers' complete journey from origin to destination. In some cases, passenger punctuality is 10 percentage points lower than train punctuality. The discrepancy between train and passenger punctuality is primarily caused by missed transfers and the fact that more trains are delayed during peak periods, where vehicles are typically more crowded.

It is concluded that being able to address passengers' preferences explicitly is the basis for a more passenger oriented railway planning. Accurate and disaggregate passenger travel data facilitates a more passenger oriented planning, especially when transfer patterns are revealed. Coupling generic optimisation approaches of railway operations with knowledge of passengers' travel behaviour (e.g. through transport models) will enhance the reliability and applicability of the results. Thereby, decreasing the gap between what railway planners provide and how it is perceived and experienced by the passengers. In the following subsections, the PhD study has tried to make this coupling on a range of different transit planning problems.

3.2 Line plan configuration

The line plan configuration of a transit network is typically a strategic planning problem. Changes can be imposed as a response to the building of new transit lines, e.g. such that feeder lines are guaranteed an acceptable transfer time. In this study, a line planning model is developed with the aim to generate a line plan configuration that accommodates travel demand in the best way possible. The model is developed with the purpose to serve as a supportive tool to the yearly timetable changes, where the structure of the timetable is kept, thus allowing a more accurate derivation of passengers' travel behaviour.

The contribution is a new optimisation tool, which has its strength in the line plan configuration optimisation and in the validity of the results since passengers' adapted travel behaviour is considered explicitly. The developed solution framework is applied to the suburban railway network in the Greater Copenhagen area. The improving algorithm yields a significant reduction in the number of transfers and the waiting time experienced when boarding and transferring.

3.3 Skip-stop optimisation

Skipping stops is a cheap way to reduce in-vehicle time between certain station pairs. Another result of changing stopping patterns could be that the heterogeneity of the railway operations is reduced. This would most likely happen when stopping patterns are changed in a network, where all-stop and express trains are operated in parallel. Allowing all trains to skip a limited number of stops, could increase the train spread, which is equivalent to adding more buffer time between trains, thus making the timetable more resistant to delay propagations.

The contribution of the study is a skip-stop optimisation algorithm explicitly taking into account passengers' adapted route choice behaviour as well as considering a network rather than only a corridor. The approach is applied successfully to the suburban railway network in the Greater Copenhagen area in Denmark. The large-scale application showed a reduction in passengers' travel time. In addition, the train spread in the corridors is increased, thus the delay resistance is improved. Consequently, railway passengers on average get faster and with less risk of delay from origin to destination.

3.4 Timetable optimisation

Minimising transfer waiting time can be done e.g. by increasing vehicle frequency. A cheaper way to obtain a similar reduction is by adapting the departure time from the initial station. In that case, it is essential to be aware of passengers' route choice. Therefore, the current study proposes a timetable optimisation approach that explicitly considers passengers' modified route choice as a reaction to timetable changes.

The contribution of the study is a new optimisation tool, which has its strengths with respect to both the timetable optimisation and the reliability perspective, since changes in the travellers' route choice decisions are considered explicitly whereby demand effects are taken into account. The approach is applied to a real large-scale transit network. The optimisation yields a reduction in weighted transfer waiting time, while only affecting the in-vehicle travel time and the generalised travel cost to a lesser extent.

3.5 Capacity degradability

Assessing the degradable capacity of a transit network can be used by the operators as an indicator of to what extent transit vehicles are being utilised efficiently. The current study assesses how many runs that could be cancelled before passengers are rejected from boarding. For the results to be as reliable as possible both supply and demand are represented at a disaggregate level.

The contribution of the study is the application of a schedule-based transit assignment model as well as the exact timetable for each individual vehicle. Second, the existence of a Braess-like paradox occurring when individual transit runs are cancelled is proven. Final contribution is on the application side, where the model is verified on a large-scale railway network from the Greater Copenhagen area in Denmark.

Summarising, the PhD study has given contributions to the state-of-the-art of several planning tasks that transit operators face at the tactical planning level. The main contribution of these models is found in the explicit emphasis on passengers' adapted travel behaviour, which is usually observed when transit operations are changed. All methodologies are tested on large-scale transit networks and proved to enhance passengers' travel experience. Thus by applying these methodologies in a real-life context, passengers get a faster and less uncertain journey from origin to destination.

3.6 Future research

In the literature review, several directions for future research are found. Apart from those addressed in the present PhD study, the following directions represent interesting paths for future research. First, passenger oriented key performance indicators, taking as many relevant attributes into account as possible, should by applied when optimising transit operations. Such exhaustive performance indicators will ensure that the optimisation of a specific attribute is not obtained on the expense of non-measured attributes, thus potentially leading to a de facto deterioration in service level. From a practical point of view, such indicators allow the planners to have tangible performance measures, which is often easier to handle.

From the operator’s perspective, fleet size minimisation is a top priority due to the large investment cost and operating costs. Based on this, a direction for future research would be to include fleet size minimisation in the objective function of the optimisation problems, thus making them bi- or multi-objective. By assigning weights to each element in the objective, different weight settings could be tested and the pareto frontier between the reduction in passengers’ travel impedance and the fleet size requirements could be found. It is expected that integrating rolling stock circulation explicitly in the heuristic algorithm will reduce the solution space. Therefore, passengers’ benefit will most likely be reduced compared to the results seen in this study. However, considering the fleet size in the objective would enhance the applicability of the model from the operator’s perspective, which makes it a topic for future research. An alternative strategy would be to impose a limitation on the number of operated vehicles rather than trying to minimise the needed fleet size.

Another aspect, which is not addressed in this study, is in-vehicle crowding. To enhance the validity of the results and make the models more robust against variations in demand, a future version of the schedule-based transit assignment should account for in-vehicle crowding, i.e. the probability of being able to board a specific train as well as passengers’ inconvenience related to on-board congestion. Thereby, it would also be possible to assess whether a specific line plan configuration leads to an unacceptable level of in-vehicle crowding or if a certain stopping pattern or a specific timetable configuration would force passengers to adapt their path choice, and thus overload particular transit runs.

In this study, travel time is modelled as deterministic. In reality, travel time is related with a large degree of uncertainty, which is also the reason that passengers are often complaining about delays when travelling by public transport. A topic for future research would be to consider travel time as stochastic rather than deterministic. This could be done either at the arc level, where certain links (e.g. roads or track segments) sre associated with a defined travel time distribution or at the line level, where specific lines or vehicle types could have a certain probability of being late related to them.

Integrating the developed approaches with each other, thus comprising a wider optimisation tool is indeed also a direction for future research. Such an extension requires a lot of work on the model formulation and solution algorithm. The biggest risk is that combining the problems would increase the size, thus making it even more intractable in terms of searching as much of the solution space as possible. From the operator’s perspective, it would, however, be interesting to integrate the line planning and skip-stop optimisation with timetable optimisation.

References

Cascetta, E. (2009). Random utility theory. *Transportation systems analysis,* 89-167. Springer US.

Ceder, A. (2007). Public transit planning and operation: theory, modeling and practice. Elsevier, Butterworth-Heinemann.

Constantin, I., & Florian, M. (1995). Optimizing frequencies in a transit network: A nonlinear bi-level programming approach. *International Transactions in Operational Research*, *2*(2), 149-164.

Friedrich, M., & Wekech, S. (2004). A schedule-based transit assignment model addressing the passengers' choice among competing connections. *Schedule-Based Dynamic Transit Modeling: theory and applications*, 159-173. Springer US.

Ibarra-Rojas, O. J., & Rios-Solis, Y. A. (2012). Synchronization of bus timetabling. *Transportation Research Part B: Methodological*, *46*(5), 599-614.

Kikuchi, S., & Vuchic, V. R., (1982). "Transit vehicle stopping regimes and spacings", *Transportation Science*, *16*(3), 311-331.

Lam, W. H., & Bell, M. G. (2002). Advanced modeling for transit operations and service planning. Emerald Group Pub Ltd.

Lee, Y. J., & Vuchic, V. R. (2005). Transit network design with variable demand. *Journal of Transportation Engineering*, 131, 1-10.

Liebchen, C. (2006). Periodic timetable optimization in public transport. *PhD dissertation,* Technical University Berlin.

Medeossi, G., Marchionna, A. & Longo, G. (2009). Capacity and reliability on railway networks: A simulative approach. *PhD dissertation*, University of Trieste.

Nielsen, O. A., & Frederiksen, R. D. (2006). Optimisation of timetable-based, stochastic transit assignment models based on MSA. *Annals of Operations Research*, *144*, 263-285.

Nielsen, O. A., Landex, A., & Frederiksen, R. D. (2008). Passenger delay models for rail networks. *Schedule-Based Modeling of Transportation Networks: Theory and applications*, *46*, 27-49, Springer.

Nuzzolo, A., & Crisalli, U. (2004). The schedule-based approach in dynamic transit modelling: A general overview. In *Schedule-Based Dynamic Transit Modeling: theory and applications,* 1-24. Springer US.

Nuzzolo, A., Crisalli, U., & Rosati, L. (2012). A schedule-based assignment model with explicit capacity constraints for congested transit networks. *Transportation Research Part C: Emerging Technologies*, *20*, 16-33.

Nuzzolo, A., Russo, F., & Crisalli, U. (2001). A doubly dynamic schedule-based assignment model for transit networks. *Transportation Science*, *35*(3), 268-285.

Oduguwa, V., & Roy, R. (2002). Bi-level optimisation using genetic algorithm. *Artificial Intelligence Systems, 2002. (ICAIS 2002). 2002 IEEE International Conference,* 322-327. IEEE.

Schmidt, M, & Schöbel, A. (2010). The complexity of integrating routing decisions in public transportation models. In *Proceedings OASIcs*, 2757.

Schmöcker, J. D., & Bell, M. G. (2009). The build-up of capacity problems during the peak hour. In *Schedule-Based Modeling of Transportation Networks,* 1-23. Springer US.

Schmöcker, J. D., Fonzone, A., Shimamoto, H., Kurauchi, F., & Bell, M. G. (2011). Frequency-based transit assignment considering seat capacities. *Transportation Research Part B: Methodological*, *45*(2), 392-408.

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

Schöbel, A. & Kratz, A. (2009). A bi-criteria approach for robust timetabling. *Robust and Online Large-Scale Optimization. Lecture Notes in Computer Science*, *5868*, 119-144.

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

Vansteenwegen, P., & Van Oudheusden, D. (2007). Decreasing the passenger waiting time for an intercity rail network. *Transportation Research Part B: Methodological*, *41*(4), 478-492.

Vromans, M. J., Dekker, R., & Kroon, L. G. (2006). Reliability and heterogeneity of railway services. *European Journal of Operational Research*, *172*(2), 647-665.

Wang, L., Jia, L. M., Qin, Y., Xu, J., & Mo, W. T. (2011). A two-layer optimization model for high-speed railway line planning. *Journal of Zhejiang University SCIENCE A*, 12(12), 902-912.

Wong, R. C., Yuen, T. W., Fung, K. W., & Leung, J. M. (2008). Optimizing timetable synchronization for Rail mass transit. *Transportation Science*, *42*, 57-69.

Zhao, F., Ubaka, L. & Gan, A. (2005). Transit network optimization: minimizing transfers and maximizing service coverage with an integrated simulated annealing and tabu search method. *Transportation Research Record: Journal of the Transportation Research Board*, 1923, 180-188.

Zhigang, L., Jinsheng, S., Haixing, W., & Wei, Y. (2007). Regional bus timetabling model with synchronization. *Journal of Transportation Systems Engineering and Information Technology*, *7*(2), 109 112.

Appendix 1: Parbo et al. (2014)

User perspectives in public transport timetable optimisation

Jens Parbo, Otto Anker Nielsen & Carlo Giacomo Prato

Technical University of Denmark, Department of Transport, Bygningstorvet 116B, 2800 Kgs. Lyngby, DK-Denmark

Published in *Transportation Research Part C: Emerging Technologies*, 48, 2014, 269-284.

Abstract

The present paper deals with timetable optimisation from the perspective of minimising the waiting time experienced by passengers when transferring either to or from a bus.

Due to its inherent complexity, this bi-level minimisation problem is extremely difficult to solve mathematically, since timetable optimisation is a non-linear non-convex mixed integer problem, with passenger flows defined by the route choice model, whereas the route choice model is a non-linear non-continuous mapping of the timetable. Therefore, a heuristic solution approach is developed in this paper, based on the idea of varying and optimising the offset of the bus lines. Varying the offset for a bus line impacts the waiting time passengers experience at any transfer stop on the bus line.

In the bi-level timetable optimisation problem, the lower level is a transit assignment calculation yielding passengers' route choice. This is used as weight when minimising waiting time by applying a Tabu Search algorithm to adapt the offset values for bus lines. The updated timetable then serves as input in the following transit assignment calculation. The process continues until convergence.

The heuristic solution approach was applied on the large-scale public transport network in Denmark. The timetable optimisation approach yielded a yearly reduction in weighted waiting time equivalent to approximately 45 million Danish kroner (9 million USD).

Keywords
Bus timetabling, public transport optimisation, passenger behaviour, waiting time, large-scale application.

1 Introduction

In a report from the Capital Region of Denmark (RH, 2009), it was estimated that 11.5 billion Danish kroner (DKK) will be lost due to travellers being delayed because of congestion in the Copenhagen Region in 2015. Furthermore, it was stated in the report that, to avoid the outlined scenario, people ought to start travelling by public transport rather than by car. The question is how this change in the market share between private and public transport is actually realised?

The present paper deals with timetable optimisation from the perspective of minimising the waiting time experienced when transferring either to or from a bus.

1.1 Literature review

Designing an attractive transit network is an important and strategic task, in the literature often referred to as the Transit Route Network Design Problem (TRNDP). Based on an existing bus network, Bielli et al. (2002) aimed at improving the performance and reducing the need for rolling stock by adapting lines and their frequency. Lee & Vuchic (2005) tried to design an optimal transit network as a compromise between minimal travel time, transit operator's profit maximisation and minimisation of social costs. Elaborating mainly on the travel time description, Fan & Machemehl (2006) considered the transit route network design problem but separated travel time into four components (walking time, waiting time, in-vehicle time and transfer cost). The TRNDP has received much attention in the literature, and its significant contribution was notably summarised in two reviews by Kepaptsoglou & Karlaftis (2009), focusing on design objectives, operating environments, and solution approaches, and Guihaire & Hao (2008), focusing on unifying the area. Regarding future developments within this area, Kepaptsoglou & Karlaftis (2009) recommended that the focus should be on transfer policies and passenger transfer related items as waiting and walking distances, while Guihaire & Hao (2008) suggested that the focus should be on privatization and deregulation, as well as integration and intermodality among transit networks by focusing on improving transfers globally instead of looking at within-mode transfers.

In the literature, several solutions have been proposed to the timetable optimisation problem with various approaches to the consideration of transfers. One of the problems that have received much attention is the Timetable Synchronisation Problem (TTSP, e.g., Ceder, 2007; Liu et al., 2007; Ibarra-Rojas & Rios-Solis, 2012), which aims at maximising the number of simultaneous arrivals at transfer stations. Wong et al. (2008) developed a timetable optimisation model trying to minimise the total passenger transfer waiting times by changing the offset of the bus lines. This approach was also used, though with different objectives, by Bookbinder & Desilets (1992), Knoppers & Muller (1995), Cevallos & Zhao (2006), Hadas & Ceder (2010) and Petersen et al., (2012). Guihaire & Hao (2010) maximised the quality and quantity of transfer opportunities. While the quantity was self-explanatory, the quality was a twofold concept: firstly, it was based on the number of passengers; secondly, it was based on an ideal transfer time, i.e. a cost function was introduced to force the transfer time to be as close as possible to the ideal one. Niu & Zhou (2013) applied a timetable optimisation approach taking into account the passengers boarding at crowded stations. The objective was to minimise passengers' waiting time at stops and also reduce the waiting time passengers who were not able to board their desired service suffered because of congestions. They applied a genetic algorithm to solve the problem for each station in a double-track corridor. De Palma & Lindsey (2001) tried to minimise schedule delay (i.e., difference between preferred and actual departure time) by choosing the best timetable among a finite set of a priori created timetables. Taking a more holistic and strategic view of the transit network, Zhao & Ubaka (2004) applied two different algorithms to find the optimal set of transit routes to maximise route directness, minimise number of transfers and maximise service coverage. Another alternative perspective was used in the study by Yan et al. (2012), where the objective was to design a reliable bus schedule for fixed bus routes with a series of control points, and the punctuality of the busses was continuously controlled for and it was intended to improve it by letting the drivers recover the schedule by speeding up in order to reach the next control stop on time. Including the requirements for different types of rolling stock, Ceder (2011) developed an extended version of the deficit function to efficiently allocate different types of rolling stock where needed to accommodate the demand on each transit line based on an existing timetable.

In all of these studies, some prior information on users' travel behaviour was used, but passengers were assumed not to change their route choice when the timetable was changed. Normally, one would expect demand to change accordingly, when the supply is changed. In this context, the supply should be seen as the transit system, hence also the timetable, while the demand reflects the transport, namely the passengers' route choice. With this in mind, it seems appropriate to look at some of the timetable optimisation approaches which have considered the balance between supply and demand. Actually, this balance was noted as missing by Zhao & Ubaka (2004) and more recently by Ibarra-Rojas & Rios-Solis (2012). An early study formulated the timetable optimisation problem as a bi-level nonlinear non-convex mixed integer programming problem (Constantin & Florian, 1995). The objective of the upper level was to minimise the total expected travel time plus the waiting time. This was done by changing frequency settings in the timetable. The lower level problem was a transit assignment model with frequencies determined by the upper level. Wang & Lin (2010) developed a bi-level model to minimise operating cost related to the size of the fleet plus the total travel cost for passengers. Here the upper level referred to the determination of service routes and the associated headways. The lower level referred to the route choice behaviour, which was found by using a deterministic Frank-Wolfe loading approach. Ma (2011) applied a bi-level approach for the optimal line frequencies in a transit network, meaning the frequencies that minimise passengers travel time plus the operating cost. The lower level problem (route choice) was solved by using a Cross Entropy Learning algorithm, which was able to find the user equilibrium in transport networks. The upper level problem (optimising line frequencies) used the Hooke-Jeeves algorithm to find improvements in the current solution. Considering the same problem of finding the optimal frequency for a bus network, Yu et al. (2009) applied a bi-level programming model with the objective to reduce passengers' total travel time. In this approach, the upper level determined the bus frequencies by a genetic algorithm while the lower level assigned transit trips to the bus route network by use of a label-marking method. The two levels were solved sequentially until convergence.

One of the first studies to consider transfer time minimisation and also how passengers adjusted their travel patterns accordingly was Feil (2005), who applied a Steepest Descent approach to find the most promising offset changes and evaluated their actual impact with a public assignment model.

1.2 Objective and contribution

In the present paper, the objective was to minimise the weighted transfer waiting time. A weight reflecting the number of passengers transferring and their actual value of time was assigned to every transfer. The weight was based on the individual passenger's trip purpose. The waiting time was the time that elapsed between the alighting of one run and the boarding of the next run.

The optimisation was performed with the view of finding the optimal departure time (offset) for each bus line to reduce passengers' waiting time when transferring. Instead of treating passengers' route choice as static and predetermined, the approach developed in this paper treated their route choice as pseudo-dynamic. This was done in an iterative process where the output of the timetable optimisation (i.e. the new timetable) served as input to the public assignment model. The output from this public assignment model (i.e. passengers' travel patterns) was then used as input in the timetable. Due to its inherent complexity, finding the optimal offset with this particular objective for the bus lines was extremely difficult analytically. Tabu Search was chosen because of its ability to avoid being trapped in local minima, and also because it had proven to be superior compared to other metaheuristics when bus timetables in Copenhagen were to be optimised (Jansen et al., 2002).

The heuristic solution approach was applied to the highly complex bus network in Denmark. Applying the heuristic solution approach to the large-scale public transport network in Denmark should preferably reduce the waiting time passengers experience when transferring, while keeping their in-vehicle time and their total generalised travel cost at a constant level.

2 Method

The current study applied a bi-level timetable optimisation approach, where the objective was to minimise the weighted transfer waiting time. Timetable optimisation (upper level) was integrated with a public assignment model (lower level) to assess how travellers change their behaviour according to the changes imposed in the timetable. This is not a constructive heuristic. Therefore, to make the developed approach work properly, it is necessary to have an existing transit network as an initial solution and being able to run a schedule-based transit assignment, i.e. having a timetable explicitly stating when every transit vehicle departs from its initial stop and when it arrives and departs from the sub sequent stops all the way to its destination.

2.1 Analytical formulation

The mathematical formulation of the timetable optimisation problem was as follows:

$$\min . \sum_{i=1}^{M} \sum_{j=1}^{M} \sum_{s=1}^{N} \omega_{ij}^{s} w_{ij}^{s} \tag{1}$$

where

$$w_{ij}^{s} = \min \left\{ \pi_j + \alpha_j^s + \beta_j^s - (\pi_i + \alpha_i^s + \delta^s) \middle| \pi_j + \alpha_j^s + \beta_j^s \geq \pi_i + \alpha_i^s + \delta^s \right\} \tag{2}$$

$$\omega_{ij}^{s} = \sum_{c} T_{ijc}^{s} \cdot VoT_c \tag{3}$$

subject to

$$\alpha_{i-1}^{s} + H_{k,i-1,i} \leq \alpha_i^s, \forall k,i,s \in K,M,N \tag{4}$$

$$\pi_k \geq 0, \forall k \in K \tag{5}$$

$$H_{k,i-1,i} \geq 0, \forall k,i \in K,M \tag{6}$$

$$\alpha_i^s \geq 0, \forall i,s \in M,N \tag{7}$$

$$\beta_i^s \geq 0, \forall i,s \in M,N \tag{8}$$

$$\delta^s \geq 0, \forall s \in N \tag{9}$$

In the objective function, w_{ij}^S is the waiting time between busses i and j at a stop s, ω_{ij}^S is a weight reflecting the importance of every single transfer, T_{ijc}^S is the number of passengers transferring between busses i and j at stop s, the index c refers to the different passengers groups, each of these with their own value of time[1]. Constraints (4) ensure that overtaking does not occur. α_{i-1}^S is the departure time from stop s for bus i-1, while $H_{k,i-1,i}$ is the headway between busses i-1 and i belonging to bus line k. Constraints (5) ensure that the departure time π_k from the initial stop of the first bus on bus line k, is positive. Constraints (6) and (7) ensure that all $H_{k,i-1,i}$ and α_i^S are positive, while constraints (8) and (9), respectively, indicate that the dwell time β_i^S of a bus i at stop s and, finally, the station-specific changing and orientation time δ^s equivalent to the minimum amount of time a passenger needs to change platform at stop s should be positive. This value is an input data applied to make more robust transfers and enhance the probability that passengers make the transfer even when services experience disruptions. Finally, the three different sets K, M and N refer to the set of bus lines, bus groups and transfer stops respectively.

2.2 Heuristic solution approach

This bi-level minimisation problem is extremely difficult to solve mathematically, since the timetable optimisation is a non-linear non-convex mixed integer problem (NP-Hard according to Nachtigall & Voget, 1996; Cevallos & Zhao, 2006), with passenger flows defined by the route choice model, where the route choice model is a non-linear non-continuous mapping of the timetable. Therefore, a heuristic solution approach was developed based on the idea of varying the offset of the bus lines. A variation in the offset value for a bus line affected the waiting time experienced at any transfer stop this bus passed.

Changing busses' offset value can be done at three levels: the most disaggregate, where the offset of every single bus is changed, a more aggregate level, where groups of busses (typically with similar characteristics) are changed together, and the line level, where every single bus belonging to a certain line are subject to an offset change. In this paper it was chosen that groups of busses from the same line were subject to offset changes. This choice was a compromise between optimisation potential (often larger when being at a more disaggregate level) and maintaining the existing structure of the timetable including typically fixed equal headway for each bus line for certain time intervals. Bus groups are created based on time of day and travel direction (e.g. in the evening period, roads are less congested, which implies that the average travel time is less than during peak hours). Due to the non-symmetric travel times (different in each direction), there is a group of busses for each direction of a bus line. Therefore, busses from a certain line have fixed headway in each direction and in each time period. The chosen approach allowed a potentially larger reduction in weighted transfer waiting time, while neither changing the headway of a bus line in the forward direction nor in the backward direction. However, the layover time might be changed. In theory this could have an impact on the fleet size. Including an evaluation of the impact on fleet size of every bus line's offset change was out of the scope of this article, but clearly a topic for future research.

[1] c = 1: Commuter trips (35.4 DKK/h). c = 2: Business trips (270 DKK/h). c = 3: Leisure trips (12.6 DKK/h).

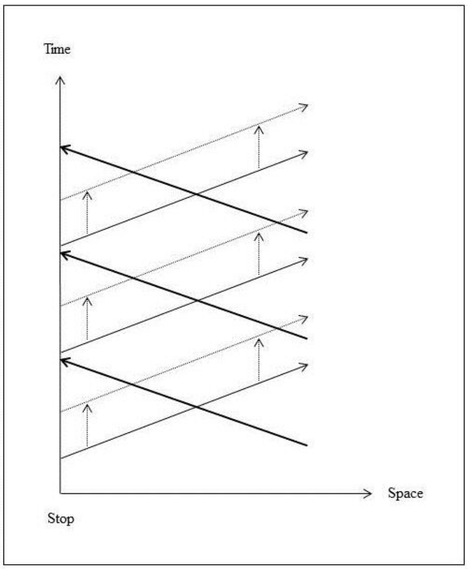


Figure 6 - Offset change

In figure 1, an example of an offset change is indicated. The thin black arrows are one group of busses running in forward direction, the dashed arrows indicate the change in offset for that group of busses, while the thick black arrows represent a bus group from the same bus line running in the opposite direction.

A Tabu Search algorithm was applied to find the appropriate bus lines and their most promising offset changes. The reason for applying Tabu Search was because of its ability to search the solution space intelligently, i.e. to escape local minima and prevent cycling through the solution space (Glover, 1990).

The explicit consideration of passengers' modified route choice as a course of changes in the timetable in the present study was done as a sequential process, where the output of the timetable optimisation (i.e. the new timetable) served as input to the public assignment model. The output from this public assignment model (i.e., the passengers' travel patterns) was then used as input in the timetable optimisation. The process was continued until the objective value began to converge.

The entire heuristic solution approach worked according to the following step-wise approach elaborated in the following.

0. **Run public assignment.**
1. **Calculate objective value.**
2. **Calculate optimisation potential for each offset change.**
3. **Impose offset changes.**
4. **If stopping criterion met, stop.**
5. **Otherwise, run public assignment and go to 1.**

2.2.1 Calculation of objective value

The aim of this optimisation was to minimise the overall weighted waiting time experienced by passengers when transferring in the transit network. After each public assignment calculation, the objective value was calculated. To do this properly, we needed to distinguish between direct transfers and transfers including walking.

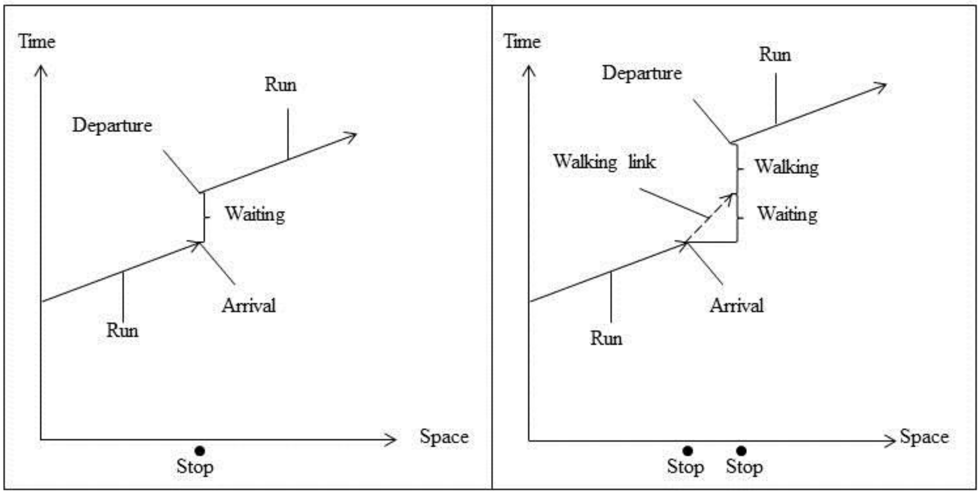


Figure 7 (a) direct transfer and (b) transfer including a walk

Figure 2(a) and (b) depicts a direct transfer and a transfer including a walk in a time-space diagram, respectively. From the figure it is evident that the waiting time is the time spent at the stop from which the next bus run departs. Calculating the weighted waiting time spent when transferring for the two types of transfers was done in the following two ways respectively according to figure 2, with the index *c* referring to the trip purpose.

$$\sum_c (Departure - Arrival) * (VoT_c * \# Pax_c) \tag{10}$$

$$\sum_c (Departure - Arrival - Walktime) * (VoT_c * \# Pax_c) \tag{11}$$

2.2.2 Optimisation potential

The decision variables in this problem were the offset of the bus lines. Therefore, it was necessary to assess how changing the offset of a given bus line affected the solution value. This impact was treated as an estimate of the improving effect on the solution value of a given offset change for a given bus line and was referred to as optimisation potential. The reason for naming it potential was due to the uncertainty in the calculations (i.e. there might be a difference between the calculated optimisation potential and the realised improvement due to passengers' adapted behaviour). Ideally, the impact on the objective value for each offset change should be assessed by a public assignment calculation. But since this would be extremely time-consuming, approximations were used instead.

To search the solution space comprehensively, a large neighbourhood needed to be considered. This was done by calculating the optimisation potential for every bus line (with at least two runs) in the interval between $[-h_{max}; +h_{max}]$[2] with increments of one minute, while maintaining the offset of all other bus lines. To ensure that all relevant passenger interactions were taken into account, and not just observed at a single transfer point, we distinguished between transfer points where passengers transfer either to (figure 3a) or from (figure 3b) the bus line of interest.

[2] H_{max} was the maximum headway for a given bus. If the maximum headway of a bus line was 5 minutes the optimisation potential was calculated for the following offset changes (-5, -4, -3, -2, -1, 0 ,1, 2, 3, 4, 5), where 0 was equal to the original offset.

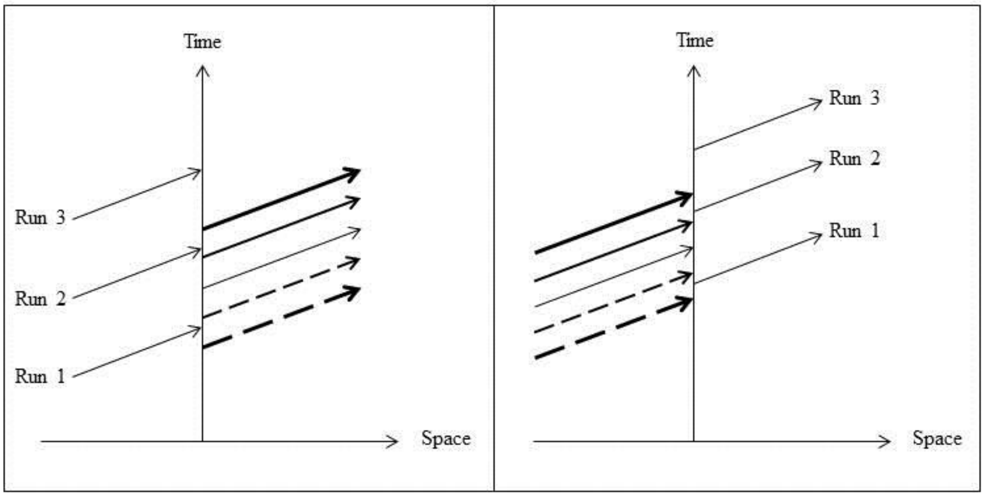


Figure 8 (a) Feeding transfer and (b) connecting transfer

Travellers' transfer patterns were revealed from the public assignment calculation. Based on this, the impact of an offset change for a bus line was assessed under the assumption that passengers' transfer patterns were unaffected by offset changes. Considering a given bus line, the first task was to identify the transfer points where passengers transferred to or from the bus line. Having identified all the transfer points of a bus line enabled us to estimate the optimisation potential.

In figure 3, possible effects (single transfer point) of offset changes can be observed. Offset changes of the bus line of interest are marked by thicker (forward) and dashed (backward) arrows in a time-space diagram. In figure 3 (a), a large forward offset change results in less transfer waiting time for passengers from run 2. On the other hand, a large backward offset change means that people from run 1 are not able to board the bus line. In such cases the given offset change was penalised heavily to avoid a situation where predicting the passenger adaptions became impossible.

Calculating the optimisation potential for feeding transfers was done as outlined below (the approach used for connecting transfers was to a large extent similar and therefore omitted). The calculations were performed for each transfer point and for every feasible offset change of the current bus line according to the results from the public assignment calculation.

1. Sort runs of current bus line according to their departure times (ascending order).
2. **Direct transfer.**
   Calculate the waiting time between feeding run *i* and run *1* (i.e. earliest departing run) of the current bus line at transfer station *s* in the following way
   $$w_{i,1}^{s} = dep_i^{s} - arr_1^{s}$$
   where dep_i^s is the time at which run *i* departs from station *s*, and arr_1^s is the time at which run *1* arrives at station *s*.
   If $w_{i,1}^S < 0$, then select the second earliest departing run of the current bus line and calculate w_{i2}^S. The process continues until $w_{i,j}^S$ turns positive or equals zero for a given *j* or until all runs of the current bus line are examined. In the latter case, penalise w_{ij}^S to avoid unpredictable offset changes.
3. **Transfer including walk.**
   Calculate the waiting time between feeding run *i* and run *1* of the current bus line at transfer station *s* in the following way

$$w_{i,1}^{s} = dep_{i}^{s} - (arr_{1}^{s1} + walk^{s,s1})$$

The parameter $walk^{s,s1}$ is the time it takes to walk from station *s* to station *s1*.
If $w_{i,1}^{S} < 0$, then select the second earliest departure and calculate $w_{i,2}^{S}$. The process continues until $w_{i,j}^{S}$ turns positive or equals zero for a given *j* or until all runs of the current bus line are examined. In the latter case, penalise $w_{i,j}^{S}$.

4. Multiply $w_{i,j}^{S}$ by the weight factor (number of passengers and their value of time) for the particular transfer.
5. The optimisation potential for the specific offset change is now equal to the difference between the value calculated in step 4 and the product of $w_{i,j}^{S}$ and $\omega_{i,j}^{S}$ calculated according to the do-nothing scenario (i.e. where offsets are not changed).

This process was repeated until both connecting and feeding transfers for all bus lines had been examined.

2.2.3 Imposing offset changes

Having calculated the optimisation potential of every feasible offset change of all bus lines, the next step was to impose a subset of these. Examining the optimisation potential of every offset change enabled us to impose the offset change with the largest optimisation potential, under the condition that the current bus line was not labelled as tabu. After imposing the most promising offset change, the bus line was labelled as tabu. We also prohibited offset changes on bus lines comprised in the sub-network of the current bus line (see section 2.2.4).
This process continued until no offset changes could be imposed without violating the sub-network constraint, and no positive optimisation potentials existed for any bus lines not labelled as tabu. The reason for labelling a bus line as tabu rather than only labelling a certain offset change as tabu was to ensure sufficient diversification, when exploring the solution space.

2.2.4 Sub-networks

Performing several offset changes based on optimisation potentials without calculating their exact impact from a public assignment calculation was based on dividing the transit system into sub-networks with no or only negligible passenger interaction. Sub-networks were not predefined and static, but simply created on the go when offset changes were imposed. Every bus line had its own sub-network comprising all bus lines crossing its trajectory and bus lines with passenger interaction (either direct or by a walking link). This meant that the order in which bus lines were chosen affected the way in which sub-networks were formed, hence prohibiting certain bus lines to having their offset changed. After imposing an offset change on a bus line, we prohibited offset changes on bus lines comprised in this bus line’s sub-network. Simply because changing the offset of two bus lines with significant passenger interaction could have a counteracting effect on the total waiting time.

The assumption about no passenger interaction between different sub-networks was legitimate, when journeys in the transit system only consisted of either one or two trips (e.g. Bus or Bus->Train). This assumption was important to impose offset changes on more than one bus line before running another transit assignment, given the large calculation time of the assignment model. However, the output data did not reveal passengers’ exact route choice, only transfer patterns were revealed, not the entire journey. Therefore, only services with direct passenger interaction were identified from the output data. From the Danish national transport survey, we know that only 5 % of all transit journeys consist of three or more trips (DTU Transport, 2013). If the second leg in a three leg journey was performed by bus (around 1 % in total according to DTU Transport (2013)), first and last legs were comprised in the sub-network. Hence, only in around 4 % of all transit journeys, a part of the passenger interaction was left unrevealed when applying the

described methodology. Consequently, the optimisation potential estimated for every offset change should be close to the one revealed from the transit assignment.
When applying this methodology to other transit networks, it is essential to have knowledge about the amount of journeys consisting of 3 or more legs. The higher the share of long-chained trips, the less certain the estimated optimisation potentials become. Therefore, under particular severe circumstances (e.g. networks where the majority of journeys are long-chained) it can be necessary to assess the optimisation potential with a transit assignment calculation. However, developing smarter strategies might be a first step e.g. creating larger sub-networks.

2.2.5 Stopping criterion

The process of imposing potential improving offset changes (upper level) continued until no feasible and improving offset changes remained. Then another public assignment calculation (lower level) was run to reveal the adapted passenger behaviour. After a public assignment calculation, all non-tabu bus lines were again subject to offset changes. The entire process (upper and lower levels) continued until the objective value converged.

2.2.6 Pseudo-code

The following pseudo-code gathers the threads from the previous sub-sections and presents the entire algorithm in a clearer way.

Initialisation, Run public assignment, T_{ijc}^{S}

<u>Upper-level problem</u>
Calculating optimisation potentials
For all non-tabu bus lines *BL*

  For all feasible offset changes *OC*

    Calculate optimisation potential *OP*

Store values *(BL, OC, OP)* in a list *L*

Imposing offset changes, w_{ij}^{S}

Continue the following until *L* is empty

  If $OP < 0$, then remove from *L*

  Impose the offset change with the largest *OP* in *L*

  Label *BL* as tabu

  Derive sub-network *sn* for *BL*

  Remove all values *sn* from *L*

Go to Lower-level problem

<u>Lower-level Problem</u>
Run public assignment, T_{ijc}^{S}
Calculate solution value
If stopping criterion is met, terminate.
Otherwise, go to upper-level.

The initialisation yielded passengers' travel behaviour. The relevant information in this context was how many passengers $T_{ijc}{}^{S}$ were transferring between services at which stations. Recall that $T_{ijc}{}^{S}$ was the number of passenger transferring from bus line *i* to bus line *j* at transfer stop *s*, while *c* referred to the different passenger groups. The minimisation problem was a bi-level minimisation problem, where the upper-level problem was the timetable optimisation. Based on the number of passengers transferring, $T_{ijc}{}^{S}$, calculated in the lower-level problem, bus lines' offset values were changed to optimise transfer waiting time, $w_{ij}{}^{S}$. The modified offset values served as input (together with the other network characteristics) for the lower-level problem, where passenger flows were derived by a route choice model. The lower level problem yielded the number of passengers transferring between two lines at a certain stop, $T_{ijc}{}^{S}$. This sequential bi-level optimisation process continued until the stopping criterion was met.

3 Data

The described optimisation approach was tested on the public transit network in Denmark on the basis of the newly developed Danish National Transport Model. This model is currently under development and the final version (2.0) is scheduled for 2015. The version used for public assignment calculations in this study was version 1.0. In this section the model and the bus network in Denmark are described.

3.1 Public assignment model

The public assignment model was schedule-based, which meant that every single run of the bus lines was described. Demand was assigned uniformly within 10 different time-of-day periods. The model applied a utility-based approach to describe travellers' perceived travel costs. The formulation of the utility function reflected the perceived cost of travelling from zone *i* to zone *j* at time *t* for passenger group *c* (i.e. generalised travel cost) as follows.

$$\begin{aligned} C_{ijtc} &= \beta_c * WaitingTime_{ij} + \beta_c * WaitInZoneTime_{ij} + \\ &\beta_c * WalkTime_{ij} + \beta_c * ConnectorTime_{ij} + \\ &\beta_c * NumberOfChanges_{ij} + \beta_c * TotalInVehicleTime_{ij}, \forall t, c \end{aligned} \tag{12}$$

In this formula, C_{ijtc} is the utility, *WaitingTime* is the transfer waiting time, *WaitInZoneTime* is the waiting time at home or in the origin zone, *WalkTime* is the walking time used when transferring, *ConnectorTime* is the time used for getting from home to the desired transit station, *NumberOfChanges* is the number of transfers during a journey, and *TotalInVehicleTime* is the time spent driving in transit vehicles. Together these parameters reflect each traveller's disutility associated with a trip in the transit system. The *β's* represent the weights of each of the 6 parameters. For each passenger group the beta values are outlined in table 1. All beta values except the ones for *ChangePenalty* are in DKK/minute. *ChangePenalty* is an impedance cost incurred for every transfer[3]. Based on this it is easy to tell that transit users are assumed to be transfer averse e.g. commuters prefer four minutes extra travel time to a journey including a transfer.

[3] The values build on the critical study of Nielsen (2000), but were recalibrated in the Danish national transport model to fit the passenger flows to observed counts.

Table 1 - beta values (VoT) and share of trips

Trip types	WalkTime	Waiting Time	Connector Time	Wait InZone	Change Penalty	Bus InVehicleTime	Share of all trips
Commuter	0.633	0.59	0.64	0.28	2.20	0.56	42.3 %
Business	4.50	4.50	4.50	2.35	18.8	4.70	2.3 %
Leisure	0.209	0.21	0.21	0.117	1.10	0.19	55.4 %

The transit fare system in Denmark is mostly OD-based, hence to a large extent independent of passengers' route choice. It was thus not the level of VoT as such that influence passengers' route choice but rather the ratio between the different time components. The fact that business travellers had significantly large time values compared to commuter trips and leisure trips did not bias the optimisation, since business trips only comprised 2.3 % of all transit trips.

Travellers' route choice behaviour was based on utility maximisation and the travellers were assumed to have complete knowledge of the entire network and timetables. Albeit this assumption seems optimistic, passenger information has reached a level with real-time information available on webpages, cell phone-apps and stations, which implies that the assumption in many ways is realistic. Theoretically, this means that passenger flows derived from SUE and UE become similar. Deriving the SUE of a transit network, passengers optimise their perceived utility from their known set of paths from origin to destination. In the UE, passengers are assumed to be familiar with all paths, and choose the one that maximise their utility. Providing the passengers with sufficient real time information on the state of the transit system, the UE and SUE coincide since the perceived utility become equivalent to the objective utility.

The network loading was done by an all-or-nothing assignment where all passengers were loaded onto the routes that maximised their utility. It could be argued that it would have been more realistic if the stochastic user equilibrium was found instead. However, due to uniformly distribution of departure time, different routes might have been used between each OD-pair at different departure times during the day. Likewise, it is extremely seldom that passengers in the Danish transit network are rejected entering coaches due to crowding, and most passengers get seats while on-board. A pure user equilibrium method would hence resemble an all-or-nothing model with very few exceptions. For the same reason vehicle capacity (i.e. also the ability to board a transit vehicle) is not modelled in the Danish national transport model. However, crowding could be built into the route choice model with a flow-dependent cost function. The passenger flows that were optimised would hence depend on the crowding function.

The method developed in the current paper was a bi-level optimisation between the timetable optimisation resulting in a better synchronisation and a route choice model deriving passengers' travel behaviour. A better synchronisation would generally yield benefits independent on crowding. For example if 100 passengers transferred from a low frequent train at time 00 to a bus service with 10 minutes frequency and a capacity per bus of 80 passengers. If the prior bus schedule ran at minutes 08 and 18, then 80 passengers had to wait 8 minutes and 20 had to wait 18 minutes. If the optimised schedule was 01 and 11, then both groups of passengers would gain 7 minutes of transfer time.

It is true though, that the political constraint that all schedules should keep their frequency – e.g. 10 minutes – may be less well in the crowded case than the uncrowded, e.g. in the example above one may run two busses at 01. This is a trickier problem though, because then you will also get a lower frequency along the route, hence more waiting time for the non-transferring passengers who arrive at the station at random. It is trivial to relax the headway restriction. If the network is recoded so each departure has its own line number, and a crowding function introduced, then all departures will be optimised independent of each other, and the crowding case mentioned above be solved. The optimisation would presumably result in larger reductions of the KPIs introduced in section 4. But the structure and memorability of the timetable will be lost, and at the same time the requirements for rolling stock may also increase. Finally, passengers arriving at their station at random will be affected by longer waiting times due to busses bunching up.

3.2 The public transport network in Denmark

In figure 4 all transit lines in Denmark are outlined. This figure shows that the transit network in Denmark consists of several smaller networks in the larger cities and a fair amount of regional and inter-city lines connecting these. The network consists of the following:

- 1,794 public lines (e.g. train, bus, metro, S-train etc.), of which 1,440 are busses
- 8,373 variants of the public lines, of which 7,877 are busses
- 22,187 stops, of which 21,396 are bus-stops
- 1077 zones, 3 trip purposes and hence app. 3.5 million OD cells

Since train schedules have many restrictions – e.g. limited overtaking possibilities at double-track lines and fixed meeting stations at single-track lines – it was decided to assume that train schedules were fixed and only bus schedules were subject to offset changes. Passengers transferring to/from other modes than busses were considered as well.

O/D-matrices for 10 different time intervals representing a single day were used to describe the demand. Within each time interval, travellers were split uniformly into 2-minutes intervals and launched within each of these. In the test of the approach, only evening period from 6 pm to 9 pm was considered since most transit lines ran with lower frequency, potentially leaving a larger potential for improvement.

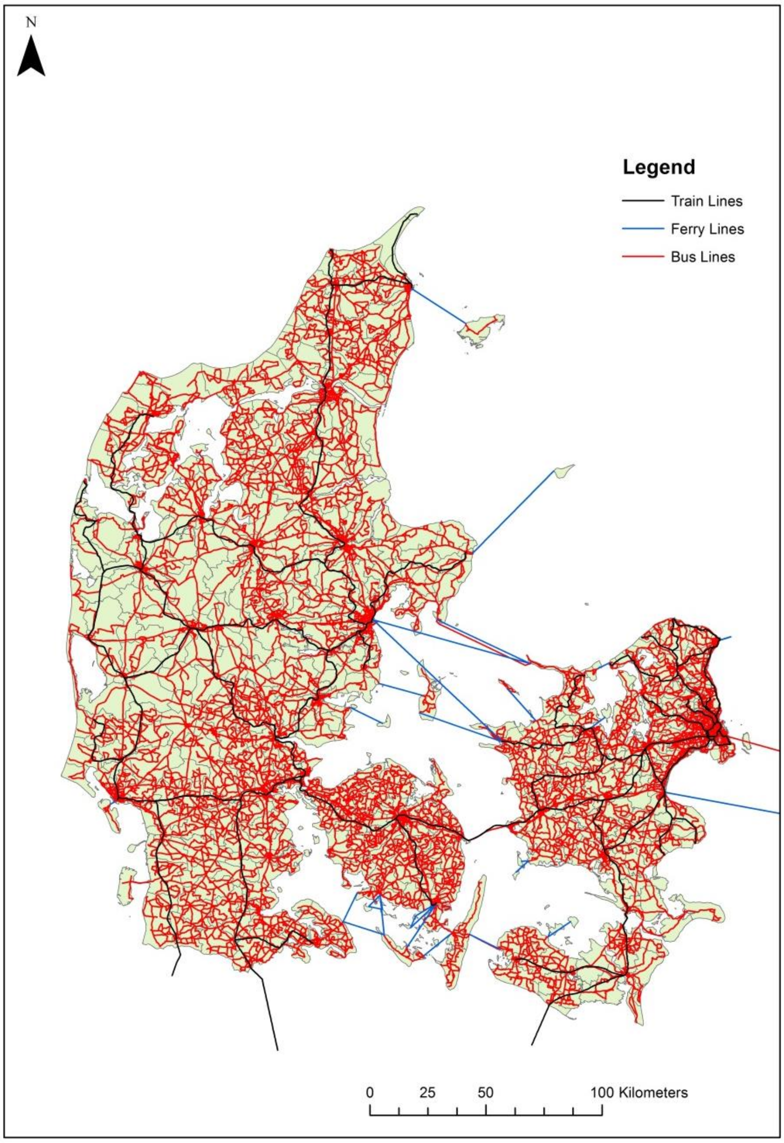


Figure 9 - Transit lines in Denmark

In Denmark, there are some provincial towns with minor bus networks. However, even in more provincial parts of the network, there are regional busses that connect towns, and it is hence difficult to extract sub-networks. Therefore, it was chosen to consider the entire bus network of Denmark as subject to the timetable optimisation. An important point for considering the entire bus network was also that the algorithm was designed to optimise under variable demands, and did not change interdependent bus lines at the same iteration. Instead, the most promising ones were modified while all crossing bus lines were locked (see section 2.2.4). Afterwards, passenger flows were recalculated with the route choice model and based on these, the objective value was updated. If all bus-routes (including dependent bus lines) were optimised in a sub-network at the same time in the inner loop, there was the risk, that flows and hence also the objective function in the next iteration deviated too much from the flows in the previous iteration, and that the algorithm thus would oscillate and converge slower. A key assumption in the algorithm was thus, that the candidate bus line was solved to optimality given the flows from the prior iteration. This made this bus line more attractive and it meant that the flows on this bus line would change (mainly increase, but also shift between departures). The transfer patterns to all crossing bus lines would then change, but how they actually changed was first revealed when the route choice model was run in the next iteration of outer bi-level problem.

To get an impression of the complexity of the bus network, figure 5 shows an example of bus lines (turquoise) that crosses bus line 200S in Copenhagen. It is clear that there are many interdependencies in the network, and if these bus lines were to be optimised on flows that were too far from the timetable in the present iteration, the algorithm would never converge.

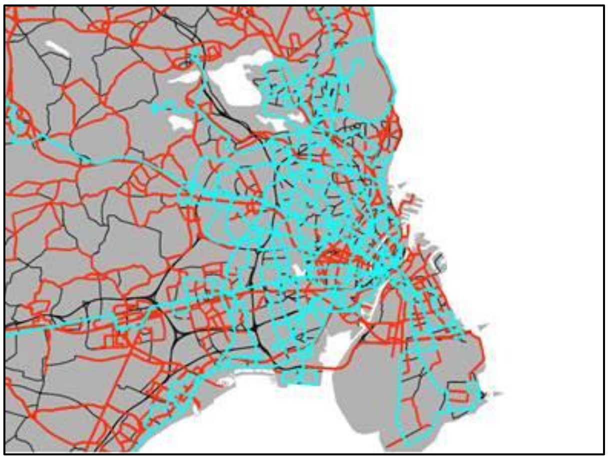


Figure 10 - Transit lines crossing bus lines 200S

4 Results and discussion

This section presents the results obtained when applying the heuristic solution approach to timetable optimisation to the Danish public transport network. It should be noted that the network contained transit lines from all over Denmark, but only bus lines were subject to offset changes. Table 2 shows that it took 5 iterations before the bi-level timetable optimisation converged. At this point, the weighted waiting time was reduced by more than 5 %, while barely affecting the other time components of the journeys. Figure 6 illustrates that the largest improvement in objective value occurred over the first couple of iterations, while the latter iterations showed that the objective value converged.

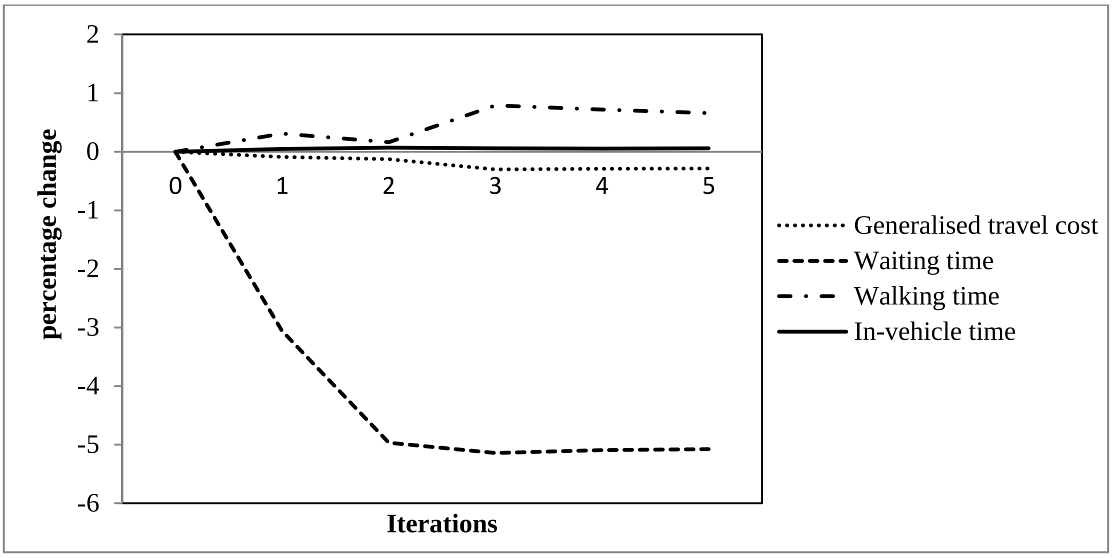


Figure 11 - Development in solution value

The generalised travel cost was reduced by 0.3 %, which meant that transit users in general were better off than before the optimisation. By showing this, it was proven that the weighted transfer waiting time was not just reduced to the detriment of something else. The generalised travel cost would reveal if offset changes implied that busses started to bunch up. In that case passengers would have to wait longer at their boarding stops and also at some non-prioritised transfer stops. In the current study, bus bunching was prevented by only allowing offset changes to be imposed on entire groups of busses at the time.

Walking time increased by 0.66 % which was explained by the increase in attractiveness of the transfer possibilities. Therefore, users in the transit system might have preferred to add an extra transfer to get from A to B, even though it included some walking. From table 1 it is evident that beta values for walking time and waiting are almost equal. This could be interpreted as passengers, who have chosen to make a transfer, were indifferent between waiting and walking. Likewise, it was seen that these beta values were close to the ones for in-vehicle travel time. Therefore, it was reasonable to assume that Danish transit users were more focused on the total travel time from origin to destination rather than whether they were walking, waiting or being in a transit vehicle.

The small increase in in-vehicle time could be a result of a minor change in the users' travel patterns due to offset changes as well as passengers shifting from long access mode (walk/bicycle) to stations to bus, if the bus was better coordinated with the train service.

Despite exhibiting monetary reductions in the different time components as outlined in table 2, the monetary reduction is not equivalent to an increase in revenue for the bus company. It should rather be seen as what passengers are willing to pay to avoid excessive waiting time, e.g. as evaluated in socio-economic cost benefit analyses (CBA). Since bus companies' operations are subsidised by the public authorities in Denmark (government, regions, municipalities), the CBA is indeed a criterion when deciding upon the subsidy.

Table 2 - Development in solution value

	0^{th} iteration			
	Generalised travel cost	Waiting time	Walking time	In-vehicle time
Total (DKK)	6 320 322.78	125 777.54	46 925.99	1 598 292.26
	1^{st} iteration			
Total (DKK)	6 314 552.19	121 932.69	47 071.99	1 599 006.58
Absolute change	5 770.59	3 844.86	-146.00	-714.32
Percentage change	-0.09	-3.06	0.31	0.04
	2^{nd} iteration			
Total (DKK)	6 312 161.07	119 527.52	47 001.78	1 599 357.19
Absolute change	8 161.71	6 250.02	-75.79	-1 064.92
Percentage change	-0.13	-4.97	0.16	0.07
	3^{rd} iteration			
Total (DKK)	6 301 220.29	119 311.98	47 297.51	1 599 200.96
Absolute change	19 102.49	6 465.56	-371.52	-908.70
Percentage change	-0.30	-5.14	0.79	0.06
	4^{th} iteration			
Total (DKK)	6 302 028.98	119 369.32	47 264.52	1 599 175.41
Absolute change	18 293.80	6 408.22	-338.53	-883.15
Percentage change	-0.29	-5.09	0.72	0.06
	5^{th} iteration			
Total (DKK)	6 302 235.64	119 394.24	47 235.01	1 599 243.69
Absolute change	18 087.14	6 383.30	-309.02	-951.43
Percentage change	-0.29	-5.08	0.66	0.06

Examining benefits at the zone level exhibited in figure 7 enabled three different conclusions to be drawn. For the four largest cities in Denmark (in terms of population) it was seen that the zones around Aalborg and Copenhagen experienced less waiting time reduction than the zones around Aarhus and Odense. This was explained by the great focus on coordination of public transport in Copenhagen and Aalborg the recent years and a strategy of less bus lines with a high frequency, so called “A-busses” or “metro-busses” respectively. However, it was encouraging to see that the applied methodology yielded an even further reduction in waiting time (better coordination). On the western part of the Zealand a lot of zones experienced a significant reduction of weighted waiting time. In this part of Zealand, the old county planned bus lines independently of the railway operations. For that reason no coordination between the two modes were planned. Furthermore, the county did not collect detailed counts and thus did not systematically evaluate flows. Therefore, transit users in these zones benefited great from such a bus timetable optimisation. A somewhat similar pattern was seen on the island of Funen around the city of Odense. The transit planning on Funen used to be split between the 32 different municipalities, which explained the lack of overall coordination compared e.g. to Copenhagen, where the public company Movia (prior HUR, and HT) organise and tender the bus operations on behalf of all municipalities. It should be noted that a recent public sector governance reform has extended Movia’s responsibility to the whole Zealand and made a similar organisation for Funen, but this has not yet led to major changes of the timetabling yet.

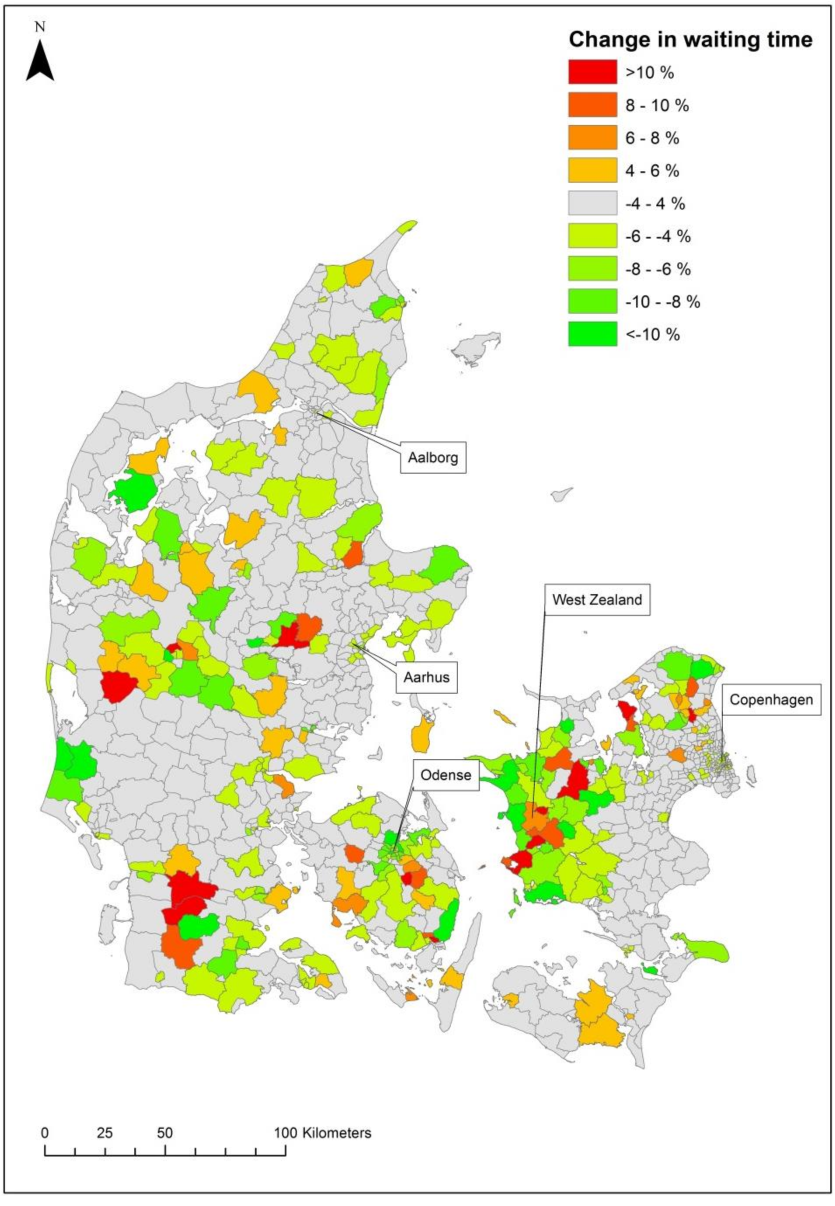


Figure 12 - Waiting time change (percentage) at zone level (Denmark)

The developed timetable optimisation approach was also applied to the transit network in the Greater Copenhagen area, where the timetable used during the morning peak hours served as initial solution. The change in weighted waiting time at the zone level is shown in figure 8, where it is seen that primarily zones lying along the train lines experience a better coordination of transit services. This was because of the large passenger flows at train stations (e.g. from bus to train or the opposite way around). In the Greater Copenhagen area train stations were often used as bus hubs, which meant that several bus lines served as feeder modes and connecting modes for the trains here. Despite the reduction during the morning peak hours for the Greater Copenhagen area, it was not possible to obtain the same large magnitude of weighted waiting time reduction as for the evening period. This was explained from the higher frequency during rush hours compared to the evening period examined in the national transport model (see figure 6).

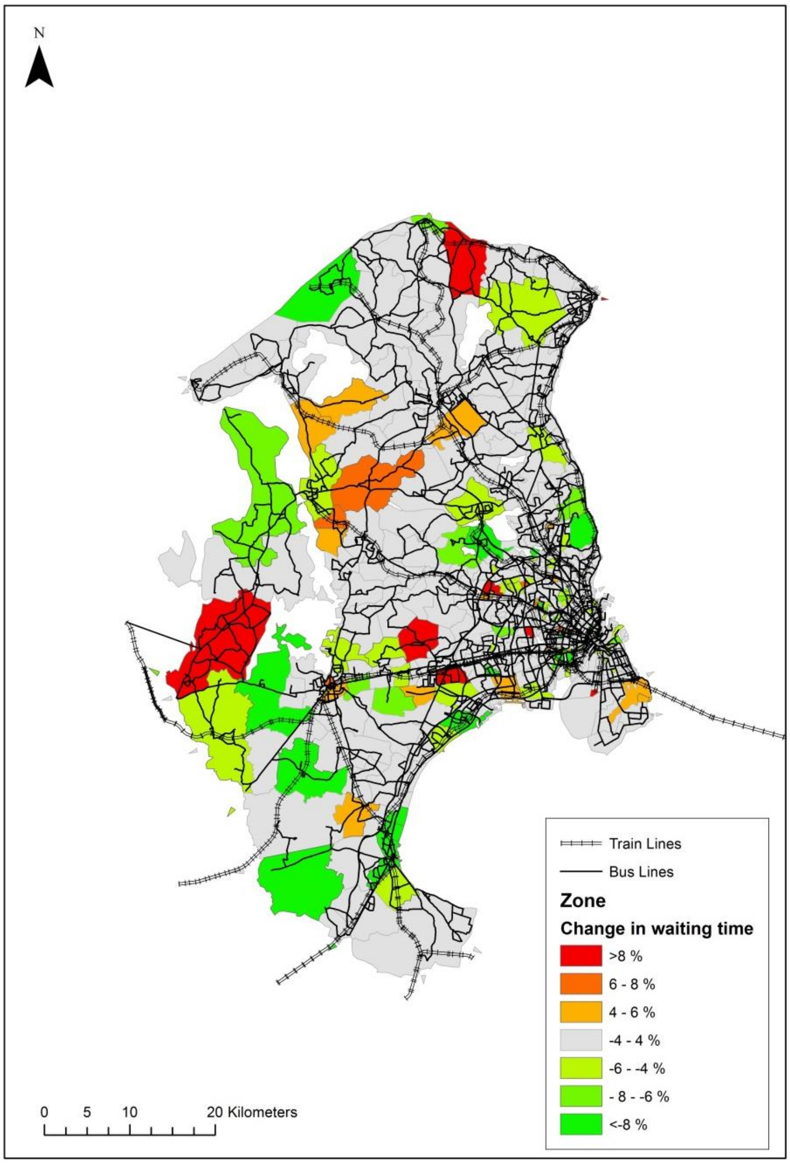


Figure 13 - Waiting time change (percentage) at zone level (Greater Copenhagen area)

The improvement of the generalised cost adds up to approximately 45 million DKK of value of time gains on a yearly basis (if summarised to a full day of operations and a full year), which is a large benefit for the society, given that it entirely comes from optimisation of the timetabling at limited extra costs. It is important to be aware that the reduction in weighted transfer waiting time that the present optimisation yielded is not equivalent to increased revenue for the bus company. Assigning monetary values to waiting time should rather be seen as an indication of how much passengers are willing to pay to avoid this extra waiting time. However, when travelling by bus becomes more attractive from a user's perspective, bus as travel mode may experience an increase in market share, hence providing larger revenue. Fortunately, an improvement in transfer waiting time as outlined in this paper can be obtained at a negligible cost. In this regard it should be mentioned that there is a gap between research and practice. Despite the promising results outlined, the bus company may have several other reasons to consider changing the timetable inappropriate.

It is important to acknowledge the deviating desires among the different stakeholders. According to one of the Danish bus operators, Movia, the fleet size required to fulfil the contractual obligations is, because of the high expenses, the main priority for bus companies. Customers, on the other hand, are mainly interested in a fast and reliable service between origin and destination (Ceder, 2007). Therefore, in the case a transfer is needed to complete a journey in the transit system, passengers prefer if the system is coordinated in such a way that waiting time is minimised. The two different desires may not always go hand in hand as minimising transfer waiting time can cause an extra need for rolling stock, while minimising the fleet size can imply longer waiting time when transferring for the timetable to be feasible.

Regarding the gap between research and practice one way of improving the real-life applicability could be to assume stochastic travel times rather than deterministic travel times as in this study. Nuzzolo et al (2012) applied a doubly dynamic schedule-based transit assignment taking into account how frequent transit users, based on a learning process, adapted their departure time, boarding stop choice and run choice. In this way, the results would be more robust towards deviations from the scheduled travel times and presumably more similar to real life. However, this would be extremely demanding in terms of computation time with such a large network as the Danish (see specifications in section 3.2). In this case, also estimates of average delays should be known and extra buffer time in the transfer times should, if necessary, be imposed. Regarding the current model, this is not a major change, but when delay data is available only minor changes in the calculation of transfer times should be made. Nevertheless, this study should be seen as a proof of concept for combining optimisation methods and assignment models in transit systems. Potential extra features would be a topic for future research.

5 Conclusion and future work

The current study proposed a timetable optimisation approach that explicitly considered passengers' modified route choice as a reaction to changes in the timetable. The approach was applied to a real large-scale transit network. The optimisation yielded a significant reduction in weighted transfer waiting time, while only affecting the in-vehicle travel time and the generalised travel journey cost to a lesser extent.

Overall, the study contributes to the literature by proposing a new optimisation tool, which has its strengths with respect to both the timetable optimisation and the reliability perspective, since changes in the travellers' route choice decisions are considered explicitly whereby demand effects are taken into account.

A topic for future research could be to use stochastic rather than deterministic travel times. Another additional feature could be to consider the problem as multi-objective, e.g. minimise fleet size and transfer waiting time. In this way, convincing bus companies of the applicability of the results might be a smoother process. Finally, the order in which offset changes were imposed could be changed from the "greedy" approach introduced in this paper, where the feasible offset change with the largest optimisation potential was chosen. Instead, a knapsack-inspired approach could be applied. Combining the bus lines (for which the offset should be changed) in a way that yields the largest total potential improvement, while still obeying the idea of sub-networks, could be an idea for future research.

References

Bielli, M., Caramia, M., & Carotenuto, P. (2002). Genetic algorithms in bus network optimization. *Transportation Research Part C: Emerging Technologies*,*10*(1), 19-34.

Bookbinder, J. H., & Désilets, A. (1992). Transfer optimization in a transit network. *Transportation Science*, *26*(2), 106-118.

Ceder, A. (2007). *Public transit planning and operation: theory, modeling and practice*. Elsevier, Butterworth-Heinemann.

Ceder, A. A. (2011). Public-transport vehicle scheduling with multi vehicle type. *Transportation Research Part C: Emerging Technologies*, *19*(3), 485-497.

Cevallos, F., & Zhao, F. (2006). Minimizing transfer times in public transit network with genetic algorithm. *Transportation Research Record: Journal of the Transportation Research Board*, *1971*(1), 74-79.

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

DTU Transport. (2013) "Transportvaneundersøgelsen 2006-2012".

Fan, W., & Machemehl, R. B. (2006). Optimal transit route network design problem with variable transit demand: genetic algorithm approach. *Journal of transportation engineering*, *132*(1), 40-51.

Feil, M., (2005). Optimisation of Public Transport Timetables with respect to Transfer. *Master thesis*

Glover, F. (1990). Tabu search: A tutorial. *Interfaces*, *20*(4), 74-94.

Guihaire, V., & Hao, J. K. (2010). Improving timetable quality in scheduled transit networks. In *Trends in Applied Intelligent Systems* (pp. 21-30). Springer Berlin Heidelberg.

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

Hadas, Y., & Ceder, A. A. (2010). Optimal coordination of public-transit vehicles using operational tactics examined by simulation. *Transportation Research Part C: Emerging Technologies*, *18*(6), 879-895.

Ibarra-Rojas, O. J., & Rios-Solis, Y. A. (2012). Synchronization of bus timetabling. *Transportation Research Part B: Methodological*, *46*(5), 599-614.

Jansen, Leise Neel; Pedersen, Michael Berliner & Nielsen, Otto Anker (2002). Minimizing Passenger Transfer Times in Public Transport Timetables. *7th Conference of the Hong Kong Society for Transportation Studies, Transportation in the information age.* Proceedings, pp.229-239. 14 December, Hong Kong.

Knoppers, P., & Muller, T. (1995). Optimized transfer opportunities in public transport. *Transportation Science*, *29*(1), 101-105.

Kepaptsoglou, K., & Karlaftis, M. (2009). Transit route network design problem: review. *Journal of transportation engineering*, *135*(8), 491-505.

Lee, Y. J., & Vuchic, V. R. (2005). Transit network design with variable demand. *Journal of Transportation Engineering*, *131*(1), 1-10.

Liu, Z., SHEN, J., WANG, H., & YANG, W. (2007). Regional bus timetabling model with synchronization. *Journal of Transportation Systems Engineering and Information Technology*, *7*(2), 109-112.

Ma, T. Y. (2011). A hybrid multiagent learning algorithm for solving the dynamic simulation-based continuous transit network design problem. *Technologies and Applications of Artificial Intelligence (TAAI), 2011 International Conference on* (pp. 113-118). IEEE.

Nachtigall, K., & Voget, S. (1996). A genetic algorithm approach to periodic railway synchronization. *Computers & Operations Research*, *23*(5), 453-463.

Nielsen, O.A. (2000). A Stochastic Traffic Assignment Model Considering Differences in Passengers Utility Functions. *Transportation Research Part B Methodological. Vol. 34B*, No. 5, pp. 337-402. Elsevier Science Ltd.

Niu, H., & Zhou, X. (2013). Optimizing urban rail timetable under time-dependent demand and oversaturated conditions. *Transportation Research Part C: Emerging Technologies*, *36*, 212-230.

Nuzzolo, A., Crisalli, U., & Rosati, L. (2012). A schedule-based assignment model with explicit capacity constraints for congested transit networks. *Transportation Research Part C: Emerging Technologies*, *20*(1), 16-33.

de Palma, A., & Lindsey, R. (2001). Optimal timetables for public transportation. *Transportation Research Part B: Methodological*, *35*(8), 789-813.

RH, (2009). Foer biltrafikken staar stille Hvad kan den kollektive transport bidrage med? *Region Hovedstaden*. Report. June 2009.

Wang, J. Y., & Lin, C. M. (2010). Mass transit route network design using genetic algorithm. *Journal of the Chinese Institute of Engineers*, *33*(2), 301-315.

Wong, R. C., Yuen, T. W., Fung, K. W., & Leung, J. M. (2008). Optimizing timetable synchronization for rail mass transit. *Transportation Science*, *42*(1), 57-69.

Yan, Y., Meng, Q., Wang, S., & Guo, X. (2012). Robust optimization model of schedule design for a fixed bus route. *Transportation Research Part C: Emerging Technologies*, *25*, 113-121.

Yu, B., Yang, Z., & Yao, J. (2009). Genetic algorithm for bus frequency optimization. *Journal of Transportation Engineering*, *136*(6), 576-583.

Zhao, F., & Ubaka, I. (2004). Transit network optimization-minimizing transfers and optimizing route directness. *Journal of Public Transportation*, *7*(1), 63-82.

Appendix 2: Parbo et al. (2015a)

Passenger perspectives in railway timetabling: A literature review

Jens Parbo, Otto Anker Nielsen & Carlo Giacomo Prato

Technical University of Denmark, Department of Transport, Bygningstorvet 116B, 2800 Kgs. Lyngby, Denmark

Accepted for Publication in *Transport Reviews*, 2015.

Abstract

When looking at railway planning, a discrepancy exists between planners who focus on the train operations and publish fixed railway schedules, and passengers who look not only at the schedules but also at the entirety of their trip, from access to waiting to on-board travel and egress. Looking into this discrepancy is essential, as assessing railway performances by merely measuring train punctuality would provide an unfair picture of the level of service experienced by passengers. Firstly, passengers' delays are often significantly larger than the train delays responsible for the passengers to be late. Secondly, trains' punctuality is often strictly related to too tight schedules that in turn might translate into knock-on delays for longer dwelling times at stations, trip delays for increased risk of missing transfer connections, and uncertain assessment of the level of service experienced, especially with fluctuating passenger demand.

A key aspect is the robustness of railway timetables. Empirical evidence indicates that passengers give more importance to travel time certainty than travel time reductions, as passengers associate an inherent disutility with travel time uncertainty. This disutility may be broadly interpreted as an anxiety cost for the need for having contingency plans in case of disruptions, and may be looked at as the motivator for the need for delay-robust railway timetables. Interestingly, passenger-oriented optimisation studies considering robustness in railway planning typically limit their emphasis on passengers to the consideration of transfer maintenance. Clearly, passengers' travel behaviour is far more complex and multi-faceted and thus several other aspects should be considered, as becoming more and more evident from passenger surveys.

The current literature review starts by looking at the parameters that railway optimisation/planning studies are focused on and the key performance indicators that impact railway planning. The attention then turns to the parameters influencing passengers' perceptions and travel experiences. Finally, the review proposes guidelines on how to reduce the gap between the operators' railway planning and performance measurement on the one hand and the passengers' perception of the railway performance on the other hand. Thereby, the conclusions create a foundation for a more passenger-oriented railway timetabling ensuring that passengers are provided with the best service possible with the resources available.

Keywords

Railway planning, Railway operations, Passengers' perceptions, Timetable robustness, Timetable optimisation.

1 Introduction

Firstly, the purpose of the present literature review is to review the vast amount of papers focusing on different aspects of designing a timetable with a train-oriented focus. Secondly, the review compares these approaches to the way passengers perceive and value railway operations. Based on the gap between passengers' perception of railway operations and the way timetables are designed, suggestions for future research directions are outlined.

A discrepancy exists between how train-oriented railway operations are planned with the main focus being on the trains and how passengers actually perceive and respond to railway performances. The ideal and simplistic vision of railway operations is that trains run according to the planned schedule. Transportation science has dedicated a great deal of attention to methods guaranteeing that trains run on time regardless of what passengers do. In reality however, disturbances to the timetable are frequent and affect passengers. Contradicting desires to an effective usage of resources between operators and passengers mean that the optimal timetable for the operators could be far from optimal for the passengers (Medeossi et al., 2009; Schöbel & Kratz, 2009). Addressing this discrepancy is crucial as passenger delays are often larger than the train delays responsible for the passengers being delayed (Vansteenwegen & Oudheusden, 2007; Nielsen et al., 2008). Empirical evidence showed that passenger on-time performance was up to 10 percentage points below train punctuality during peak hours, with the reasons being cancelled trains, missed transfers and/or route choice adaptations (Nielsen et al., 2008).

For the passengers to rely on the timetable, travel time should be stable and kept at a minimum. Typically, contracts with railway operators have explicit targets for trains' on-time performance and railway operators not providing punctual service face financial penalties (Noland & Polak, 2002). To meet the punctuality goals, planners and researchers have tried to make railway operations more robust. Several definitions of robustness supporting timetable planning have been proposed, although a general definition of timetable robustness was recently noted as missing (Dewilde et al., 2011; De-Los-Santos et al., 2012). Early robustness definitions focused on the ability to absorb minor disruptions and the recoverability of the schedule (Vromans, 2005; Bush, 2006; Salido et al., 2008; Cacchiani et al., 2009; Medeossi et al., 2009). Later definitions also considered the trade-off between having a very tight (nominal) timetable, with only minimum safety headway between subsequent trains, and having a (robust) timetable that could absorb minor delays, thus reducing travel time uncertainty (Schöbel & Kratz, 2009; Dewilde et al. 2011 & 2013). The difference between a nominal and a robust timetable was defined as the *price of robustness* (Schöbel & Kratz, 2009).

On-time performance of the trains is not always a major passenger concern and it is highly dependent on the network characteristics. In low frequency networks, passengers perceive on-time performance to be the most important service characteristic (Milan, 1996). However, in high frequency networks, regularity (i.e., the ability to keep equal headways between trains) is perceived as more important (Weston et al., 2006; Sun & Xu, 2012). Although on-time performance is known to be among the most important factors influencing mode choice (Carrasco, 2012), transit users are very unlikely to change mode choice in response to delayed trains (Batley et al., 2011). Instead, passengers start adapting the route choice or departure time choice accordingly (Benezech & Coulombel, 2013).

Passengers' perceptions of railway performances affect their travel behaviour (Nielsen, 2000) and hence addressing all attributes perceived as important by passengers is essential to improve railway timetables. Failing to do so and considering only a subset of attributes could potentially result in an only apparent timetable enhancement obtained at the expense of non-measured attributes highly relevant to passengers. In fact, it has been shown that when travellers experience transit vehicles leaving early one out of ten times, they tend to perceive the probability of transit vehicles leaving early as larger than the de facto 10% (Rietveld, 2005). A reason that travellers often underestimate the experienced quality is that low quality often coincides with peak periods, and they typically emphasise punctuality more during peak periods due to less flexible arrival times. Treating peak and non-peak delays equally is thus not fair when looking at the number of passengers affected and their arrival time flexibility (Vansteenwegen & Oudheusden, 2007). Planners designing timetables often focus on minimising travel time. Passengers are taken into account by using passenger counts (e.g. Liebchen et al., 2007; Schöbel & Kratz, 2009), but in the design process, passengers' adapted travel behaviour as a result of a changed timetable is generally neglected. Therefore, using old passenger counts yields an inaccurate picture of what can be expected by the passengers. Passengers' travel behaviour should be considered explicitly in order to assess what impacts passengers can expect from future timetables. Recent studies have emphasized that exhaustive passenger-oriented measurements are needed to avoid that passengers are given a low priority in railway planning (Medeossi et al., 2009; Carrasco, 2012; Andersson et al., 2013). For example, measurements of passengers' average delay would ensure that cases where some trains are delayed on purpose will be avoided if they imply reductions in the level of service.

The current review is structured as follows. Section 2 presents research focused on improving robustness-related attributes in railway planning from the planners' perspective. Section 3 first outlines passenger-oriented optimisation of railway operations and then examines how passengers actually perceive railway performances. Finally, section 4 summarises the gap between the planners' and the passengers' perspectives and proposes directions for future research within passenger-oriented railway planning.

2 Train-oriented railway timetabling

Traditionally, the focus in railway planning is train-oriented. To meet the on-time requirements of the trains, railway plans need to be robust against delays. Since future delays are unknown, planners try to approximate them, e.g. by using historical data (Hooghiemstra & Teunisse, 1998; Tsuchiya et al., 2006; Takeuchi et al., 2007; Kroon et al., 2008; Benezech & Coulombel, 2013). However, Sels et al. (2012) suggested that only historical delays that are expected to be present in a future timetable should be included. Yaghini et al. (2013) used a neural network model and applied it to the Iranian railways. This delay prediction made it easier for operators to create suitable timetables minimising delays, errors and problems for future railways. Milinkovic et al. (2013) developed a model based on a fuzzy petri net model aimed at estimating delays through train dispatchers' experience for railway networks where delays were not recorded regularly.

Planners have certain "tools" to make timetables robust against delays, e.g. adding time supplements, lowering heterogeneity (i.e. similarity in stopping patterns and headways), finding optimal speed and reducing interdependencies between trains (Schöbel & Kratz, 2009; Goverde, 2010; Salido et al., 2012). In the following, the usage of these "tools" is described in detail as well as their impact on operations and performance measurements. Table 1 provides a bird's eye overview of which "tools" have been used in the literature, and for each "tool" details whether it is explicitly considered, ✓, implicitly considered (✓), or not considered at all, ×. In table 1, it is distinguished whether the particular study applies an optimisation approach or the aim is more practice-oriented and descriptive. The distinction should make it easier to choose which reference to consult based on whether the reader is interested in thorough descriptions of certain characteristics or particular algorithmic aspects.

2.1 Time supplements

Time supplements are the extra time trains are assigned to the running time between two stations, the dwelling time at intermediate stations or the layover time at the terminal station. Time supplements are not to be confounded with buffer time, which is the time added to the minimum headway between subsequently running trains (see figures 1 and 2).

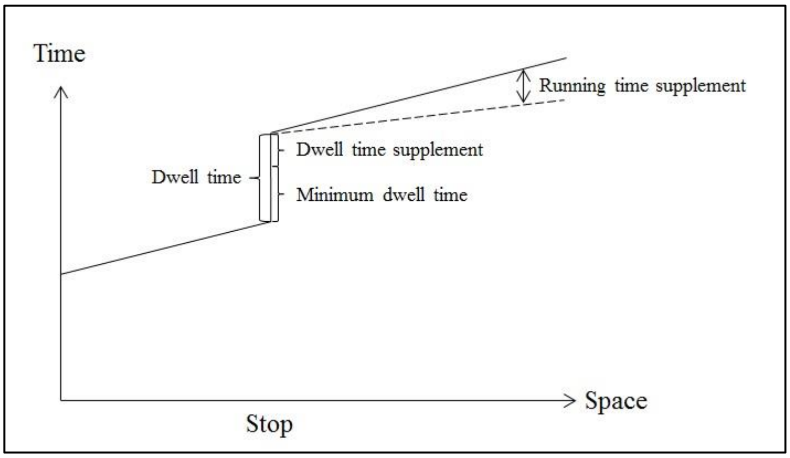


Figure 14 - Time supplements

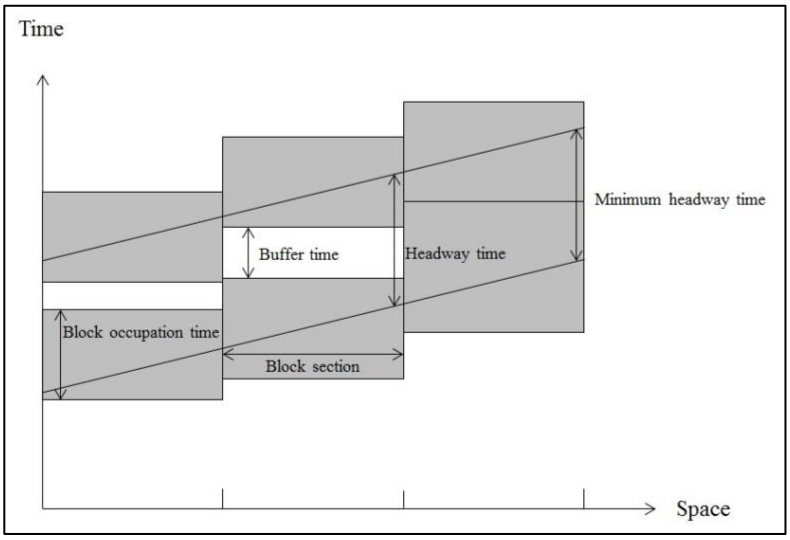


Figure 15 - Buffer time

2.1.1 Travel time determination

Determining the published travel time is not straightforward. In practice, running a certain trip would never result in the same travel time. Bush (2006) advocated for the use of the 85th percentile (according to historical travel time data) for the scheduled travel time. This specific percentile was chosen because the railway company in question was said to provide a decent level of service when being on-time in 85% of all trips. Choosing a sufficiently high percentile is equivalent to implicitly adding running time supplements. Oort (2011) suggested using the 35th percentile in low frequency systems to minimise passengers' travel time. The reason that two different recommendations were given was a result of different objectives: Bush (2006) focused on providing a certain mean level of service in high frequency networks, while Oort (2011) tried to minimise passengers' total travel time for long-headway services in a system where small delays did not impact other trains.

Table 3 - Overview of reviewed literature on train-oriented railway planning

Studies	Purpose	Measure/definition developed	Train related service characteristics addressed					
			Buffer time	Time supplements	Heterogeneity	Speed	Inter-dependencies	Flexibility
Optimisation								
Higgins et al. (1997)	Reducing interdependencies		✓	×	×	×	✓	(✓)
Burkolter et al. (2005)	Routing and train precedence		×	×	×	×	✓	×
Sun & Hickman (2005)	Recovery strategies		(✓)	×	(✓)	(✓)	×	(✓)
Liebchen et al. (2007)	Allocation of time supplements	Price of Robustness & Ratio of Delay	✓	✓	×	×	×	×
Salido et al. (2008)	Robustness of timetable	Robustness indicator	✓	✓	(✓)	×	(✓)	×
Cacchiani et al. (2009)	Allocation of time supplements		✓	✓	×	×	(✓)	×
Fischetti & Monaci (2009)	Allocation of time supplements	Light robustness	✓	✓	×	×	×	×
Schöbel & Kratz (2009)	Delay absorption	Price of robustness	✓	✓	×	×	×	×
Caprara et al. (2010)	Recovery strategies		✓	×	×	×	✓	✓
Goverde (2010)	Delay propagation	Delay propagation of initial delays	✓	✓	×	×	✓	×
Dewilde et al. (2011)	Delay absorption		✓	×	×	×	×	×
Armstrong et al. (2012)	Improve reliability		✓	×	×	×	×	×

De-Los-Santos et al. (2012)	Stability against failures	Network stability against intentional random infrastructure attacks	×	×	×	×	✓	✓
Gestrelius et al. (2012)	Reducing interdependencies		×	×	×	×	✓	×
Sels et al. (2012)	Assigning trains to platforms		×	×	×	×	✓	(✓)
Dewilde et al. (2013)	Train spread	Spread of trains	✓	✓	×	×	✓	(✓)
Planning & Describing characteristics								
Kikuchi & Vuchic (1982)	Skipping stops		×	×	✓	(✓)	(✓)	×
Carey (1999)	Train spread		✓	✓	×	×	×	×
Huisman & Boucherie (2001)	Punctuality	Mean train delay and train delay probability	✓	✓	×	×	×	×
Studies	Purpose	Measure/definition developed	Train related variables addressed					
			Buffer time	Time supplements	Heterogeneity	Speed	Inter-dependencies	Flexibility
Vromans (2005)	Heterogeneity	SSHR and SAHR	✓	×	✓	×	×	×
Bush (2006)	On-time performance		(✓)	×	×	✓	×	×
Hofman et al. (2006)	Recovery strategies		✓	×	×	×	(✓)	(✓)
Kroon et al. (2007)	Allocation of time supplements	Weighted Average Distance of the allocated time supplement	✓	✓	×	×	×	×

Yuan & Hansen (2007)	Delay absorption		✓	✓	×	×	✓	×
Fischetti et al. (2009)	Allocation of time supplements		✓	✓	×	×	×	×
Flier et al. (2009)	Detecting delays		×	×	×	×	✓	×
Medeossi et al. (2009)	Punctuality	Delay Frequency Index	✓	✓	×	×	×	×
Goerigk & Schöbel (2010)	Allocation of time supplements		✓	✓	×	×	(✓)	×
Andersson et al. (2011)	Allocation of time supplements		✓	✓	×	×	×	×
Oort (2011)	Travel time minimisation		×	×	×	✓	×	×
Forsgreen et al. (2012)	Adding flexibility		(✓)	✓	×	×	×	✓
Salido et al. (2012)	Delay absorption		✓	✓	✓	✓	×	×
Yamamura et al. (2012)	Detecting delays		×	×	×	×	✓	×
Goverde & Hansen (2013)	Defined timetabling levels	Definitions of 4 different timetable quality levels	✓	✓	(✓)	×	✓	✓
Andersson et al. (2013)	Robustness of timetable	Robustness in Critical Points	✓	✓	×	×	✓	✓

2.1.2 Robust optimisation

One paradigm in the optimisation of timetable robustness (focused on adding time supplements) is robust optimisation. Robust optimisation derives a feasible timetable including time supplements large enough to absorb the most frequently occurring delays. The amount of time supplements to allocate is a compromise between travel time extension and delay propagation risk. Having determined the total amount of time supplements, the question is where to allocate it. A proportional allocation of the running time supplements does not minimise average delay (Rietveld, 2005). Supplements on the earliest and in particular the last part of the line should be below the average supplement size. Delays incurred in the beginning are often relatively small, thus running time supplements are lost. Placing less time supplements in the end of a line may be done, because only delays on the last stations could be relieved from this, hence not affecting as many stations as possible (Vromans, 2005). Andersson et al. (2011) outlined the importance of being familiar with frequently occurring delays and, by comparing different train types with different distributions of allocated time supplement (early departures prohibited), they recommended allocating time supplements shortly after the points where the most frequent delays occurred.

Generally, robust optimisation had a tendency to add too many time supplements, resulting in a significant reduction of capacity utilisation (Fischetti & Monaci, 2009). With the typical definition of robustness focusing on absorbing minor delays, most definitions failed to describe the size of “minor” delays (e.g. Vromans, 2005; Bush, 2006; Salido et al., 2008; Cacchiani et al., 2009; Medeossi et al., 2009). Schöbel & Kratz (2009) applied robust optimisation and developed dynamic measures, which in theory could absorb even large delays. They measured the maximum number of passengers missing a connection and the sum of delays when all delays were below a certain threshold value. To account for the conservative aspects of robust optimisation, Fischetti & Monaci (2009) introduced light robustness. From the nominal timetable (without time supplements), a maximum deterioration of the objective function was fixed, thereby limiting the amount of added time supplements. Goerigk & Schöbel (2010) treated the robust timetabling problem differently. Having generated several feasible timetables with different amounts of time supplements, each of these was tested against different delay scenarios.

2.1.3 Measurements

To overcome the limitations of discrete service measurements (e.g., measurements of whether the train is punctual or not), continuous measurements are typically better to reflect the level of service passengers experience. Huisman & Boucherie (2001) introduced two performance measures related to punctuality: the mean delay of a train and the delay probability of a train. Although being measured only at the final station, the mean delay reflects the actual performance of the train more accurately than a discrete punctuality (on-time) measure. Instead, the delay probability of a train reflects the risk of a particular train being delayed.

To provide guidelines on how different distributions of time supplements impacted the level of service, Kroon et al. (2007) developed a measure to assess at which part of the line time supplements were added:

$$WAD = \sum_{t=1}^{N} \frac{2t-1}{2N} * s_t \qquad (12)$$

where *WAD* is the Weighted Average Distance of the allocated time supplement from the starting point, *N* is the number of consecutive trips *t* between stations, and s_t is the amount of time supplement on a particular trip *t*. This measure takes a value between 0 and 1, indicating at which part of the line time supplements are allocated, with 0.5 indicating a uniform allocation. This measure was later used by Fischetti et al. (2009) who focused on minimising cumulative delays and found that allocating the majority of time supplements towards the first part of the railway line minimised delay propagation because it allowed trains to use it throughout the entire trip, thus increasing the chance of being on-time at all subsequent stops (early departures allowed).

In Italy, the punctuality level was solely measured at terminal stations (Medeossi et al., 2009). Consequently, large time supplements were placed on track segments leading up to that station. Additionally, the aggregated effect of infrastructure failures and failures caused by operating companies became indistinguishable. An underway delay measure was thus more useful when investigating performance on certain track sections (Nyström & Söderholm, 2005; Andersson et al., 2011). To make up for the weaknesses outlined, Medeossi et al. (2009) proposed a continuous punctuality measure “Delay Frequency Index” taking into account both running time deviation and whether or not trains were on time at the final station:

$$F = \sum_{i=1}^{n} (N_i / N * D_i / P * f) \qquad (12)$$

where *N* is the number of trains, N_i is the number of trains arriving in delay interval *i*, D_i is the magnitude of the delay in interval *i*, *P* is a threshold value indicating whether a train is on-time, *f* is a weight coefficient, and *F* is a percentage indicator.

2.2 Stability

Stability is defined as the inherent ability of a timetable to limit propagation of minor delays. The most common approach to enhance timetable stability is to add buffer time between subsequently running trains. Buffer time is often imposed similarly as adding time supplements (see section 2.1). Adding buffer time between trains is a trade-off between capacity utilisation and delay propagation. The mean knock-on delay of all trains passing a station was shown to increase exponentially with the decreasing amount of scheduled buffer time between train paths; on the other hand, allocating too much buffer time could remain unused, thus increasing passengers’ travel time (Yuan & Hansen, 2007).

Armstrong et al. (2012) tried to adapt the spread of trains running in a corridor with the aim to maximise reliability and potentially release capacity for additional railway services.

Forsgreen et al. (2012) addressed timetable stability by solely publishing a subset of the arrival and departure times, thus imposing flexibility by allowing traffic managers to re-distribute buffer time during operations. The same idea was addressed by Goverde & Hansen (2013) who elaborated on different levels of delay resistance for timetables. The highest level was a resilient timetable which was robust against delays and flexible enough to handle disturbances.

Another approach aiming at maximising the spread of trains is rerouting trains, assigning alternative platforms or changing schedules (Dewilde et al., 2013). In high frequency networks, the aim would be to increase the smallest headway as much as possible (Carey, 1999). Sels et al. (2014) extended the idea behind assigning alternative platforms. Assigning trains to alternative platforms could create inconsistency and cause longer walking distances for transferring passengers (Dollevoet, 2013). Dewilde et al. (2013) and Sels et al. (2014) did not ensure that frequency was maintained and consistent when assigning alternative platforms, although, it was known to impact customer satisfaction (Sun & Xu, 2012).

2.3 Heterogeneity

Heterogeneity refers to the dissimilarity in the way different trains are operated. Having low heterogeneity, namely a high degree of dissimilarity of trains' stopping patterns, headway, and speed, allows more trains to run on the tracks when overtaking is prohibited. Heterogeneous operations are typically seen on tracks where different types of trains are run. Kikuchi & Vuchic (1982) covered the impacts of imposing skip-stop services, namely changes in access time, in-vehicle time, fleet size requirements and operating costs. Access time increases on stops experiencing a less frequent service. System area coverage is reduced, thus affecting the number of potential customers. Riding time decreases since dwell time, acceleration and deceleration time are saved on skipped stops. Reduced in-vehicle time implies shorter round trip time, reducing requirements for rolling stock, hence also the operating costs. To compare heterogeneity of railway operations in a corridor, Vromans (2005) developed the following two analytical measures:

$$SSHR = \sum_{i=1}^{n} \frac{1}{h_i^-} \tag{12}$$

$$SAHR = \sum_{i=1}^{n} \frac{1}{h_i^A} \tag{12}$$

When measuring Shortest Headway Reciprocals (*SSHR*), h_i is the smallest scheduled headway between trains *i* and *i+1* on a track segment. This measure turned out to fall short because the departure headway was considered as equally important as the arrival headway. Vromans (2005) found arrival headways more important when trains were blocking each other. To cope with this, the Sum of Arrival Headway Reciprocals (*SAHR*) only accounted for arrival headways between subsequent trains, with h_i^A being the headway between train *i* and train *i+1* at the arrival station. *SAHR* is a single point measure not taking into account the entire track segment. Therefore, a more elaborate heterogeneity measure averaging *SSHR* and

SAHR was suggested (Vromans et al., 2006). A straightforward way to change heterogeneity is to change stopping patterns on railway lines (see Parbo et al., 2014 for a list of references on changing stopping patterns).

2.4 Network interdependencies

Network interdependencies refer to train movements where precedence constraints are enforced. Precedence constraints state the train order and whether trains should be held to maintain connections. Network interdependencies, although important for network connectivity, increase the risk of propagating delays. The more constraints imposed on trains, the more vulnerable to disruptions the system becomes. Although facilitating smooth transfers between several trains, synchronised networks create "snowballing" delays, when trains are held to maintain connections (Vansteenwegen & Oudheusden, 2007). Consequently, synchronised networks experience larger increases in passengers' travel time when connections are not maintained (Finger et al., 2014).

2.4.1 Changing network layout

The intuitive way to reduce network interdependencies is to avoid conflicting train movements. A "cheap" approach towards reducing the interdependencies for a single track network without extending it fully to double track lines is to position a set of sidings to limit conflicting train movements (Higgins et al., 1997). Also, running railway lines independently of each other, building bridges, crossovers, side tracks or extending the network layout from single track to double track will reduce the number of interdependencies (Landex, 2008; Gestrelius et al., 2012). Besides the limited capacity, delays propagated easier on single track lines because of interdependent train movements both from ahead and behind (Landex, 2008).

2.4.2 Planning

Burkolter et al. (2005) proposed a two level method determining train precedence constraints and routing, respectively. The higher level created a tentative dense timetable by applying Petri Nets modelling on an aggregated track topology. Afterwards, the lower level verified the tentative timetable on local exact topologies.

Salido et al. (2008) suggested using the following robustness indicator *R(x)* on double track lines explicitly considering interdependencies between trains. The following robustness value was assigned to the timetable:

$$R(x) = \sum_{T=1}^{NT} \sum_{S=1}^{NS} Buff_{TS} * \%Flow_{ST} * TT_S * NSucT_T * (NS - S) / NS \qquad (12)$$

where $Buff_{TS}$ is the buffer time a given train T has on a given station S, $\%Flow_{ST}$ is the percentage of passenger flow in train T and station S, TT_S is the percentage of tightness of track between stations S and $S+1$, $NSucT_T$ is the number of trains that may be disrupted by train T, NS and NT is the number of stations and trains, respectively. The robustness measure was developed to compare timetable quality among different timetables for the same track layout (Salido et al., 2008).

Flier et al. (2009) presented efficient algorithms to detect the interdependencies that occurred due to precedence constraints and due to maintained connections. Yamamura et al. (2012) developed an algorithm based on daily recorded traffic data to identify frequently occurring and widely influential delays. A backwards tracing algorithm was applied to find the primary delay causing the secondary delays. The approach identified notorious delays as well as delays that had not been recognised by the timetable planners.

De-los-Santos et al. (2012) addressed interdependencies by developing two robustness measures where the stability of the network was considered when random failures and intentional attacks, respectively, were imposed on a track segment in the network. The two measurements were as follows:

$$\delta R(N,E) = \frac{T(K_{|N|})}{\max_{\bar{e} \in E} DT((N,E),\bar{e})} \quad (12)$$

$$\mu R(N,E) = \frac{T(K_{|N|})}{\sum_{\bar{e} \in E} DT((N,E),\bar{e}) / |E|} \quad (12)$$

where δR is the ratio between the total travel time in the complete network $T(K_{|N|})$ and the total travel time $DT(*)$ when the track segment $\bar{e}$ of the network (N,E) that increases the overall travel time the most is blocked, and μR is the ratio between total travel time in the complete network and the average total travel time in case of blocked track segments. This way of considering stability goes well in hand with the statistical definition of robustness and in particular the derived reliability and resilience. In this context, reliability is the probability that the network is connected given a failure probability for every edge and resilience is the probability that the network disconnects after exactly *i* failures (Klau & Weiskircher, 2005).

Andersson et al. (2013) developed robustness indicators deducible from the timetable before operations had taken place, Robustness in Critical Points (RCP). Critical points were points in the network with interdependencies between trains. RCP reflected the flexibility determined by the time supplements as well as the buffer time. The more time supplements and buffer time, the larger the flexibility to reschedule the trains in the case of delays.

2.5 Recovery strategies

Having planned the timetable to be robust against delays does not guarantee 100% punctuality. The idea behind recovery strategies is to recover from the disruptions to the planned schedule fast and smoothly, deteriorating the service as little as possible. If time is not a limitation, planners could reset and restart the system. The challenge, though, is to recover while maintaining a proper level of service for the passengers.

Sun & Hickman (2005) formulated a nonlinear 0-1 integer programming skip-stop problem as a real-time decision support tool with the binary integer variables representing whether or not to skip a stop. The idea was to allow vehicles to catch up on their delay by skipping stops. Sun & Hickman (2005) outlined that skipping stops in real-time should be imposed with care, infrequently and never on subsequent trains from the same line.

Hofman et al. (2006) considered the following recovery strategies:

- Platform changes
- Allowing overtaking
- Skipping stops
- Early turning
- Reducing dwell time, headway and running time to a minimum
- Changing train status
- Inserting trains
- Cancellation of lines

Simulation was used to test the recovery strategies. Hofman et al. (2006) found that strategies yielding a large increase in headways resulted in the largest punctuality increase. When disruptions occurred and no recover strategies were imposed, the capacity utilisation had a huge impact on the regularity. Liebchen et al. (2007) created a two stage model, which in the first stage computed a robust timetable and in the second stage solved the delay management problem. The timetable quality was evaluated through simulation of railway operations. Caprara et al. (2010) presented a robust scheduling approach in combination with three different recovery strategies to mitigate knock-on delays. Firstly, they let delays propagate by keeping the nominally assigned train order. Secondly, they chose alternative platforms, i.e. exploited non-utilised resources. Thirdly, they tried all possible strategies e.g. allowing completely different strategies than nominally assigned. The results indicated that incorporating robustness considerations into the train routing problem, together with appropriately chosen online re-scheduling algorithms, led to better train punctuality. However, neither of the studies considered the effect of recovery strategies on the passengers.

2.6 Summary

Studies addressing robustness against delays of railway operations by improving train related service characteristics show that several approaches have been developed. Despite the vast amount of different approaches, the basic idea behind enhancing robustness can be boiled down to reducing the risk that delays affect subsequently running trains by using the following “tools”:

- Adding time supplements
- Adding buffer time
- Cancelling trains
- Reducing network interdependencies
- Lowering heterogeneity

None of the studies reviewed in section 2 takes the passengers explicitly into account. The “tools” are in all the reviewed studies applied with the aim to ensure the on-time performance of the trains without knowing or monitoring the exact impact on the passengers’ travel experience.

3 Passenger perspectives

While planned railway schedules are published and fixed, passengers’ planned itineraries are private and affected by their perception of several different attributes. Accordingly, measuring passengers’ on-time performance is much more complex than evaluating whether trains are on time. At the same time, neglecting passengers in the planning may result in suboptimal railway plans. Table 2 outlines, an overview of passenger-oriented railway planning studies. For each study, the passenger-related attributes emphasised are outlined in their being explicitly considered, ✓, implicitly considered (✓) or not considered at all, ✗. This section starts by reviewing the passenger-oriented railway optimisation studies focusing on robustness and then examines the way passengers’ perceive railway operations and how their perception affects their travel behaviour.

Table 4 - Overview of reviewed literature on passenger-oriented railway planning

Studies	Purpose	Measure developed	Passenger related variables addressed					Data source used to reveal passenger demand
			In-vehicle time	Waiting time	Transfers/ Connections	Exact route choice	Delays	
Optimisation & Planning								
Vansteenwegen & Oudheusden (2006)	Emphasising passenger heavy trains in the planning		✓	×	(✓)	×	×	Passenger counts
Weston et al. (2006)	Successful transfers through dispatching		✓	✓	✓	✓	✓	Simulating operations
Network (2008)	Redistribute trains to maximise passenger satisfaction		✓	✓	(✓)	×	×	Passenger counts
Oort et al. (2010)	Dispatching focusing on holding times to ensure successful transfers		✓	✓	✓	(✓)	×	Applied to a tram line
Goerigk et al. (2011)	Successful completion of transfers		×	(✓)	✓	×	✓	Approximated
Kanai et al. (2011)	Determining whether to keep or drop connections when delays occur		×	×	✓	✓	(✓)	Simulating operations
Corman et al. (2012)	Delay management		×	×	✓	(✓)	✓	Approximated
Dollevoet et al. (2012)	Dynamic delay management		✓	✓	✓	(✓)	✓	Approximated
Sels et al. (2012)	Minimise passengers' travel time		✓	✓	✓	(✓)	×	Deterministic Shortest Path
Dewilde et al. (2013)	Reduce delay propagation	Normalised robustness	✓	✓	✓	×	(✓)	Passengers' route choice calculated by shortest path algorithm

Transport models & passenger surveys								
Noland & Small (1995)	Addressing delay uncertainty		✓	(✓)	(✓)	✗	✓	Passenger survey
Nielsen (2000)	Accounting for seat availability in assignment models		✓	✓	✓	✓		Passenger counts
Bates et al. (2001)	Estimating value-of-time		✓	(✓)	(✓)	✗	✓	Passenger survey
Nuzzolo et al. (2001)	Examining passengers travel behaviour as a result of historical delays	day-to-day learning process	✓	✓	✓	✓	✓	Transit assignment model
Chen et al. (2002)	Estimating value-of-time		✓	(✓)	(✓)	✗	✓	Passenger survey
Noland & Polak (2002)	Passengers' perception of travel time vs. reliability	Reliability ratio	✓	✓	✗	✗	✓	Passenger survey
Studies	Purpose	Measure developed	Passenger related variables addressed					Data source used to reveal passenger demand
			In-vehicle time	Waiting time	Connections/ Transfers	Exact route choice	Delays	
Rietveld (2005)	Passengers' overestimation of train delays		(✓)	✗	✓	✗	✓	Passenger survey
Tsuchiya et al. (2006)	Passenger information systems used to guide passengers when delays occurred		✓	(✓)	(✓)	(✓)	✓	Passenger survey
Takeuchi et al. (2007)	Describing passengers' perceived level of service		✓	✓	✓	✓	✓	Passenger survey
Nielsen et al. (2008)	Distinguishing train and passenger delays		(✓)	(✓)	(✓)	✓	✓	Transit assignment model

Sumalee et al. (2009)	Examining passengers travel behaviour as a result of historical seat availability		✓	✓	(✓)	✓	✓	Transit assignment model
Börjesson & Eliasson (2011)	Assessing average delay as reliability measure		✓	✓	✗	(✓)	✓	Passenger survey
Hensher et al. (2011)	Addressing risk and uncertainty of delays		✗	✓	✓	✗	✓	Passenger survey
Sels et al. (2011)	Derive passenger flows from ticket sales data		✓	(✓)	(✓)	(✓)	✗	Ticket sales data
Sun & Xu (2012)	Estimating passengers' route choice from travel card data		(✓)	(✓)	(✓)	(✓)	✗	Travel card data
Börjesson et al. (2012)	Estimating value-of-time		✓	✓	(✓)	✗	✓	Passenger survey
Jiang et al. (2012)	Distinguishing train and passenger delays		(✓)	(✓)	(✓)	✓	✓	Transit assignment model
Nuzzolo et al. (2012)	Accounting for seat availability in schedule-based assignment models		✓	✓	✓	✓	✓	Passenger counts
Shi et al. (2012)	Assess how transfers impact route choice		✗	(✓)	✓	✓	(✓)	Transit assignment model
Wardman et al. (2012)	Meta study on value-of-time studies		✓	✓	✓	✗	✓	Passenger survey

3.1 Optimisation & Planning

3.1.1 Transfer maintenance

From the passengers' perspective, a crucial part of a well-functioning railway system is maintaining transfers. In Denmark, every third transit trip includes at least one transfer (DMC, 2015). Maintaining transfers is a compromise between passengers' travel time and delay propagation. Holding trains to maintain transfers increases the risk of propagating train delays and increases the travel time for on-board passengers, while cancelled or missed connections impose extra delays on transferring passengers. In this regard, distinguishing trains as either punctual or not based on an (to some extent) arbitrary threshold can distort the picture of the service provided. Passengers on trains delayed by less than the threshold value can also miss a transfer (Carrasco, 2012). Consequently, allocating time supplements on stations with negligible transfer loads is ineffective (Oort et al., 2010; Dewilde et al., 2013). An approach emphasising the passengers is to determine paths on which connections will be maintained even when delays occur. This can be done by allocating additional time supplements on stations or by modifying the timetable, so that subsequently running trains are not affected if minor delays occur (Vansteenwegen & Oudheusden, 2006).

In real-time, missed transfers can be reduced through dispatching. Weston et al. (2006) compared actual arrival time at the destination station to the planned arrival time. Afterwards, it was checked if the transfer was completed. A microscopic simulation of a congested part of the rail network in U.K. was performed. Weston et al. (2006) concluded that, due to missed connections, minimising train delays did not necessarily minimise passenger delays.

The problem of maintaining transfers is sometimes referred to as the timetable information problem, where light and strict robustness may be distinguished. A strict robust path is defined as a path where all transfers are maintained under every delay scenario. Similar to the timetabling problem, light robustness was found superior to strict robustness since it ensured a modest level of robustness while only deteriorating passengers' travel time marginally (Goerigk et al., 2011).

Kanai et al. (2011) developed a delay management plan minimising passengers' dissatisfaction (the trade-off between additional in-vehicle time and extra waiting time due to missed transfers) by combining simulation and optimisation. Passengers behaved as if trains were on time. The decision of keeping or dropping connections was solved by a tabu search algorithm and evaluated by a passenger simulation model. Kanai et al. (2011) found that travellers were rarely willing to wait until the next service arrived, but would rather seek alternative routes.

Corman et al. (2012) developed a bi-objective delay management strategy. The aim was to minimise the number of missed connections (weighted by the number of passengers) and avoid train conflicts when re-scheduling trains by using a heuristic algorithm (Corman et al., 2012). No recommendations on how to handle these issues were given; instead it was proposed that different stakeholders should agree on a "practical optimum" between maintaining and cancelling connections

3.1.2 Emphasising travel behaviour

In transit planning, the impacts of timetable changes on passengers' travel behaviour should be considered explicitly to accurately quantify the derived impacts (Sun et al., 2014). One way to obtain passengers' route choice is by ticket sales data. However, not all ticket sales reveal ending and transfer points, thus different approximation methods are needed. Sels et al. (2011) distributed passengers on stations according to existing demand data and calculated passenger flows by shortest path algorithms. Sun & Xu (2012) described how travel card data from the Beijing metro was used to analyse travel time variability and estimate passengers' route choice behaviour.

Sels et al. (2012) applied FA-PESP (Flow Allocation Periodic Event Scheduling Problem) introducing a feedback loop between (passenger) flow allocations and timetabling. The aim was to minimise passengers' travel time. Dollevoet et al. (2012) elaborated the flow allocation and proposed a delay management model allowing passengers to change their route choice when disruptions occurred.

Dewilde et al. (2013) addressed the trade-off between delay propagation and passenger travel time explicitly through *Normalised Robustness* (*NR*):

$$NR = \frac{TT + WaitCostEx}{NomTT} \tag{12}$$

where *TT* is passengers' realised travel time, *WaitCostEx* is passengers' perceived extra waiting cost, and *NomTT* is passengers' nominal travel time. *WaitCostEx* addresses the unused time supplements, which are assumed to be an annoyance factor when trains run according to the planned schedule.

3.2 Passengers' perspectives

When passengers plan a journey, their perception of quantifiable attributes (e.g., valuation of in-vehicle, waiting time, transfers), non-quantifiable attributes (e.g., aversion towards being late, comfort perception) and whether they are frequent or occasional users have a huge impact on their travel behaviour (Takeuchi et al., 2007; Carrasco, 2012; Vij et al., 2013). In transport models, a generalised travel cost function may be used to reflect the impedance travellers face when travelling by public transport, thus enabling the estimation of travellers' route choice. One example was presented by Nielsen (2000):

$$GenCost = Cost_a + (\beta_{(t)} + \xi_{(t)}) \cdot (\sum_t \frac{\beta_{(t)i}}{\beta_{(t)}} \cdot t_{ai}) + Seat_i + \varepsilon_a \tag{12}$$

where T_{ai} is the time spent on link *a* when having boarded line *i*, $Cost_a$ is the cost on link a, $Seat_i$ is the disutility associated with not being able to get a seat on line *i* (Nielsen, 2000), ξ is an error component capturing taste variation for different attributes among passengers, and ε_a captures the differences in passengers' perception of different routes. The β's were related to access/egress time, waiting + transfer time, in-vehicle time, headway (hidden waiting time), and delay (simulated based on discrete distributions), respectively.

According to the generalised travel cost function outlined above, both the hidden waiting time and the delays impact passengers' perception of railway performance. Hypothetically, an operator could be interested in optimising railway operations for the passengers by minimising knock-on delays, for example by adding more buffer time. Despite increased punctuality, this might not be unambiguously good, since hidden waiting time has now increased. Similar interdependencies between other attributes mean that exhaustive performance measurements are required to reveal the actual impacts of certain initiatives (Yuan & Hansen, 2007).

One approach to investigate the difference between passenger delays and train delays is to use a traffic assignment model. Assignment models consider quantifiable attributes that impact passengers' route choice, hence based on a certain plan, the outcome of the assignment model reveals how passengers are assumed to travel. Nielsen et al. (2008) analysed this and found that passengers' on-time performance was significantly lower than that of the trains.

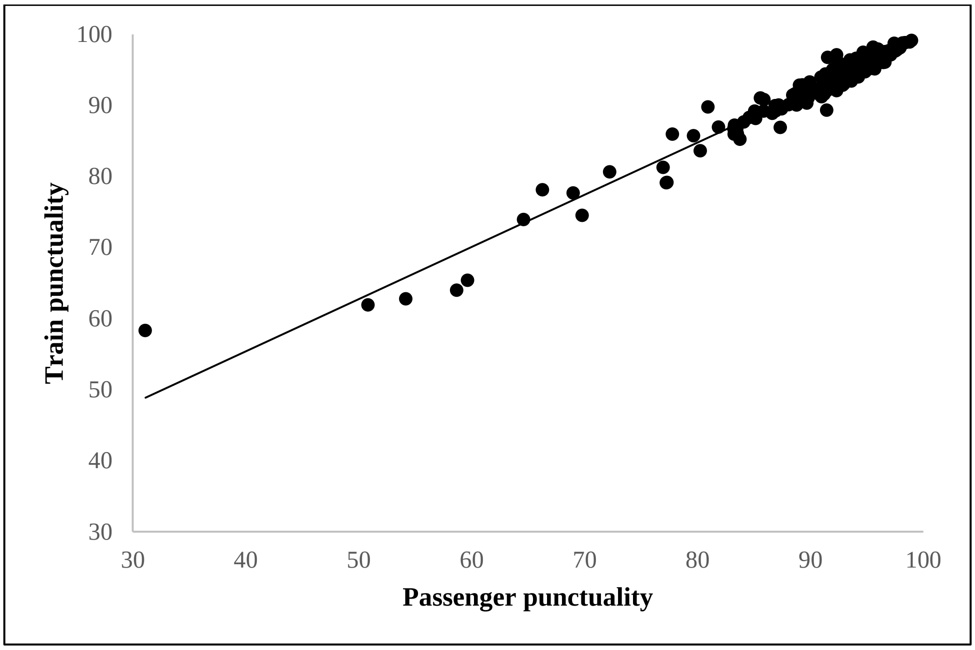


Figure 16 - Passenger punctuality vs. train punctuality on the suburban railway network in the Greater Copenhagen area (Weekly data from 2010 - 2014)

As outlined in figure 3, also more recent data show that on average train punctuality is higher than passenger punctuality. Passenger punctuality is defined as the number of passengers reaching their destination within a certain time threshold. Passenger punctuality evaluates the on-time performance of passengers' entire journey from origin to destination including transfers, while train punctuality only evaluates the on-time performance of individual train trips. The fact that passenger punctuality was generally worse than the train punctuality was primarily explained by passengers missing transfer connections and the fact that peak hour trains (having larger passenger loads) more often were delayed. Although passengers were significantly more delayed than the trains responsible for their delays, minor train delays did not necessarily cause passenger delays. In fact, some passengers were able to take a connecting train earlier than planned (Jiang et al., 2012).

3.2.1 Travel time variability

Passengers' preferences are generally obtained by either observing behaviour or asking hypothetical questions. The first approach is referred to as revealed preferences (RP) and the latter as stated preferences (SP), and the two approaches can also be combined. Regarding travel time variability, RP surveys are hardly applicable since observations are needed for long periods to deduce travellers' reactions to disruptions. Consequently, studies considering travel time variability traditionally use SP approaches.

Bates et al. (2001) used SP to examine travellers' departure time accounting for the relevant variables. According to the SP-study, travellers departed earlier when trips were important and travellers were risk-averse. Below, an example of a utility function from Bates et al. (2001) is shown:

$$E[U(t_h)] = \alpha E[T(t_h)] + \beta E[SDE(t_h)] + \gamma E[SDL(t_h)] + \theta p_L(t_h) \tag{12}$$

where the expected utility $U(*)$ is dependent only on the departure time t_h, $E[*]$ is the expected value, $T(*)$ is the travel time, $SDE(*)$ is the early deviation from the planned arrival time, $SDL(*)$ is the late deviation from the planned arrival time, and p_L is the probability of arriving later than the planned arrival time. α, β, γ and θ are parameters to be estimated under the assumption that travellers are utility maximisers.

Noland & Polak (2002) measured the importance of reliability relative to travel time by defining a reliability ratio *RR*:

$$RR = \frac{\beta}{\alpha}\ln(1+\frac{\gamma}{\beta}) \qquad (12)$$

where α is a parameter associated with in-vehicle time, β relates to *SDE* and γ to *SDL*. Due to the discrete nature of public transport services, the disutility associated with low reliability is large.

Examining how the extent of train delays was perceived by the travellers, a few large delays proved to be more hurtful than several minor delays (same total extent) (Vromans et al., 2006). This was explained by the ability for several trains to utilise their buffer time to catch up on smaller delays, while one large delay was more likely to affect subsequent trains. Using average delay as performance indicator was thus misleading. A much larger disutility was associated with a 2% risk of being 50 minutes late than a 10% risk of being 10 minutes late, thus forcing risk-averse travellers to take an earlier train (Börjesson & Eliasson, 2011). Although disturbances are highly disregarded, passengers generally consider minor delays acceptable. In a survey among transit users from Britain, 87% of the passengers were satisfied with a five minute delay at their departure station. This number fell to 77% when delays were between six and nine minutes (Transport focus, 2014).

The ratio between in-vehicle time and reliability for public transport was found by Bates et al. (2001) to be above the ones for car (app. 1.3) and below 2. In a SP-survey from the Netherlands this value was estimated to 1.4, i.e. passengers considered a 1 minute reduction in travel time variability 1.4 times higher than a 1 minute of travel time reduction (Oort, 2011). From a meta-analysis of European studies (conducted between 1963 and 2011 by Wardman et al., 2012), it emerged that the four variables used to reflect passengers' perception of travel time variability were:

- Schedule Delay Early (SDE)
- Schedule Delay Late (SDL)
- Late arrival
- Standard deviation of travel time (StdDev)

Relative to in-vehicle time, the values for *SDE*, *SDL*, *Late arrival* and *StdDev* were on average 0.8, 1.68, 3.29 and 0.66, respectively. The uncertainty related to late arrival and standard deviation of travel time was thus considered significantly less attractive than using one minute in a vehicle. The travel distance did not affect the variables remarkably and time multipliers were quite similar between the different studies (Wardman et al., 2012).

3.2.2 Delay uncertainty and delay risk

When travelling, passengers have a perception (from past experiences) of the travel time based on their route, mode and departure time choice, respectively. Travel time deviations can be either expected (risk) or unexpected (uncertainty) by the travellers (Hensher et al., 2011). Noland & Small (1995) added an extra disutility term to the travel time uncertainty; the discrete lateness penalty emphasising passengers aversion towards being late. This was refined by Bates et al. (2001) and Chen et al. (2002) who considered travellers' perceived uncertainty distinguishing between risk-averse and risk seeking travellers. Conclusions were identical: risk-seeking travellers were less stressed by travel time uncertainty than risk-averse travellers, and they did not consider reliability, but rather chose the path with minimum travel time, while risk-averse passengers chose the most reliable route (Finger et al., 2014).

Tsuchiya et al. (2006) examined passengers' perception of a support system informing about optimal routes in case of disruptions. The information was based on predicted resumption time from the disturbance estimated from historical data. The information helped passengers decide whether to wait for resumption or not and, if not, which detour to choose: 94% preferred to have this piece of information as soon as possible, although subject to uncertainty, rather than waiting until the information was certain. Additionally, passengers appreciated being informed about the cause of the delay. When delays were caused by external factors travellers' negative emotions were alleviated compared to the situation where the operator was responsible for the delay (Transport focus, 2014; Cheng & Tsai, 2014).

Börjesson et al. (2012) studied passengers' response to travel time variations by testing the equivalency between scheduling models and reduced-form models. Passengers' valuation of expected delay was significantly larger for reduced-form models. In scheduling models, information about being late was associated with less uncertainty than in reduced-form model. An inherent disutility (interpreted as an anxiety cost or a cost associated with the need for contingency plans) was associated with the uncertainty (Börjesson et al., 2012). Passengers disliked the stress they felt, when onward planned connections looked doubtful. However, if the onward train was part of a frequent service, missed connections were less damaging (Weston et al., 2006; Cheng & Tsai, 2014). Furthermore, passengers reported that they were more tolerant towards delays when being in the vehicle, because they could typically see the cause of the problem causing the delay (Transport focus, 2014). This finding supports that it is in particular the uncertainty aspect that passengers dislike.

3.2.3 Recovery strategies

When passengers face disruptions, they seek to find alternative routes. If disruptions start occurring frequently, travellers start adapting their departure time or mode choice. To examine passengers' recovery strategies, Nuzzolo et al. (2001) developed a doubly dynamic schedule-based assignment model. Doubly dynamic refers to the explicit consideration of within-day and day-to-day variations in travel behaviour. Nuzzolo et al. (2001) formulated the day-to-day learning process mathematically:

$$X_t^{fo} = \gamma \cdot X_{t-1}^{ex} + (1-\gamma) \cdot X_{t-1}^{fo} \tag{12}$$

where the forecasted value X_t^{fo} for an attribute on day *t* is expressed by a convex combination of the attribute forecast X_{t-1}^{fo} and the value realised X_{t-1}^{ex} from the day before (*t-1*). γ is a weight between 0 and 1. Nuzzolo et al. (2001) found that the number of transfers during a journey did not have a significant impact on short term adaptations.

Sumalee et al. (2009) developed a dynamic stochastic assignment model taking seat availability into account. Information about seat availability prior to their trip made travellers adapt their departure time and/or route choice to maximise their probability of getting a seat.

Shi et al. (2012) developed an equilibrium-based rail passenger flow model explicitly taking the probability of successful transfers into account. Passengers tended to choose a path, where the probability of a successful transfer was higher. Additionally, Shi et al. (2012) found that waiting time mostly affected within-day variations in travel behaviour. Attributes affecting day-to-day variations in travel behaviour were in-vehicle time, transfer time and comfort level.

Nuzzolo et al. (2012) developed a schedule-based dynamic assignment model for transit networks that jointly simulated departure time, boarding stop and run choices on the basis of mixed pre-trip/en-route choice behaviour. Pre-trip choices concerned departure time and boarding stops, since they were mainly influenced by past experiences of congestion, while en-route choices occurred at boarding stops and concerned the decision whether to board a specific vehicle. Vehicle capacity impacted passengers' departure time choice significantly. Passengers tried to board an earlier vehicle to reduce failure-to-board probability.

Van der Hurk (2015) developed a model that, in the case of large disruptions, personalised passenger information on alternative routes. The model took into account the probability of boarding and the uncertainty in the duration of the disruption. For all tested disruption scenarios, the conclusion was that when passengers are provided personalised information on alternative routes taking into account the probability of boarding, the average expected delay a passenger may experience is reduced significantly compared to the case where information is not personalised or the case where information is personalised but does not account for boarding probability.

3.3 Summary

To reduce the impact of train disruptions on passengers, maintaining transfers is often the main concern among optimisation studies. Passenger loads from the existing system are used as weights in order to prioritise specific transfer connections. Reducing the risk of missing transfer connections is typically done by adding additional time supplements to the involved stations. Doing so is a trade-off between delays passing on to subsequently running trains and imposing additional passenger travel time and should thus be imposed with care.

From the body of literature, it is clear that several quantifiable attributes impact passengers' travel behaviour. Unfortunately, all these attributes are rarely considered at the same time when planning railway operations. Comparisons between train and passenger on-time performance, respectively, reveal that due to e.g. missed transfers and demand variability, train punctuality is often significantly higher. These studies highlight the importance of having delay resistant timetables, thus minimising uncertainty and providing a punctual service to the passengers. Travel time uncertainty is highly disliked by travellers, not only when disruptions occur, but also the need for having contingency plans. Consequently, travellers value reduced travel time uncertainty higher than reduced travel time.

4 Conclusions and future research

The present literature review provides an analysis of the work that, on the one hand, has aimed at enhancing operational characteristics related to the robustness of railway operations and, on the other hand, has focused on passengers' perception of railway performance. Table 1 reveals that the service characteristics considered most often regarding robustness are time supplements and buffer times. Although all studies agreed on lowering capacity utilisation to improve robustness (delay resistance), their recommendations vary a lot based on the network layout, how performance is measured and whether or not trains are allowed to depart earlier than scheduled.

When focusing explicitly on the passengers, table 2 shows that maintaining transfers has been the objective pursued most frequently. From the review of passengers' perspectives in railway operations it is evident that their perception of the service level affects their travel behaviour. Passengers' travel behaviour (i.e. mode choice, route choice, departure time choice) depends on several attributes, e.g. in-vehicle time, transfer time, waiting time, access/egress time, crowding level and delays. Using on-time performance of trains as performance measurement thus turns out to be inadequate. Taking passengers' travel patterns into account shows that passengers' on-time performance in some cases is 10 percentage points lower than for the trains. To close the gap between how railway planning is performed and measured and, on the other hand, passengers' perception of railway performance and their actual experiences, the following directions for future research are identified.

Understanding passengers' preferences and being able to address these preferences explicitly in the planning is the basis for a more passenger-oriented railway planning. Accurate and disaggregate passenger travel data facilitates a more passenger-oriented planning, especially when transfer patterns are revealed. Passenger-oriented KPIs (taking all relevant attributes into account) should be applied by researchers and operators. Exhaustive performance measurements ensure that enhancing a single parameter is not realised at the expense of non-measured parameters, thus leading to a de facto deterioration or status quo in service level.

Regarding travel time variability, studies show that passengers rate schedule adherence higher than travel time reductions. Therefore, a definition of robustness and related measurements need to capture the system performance as well as the efficiency, i.e. travel time and capacity utilisation. When the *price of robustness* is disregarded, enhanced robustness may be achieved by allocating disproportionally large time supplements and buffer times.

Since robustness is a trade-off between several attributes, improving railway operations should take a more holistic and generic view, thus addressing all relevant attributes rather than only a subset. The same goes for the performance measurements. Coupling generic optimisation approaches and planning of railway operations with knowledge of passengers' travel behaviour (e.g. through transport models) will enhance the reliability and applicability of the results, thereby decreasing the gap between what railway planners provide and how it is perceived and experienced by the passengers.

Finally, providing real-time information to the passengers about their itinerary when delays occur is shown to have a stress-relieving effect. There is a huge potential in individualising information on the current state of the transit system, e.g. travel time uncertainty, delayed and cancelled services. Providing passengers with information about route alternatives in case of large disruptions taking into account seat availability has a significant impact on the average expected delay and thus also on their travel experience.

References

Andersson, E.V., Peterson, A., & Krasemann, J.T. (2013). Quantifying railway timetable robustness in critical points. *Journal of Rail Transport Planning & Management*, *3*(3), 95-110.

Andersson, E., Peterson, A., & Krasemann, J.T. (2011). Robustness in Swedish Railway Traffic Timetables. In *Railrome 2011: Book of Abstracts 4th International Seminar on Railway Operations Modelling and Analysis*.

Armstrong, J., Preston, J., Potts, P., Paraskevopoulos, D. & Bektas, T. (2012). Scheduling trains to maximize railway junction and station capacity. *Proceedings of 12th Conference on Advanced Systems for Public Transport (CASPT)*, Santiago, Chile.

Bates, J., Polak, J., Jones, P., & Cook, A. (2001). The valuation of reliability for personal travel. *Transportation Research Part E: Logistics and Transportation Review*, *37*(2), 191-229.

Batley, R., Dargay, J., & Wardman, M. (2011). The impact of lateness and reliability on passenger rail demand. *Transportation Research Part E: Logistics and Transportation Review*, *47*, 61-72.

Benezech, V., & Coulombel, N. (2013). The value of service reliability. *Transportation Research Part B: Methodological*, *58*, 1-15.

Börjesson, M., & Eliasson, J. (2011). On the use of “average delay” as a measure of train reliability. *Transportation Research Part A: Policy and Practice*, *45*(3), 171-184.

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

Burkolter, D., Herrmann, T., & Caimi, G. (2005). Generating dense railway schedules. *Advanced OR and AI Methods in Transportation*, 290-297.

Cacchiani, V., Caprara, A., & Fischetti, M. (2009). Robustness in train timetabling. *8th Cologne-Twente Workshop on Graphs and Combinatorial Optimization*, pp. 171-174.

Caprara, A., Galli, L., Kroon, L. G., Maróti, G., & Toth, P. (2010). Robust train routing and online re-scheduling. In *10th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems*, pp. 24-33.

Carrasco, N. (2012). Quantifying reliability of transit service in Zurich, Switzerland. *Transportation Research Record*, *2274*, 114-125.

Carey, M. (1999). Ex ante heuristic measures of schedule reliability. *Transportation Research Part B: Methodological*, *33*(7), 473-494.

Chen, A., Ji, Z., & Recker, W. (2002). Travel time reliability with risk-sensitive travelers. *Transportation Research Record*, *1783*, 27-33.

Cheng, Y.H., & Tsai, Y.C. (2014). Train delay and perceived-wait time: passengers' perspective. *Transport Reviews*, *34*(6), 710-729.

Corman, F., D'Ariano, A., Pacciarelli, D., & Pranzo, M. (2012). Bi-objective conflict detection and resolution in railway traffic management. *Transportation Research Part C: Emerging Technologies*, *20*, 79-94.

DMC (2015). The Danish National Travel Survey. Available from: <http://www.modelcenter.transport.www7.sitecore.dtu.dk/english/TU>. [Accessed 30 June 2015].

Dewilde, T., Sels, P., Cattrysse, D., & Vansteenwegen, P. (2013). Robust railway station planning: An interaction between routing, timetabling and platforming. *Journal of Rail Transport Planning & Management*, *3*(3), 68-77.

Dewilde, T., Sels, P., Cattrysse, D., & Vansteenwegen, P. (2011). Defining robustness of a railway timetable. In *Railrome 2011: Book of Abstracts 4th International Seminar on Railway Operations Modelling and Analysis*.

Dollevoet, T.A.B. (2013). Delay management and dispatching in railways. *Doctoral dissertation*, Rotterdam School of Management, Erasmus University.

Dollevoet, T., Huisman, D., Schmidt, M., & Schöbel, A. (2012). Delay management with rerouting of passengers. *Transportation Science*, *46*, 74-89.

Finger, M., Haller, A., Martins, S.S., & Trinkner, U. (2014). Integrated Timetables for Railway Passenger Transport Services. *Competition and Regulation in Network Industries*, *15*(1), 78-107.

Fischetti, M., & Monaci, M. (2009). Light robustness. *Robust and online large-scale optimization*, 61-84. Springer Berlin Heidelberg.

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

Forsgren, M., Aronsson, M., Gestrelius, S., & Dahlberg, H. (2012). Using timetabling optimization prototype tools in new ways to support decision making. *Computers in Railways XIII: Computer System Design and Operation in the Railway and Other Transit Systems*, *127*, 439.

Gestrelius, S., Aronsson, M., Forsgren, M. & Dahlberg, H. (2012). On the delivery robustness of train timetables with respect to production replanning possibilities. *2nd International Conference on Road and Rail Infrastructure in* Dubrovnik, Croatia.

Goerigk, M., Knoth, M., Müller-Hannemann, M., Schmidt, M., & Schöbel, A. (2011). The price of robustness in timetable information. *OASIcs-OpenAccess Series in Informatics*, *20*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

Goerigk, M., & Schöbel, A. (2010). An empirical analysis of robustness concepts for timetabling. *OASIcs-OpenAccess Series in Informatics*, 14. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

Goverde, R.M. (2010). A delay propagation algorithm for large-scale railway traffic networks. *Transportation Research Part C: Emerging Technologies*, *18*(3), 269-287.

Goverde, R.M., & Hansen, I. (2013). Performance indicators for railway timetables. *Intelligent Rail Transportation (ICIRT), 2013 IEEE International Conference on*, 301-306.

Hensher, D.A., Greene, W.H., & Li, Z. (2011). Embedding risk attitude and decision weights in non-linear logit to accommodate time variability in the value of expected travel time savings. *Transportation Research Part B: Methodological*, *45*(7), 954-972.

Higgins, A., Kozan, E., & Ferreira, L. (1997). Modelling the number and location of sidings on a single line railway. *Computers & Operations Research*, *24*(3), 209-220.

Hofman, M., Madsen, L., Jespersen Groth, J., Clausen, J., & Larsen, J. (2006). Robustness and recovery in train scheduling-a case study from DSB S-tog a/s. *OASIcs-OpenAccess Series in Informatics,* 5. Schloss Dagstuhl-Leibniz-Zentrum für Informatik.

Hooghiemstra, J.S., & Teunisse, M.J. (1998). The use of simulation in the planning of the Dutch railway services. *IEEE Simulation Conference Proceedings*, *2*, 1139-1145.

Huisman, T., & Boucherie, R.J. (2001). Running times on railway sections with heterogeneous train traffic. *Transportation Research Part B: Methodological*, *35*(3), 271-292.

van der Hurk, E. (2015). Passengers, information, and disruptions . *PhD dissertation,* Rotterdam School of Management, Erasmus University.

Jiang, Z.B., Li, F., Xu, R.H., & Gao, P. (2012). A simulation model for estimating train and passenger delays in large-scale rail transit networks. *Journal of Central South University*, *19,* 3603-3613.

Kanai, S., Shiina, K., Harada, S., & Tomii, N. (2011). An optimal delay management algorithm from passengers' viewpoints considering the whole railway network. *Journal of Rail Transport Planning & Management*, *1*, 25-37.

Kikuchi, S., & Vuchic, V.R. . (1982). Transit vehicle stopping regimes and spacings. *Transportation Science*, *16*(3), 311-331.

Klau, G. W., & Weiskircher, R. (2005). Robustness and resilience. *Network Analysis,* 417-437, Springer Berlin Heidelberg.

Kroon, L., Dekker, R., & Vromans, M. (2007). Cyclic railway timetabling: A stochastic optimization Approach. *Algorithmic Methods for Railway Optimization*, 41-66.

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

Landex, A. (2008). Methods to estimate railway capacity and passenger delays. *PhD dissertation,* Technical University of Denmark.

Liebchen, C., Schachtebeck, M., Schöbel, A., Stiller, S., & Prigge, A. (2010). Computing delay resistant railway timetables. *Computers & Operations Research*, *37*(5), 857-868.

Medeossi, G., Marchionna, A. & Longo, G. (2009). Capacity and reliability on railway networks: A simulative approach. *PhD dissertation*, University of Trieste.

Mesa, J.A., Ortega, F.A., & Pozo, M. A. (2009). Effective allocation of fleet frequencies by reducing intermediate stops and short turning in transit systems. *Robust and Online Large-Scale Optimization*, 293-309, Springer Berlin Heidelberg.

Milan, J. (1996). The Trans European railway network: Three levels of services for the passengers. *Transport Policy*, *3*(3), 99-104.

Milinković, S., Marković, M., Vesković, S., Ivić, M., & Pavlović, N. (2013). A fuzzy Petri net model to estimate train delays. *Simulation Modelling Practice and Theory*, *33*, 144-157.

Network, B.P.T. (2008). More passengers and reduced costs—the optimization of the Berlin public transport network. *Journal of Public Transportation*, *11*(3), 4.

Nielsen, O.A. (2000). A stochastic transit assignment model considering differences in passengers utility functions. *Transportation Research Part B: Methodological*, *34*(5), 377-402.

Nielsen, O.A., Landex, A., & Frederiksen, R.D. (2008). Passenger delay models for rail networks. *Schedule-Based Modeling of Transportation Networks: Theory and applications*, *46*, 27-49, Springer.

Noland, R.B., & Polak, J.W. (2002). Travel time variability: a review of theoretical and empirical issues. *Transport Reviews*, *22*, 39-54.

Noland, R.B., & Small, K.A. (1995). Travel-time uncertainty, departure time choice, and the cost of morning commutes. *Transportation Research Record*, *1493*, 150-158.

Nuzzolo, A., Crisalli, U., & Rosati, L. (2012). A schedule-based assignment model with explicit capacity constraints for congested transit networks. *Transportation Research Part C: Emerging Technologies*, *20*, 16-33.

Nuzzolo, A., Russo, F., & Crisalli, U. (2001). A doubly dynamic schedule-based assignment model for transit networks. *Transportation Science*, *35*(3), 268-285.

Nyström, B. & Söderholm, P. (2005) Improving railway punctuality by maintenance. *Research project of Luleå University of Technology.*

Oort, N.v. (2012). Quantifying benefits of enhanced service reliability in public transport. *Proceedings of 12th Conference on Advanced Systems for Public Transport (CASPT)*, Santiago, Chile.

Oort, N.v., Wilson, N.H., & Van Nes, R. (2010). Reliability improvement in short headway transit services. *Transportation Research Record*, *2143*, 67-76.

Oort, N.v. (2011). Service reliability improvement by enhanced network and timetable planning. *Proceedings of 12th Conference on Advanced Systems for Public Transport (CASPT),* Santiago, Chile.

Parbo, J., Nielsen, O.A., & Prato, C.G. (2014). Adapting stopping patterns in complex railway networks to reduce passengers' travel time. *Submitted to Transportation Research Part C: Emerging Technologies.*

Rietveld, P. (2005). Six reasons why supply-oriented indicators systematically overestimate service quality in public transport. *Transport Reviews*, *25*(3), 319-328.

Salido, M.A., Barber, F., & Ingolotti, L. (2012). Robustness for a single railway line: Analytical and simulation methods. *Expert Systems with Applications*, *39*(18), 13305-13327.

Salido, M.A., Barber, F., & Ingolotti, L. (2008). Robustness in railway transportation scheduling. *Intelligent Control and Automation, 2008. WCICA 2008. 7th World Congress on* (pp. 2880-2885). IEEE.

De-Los-Santos, A., Laporte, G., Mesa, J.A., & Perea, F. (2012). Evaluating passenger robustness in a rail transit network. *Transportation Research Part C: Emerging Technologies*, *20*, 34-46.

Schöbel, A. & Kratz, A. (2009). A bi-criteria approach for robust timetabling. *Robust and Online Large-Scale Optimization. Lecture Notes in Computer Science*, *5868*, 119-144.

Sels, P.H.A., Dewilde, T., Vansteenwegen, P., & Cattrysse, D. (2012). Automated, passenger time optimal, robust timetabling, using integer programming. *Proceedings of the 1st International Workshop on High-Speed and Intercity Railways*, 87-92. Springer Berlin Heidelberg.

Sels, P., Dewilde, T., Cattrysse, D., & Vansteenwegen, P. (2011). Deriving all passenger flows in a railway network from ticket sales data. In *Railrome 2011: Book of Abstracts 4th International Seminar on Railway Operations Modelling and Analysis*.

Sels, P., Vansteenwegen, P., Dewilde, T., Cattrysse, D., Waquet, B., & Joubert, A. (2014). The train platforming problem: The infrastructure management company perspective. *Transportation Research Part B: Methodological*, *61*, 55-72.

Shi, F., Zhou, Z., Yao, J., & Huang, H. (2012). Incorporating transfer reliability into equilibrium analysis of railway passenger flow. *European Journal of Operational Research*, *220*(2), 378-385.

Sumalee, A., Tan, Z., & Lam, W.H. (2009). Dynamic stochastic transit assignment with explicit seat allocation model. *Transportation Research Part B: Methodological*, *43*(8), 895-912.

Sun, L., Jin, J.G., Lee, D.H., Axhausen, K.W., & Erath, A. (2014). Demand-driven timetable design for metro services. *Transportation Research Part C: Emerging Technologies*, *46*, 284-299.

Sun, Y., & Xu, R. (2012). Rail transit travel time reliability and estimation of passenger route choice behavior: Analysis using automatic fare collection data. *Transportation Research Record*, *2275*, 58-67.

Takeuchi, Y., Tomii, N., & Hirai, C. (2007). Evaluation method of robustness for train schedules. *Railway Technical Research Institute, Quarterly Reports*, *48*(4).

Transport Focus (2014). How late is late – What bus passengers think about punctuality and timetables. *Available from:* http://www.transportfocus.org.uk/research/publications/how-late-is-late-what-bus-passengers-think-about-punctuality-and-timetables-full-report. *[30 June 2015]*.

Tsuchiya, R., Sugiyama, Y., Yamauchi, K., Fujinami, K., Arisawa, R., & Nakagawa, T. (2006). Route-choice support system for passengers in the face of unexpected disturbance of train operations. *Computers in Railways X.* In *Proceedings of the tenth international conference.* WIT, Boston.

Vansteenwegen, P., & Van Oudheusden, D.V. (2007). Decreasing the passenger waiting time for an intercity rail network. *Transportation Research Part B: Methodological*, *41*(4), 478-492.

Vansteenwegen, P., & Oudheusden, D.V. (2006). Developing railway timetables which guarantee a better service. *European Journal of Operational Research*, *173*(1), 337-350.

Vij, A., Carrel, A., & Walker, J.L. (2013). Incorporating the influence of latent modal preferences on travel mode choice behavior. *Transportation Research Part A: Policy and Practice*, *54*, 164-178.

Vromans, M. (2005). Reliability of railway systems. *PhD dissertation*, Rotterdam School of Management, Erasmus University.

Vromans, M.J., Dekker, R., & Kroon, L.G. (2006). Reliability and heterogeneity of railway services. *European Journal of Operational Research*, *172*(2), 647-665.

Wardman, M., Chintakayala, P., de Jong, G., & Ferrer, D. (2012). European wide meta-analysis of values of travel time. *Report prepared for the European Investment Bank*.

Weston, P., Goodman, C., Takagi, R., Bouch, C., Armstrong, J., & Preston, J. (2006). Minimising train delays in a mixed traffic railway network with consideration of passenger delay. *7th World Congress on Railway Research*, Montréal, Canada.

Yaghini, M., Khoshraftar, M.M., & Seyedabadi, M. (2013). Railway passenger train delay prediction via neural network model. *Journal of Advanced Transportation*, *47*(3), 355-368.

Yamamura, A., Koresawa, M., Adachi, S., & Tomii, N. (2012). Identification of causes of delays in urban railways. *Computers in Railways XIII*. In *Proceedings of the thirtenth international conference*. WIT, Boston.

Appendix 3: Parbo et al. (2015b)

Adapting stopping patterns in complex railway networks to reduce passengers' travel time

Jens Parbo, Otto A. Nielsen, Carlo G. Prato

Technical University of Denmark, Department of Transport, Bygningstorvet 116B, 2800 Kgs. Lyngby, Denmark

Resubmitted after third round of review to *Transportation Research Part C: Emerging Technologies*

Abstract
Travel time reductions in railways are typically costly and achieved through investments in rolling stock or infrastructure. Skipping stops, on the other hand, is a cost-effective way to reduce in-vehicle travel time for passengers on-board the train. The present paper deals with skip-stop optimisation for railway lines with the aim to reduce passengers' travel time and at the same time the heterogeneity of the railway operations in all corridors of a metropolitan railway network.

The problem is solved as a bi-level optimisation problem, where the lower level is a schedule-based transit assignment model and the upper level is a skip-stop optimisation model. The transit assignment yields passenger flows, which serve as input to the skip-stop optimisation. The updated stopping patterns and the correspondingly reduced in-vehicle times then serve as input in the subsequent route choice calculation. This bi-level minimisation problem is extremely difficult to solve mathematically, since skip-stop optimisation is a mixed-integer problem, whereas the route choice model is a non-linear non-continuous mapping of the timetable. Consequently, a heuristic approach is developed in this study and applied to a real-world case-study.

The approach is applied to the suburban railway network in the Greater Copenhagen area (Denmark). The optimisation yielded a reduction in railway passengers' in-vehicle time of 5.48 % and a reduction in passengers' generalised travel cost by 1.81 %. The reduction in in-vehicle time, number of transfers and waiting time is equivalent to 75 million DKK (about 10 million EUR) compared to existing stopping patterns.

Keywords
Railway timetabling, public transport optimisation, passenger behaviour, skip-stop services, large-scale application.

1 Introduction

Transit timetables are basically a contract between the operator and the passengers. It is essential to have a passenger-oriented railway system to maximise the potential patronage. This study deals with skip-stop optimisation in railway networks from the passengers' perspective at the tactical planning level. The aim is to minimise passengers' travel time and the heterogeneity of the railway operations, thus reducing the risk that delays propagate to subsequent trains.

When timetables are updated, the operator has to compromise with desires from various stakeholders, e.g. the passengers, who want a robust service with high speed and high frequency, and the operator, who is usually more concerned about maximising profit by e.g. reducing fleet size (Ceder, 2007). In this study we develop a tool for the planners, so that they can extend the yearly timetable update to include the stopping pattern configuration. The idea is that changes in demand over the years can outdate old stopping patterns in terms of minimising passengers' travel time. In this regard, this tool can be very valuable, since it has the ability to improve the stopping patterns while keeping the timetable structure unchanged, i.e. the number of departures for each railway line is kept.

When railway operations are changed, passengers try to adapt their travel behaviour accordingly. This interaction is rarely captured in optimisation problems. Consequently, the estimated impacts of the optimisation experienced by the passengers may be inaccurate when compared to reality. The present study explicitly considers the interaction between changes in railway operations and passengers' response (i.e. their changed travel behaviour). A bi-level heuristic is developed, where skipping stops and deriving passengers' adapted travel behaviour are done sequentially. A heuristic algorithm is applied to the skip-stop optimisation, while a public assignment model reveals passengers' adapted route choice.

Existing literature is reviewed in section 2. In sections 3 and 4, the skip-stop optimisation model and the public assignment model, respectively, are explained. In section 5, the methodology is applied to a case network. Section 6 provides the results and a sensitivity analysis, while section 7 concludes the paper and outlines directions for future research.

2 Literature Review

Kikuchi & Vuchic (1982) pioneered in explaining the impacts on passengers when introducing skip-stop services, namely changes in access time, waiting time and in-vehicle time. The operator, on the other hand, is mostly concerned about fleet size requirements and operating costs. Access time and waiting time increase on stops experiencing a less frequent service and at the same time system area coverage decrease. In-vehicle time is reduced since dwell time, acceleration and deceleration time are saved for skipped stops.

Changing stopping patterns also affects the heterogeneity of the railway operations. Homogeneous (solid lines) and heterogeneous (dashed lines) operations, respectively, are outlined for one direction of a double lined corridor in figure 1.

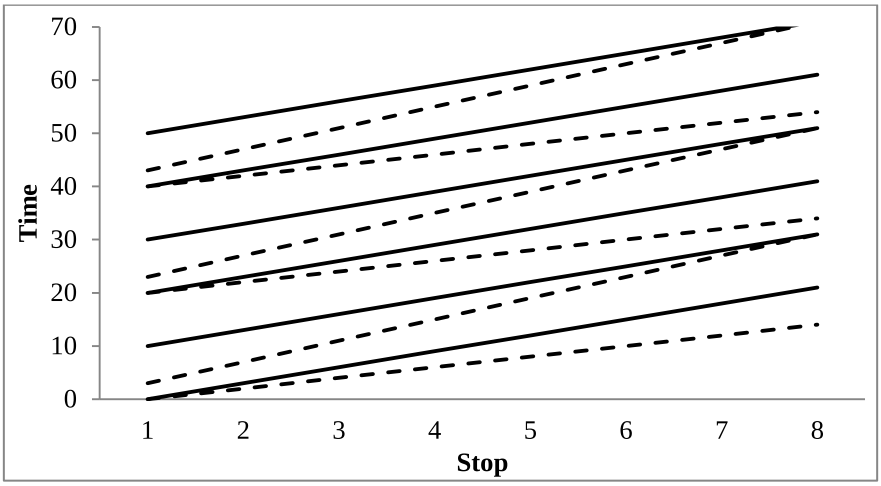


Figure 17 – Heterogeneous (dashed lines) and homogeneous (solid lines) operations

Heterogeneity refers to the difference in stopping patterns, headways and speed among trains on parallel tracks. Reducing heterogeneity is equivalent to lowering capacity utilisation, i.e. implicitly adding buffer time between subsequently running trains. Fewer trains can be run under heterogeneous operations within the same time period compared to homogeneous operations. Vromans et al. (2006) developed two analytical measures to compare heterogeneity among timetables.

$$SSHR = \sum_{i=1}^{n} \frac{1}{h_i^-} \qquad SAHR = \sum_{i=1}^{n} \frac{1}{h_i^A}$$

For the Sum of Shortest Headway Reciprocals (*SSHR*), h_i is the smallest scheduled headway between train *i* and train *i+1* on a track segment. h_i^A is the headway between trains *i* and *i+1* at the arrival stop. *SSHR* had limitations because departure headway was considered as important as arrival headway. However, arrival headways were more important, when trains were blocking each other (Vromans et al., 2006). Therefore, Sum of Arrival Headway Reciprocals (*SAHR*) was developed, only accounting for the arrival headways between subsequent trains.

2.1 Methodologies

Skip-stop optimisation in transit networks has attracted an increasing amount of attention within the last years. All the reviewed studies used a railway corridor as test network for their skip-stop optimisation. The objective was either minimising passengers' travel time (e.g. Suh et al. (2002); Mesa et al. (2009); Jong et al. (2012); Lee (2012); Sogin et al. (2012); Feng et al. (2013); Jamili et al. (2014); Katori et al. (2014)) or a combination of minimising passengers' travel time and operating costs (e.g. Leiva et al. (2010); Freyss et al. (2013); Chen et al. (2014); Lin & Ku (2014)). The problem was typically formulated as a mixed integer linear problem, which is computationally hard to solve for real-world problems. Consequently, heuristic methods were applied, typically genetic algorithms (e.g. Jong et al. (2012); Lee (2012); Sogin et al. (2012); Feng et al. (2013); Chen et al. (2014); Lin & Ku. (2014))

Suh et al. (2002) and Mesa et al. (2009) approached the skip-stop optimisation problem by skipping smaller stops. Leiva et al. (2010) designed an optimisation method minimising user costs. Railway lines were separated in all-stop services, express services (origin and destination stops served) and skip-stop services (a subset of stops served). First, a local minimum was found by solving the problem analytically, afterwards, the fitness of the solution was proved by a heuristic. Freyss et al. (2013) applied a continuous approximation method to find the optimal distance between visited stops, implicitly also the best performing stopping patterns. Jamili et al. (2014) applied a fuzzy approach to reach a more flexible solution. The aim was to minimise travel time and maximise the train spread. Headway constraints were relaxed to increase the size of the solution space. Katori et al. (2014) applied dynamic programming to find optimal stopping patterns and potential overtaking stops. Local timetables were formed from train diagrams.

In table 1, the reviewed skip-stop optimisation papers are split in two dimensions. The first dimension outlines how passengers' travel behaviour is taken into account and the second dimension exhibits the solution approaches. As outlined in e.g. Lin & Ku (2014), solving the skip-stop problem for real-world networks to optimality takes forever. Consequently, most papers develop heuristic solution approaches that are able to find good solutions fast. Static passenger demand neglects adaptations in travel behaviour, while responsive passenger demand explicitly considers adapted travel behaviour when skipping stops. Within the skip-stop optimisation field, there is a gap in the literature as regards methods considering passenger demand as responsive.

Several of the reviewed studies applied static demand data based on passenger counts or ticket sales (e.g. Suh et al. (2002); Mesa et al. (2009); Leiva et al. (2010); Lee (2012); Jong et al. (2012); Feng et al. (2013); Jamili et al. (2014); Katori et al. (2014); Lin & Ku (2014)). Other studies did not have such data and were forced to use approximations. Freyss et al. (2013) assumed that destinations of all boarding passengers were evenly distributed on all remaining stops. The assumption undermines the potential benefits, since skipping a particular stop usually is beneficial when on-board passengers' benefit (i.e. reduced travel time) exceed the cost experienced by boarding and alighting passengers (e.g. lower frequency). Consequently, introducing skip-stop services is mostly beneficial in systems with large variations in passenger loads between stations in a corridor (Chen et al., 2014). Sogin et al. (2012) and Chen et al. (2014) determined the passenger split on trains going between stop pairs solely based on the travel time. According to Nielsen (2000) and Raveau et al. (2014), transfers and waiting times are important factors describing passengers' route choice. Disregarding relevant factors can result in inaccurate passenger flows (Nielsen, 2000).

Table 1 – Overview of skip-stop optimisation approaches

		Passenger demand		
		Static		**Responsive**
		Predetermined	**Approximated**	
Solution Method	**Heuristic**	Leiva et al. (2010) Lee (2012) Jong et al. (2012) Feng et al. (2013) Lin & Ku (2014)	Sogin et al. (2012) Freyss et al. (2013) Chen et al. (2014)	Present study
	Dynamic Programing / Other	~ Suh et al. (2002) Mesa et al. (2009) Jamili et al. (2014) Katori et al. (2014)		

Skip-stop services should be applied with care and only when trains run at a high frequency and capacity utilisation is moderate to low. Thereby, reduced service frequency would not distract passengers and the capacity utilisation would not exceed the maximum as a result of increased heterogeneity when stops are skipped (Freyss et al., 2013).

Passengers' adapted route choice should be considered explicitly to quantify the impacts of changes in railway operations accurately (Sun et al., 2014). So far no skip-stop optimisation studies considered responsive passenger demand (table 1). Pioneers have begun to account for passengers' response (i.e. adapted travel behaviour) to changed transit operations. Constantin & Florian (1995) considered demand and supply balance in a bi-level frequency optimisation problem. The objective was to minimise expected travel and waiting time by changing frequency settings. The lower level problem was a transit assignment model with frequencies determined by the upper level. Wang & Lin (2010) developed a bi-level model to minimise transit operating cost related to the fleet size and travel cost for passengers. The upper level determined the routes and the associated headways. The lower level was a transit assignment. Parbo et al. (2014) minimised passenger waiting time in a large-scale transit network by optimising departure times for buses. Passengers' adapted route choice was derived by a transit assignment.

2.2 Objectives and Contribution

The objective of the skip-stop optimisation in the present study is to reduce the passengers' travel time and the heterogeneity of the railway operations. A sequential bi-level skip-stop optimisation/route choice model approach is developed. The upper level solves the skip-stop optimisation problem. The new stopping patterns serve as input to the lower level public assignment calculation, where passengers' responses to changes in stopping patterns (i.e. their adapted route choice behaviour) are calculated. The output from the public assignment calculation (i.e. passengers' route choice) is then used as input to the skip-stop optimisation. The iterative bi-level process continues until no stops can be skipped with a positive effect on the objective value.

The novelty in this approach lies primarily in the level of detail in which passengers' adapted path choice is being derived and incorporated explicitly in the optimisation by use of a schedule-based transit assignment model. This allows a derivation of the passenger loads on individual trains on each of the track segments traversed. We compare our model with a similar model, where passengers are assumed to maintain their path choice when stops are skipped. This comparison outlines the benefits of the demand responsive approach, where passengers adapt their path choice when stops are skipped. Another important contribution of the current paper is the application to a real-world problem consisting in a suburban large-scale railway network.

3 Method

For a railway line with *n* stops, *n-2* stops are subject to the skip-stop optimisation. Initial and terminal stops are not skip-stop candidates. The number of different stopping patterns considered for each railway line is thus $P = 2^{n-2}$. In a corridor, the number of different stopping patterns to consider is found by using the binomial coefficient.

$$\binom{P}{TL} = \frac{P!}{TL!*(P-TL)!} \tag{13}$$

where *P* is the number of different stopping patterns calculated as outlined above, and *TL* is the number of railway lines operated. To ease the understanding, consider the following numerical example where two lines ($TL = 2$) are operating in a corridor with 4 stops. The number of different stopping patterns *P* for each railway line is 4 ($=2^{(4-2)}$). Operating two lines with possibly 4 different stopping patterns each allows the two railway lines to be operated in 16 different ways. However, when two railway lines are operating with similar stopping patterns it is equal to the situation where a single line is operating with higher frequency, thus 4 combinations out of 16 can be disregarded. Furthermore, when railway line 1 is operating with stopping pattern A and railway line 2 is operating with stopping pattern B, passengers do not perceive it different from the situation where railway lines 1 and 2 and operating with stopping patterns B and A respectively, thus the six mirrored stopping patterns are also disregarded. Practically speaking, only 6 different stopping patterns are considered for the corridor. The same number you get by inserting the values ($P=4$, $TL=2$) into equation 1.

The notation used to formulate the skip-stop optimisation problem mathematically is outlined in table 2.

Table 2 – Notation for the analytical formulation of the skip-stop optimisation problem

i	Origin stop
j	Destination stop
c	Trip purpose (1: Commuter trips. 2: Business trips. 3: Leisure trips)
k	Train run
l	Railway line
$H_{s,k,k+1}$	Actual headway on stop s between subsequently running trains k and $k+1$
H	Minimum safety headway
s	Stop
M	Large positive constant
ω_1, ω_2	Weights for the terms in the objective function
p	Stopping pattern
$d_{i,j,c}$	Passenger demand between i and j for trip purpose c
TT^p_{ij}	Travel time between i and j when p is operated
TT_{ij}	Minimum travel time from i to j among all operated stopping patterns
a^p_{ij}	Parameter equal to 1 if p visits both i and j, 0 otherwise
TL	Number of railway lines running operating with different stopping patterns in a corridor
x^p_l	Binary variable equal to 1 if railway line l is operated with stopping pattern p, 0 otherwise
$b^{k,p}$	Departure time for k from initial stop when p is operated
$t^{k,p}_s$	Travel time from the initial stop to stop s for k when p is operated

Minimise

$$\omega_1 \sum_i \sum_j \sum_c TT_{ij} * d_{ijc} + \omega_2 \sum_s \sum_k \frac{1}{H_{s,k,k+1}} \quad (14)$$

Subject to

$$H_{s,k,k+1} = ((b^{k+1,p} + t_s^{k+1,p}) - (b^{k,p} + t_s^{k,p})) * x_l^p, \forall l,k,p,s \quad (15)$$

$$H_{s,k,k+1} \geq H, \forall s,k \quad (16)$$

$$TT_{ij} \geq TT_{ij}^p + M*(1 - a_{ij}^p * x^p{}_l), \forall p,l,i,j \quad (17)$$

$$\sum_p x^p{}_l = 1, \forall l \quad (18)$$

$$x^p{}_l \in \{0,1\}, \forall p,l \quad (19)$$

$$t_s^{k,p} \geq 0, \forall k,p,s \quad (20)$$

$$b^{k,p} \geq 0, \forall k,p \quad (21)$$

$$TT_{ij} \geq 0, \forall i, j \qquad (22)$$

$$TT_{ij}^{p} \geq 0, \forall p, i, j \qquad (23)$$

The objective (2) is to minimise the passengers' travel time (divided into groups *c* based on their trip purpose) as well as the heterogeneity of the train operations. The heterogeneity measure is equivalent to the *SSHR* from Vromans et al. (2006) and gives an indication of the change in the capacity utilisation and the amount of buffer time between trains when comparing the existing and the optimised stopping patterns in corridors. The weights added to the two terms in the objective functions are applied so that the operator can put more or less emphasis on either part. Also, they are necessary since the units of the two parts are not equivalent. Equation (3) derives the headway between consecutively running trains runs *k* and *k+1* on every *s* for each *p*. Constraint (4) ensures that the derived headways are not below the minimum safety headway. The objective function (2) and the constraints (3) and (4) are developed in the present study in order to take into account explicitly the heterogeneity of the train operations in the objective function. The remaining formulation is based on the MILP formulation by Jong et al. (2012). Equation (5) derives the minimum travel time between *i* and *j* over all *p*. If *i* and *j* are not visited for *p*, the travel time is set to a large number, thus it is indirectly ensured that all stops are visited. Otherwise, the objective will be penalised by *M* multiplied by the demand d_{ijc}. Constraint (6) ensures that each line is operated with one and only one stopping patterns. Constraints (7) – (11) are domain setting constraints.

The aim of the current study is to optimise stopping patterns in railway networks while explicitly accounting for passengers' adapted route choice behaviour. Optimising stopping patterns is computationally very complex (e.g. Sogin et al. (2012); Jong et al. (2012)). Passengers' travel behaviour is derived by a route choice model, a non-linear non-continuous mapping of the timetable. Due to the complexity of the two problem instances, a bi-level heuristic solution approach solving the two problems sequentially is developed. The skip-stop optimisation is solved by a heuristic algorithm (upper level) and integrated with a public assignment model (lower level) to assess how passengers adapted their travel behaviour according to the modified stopping patterns.

3.1 Prerequisites and Assumptions

An existing transit network is required as a starting point including a timetable explicitly stating when transit vehicles arrive and depart from all stops. The transit assignment model needs to be schedule-based, i.e. revealing passenger flows on every train, thus also the number of boarding, alighting and through-going passengers for each train's stop. Only railway lines running in corridors are subject to the skip-stop optimisation. Corridors are formed based on the network structure. For the railway lines in a corridor to be subject to the skip-stop optimisation the following requirements needs to be satisfied.

- At least two railway lines should run on the same tracks.
- Railway lines should run at high frequency.
- Capacity usage in the corridor should be below maximum.

The first point ensures that all stops are served by at least one railway line. In the case that only one railway line is operating in a corridor it has to be an all-stop line. When changing stopping patterns for the railway lines operating in the corridor, the number of railway lines operating with different stopping patterns has to be the same before and after the optimisation. This requirement on maintaining the structure reduces the confusion passengers may experience when the stopping patterns are updated.

A high frequent system is necessary for a successful implementation of skip-stop services in order for the passengers not to experience excessive waiting time (Freyss et al., 2013). Skipping stops can force trains to slow down when overtaking is not possible (e.g. because of the track layout, non-existing side tracks) due to trains ahead when capacity utilisation is close to maximum. Consequently, capacity utilisation should be at a level ensuring that minimum safety headway constraints are not violated when changing stopping patterns.

Within corridors, guidelines on which stops to consider are made as follows:

- Skipping transfer stops is prohibited.
- Skipping initial/terminal stops is prohibited.

Closing down transfer stops could potentially have a severe impact on the affected passengers' travel experience since they are forced to find alternative routes. Skipping large transfer stops is prohibited due to the inherent uncertainty in predicting passengers' route choice adaptations. Skipping terminal stops does not directly affect passengers' in-vehicle time since there are no subsequent stops. Therefore, stopping is made "for free" regarding in-vehicle time, number of transfers and waiting time. However, the round trip time can be reduced by turning the train at an earlier stop.

Overtaking and Heterogeneity
Although capacity utilisation is below the maximum, if a train skips several stops, at some point it will have to slow down and wait for the train in front of it. Consequently, the travel time savings obtained from skipping stops would be lost. The way overtaking is addressed in the present study is inspired by Lee (2012). Considering the stops in the corridor sequentially from one end to the other, the difference in the number of skipped stops between parallel railway lines needs to be limited. Thereby, the headway is always kept larger than the sum of the safety distance and the travel time savings obtained from skipping a certain number of stops. This requirement ensures that minimum safety headway constraints are not violated and no trains are forced to wait behind or overtake other trains. Skipping stops on one railway line could potentially impact trains running on other parts of the railway network. One way to overcome this would be to adjust the departure time and/or add extra buffer time between trains.

3.2 Heuristic Solution Approach

Skip-stop optimisation is combined with a passenger assignment model in order to account explicitly for passengers' adapted travel patterns. These two models are combined in a bi-level heuristic optimisation model, where skip-stop optimisation forms the upper level and the transit assignment the lower level. The two levels are solved sequentially, where the output of one level serves as input for the other level and vice versa, i.e. the updated stopping patterns from the upper level serve as input for the passenger assignment model in the lower level, which then yields the new passenger flows to be used as input in the next upper level calculation. The entire heuristic solution approach works according to the step-wise approach (outlined below) elaborated in the following subsections.

1. Run public assignment where all lines are initialised to all-stop lines. Make a list *F* comprising all stops *s* subject to the skip-stop optimisation.
2. Derive pre-optimisation potential for each *s* in *F*.
3. Derive elaborate optimisation potential for the most promising stop *s in *F*.**
4. If *s has a positive optimisation potential, skip *s**, and remove the sub-network of *s** from *F*. Otherwise, remove only *s** from *F* and go to (2).**
5. If *F* is not empty, go to (2). Otherwise, continue.
6. If stopping criterion is met, stop. Otherwise, run public assignment and go to (1).

In step 0, the public assignment model (first lower level calculation) is run for the case where all railway lines are initialised as all-stop lines. The idea behind the initialisation is to avoid that existing stopping patterns bias the formation of new stopping patterns. Existing stopping patterns may be a result of old demand patterns. However, in this study, the methodology aims at forming demand responsive stopping patterns. Steps 1 and 2 derive the potential for skipping a certain stop on a particular railway line. In step 3, skipping stop s^* is effectuated and all stops related to the stop s^* are removed from the list F containing the stops, which are allowed to be skipped. Skipping stops (i.e. the upper level calculations) continues until the list F is empty (step 4). Then another iteration of the bi-level heuristic is run starting with a lower level passenger assignment calculation. As indicated in step 5, the process continues until the stopping criterion is met.

Optimisation Potential

A pre-optimisation potential is calculated for every stop. The most promising stops are selected, and for those an elaborate optimisation potential is calculated. The decision on whether or not to consider a certain stop on a particular railway line as a preliminary skip-stop candidate relies on the following quantitative assessment:

$$POP_{co,l,s} = \frac{\#TGPax_{co,l,s}}{\#DEPax_{co,l,s} + \#EPax_{co,l,s}}, \forall co, l, s \qquad (24)$$

where $POP_{co,l,s}$ is the pre-optimisation potential for each stop s on each railway line l running in each of the corridors co, $TGPax_{co,l,s}$ is the number of through-going passengers at stop s on line l in corridor co, $DEPax_{co,l,s}$ and $EPax_{co,l,s}$ are the number of passengers respectively disembarking and embarking at stop s from line l in corridor co.

Figure 2a shows the 4-dimensional used calculation graph outlining the data requirements for the optimisation potential. In each corridor, every railway line and its arrival/departure times for all stops should be known (supply data). For each stop, the number of through-going passengers as well as the sum of embarking and disembarking passengers should be determined.

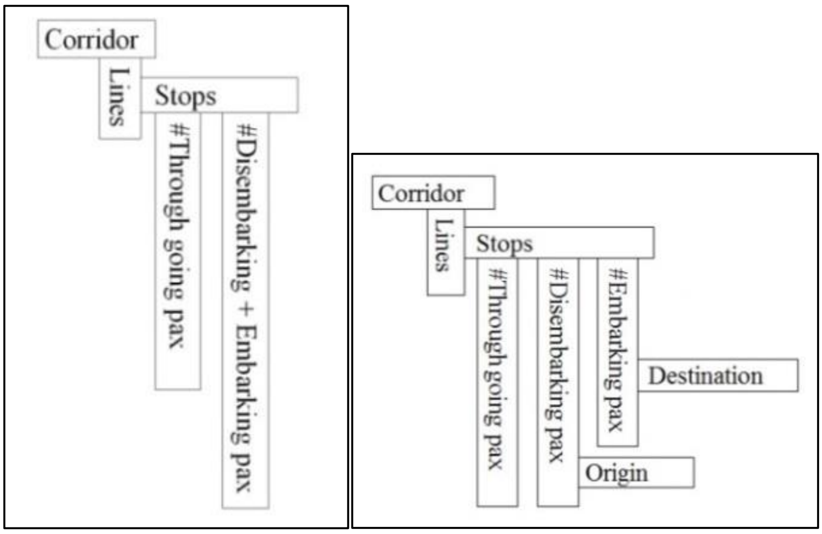


Figure 18a and b - Calculation graphs outlining the data requirements – preliminary and elaborate optimisation potential

Based on the pre-optimisation potential, the most promising stops to be skipped for all railway lines in each corridor are identified. To make a more comprehensive assessment of the benefits and costs of skipping a particular stop on a line, an elaborate optimisation potential is derived. The elaborate optimisation potential requires disaggregate data on passengers' route choices to derive the expected impacts on in-vehicle time, number of transfers and waiting time when skipping stops. The extended calculation graph is outlined in figure 2b, with the notable difference of the increase from 4 dimensions to 5 dimensions. The extension concerns information about disembarking passengers' origin stations and embarking passengers' destination stations. Based on this knowledge, it is possible to approximate how skipping the considered stop will impact the affected passengers' travel cost, since O-D information gives us the ability to assess relevant path alternatives for the affected passengers. In each corridor, every railway line and its arrival/departure times for all stops should be known (supply data). The number of through-going passengers, disembarking passengers (and their origin) and embarking passengers (and their destination), has to be known for each railway line's stop.

Route Choice Adaptations – elaborate optimisation potential

Passengers travelling within a corridor are affected in three different ways when their desired boarding/alighting stop is skipped (trip types 1, 2 and 3). Passengers travelling between stops where one stop is in the corridor and the other is outside the corridor are treated differently (trip type 4). In the following, explanations of passenger route choice adaptations for trips destined to a stop which is a skip-stop candidate are provided. The approach for calculating the effect on passenger trips originating from the stop which is a skip-stop candidate is similar, thus omitted. Trip type adaptations 1, 2 and 3 (within corridors) are outlined in figure 3. The skip-stop candidate is stop 4 on railway line B. Passengers' trip type adaptations are defined as follows.

- Trip type 1 – At least one other railway line in the corridor serves the affected passengers' boarding and alighting stops.
- Trip type 2 – Boarding and alighting stops are not served by any railway line in the corridor. But there exists a transfer stop in-between where two railway lines serving the boarding and alighting stops, respectively, meet.
- Trip type 3 – Boarding and alighting stops are not served by any railway line in the corridor and there are no transfer possibilities in-between. The affected passengers at some point have to travel in the opposite direction.

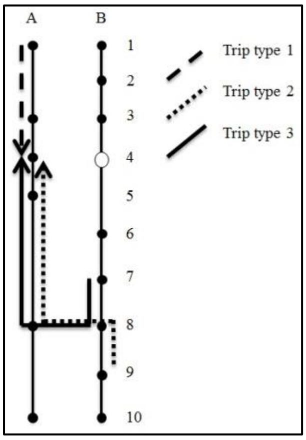


Figure 19 - Trip types within corridors

Trip type 1
Depending on the arrival time at the boarding stop, passengers might be unable to board the first departing train if it does not serve both their boarding and alighting stops. The additional generalised travel cost is estimated as the extra waiting time at the boarding stop. Passengers are assumed to arrive randomly at the boarding stop due to the high frequency of the railway lines operated in the corridor. Consequently, passengers' additional generalised travel cost is set to be equal to the increase in headway, h_new_{ijl} minus h_{ijl}, for railway lines l serving both stops i and j multiplied by the number of passengers pax_{ijlc} travelling between the two stops i and j and their VoT.

$$Cost1_{ijlc} = \frac{1}{2} * (h\_new_{ijl} - h_{ijl}) * (VoT_c * Pax_{ijlc}), \forall i, j, l, c \qquad (25)$$

Trip type 2
Depending on the arrival time at the boarding stop, passengers might have to board one of the subsequent trains. Additionally, passengers have to make a transfer to reach their desired destination. The transfer stop has to be served by two railway lines, where either of the railway lines also serves the boarding stop and the other railway line serves the skip-stop candidate. The additional generalised travel cost is equal to the waiting time at the boarding stop (i.e. ½ headway h_{ijl} of railway lines serving the boarding stop i and transfer stop j) and the transfer time conservatively set to the headway h_{jml} of railway lines l serving the transfer stop j and the skip-stop candidate. Transfers are penalised by the transfer penalty tp_c.

$$Cost2_{ijlc} = (\frac{1}{2} * h_{ijl} + h_{jml} + tp_c) * (VoT_c * Pax_{ijlc}), \forall i, j, l, c \qquad (26)$$

Trip type 3
Completing the trip requires travelling in the opposite direction and an extra transfer. This could discourage passengers from choosing the path due to reluctance to travel on routes diverging significantly from the direct one (Raveau et al., 2014). Therefore, skipping the stop is penalised heavily. The cost is set to M (a large number) multiplied by the number of passengers affected. Consequently, type 3 trips are allowed only when a negligible amount of passengers travels between the stops i and j.

$$Cost3_{ijlc} = M * Pax_{ijlc}, \forall i, j, l, c \qquad (27)$$

Trip type 4 (from outside the corridor)
Passengers travelling from outside the corridor to the skip-stop candidate are assumed to be affected only by the reduced frequency of railway lines serving the skip-stop candidate. The reduced frequency is derived by subtracting the old headway h_{ijl} from the new headway h_new_{ijl}. Passengers wait for a train serving both boarding and alighting stops rather than boarding the first arriving train and then making a transfer. The increase in generalised travel cost is set equal to the headway increase for railway lines *l* serving stops *i* and *j* multiplied by the number of passengers affected and their *VoT*.

$$Cost4_{ijlc} = \frac{1}{2} * (h\_new_{ijl} - h_{ijl}) * (VoT_c * Pax_{ijlc}), \forall i, j, l, c \qquad (28)$$

Total additional cost
For each O/D-pair where either the origin or destination stop is a skip-stop candidate (i.e. $i = s \vee j = s$), it is assessed whether skipping the candidate stop leads to a type 1, 2, 3 or 4 trip. The cost is estimated by summing equations (13) - (16).

$$x_1 * Cost1_{ijlc} + x_2 * Cost2_{ijlc} + x_3 * Cost3_{ijlc} + x_4 * Cost4_{ijlc}, \forall l, i = s \vee j = s, c \qquad (29)$$

The total additional cost for skipping stop *s* on railway line *l* is derived as the sum of costs determined by which trip type skipping stop *s* leads to between stopping pairs, where either stop *s* is origin or destination stop (i.e. $i = s \vee j = s$). In the formula, x_r's are binary variables equal to one if stops *i* and *j* are connected by a type *r* trip, zero otherwise.

Benefit

As a consequence of saved acceleration, deceleration and dwell time on the skipped stops, passengers' in-vehicle time is reduced. The reduction in in-vehicle travel time is found by subtracting the original in-vehicle travel time IVT_{ijl} between stops *i* and *j* when travelling on railway line *l* from the new and reduced in-vehicle travel time IVT_new_{ijl}. Passengers travelling on railway line *l* between stop *i* and *j* with *s* (skip-stop candidate) as intermediate stop benefit from the in-vehicle time reduction.

$$(IVT\_new_{ijl} - IVT_{ijl}) * Pax_{ijlc} * VoT_c *, \forall l, i < s < j \vee i > s > j, c \qquad (30)$$

Skipping stops

Based on the assessment of total benefits and costs, it is determined if the railway line's stop should be skipped. When the elaborate optimisation potential shows that the benefits exceed total additional costs, the stop is skipped.

Sub-networks
Ideally, the assessment of whether to skip a stop or not is derived by a passenger assignment calculation to be as accurate as possible. However, doing so would be extremely time-consuming. On the other hand, if all stops are subject to being skipped at the same time, the elaborate optimisation potential outlined above would include a large degree of uncertainty. By limiting the number of stops to be skipped in the upper level, only the most promising stops are skipped while all stops in the sub-network of the skipped stops are prohibited from being skipped. Afterwards, passenger flows are recalculated by a passenger assignment calculation in

the lower level. If all stops are subject to being skipped at the same time in the upper level, there is the risk that the flows, and hence also the objective function in the next iteration, deviate too much from the flows in the previous iteration, and that the algorithm thus would oscillate and converge either slowly or not at all.

As a compromise, sub-networks with only minor passenger interaction are developed for each stop. The idea is to speed up the calculation time by reducing the number of lower level calculations without compromising the accuracy of the elaborate optimisation potential. This is achieved by allowing several (not directly dependent) stops to be skipped in one upper level calculation. Sub-networks are created as stops are skipped. After skipping a stop, it is prohibited to skip stops from the particular stop's sub-network. The sub-network of a particular stop consisted of the following two parts:

1. For the affected railway line, the remaining stops in the corridor.
2. For the other railway lines in the corridor, the skipped stop.

Prohibiting stops from the sub-network to be skipped is done because of the difficulty in predicting passengers' changes in travel patterns when these stops are skipped. When skipping a stop leads to a large change in passenger share between the railway lines operating in the corridor, it becomes more complex, on top of the previous changes, to predict the impact of skipping additional stops on the same railway line.

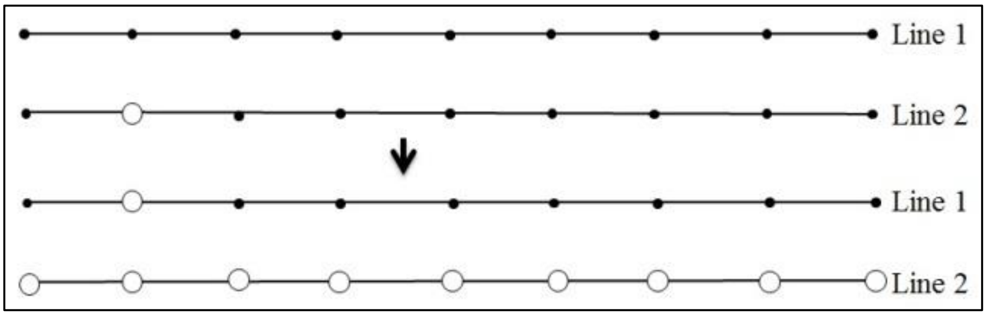


Figure 20 - Forming sub-networks in a corridor

Figure 4 outlines an artificial line diagram for two parallel railway lines in a corridor. A stop on Line 2 is skipped (white hollow). Below, the sub-network is formed (all white hollow stops). The current form of sub-network is a compromise between being able to predict the passengers' adapted travel behaviour and allowing as many stops as possible to be skipped before running another public assignment.

Stopping Criterion

Skipping stops from the list, *F*, comprising stops that are subject to being skipped (refer to the step-wise approach introduced in the beginning of section 3.2) continues until *F* is empty. Then, a transit assignment calculation is run and all non-skipped stops are loaded into *F*. This bi-level skip-stop optimisation/transit assignment approach is terminated when no stops has an elaborate optimisation potential where benefits exceed costs.

Pseudo-code

The pseudo-code gathers the threads from the previous sub-sections.

Initialisation, Run public assignment, T_{lc}^{S}

<u>Upper-level problem</u>

Calculating optimisation potentials of skipping stops

For each Corridor *Co*

For each railway line *L*

For each Stop *s*

Calculate optimisation potentials *OP*

If $OP < 0$, discard *s*

Store values *(Co, L, s, OP)* in a list *F*

If *F* is empty, terminate entire process

Otherwise, proceed

Skipping stops

Continue the following until *F* is empty

Select the stop *s** with the largest *OP* in *F*

Calculate the elaborate optimisation *EO* potential for *s**

If $EO > 0$, and no exceptions are violated, skip stop *s**

Otherwise, remove *s** from *F* and select new *s**

Derive sub-network *sn* for *s**

Remove all values *sn* from *F*

Go to Lower-level problem

<u>Lower-level Problem</u>

Run public assignment, T_{lc}^{s}

Calculate solution value

Go to the upper-level.

The initialisation yields passengers' route choice. The relevant information in this context is how many passengers (divided by trip purpose *c*) T_{lc}^{s} are through-going, respectively disembarking and embarking on each railway line *l*'s stop *s*. In this iterative bi-level minimisation problem, the upper-level problem is a skip-stop optimisation problem. The output of the skip-stop optimisation is a timetable with modified stopping patterns. The new timetable serves as input to the lower-level problem (transit assignment), where passenger flows are derived by a route choice model.

4 Public Assignment Model

The lower level of the developed bi-level heuristic solution approach is a public assignment model. Each time stopping patterns are changed in the upper level, the impact on passengers' travel cost and behaviour is assessed by a schedule-based transit assignment model comprising all transit lines. By applying a schedule-based transit model we overcome the common lines problem which can occur when frequency-based transit assignment models are used. At the same time, frequency-based models limit all transit lines to be operated with equal headway, while there is no problem in transit lines having uneven headway when applying the scheduled-based approach (Nuzzolo et al., 2012). In figure 5, the transit assignment procedure is outlined.

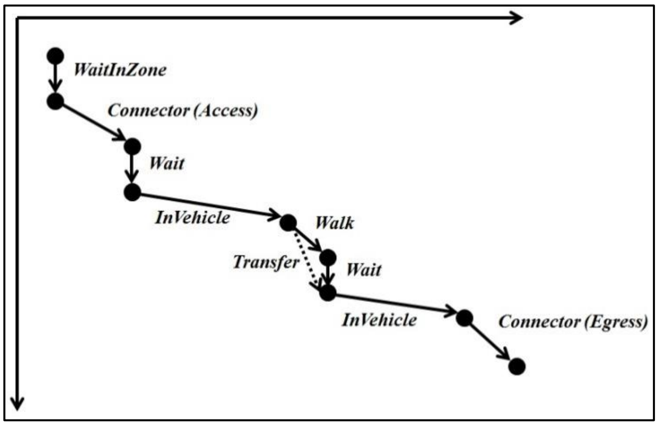


Figure 21 - Transit assignment procedure

The model applies a utility-based approach to describe passengers' travel behaviour. The utility function reflects the generalised travel cost on a path π from origin station i to destination station j for passenger group c. Each path π is uniquely defined in a disaggregate time-space graph, from which the passenger load on each individual run can be observed.

$$\begin{aligned} C_{ij\pi c} &= \beta_c * WT_{ij\pi} + \beta_c * WZT_{ij} + \beta_c * WalkT_{ij\pi} + \\ &\beta_c * ConnT_{ij} + \beta_c * \#Transfers_{ij\pi} + \beta_c * TotalIVT_{ij\pi} \end{aligned} \tag{31}$$

$C_{ij\pi c}$ is the utility for travelling between stations i and j on path π divided into trip purposes c, $WT_{ij\pi}$ is the transfer waiting time when travelling on path π between stations i and j, WZT_{ij} is the waiting time at home or in the origin zone for travellers between stations i and j, $WalkT_{ij\pi}$ is the walking time used when transferring on path π between stations i and j, $ConnT_{ij}$ is the time used getting from home to the desired boarding stop when travelling from station i to station j, $\#Transfers_{ij\pi}$ is the number of changes on path π between stations i and j, $TotalIVT_{ij\pi}$ is the total in-vehicle time on path π between stations i and j. Together, these parameters reflect passengers' disutility associated with a journey between two stations in the transit system (i.e. the generalised travel cost). For each trip purpose, beta values (*VoT)* are outlined in table 3 and are retrieved from Nielsen & Frederiksen (2006). All beta values except for *TransferPenalty* (DKK/transfer) are in DKK/min.

Table 3 - Beta values

	WalkTime	*Waiting Time*	*Connector Time*	*WaitInZone Time*	*Transfer Penalty*	*Train InVehicleTime*
Commuter	0.633	0.633	0.75	0.28	8.8	0.45
Business	4.50	4.50	4.50	1.217	64	3.783
Leisure	0.467	0.467	0.33	0.117	4	0.15

Only *WaitingTime*, *TransferPenalty* and *TotalInVehicleTime* are addressed explicitly in the heuristic skip-stop optimisation. Changes in generalised travel cost are continuously tracked so that passengers are not worse off than before the optimisation. However, increased generalised travel cost could be an overestimation of the unaddressed parameters (i.e. *WaitInZoneTime, ConnectorTime* and *WalkTime*) compared to reality. In the transit assignment, passengers are loaded uniformly every minute, which is similar to assuming that passengers do not time their arrival at boarding stops. For some passengers, this might be wrong, thus waiting time at the boarding stop is overestimated.

5 Case Study

The heuristic skip-stop optimisation approach is tested on the public transport network in the Greater Copenhagen area in Denmark (figure 6) with demand from the morning peak hours (7 am to 9 am). Only trains operated in the suburban railway network are subject to the skip-stop optimisation. In table 4, the origin and destination station, respectively, are outlined for each of the five suburban railway corridors shown in figure 6. Passengers are eligible to choose the mode they find most attractive, e.g. if skipping a certain stop leads to a deterioration of the chosen railway service, passengers may start travelling by bus.

Currently, trains in operation are either skip-stop trains or all-stop trains with a 20 minute cycle time, i.e. the timetable is repeated three times hourly. This cyclic structure is maintained throughout the skip-stop optimisation. The stopping patterns are formed individually for each railway line. However, there are limitations on the number of stops that can be skipped on a line depending on the way stopping patterns are formed for the other lines in the particular corridor.

Table 4 - Corridors

Corridor no.	Origin station	Destination station
1	Dybbølsbro	Køge
2	Danshøj	Høje Taastrup
3	Flintholm	Ballerup
4	Ryparken	Farum
5	Hellerup	Holte

Figure 22 - Suburban railway network (www.DSB.dk)

Train conflicts in the corridors are addressed by limiting the number of skipped stops between parallel railway lines. The upper limit is chosen based on the number of lines operated in the corridor and their frequency, respectively. The difference in running time between subsequently running trains is not allowed to violate the headway constraint (constraint 4). If the maximum number of consecutive stops is skipped on one railway line, then a stop has to be skipped on the other railway lines in the corridor before an extra stop can be skipped on the first railway line. Thereby, conflicts should be avoided for the optimised stopping patterns and heterogeneity should be reduced or at least maintained.

The Danish Transport Authority (2014) estimated that skipping a stop results in travel time savings of around 2-3 minutes. Their assessment involved Danish IC, regional and suburban railway lines. Suburban trains typically accelerate/decelerate faster and dwell less than IC trains, thus a 2 minute travel time saving was chosen when skipping stops.

5.1 Passengers' Route Choice

Deciding whether to skip a particular stop on a particular railway line is based on an assessment of the elaborate optimisation potential. Deriving the elaborate optimisation potential ideally requires the exact demand on each railway line's stop, and even more importantly, information about which stops passengers originate from/are destined for (i.e. their route choice). For the considered test case, passengers' exact route choice is not directly available and therefore, approximations are made.

Passengers often consider more than one path between O/D-pairs. Consequently, stop-to-stop loads are not sufficient to determine their route choice. Instead of assuming an equal distribution of passengers on stops, the following method is applied to approximate the adapted route choice of embarking and disembarking passengers when skipping a stop. The share of passengers originating from a certain stop i to the stop j which is the skip-stop candidate s by line l, is found by dividing the demand from that particular stop to the candidate stop, Pax_{ijlc}, by the total demand going to the candidate stop. This value is multiplied by the number of disembarking passengers from railway line l on the candidate stop ($j=s$), $DEPax_{jlc}$. This is done for every railway line l and for all trip purposes c.

$$Pax_{ijlc} = DEPax_{jlc} * \frac{Pax_{ijc}}{\sum_{i} Pax_{ijc}}, \forall i, j = s, l, c \qquad (32)$$

Pax_{ijkc} is used as input in the calculations of the additional generalised travel cost based on the applicable trip type between O/D-pair i to j (equations (17) – (22)). Afterwards, similar calculations are made for embarking passengers as well.

6 Results

It took six iterations of the sequential bi-level optimisation approach until no stops could be skipped beneficially. Table 5 shows the percentage change in railway passengers' generalised travel cost when comparing the optimised skip-stop patterns with the existing, respectively.

Table 5 - Percentage change in railway passengers' generalised travel cost

	Transfers	In-vehicle Time	Gen. Travel Cost	Waiting Time
Opt. vs. Exist.	1.38 %	-5.48 %	-1.81 %	1.60 %

The in-vehicle time decreased, while the waiting time and the number of transfers increased compared to the stopping patterns of the existing railway services. Railway passengers' generalised travel cost was also reduced significantly.

The optimisation yielded a yearly reduction in in-vehicle time, number of transfers and waiting time equivalent to 75 million DKK (about 10 million EUR) compared to existing stopping patterns. The savings are not equivalent to an increase in revenue, but rather an indication of how much transit passengers are willing to pay to obtain such improved services. In figure 7, the percentage change in generalised travel cost compared to the results from the existing stopping patterns is exhibited.

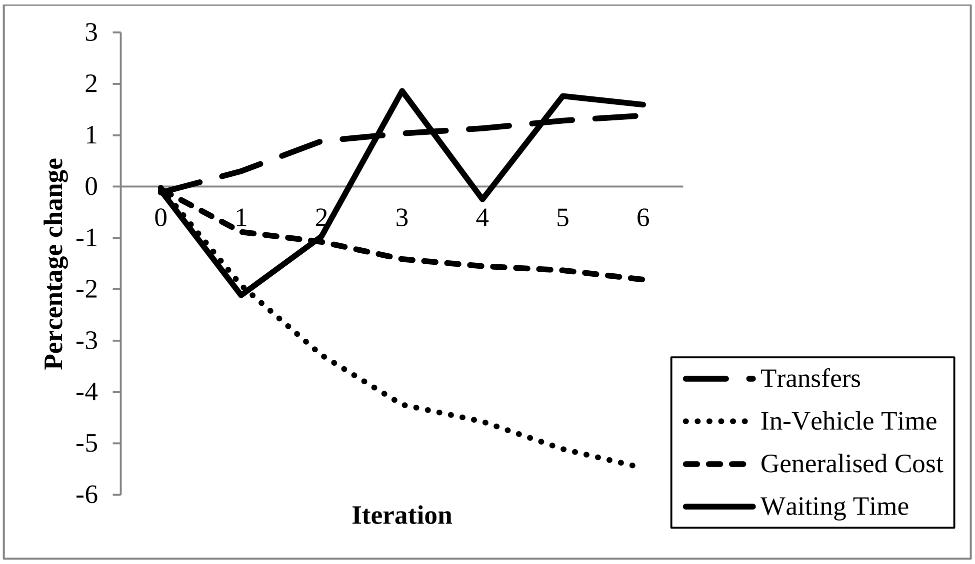


Figure 23 - Percentage change in railway passengers' generalised travel cost

In-vehicle time, generalised travel cost and the number of transfers changed most in the first iterations and then flattened more out. Due to the complexity in predicting passengers' adapted route choice without running a transit assignment model to assess the potential benefit of skipping every single stop, the development in waiting time was harder to predict. When skipping a stop, it was assumed that passengers waited for another train at their boarding stop and then travelled along the same path to their destination. However, passengers might have chosen other paths.

Besides improving the level of services experienced by the passengers, the objective of the current approach was to reduce the heterogeneity of the railway operations in the corridors. In table 6, the heterogeneity-value (*SSHR*) is presented for each of the five corridors. The percentage change between the existing and the optimised stopping patterns is shown in parentheses.

Table 6 - *SSHR* values in the corridors

Corridor No.	1	2	3	4	5
Existing	5	1.5	1.7	0.9	6
All-stop	2.4	1.35	1.35	0.6	2.4
Optimised	3.5 (-30%)	1.43 (-5%)	1.43 (-16%)	0.6 (-33%)	2.55 (-55%)

For the all-stop scenario train operations are homogeneous, thus these values serve as a lower bound on the *SSHR* in each of the corridors. A significant reduction in heterogeneity (up to 55%) was obtained compared to the existing operations. The *SSHR* values in table 6 primarily depend on the headway between subsequent trains determined by the number of railway lines operated in the corridor, but also the difference in the number of skipped stops between railway lines in the corridor.

In table 7, the average number of skipped stops is compared between existing and optimised stopping patterns. In corridors 1 and 2 more stops were skipped when comparing the optimised stopping patterns with the existing ones. The opposite was seen in corridors 4 and 5. The number of stops skipped in corridor 3 was the same in the existing and optimised stopping patterns.

Table 7 - Skipped stops statistics for every corridor

Corridor No.	**1**		**2**		**3**		**4**		**5**	
Railway lines operating	4		3		3		2		4	
Stopping pattern	Exist.	Opt.	Exist.	Opt.	Exist.	Opt.	Exist.	Opt.	Exist.	Opt.
Average	4.25	4.75	1.00	1.33	1.33	1.33	2.50	1.00	2.50	2.25
Standard deviation	3.50	0.96	1.73	0.58	2.31	0.58	3.54	0.00	2.89	0.50

Examining the standard deviation on the number of skipped stops per railway line in each corridor reveals a significant reduction when comparing the existing and optimised stopping patterns. This was a result of the limit on the difference in the number of skipped stops between railway lines operated in the same corridor.

Figure 8 visualises the optimised and existing stopping patterns in corridor 3. For the existing stopping patterns, trains were operated as either all-stop or express trains, while for the optimised stopping patterns all trains were operated as skip-stop trains.

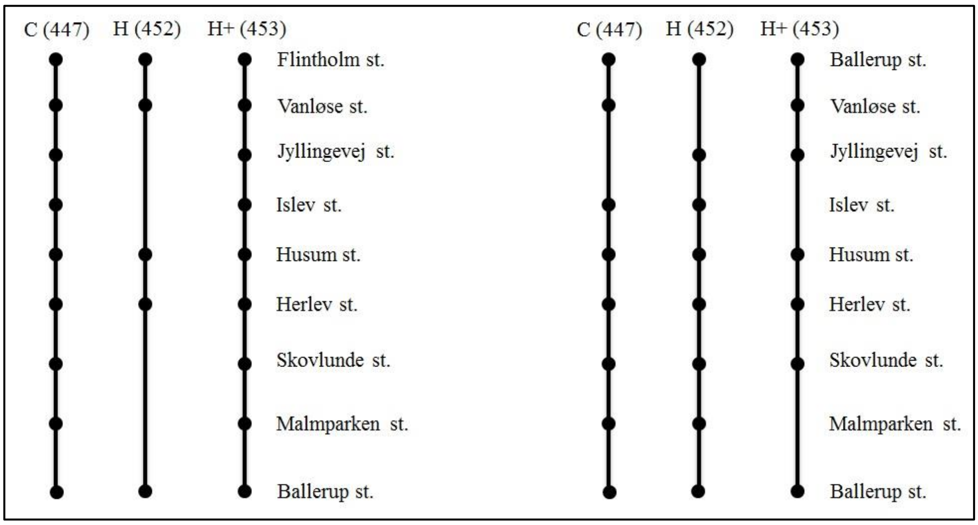


Figure 24 - Line diagram – Corridor 3 (existing (left) vs. optimised (right))

Figure 9 outlines a rough sketch of the time-space diagram for the existing and the optimised stopping patterns respectively for corridor 3. Optimising the stopping patterns yields a wider train spread. Consequently, minor delays are now more likely to be absorbed by the buffer time rather than propagating to subsequently running trains.

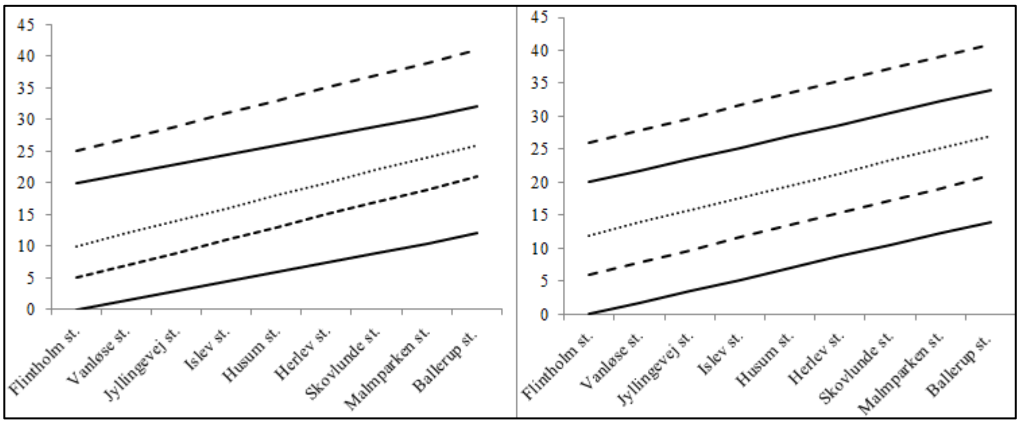


Figure 25 - Time space diagram - Corridor 3 (existing (left) vs optimised (right))

6.1 Pareto frontier

To test the effect of including heterogeneity in the objective function (eq. 2), a sensitivity analysis with ten different weight combinations is undertaken. The weight combinations are outlined in table 8.

Table 8 - Weight combinations for sensitivity analysis

Instance #	ω_1 (travel time)	ω_2 (heterogeneity)
1	0.0	1.0
2	0.1	0.9
3	0.2	0.8
4	0.3	0.7
5	0.4	0.6
6	0.5	0.5
7	0.6	0.4
8	0.7	0.3
9	0.8	0.2
10	0.9	0.1

In figure 10, the average heterogeneity of the railway operations in the five corridors is plotted on the horizontal axis, while passengers' travel time is plotted on the vertical axis. Thereby, it is possible to assess at what "cost" the reduced heterogeneity is obtained. Comparing instance 1 (favouring heterogeneity) and instance 10 (favouring travel time) shows that a 7% travel time reduction implies an increase in average heterogeneity from 1.6 to 3.6. Also, it is seen that for the first five instances the reduction in passenger travel time only implies a marginal increase in heterogeneity. On the other hand, for the instances 6-10, small reductions in passenger travel time implies relatively large increases in heterogeneity.

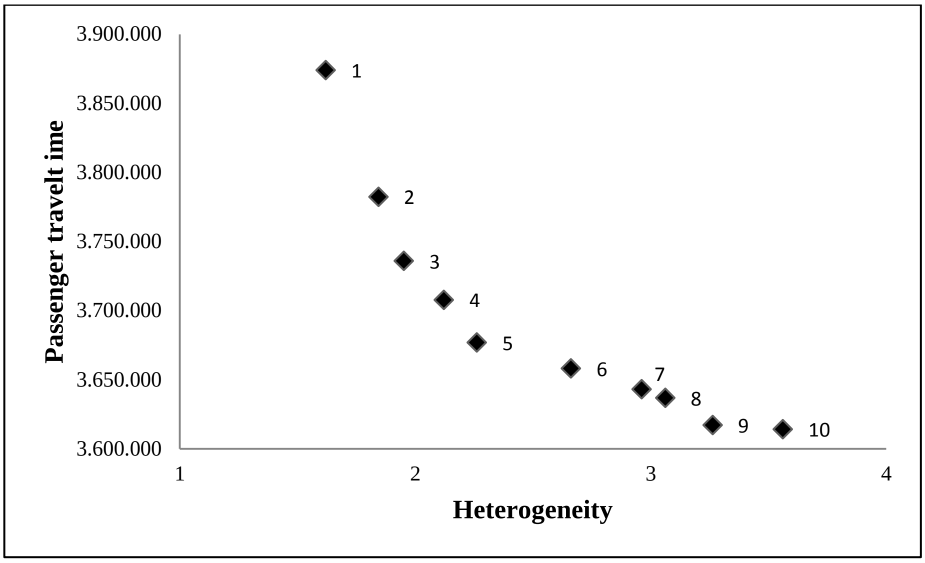


Figure 26 - Pareto frontier (including instance#)

6.2 Demand representation

To test how the model stands out from the remaining models in the literature, we explore the impacts of how demand is represented. In our model, passengers' are assumed to change their route choice as stops are skipped. We refer to this model as demand responsive. In the literature, changes in passengers' route choice are not considered. Instead, passengers' existing route choice is used to assess the impacts of changes in stopping patterns as well as serving as weights for selecting which stops to skip. The results are outlined in table 9.

Table 9 – Responsive vs. fixed demand (results)

	Transfers	InVehicle time	GenCost	Waiting time
Responsive demand	82 952,30471	3 663 410,975	102 270 093	196 486,8862
Fixed demand	82 598,76111	3 725 320,675	102 851 621	196 341,4339
Fixed relative to responsive	-0,4%	1,7%	0,6%	-0,1%

The relatively larger in-vehicle time obtained with a fixed demand representation for this network is caused by the fact that passengers' adapted route choice behaviour is not considered explicitly. Based on these results, it can be concluded that it is naïve to expect passengers do not change their route choice in response to changes in stopping patterns. Passengers' travel behaviour is uncertain when transit operations are changed. Here, it is shown that not considering passengers' travel behaviour explicitly implies that the stopping patterns become less attractive since the configuration of the stopping patterns is based on passengers' route choice in the old network. To obtain the full benefit of out skip-stop optimisation mode, it is therefore important explicitly to consider passengers' route choice behaviour responsively, i.e. when stops are skipped, passengers' route choice behaviour is reassessed such that skip-stop the optimisation can continue with the most up-to-date passenger path choice data.

7 Conclusions and Future Work

The current study proposed a bi-level skip-stop optimisation approach explicitly considering passengers' modified route choice as response to changes in railway lines' stopping patterns. The approach was applied successfully to the suburban railway network in the Greater Copenhagen area in Denmark. The optimisation yielded a significant reduction in passengers' travel time. Additionally, the spread of trains in the corridors was increased (i.e. the heterogeneity was reduced), thus the risk that minor delays propagate to subsequent trains was reduced. Consequently, railway passengers on average got faster and with a smaller risk of delay from A to B.

The current study contributes to the literature within skip-stop optimisation by developing an approach explicitly taking into account passengers' adapted route choice behaviour as well as considering a network rather than one corridor. A comparison shows that not considering passengers' travel behaviour explicitly implies that the stopping patterns become less attractive, simply because the configuration of the stopping patterns is based on passengers' route choice in the network with the old stopping patterns.

The idea behind adapting the stopping patterns is to update these as a response to the changes in station-to-station demand occurring during the past year. Timetable adaptation (as e.g. the one from Parbo et al., 2014)), on the other hand, serves the purpose to make the timetable conflict-free and at the same time minimise an objective function (e.g. minimising the transfer waiting time). Since the two models aim both at improving the transit operations with regards to the passengers, it would be interesting to explore the optimisation potential of combining these two models in a sequential manner. In the present paper, only the skip-stop optimisation is performed. This optimisation yields a 5.48 % reduction of the in-vehicle time, while the transfer waiting time increased by 1.60 %. The timetable adaptation (Parbo et al., 2014) yielded a 5.08 % reduction in transfer waiting time. It is therefore expected that an integration of the two "modules" into one running in a sequential manner, would yield a larger total benefit for the passengers. Furthermore, integrating timetable optimisation in the skip-stop optimisation by changing the offset values of the different railway lines could be done with the aim to e.g. improve headway evenness (arrival regularity) in each corridor.

To enhance the validity of the results and make the model more robust against variations in demand, a future version of the lower level should take into account in-vehicle crowding and the probability of being able to board specific trains. Another topic for future research is to integrate the developed approach with an optimisation of train departure times to facilitate a minimisation of the waiting time and maximising the spread of trains. Niu et al. (2015) developed an optimisation model to determine arrival and departure times to and from stations, thus also the time supplement and buffer time, so that passengers' waiting time were minimised. Since they considered varying demand and different stopping patterns for a corridor, future research could try to combine those models in order to optimise stopping patterns and timetables simultaneously under varying demand. Finally, it would be interesting to integrate skip-stop optimisation and fleet size optimisation. Skipping stops reduce the round trip time, which provides potential for fleet size reduction.

Acknowledgements

We gratefully acknowledge the financial support of the Danish Council for Strategic Research for the project "RobustRailS" that this study is part of. We also thank three anonymous reviewers for the insightful comments that helped improving a previous version of the paper.

References

Ceder, A., 2007. Public transit planning and operation: theory, modeling and practice. *Elsevier, Butterworth-Heinemann*.

Chen, X., Hellinga, B., Chang, C., & Fu, L., 2014. “Optimization of headways with stop -skipping control: a case study of bus rapid transit system”, *Journal of Advanced Transportation*.

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

Danish Transport Authority, 2014. “Optimisation of station structure”, *Report from March 2014.*

Feng, S., Wen-tao, Z., Ying, Y., & Dian-hai, W., 2013. “Optimal Skip-Stop Schedule under Mixed Traffic Conditions for Minimizing Travel Time of Passengers”, *Discrete Dynamics in Nature and Society*.

Freyss, M., Giesen, R., & Muñoz, J. C., 2013. “Continuous approximation for skip-stop operation in rail transit”, *Transportation Research Part C: Emerging Technologies*, *36*, 419-433.

Jamili, A., Ghannadpour, S. F., & Ghorshinezhad, M., 2014. “The optimization of train timetable stop-skipping patterns in urban railway operations”. *Computers in Railways XIV: Railway Engineering Design and Optimization*, *135*, 603.

Jong, J. C., Suen, C. S. J., & Chang, S. K. J., 2012. “Decision support system to optimize railway stopping patterns”, *Transportation Research Record: Journal of the Transportation Research Board*, *2289*(1), 24-33.

Katori, T., Kabumoto, H., & Izumi, T., 2014. “Shortening average trip times by adjusting stopping and overtaking train stations”. *Computers in Railways XIV: Railway Engineering Design and Optimization*, *135*, 577.

Kikuchi, S., & Vuchic, V. R., 1982. “Transit vehicle stopping regimes and spacings”, *Transportation Science*, *16*(3), 311-331.

Lee, Y. J., 2012. “Mathematical modeling for optimizing skip-stop rail transit operation strategy using genetic algorithm”, *National Transportation Center Research Report, Morgan State University, Available at: http://www. trb. org/Main/Blurbs/166899. aspx, accessed March, 26*.

Leiva, C., Muñoz, J. C., Giesen, R., & Larrain, H., 2010. “Design of limited-stop services for an urban bus corridor with capacity constraints”, *Transportation Research Part B: Methodological*, *44*(10), 1186-1201.

Lin, D. Y., & Ku, Y. H., 2014. Using genetic algorithms to optimize stopping patterns for passenger rail transportation. *Computer‐Aided Civil and Infrastructure Engineering*, *29*(4), 264-278.

Mesa, J. A., Ortega, F. A., & Pozo, M. A., 2009. “Effective allocation of fleet frequencies by reducing intermediate stops and short turning in transit systems”, In *Robust and Online Large-Scale Optimization* (pp. 293-309), Springer Berlin Heidelberg.

Nielsen, O. A., 2000. “A stochastic transit assignment model considering differences in passengers’ utility functions”, *Transportation Research Part B: Methodological*, *34*(5), 377-402.

Nielsen, O. A., & Frederiksen, R. D., 2006. Optimisation of timetable-based, stochastic transit assignment models based on MSA. *Annals of Operations Research, 144*(1), 263-285.

Niu, H., Zhou, X., & Gao, R., 2015. Train scheduling for minimizing passenger waiting time with time-dependent demand and skip-stop patterns: Nonlinear integer programming models with linear constraints. *Transportation Research Part B: Methodological*, *76*, 117-135.

Nuzzolo, A., Crisalli, U. & Rosati, L. (2012). A schedule-based assignment model with explicit capacity constraints for congested transit networks. *Transportation Research Part C: Emerging Technologies*, *20*(1), 16-33.

Parbo, J. Nielsen, O.A. Prato, C.G., 2014. "User perspectives in Public Transport Timetable Optimisation", *Transportation Research Part C, 48,* 269-284.

Raveau, S., Guo, Z., Muñoz, J. C. & Wilson, N. H., 2014. "A behavioural comparison of route choice on metro networks: Time, transfers, crowding, topology and socio-demographics", *Transportation Research Part A: Policy and Practice*, *66,* 185-195.

Sogin, S. L., Caughron, B. M. & Chadwick, S. G., 2012. "Optimizing Skip Stop Service in Passenger Rail Transportation", *2012 Joint Rail Conference* (pp. 501-512). American Society of Mechanical Engineers.

Suh, W., Chon, K. S., & Rhee, S. M., 2002. "Effect of skip-stop policy on a Korean subway system", *Transportation Research Record: Journal of the Transportation Research Board*, *1793*(1), 33-39.

Sun, L., Jin, J. G., Lee, D. H., Axhausen, K. W., & Erath, A., 2014. "Demand-driven timetable design for metro services", *Transportation Research Part C: Emerging Technologies*, *46*, 284-299.

Vromans, M. J., Dekker, R., & Kroon, L. G., 2006. "Reliability and heterogeneity of railway services", *European Journal of Operational Research,172*(2), 647-665.

Wang, J. Y., & Lin, C. M., 2010. "Mass transit route network design using genetic algorithm", *Journal of the Chinese Institute of Engineers*, *33*(2), 301-315.

Appendix 4: Parbo et al. (2015c)

Improving passenger oriented line planning for high frequent railway networks

Jens Parbo, Otto Anker Nielsen & Carlo Giacomo Prato

Technical University of Denmark, Department of Transport, Bygningstorvet 116B, 2800 Kgs. Lyngby, Denmark

Submitted to the special issue: *"Integrated optimization models and algorithms in rail planning and control" of Transportation Research Part C: Emerging Technologies*

Abstract.
The objective of the current study is to optimise the line configuration of a railway system so that passengers are accommodated in a way that minimises their number of transfers as well as their waiting time experienced at boarding and transfer stations, respectively. To this end, an improving algorithm is developed to solve the line planning problem with explicit consideration of passengers' travel behaviour.
The developed algorithm is a bi-level algorithm where optimising the railway line configuration and deriving the passengers' adapted travel behaviour are done sequentially. Due to the inherent complexity of the line planning problem with explicit consideration of passengers' travel behaviour, a heuristic solution approach is developed. The approach is based on swapping the first or last part of a railway line with the first or last part of another line at a station where the two lines meet. With the aim of searching the solution space intelligently, a tabu search framework is applied to the line planning problem, while a transit passenger assignment model derives passengers' adapted route choices.
The bi-level algorithm is validated on the suburban railway network operating in the Greater Copenhagen area in Denmark. Applying the improving bi-level passenger-oriented line planning algorithm to this network yields a reduction of 3.83% in railway passengers' number of transfers and 3.88% in their waiting time.

Keywords: Railway Timetabling · Public Transport Optimisation · Passenger Behaviour · Line Planning · Large-Scale Application

1 Introduction

Statistics published by the United Nations (2014) show that the number of people living in urban areas around the world is 54%. The same report announces that this number is expected to increase to 66% by 2050. One result of the increased urbanisation is a change in travel demand for urban transit systems. Not only the overall level of travel demand will change, but the spread of the travel demand may also change due to changes in residential location. To accommodate such yearly changes in travel demand, the objective of this paper is to adapt the line plan configuration of a railway network at the tactical planning level so that passengers' overall travel cost is minimised. Due to the planning perspective (tactical planning level is usually one year ahead) and the strict economic pressure most operators are operating under, the current model disregards building new infrastructure or changing the structure of the timetable, i.e. the frequency of runs as well as the service level on all edges is kept.

When timetables are updated, the operator has to balance wishes from various stakeholders, e.g. the passengers who want a reliable, fast and high frequent service and the operator who wants to maximise profit by e.g. minimising fleet size (Ceder, 2007). In this paper, we develop a tool for the planners enabling them to change the line plan configuration with an explicit passenger focus. The assumption is that changes in demand over the years can make line plans out of date in terms of providing direct lines to as many passengers as possible. To accommodate passengers when changing the line configuration of a transit network, it is essential not only to examine the station-to-station demand, but also how passengers actually adapt their travel behaviour when line plans are changed. To do so, a bi-level heuristic is developed, where the upper level optimises the line plans while the lower level derives passengers' adapted route choice behaviour.

The paper is structured as follows. In section 2, the existing literature on the line planning problem is reviewed. Section 3 presents the notation and the mathematical formulation. In section 4, the bi-level methodology developed in this study is explained. Section 5 outlines the application of the proposed methodology to the suburban railway network in the Greater Copenhagen area in Denmark together with the results obtained and a discussion of related issues. Finally, section 6 concludes the paper and outlines directions for future research.

2 Literature review

The Line Planning Problem is a well-known problem in transit optimisation. Line planning is a sub problem of the more general Transit Route Network Design Problem (TRNDP) consisting of the following five elements of which line planning is concerned only with (1) and (2).

1. Designing routes
2. Setting frequencies
3. Building timetables
4. Scheduling vehicles
5. Scheduling drivers

Route configuration is supposed to satisfy the chosen objective(s) while the frequencies of vehicles on these routes ensure sufficient vehicle capacity (Schöbel & Scholl, 2005; Kepaptsoglou & Karlaftis, 2009).

From the review of different line planning models in (Schöbel & Scholl, 2005; Ceder, 2007), it is evident that the objectives of the line planning problem are separated into three categories.

1. Passenger-oriented
2. Operator-oriented
3. System-oriented

The passenger-oriented objectives include minimising travel time, maximising route directness, maximising service area coverage, minimising waiting cost or a combination of these (e.g. Lee & Vuchic, 2005; Zhao et al., 2005; Schmidt & Schöbel, 2010; Wang et al., 2011; Schöbel, 2012). Combining these could lead to contradicting objectives. For example, focusing solely on minimising the number of transfers could lead to a route configuration with very long lines. Notwithstanding, serving all OD-pairs directly, the travel time could become unnecessarily high (Schmidt & Schöbel, 2010). The operator-oriented objectives include minimising operating cost, maximising profit, minimising fleet size or a combination hereof (e.g. Wan & Lo, 2003; Bussieck et al., 2004; Goosens et al., 2004; Lee & Vuchic, 2005). Finally, the passenger-oriented and operator-oriented objectives can be combined into so-called system or welfare oriented objectives (e.g. Mandl, 1980; Lee & Vuchic, 2005; Fan & Machemehl, 2006; Borndörfer et al., 2007; Nachtigall & Jerosch, 2008).

The next task is to determine the line configuration by choosing the optimal set of lines to be operated and the frequency needed to accommodate the passenger demand. Attractive lines are formed dynamically based on shortest path calculations and a maximum deviation from the shortest path (e.g. Lee & Vuchic, 2005; Zhao et al., 2005; Fan & Machemehl, 2006; Borndörfer et al., 2007; Nachtigall & Jerosch, 2008; Wang et al., 2011; Schöbel, 2012) or simply selected from a pool of all feasible lines (e.g. Bussieck et al., 2004; Fan & Machemehl, 2006; Schmidt & Schöbel, 2010).

Selecting which lines to apply in daily operation could be done in several ways. Mandl (1980), Bussieck et al. (2004) and Lee & Vuchic (2005) solved the problem algorithmic, Wan & Lo (2003) formulated the TNDP as a mixed integer problem, where the aim was to find the optimal route and frequency settings. Goosens et al. (2004) formulated an integer problem which was solved by a branch-and-cut algorithm with valid inequalities. Zhao et al. (2005) and Wang et al. (2011) developed a meta-heuristic, ISATG, which was a combination of Simulated Annealing, Tabu Search and Greedy Search. Goosens et al. (2006) and Borndörfer et al. (2007) solved the line planning problem as a multi-commodity flow problem, Fan & Machemehl (2006) and Schöbel (2012) used a Genetic Algorithm, Nachtigall & Jerosch (2008) applied Column Generation, while Schmidt & Schöbel (2010) applied a Dantzig-Wolfe decomposition algorithm.

2.1 Passenger demand

Passenger-oriented line planning models require data on passengers' travel patterns, e.g. O/D-demand data. These data are used to assign the necessary amount of vehicles to different lines as well as to find the optimal line configuration when minimising the passenger-oriented objective, e.g. the number of transfers. In Schmidt & Schöbel (2010) the influence of integrating O/D-demand data in the line planning problem was shown to be NP-hard even in simplified cases.

The share of passengers travelling by a certain transit vehicle and their route choice is highly dependent on the configuration of transit lines (Lee & Vuchic, 2005; Schöbel & Scholl, 2005; Fan & Machemehl, 2006; Borndörfer et al., 2007; Ceder, 2007; Guihaire & Hao, 2008; Kepaptsoglou & Karlaftis, 2009). Consequently, the line planning problem should ideally include mode choice as well as route choice calculations in order to account for adaptations in passengers' travel behaviour (Schöbel & Scholl, 2005;

Ceder, 2007; Guihaire & Hao, 2008). However, due to its inherent complexity, doing so makes the problem intractable for larger problem instances (Schmidt & Schöbel, 2010).

Regarding an explicit consideration of passenger demand, the bi-level programming paradigm appears as suitable to treat transit network design problems and in particular studies concerned with route and frequency configurations. Bi-level programming explicitly takes into account how passengers adapt their route choice to new frequency and route settings. Changing these settings and deriving passengers' adapted route choice is done sequentially in two separate optimisation problems, therefore the name bi-level. The output from the upper-level serves as input for the lower-level and vice versa.

Among recent bi-level studies optimising route and frequency settings, Szeto & Jiang (2014) applied a hybrid artificial bee colony algorithm to solve the route and frequency settings for a bus network as a bi-level problem, where the routes and their frequencies were determined at the upper level, while passenger route choice was derived at the lower level. Since the model was developed for strategic planning purposes, passengers' route choices were modelled through a frequency-based transit assignment model where in-vehicle travel time and transfers were used as performance indicators. Cancela et al. (2015) applied a bi-level approach for the TNDP for strategic planning purposes considering passengers' travel time and their transfers explicitly. Since they applied an analytical mathematical problem formulation, they were only able to solve the TNDP on a small test network due to the computational complexity. Cancela et al. (2015) pointed out that no study has succeeded in applying mathematical programming to large-scale problem instances of the Transit Network design problem. This is primarily due to the computation time, which becomes very large when network size increases and the number of potential routes, accordingly, grows extremely large. Finally, Fu et al. (2015) determined line frequencies and stopping patterns on long distance high speed railway networks in China by formulating a bi-level optimisation problem, which was solved by a greedy heuristic. Due to the long-distance context and the fact that the model was made for strategic planning purposes, only riding time and transfers were considered in the passenger-oriented objective function. The long distance context allowed the route choice to be more simplified than for e.g. suburban networks. In the long distance context, travel time to the boarding (alighting) stop and waiting time at the boarding (alighting) stop were ignored. Only in-vehicle travel time and transfers were considered to have an impact on passengers' long distance route choice.

2.2 Objectives and contribution

The objective of the present study is to improve the line plan of an existing railway network so that passengers' transfers, their waiting time at boarding stops and at transfer stops, respectively, is minimised. The contribution of this paper is an explicit and elaborate consideration of passengers' route choice (by the use of a schedule-based transit assignment model) combined with an improving line planning model, which is applicable at the tactical planning level. The shorter planning period (compared to line planning models at the strategic planning level) allows the model to be more demand responsive, since short-term demand can be predicted with larger certainty than long-term demand. Additionally, based on the shorter planning period, we make use of the existing timetable structure in order also to be able to assess the impact on passengers' waiting time at the boarding stops and transfer stops. Adopting the existing timetable structure, it is assumed that O/D-travel demand, departure time choice, boarding stop choice as well as passengers' mode choice are fixed meaning that only passengers' route choice is adaptable when the new line plan is implemented.

To take passengers' travel behaviour adaptations into account, a bi-level optimisation model is developed. At the lower level, passengers' travel behaviour is derived and at the upper level, the line configuration (route and frequency) is adapted to meet the demand. At the lower level problem, a schedule-based transit assignment model is applied which better captures passengers' route choice and the occupancy of each individual vehicle compared to the frequency-based model (which is the one typically applied to this problem) where only average vehicle occupancy rates are revealed. Furthermore, it allows a structure of the timetable which might involve uneven headways. This is appropriate when demand varies throughout time periods. It also makes the derivation of passengers' transfer waiting and boarding waiting time more accurate.

Finally, a contribution lies in the application of the model to a real large-scale network, which proves the applicability of the model developed in the current paper.

3 Mathematical formulation and notation

The current study applies a bi-level optimisation approach, where the objective is to minimise passengers' in-vehicle time, their number of transfers, and the cost for operating the train lines at a given frequency. The mathematical formulation of the line planning problem uses the notation from table 1.

Table 5 - Notation

i	Origin station.
j	Destination station.
c	Passenger group.
β_c	Value of time for different passenger groups c.
bs	Boarding stop.
s	Transfer stop.
p_{ij}	Path between i and j.
$\delta^s{}_p$	Binary parameter equal to one if transfer stop s is on the path p.
$\rho^p{}_{ij}$	Proportion of the demand between i and j choosing path p.
e	Edges (directed).
foc^l	Frequency operating cost for line l.
l	Train line.
f^l	Frequency variable of line l.
vc	Vehicle capacity.
tp	Transfer penalty.
$d_{i,j,c}$	Passenger demand between origin i and destination j.
TT_e^l	Travel time on edge e using line l.
x^l	Binary variable equal to one if line l is operated, zero otherwise.
t_{ij}	The minimum required number of transfers between i and j based on the operated lines l.
K	Pool of lines l.

The input data are an O/D matrix revealing passengers' travel demand d_{ijc} from station i to station j, divided into passenger groups c based on three travelling purposes (commuter, business or leisure trips), each with their own value-of-time. The network of railway lines is a directed graph $G = (V, E)$ with vertices V and directed edges E on which lines are supposed to be selected. Travel times are given for all edges and are dependent on which line the passenger is travelling on. Passengers choose the route that is expected to maximise their utility. Route choice relies on several factors, such as availability (frequency) of the lines, travel time, number of transfers etc. (see description of attributes affecting passengers' route choice in section 4.4). In the mathematical model, two decision variables are used: (i) a binary variable x^l that indicates whether a line is operated or not, and (ii) a continuous variable f^l that is the frequency of the operated line l. Lines are selected from the set of all feasible lines K. The analytical formulation is as follows:

$$Minimise\ w_1 \sum_i \sum_j \sum_c \left(TT_{ij} + tp * t_{ij}\right) * d_{ijc} + w_2 \sum_l foc^l * f^l \qquad (1)$$

$$TT_{ij} \geq \sum_{e \in i,j} TT_e^l + M * \left(1 - x^l\right), \forall l \qquad (2)$$

$$\sum_{l:e \in l} vc * f^l * x^l \geq d_e, \forall e \qquad (3)$$

$$\sum_{l:e \in l} f^l \leq f_e, \forall e \qquad (4)$$

$$x^l \in \{0,1\}, \forall l \qquad (5)$$

$$f^l \geq 0, \forall l \qquad (6)$$

The objective is to minimise passengers' travel time, TT_{ij}, the number of transfers, t_{ij}, multiplied by a transfer penalty, tp. Both terms are multiplied by the demand d_{ijc}. The operator's perspective is represented in the objective function by the frequency cost, foc^l, multiplied by the frequency of all lines, f^l, which should ensure sufficient passenger capacity on all vehicles. Each of the two parts of the objective function is multiplied by an operator-defined weight, w_1 and w_2 respectively. Constraints (2) set the travel time, TT_{ij}, between stations i and j based on the chosen line configuration, where TT_e^l denotes the travel time on the directed edge e when travelling on line l. When summing over all edges e included in a path between stations i and j, a lower bound for TT_{ij} is derived. When certain lines are not operated ($x^l = 0$), their travel time is set to TT_e^l plus M, which is a large positive number. Constraints (3) ensure that the demand on every edge, d_e, can be accommodated by the vehicle capacity, vc, on all operated lines ($x^l = 1$), with the outlined frequency f^l. Constraints (4) limit the frequency f^l (or capacity utilisation) on particular track section (edges e) based on the minimum required safety headway between subsequently running trains f_e. (5) and (6) are domain setting constraints for the two decision variables.

The objective function has two competing parts, namely the travel time determined by which lines are operated and the operating cost determined by the frequency of the operated lines. If only the number of transfers were considered, the line configuration would consist of few very long lines and, as a consequence, passengers would experience a longer and more indirect journey (Schöbel & Scholl, 2005). The particular objective function in this paper ensures a balance between route directness, number of transfers and operating costs.

4 Bi-level heuristic solution approach

The influence of integrating passenger demand data makes the line planning problem NP-hard (Schmidt & Schöbel, 2010). Therefore, a bi-level heuristic solution approach is developed in this study in order to be able to apply the approach to a real-size railway network.

4.1 Bi-level framework

To solve the line planning problem introduced in the previous sections, this paper develops a bi-level heuristic solution approach, where the upper level determines which railway lines to swap at which intersection stations, while the lower level derives how passengers adapt their route choice behaviour to the updated line plan configuration. To improve the configuration of the railway lines, we develop a heuristic solution approach that derives an optimisation potential, which is the expected impact a certain swap has on passengers' travel cost. We apply a heuristic algorithm based on a tabu search framework. The core principle in tabu search is to diversify the search of the solution space and avoid ending up in local minima (Glover, 1990). As a consequence of the changed railway line configuration, passengers may adapt their route choice behaviour. To take this into account, the output of the improving line planning heuristic (i.e. the updated line plans) serves as input to a passenger assignment model. The output of the passenger assignment model (i.e. passengers' adapted route choices) then serves as input for the next iteration of the improving line planning heuristic. This bi-level algorithm works iteratively according to the following stepwise approach.

0. Run Passenger Assignment (sec. 4.4).
1. Calculate optimisation potential for change in line configuration (sec. 4.2).
2. Run Tabu Search Algorithm (changing line configuration) (sec. 4.3).
3. If stopping criterion is met, terminate.
4. Otherwise, go to 0.

The entire bi-level heuristic optimisation algorithm is terminated when no improving swaps are imposed by the upper-level algorithm. Then the updated demand responsive line plan then serves as the outcome of the model.

In the following three subsections, the derivation of the optimisation potential, the tabu search framework and the passenger assignment calculation, respectively, are explained.

4.2 Upper level – Optimisation potential

The idea behind this improving heuristic solution approach for the passenger-oriented line planning problem is to adapt the line plan configuration by swapping parts of lines at intersection stations where these lines cross. In figure 1 an example of one such swap is outlined. On top, the existing line plan is shown, where the dashed line l_1 and the solid line l_2 represent, respectively, two different railway lines intersecting at the station *s*. Below is shown how a swap changes the configuration of the lines: the two lines still meet at the station, but now the configuration of the two lines is changed. After imposing the swap, passengers travelling from A to D no longer have to make a transfer, as they can simply stay on l_2 all the way. Whether to impose a swap or not depends on the potential effect that the change in line configuration has on the objective.

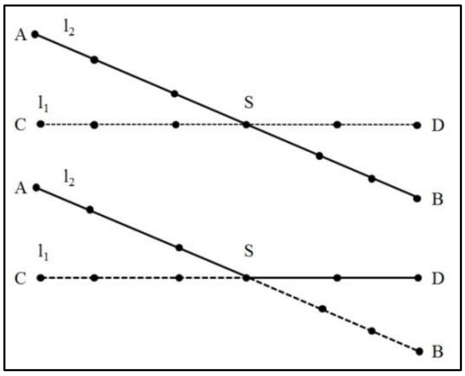


Figure 27 – Swapping lines

The decision variables in the analytical formulation of the line planning problem are which lines to operate and their frequencies. In the heuristic solution algorithm we find the railway lines to operate by swapping parts of the existing lines with each other. Ideally, the impact of every swap should be assessed by a passenger assignment calculation. However, since this would be extremely time-consuming, an optimisation potential approximating this impact is derived. The optimisation potential considers how a particular swap is expected to affect passengers' transfers as well as the waiting time passengers experience when boarding a train and transferring between two trains.

The optimisation potential of swapping two lines l_1 and l_2 at an intersecting station *s* represents the expected change every feasible swap has on the number of transfers and the waiting time passengers experience when boarding and transferring. It is calculated by summing the values from equations (7), (8) and (10) for the line configuration as it was before and after the swap, respectively. This is done for all feasible swap combinations of l_1, l_2 and *s*. By including the multiplication of the value-of-time, each equation yields a monetary value which makes the three equations comparable when summing the values.

Whenever the line configuration is changed, it is assumed that only passengers travelling on the swapped lines are affected. Since the remaining network structure is unchanged, passengers on the non-affected lines are assumed to maintain their travel behaviour. The notation used in the equations for the optimisation potential is similar to the notation used for the analytical problem formulation (see table 1).

Transfers

Swapping lines and thereby changing the line configuration means that some passengers who used to have a direct train service between origin and destination are now forced to make a transfer. On the other hand, some people who used to transfer now have one less transfer or no transfers.

To estimate the change in the number of transfers, we explicitly express the number of passengers involved in a transfer and their value-of-time based on their trip purpose. The passenger weighted transfer cost is derived as follows:

$$\sum_c t_{ij} * d_{ijc} * \beta_{transfer,c}, \forall i, j \qquad (7)$$

Imposing a swap in the heuristic solution algorithm is similar to changing four x^l variables in the analytical problem formulation. As in the analytical problem formulation, the values of the x^l variables determines the value of the t_{ij} variables (the minimum number of transfers required between station pairs *i* and *j*). Due to the inconvenience when transferring, passengers are assumed to minimise the number of transfers when travelling from origin to destination. To assess the impact a certain swap has on the number of transfers, equation (7) is derived for the updated x^l values and compared to the value of equation (7) using the original x^l values.

Transfer waiting time

By keeping the structure of the timetable, we keep the departure times of the different railway lines. Therefore, we are also able to assess transferring passengers' waiting time. To assess the potential change in passengers' transfer waiting time, the principle is basically the same as when assessing the change in the number of transfers. However, now also an estimate is needed of the number of passengers using the railway lines after being swapped and how the swap affects the transfer waiting time. To ease the understanding of equation (8), parentheses are put around the passenger demand and the transfer waiting time, respectively.

$$\sum_{l:s\in l}\sum_{p}\sum_{s}\sum_{c}\left(\rho_{ij}^{p} * d_{ijc}\right) * (\delta_{p}^{s} * ½ * h_{l,s}) * \beta_{wait,c}, \forall i,j \qquad (8)$$

We sum over all lines *l* passing by transfer stop *s*, all paths *p*, all transfer stops *s* and all three trip purposes *c*. In the first parenthesis, the passenger demand on path *p* between stations *i* and *j* is derived. In the second parenthesis, the transfer waiting time is derived as half the headway of the connecting line *l* stopping at transfer stop *s*, when this stop is part of the path. Using half of the headway as the expected transfer waiting time is done under the assumption that nothing is done to coordinate the transfers in the transit network. Only passengers choosing path *p* going by transfer stop *s* will experience this particular transfer waiting time. Finally, the passengers' weighted transfer waiting time is multiplied by the value of waiting time.

The headway $h_{l,s}$ used in equation (8) is derived from the well-known relationship outlined in equation (9). Here f^{ls} is the frequency of the attractive transfer lines stopping a transfer station *s*.

$$h_{l,s} = \frac{60\ min}{f^{ls}} \qquad (9)$$

Initial waiting time

When deriving the expected change in initial waiting time, it is assumed that passengers will use the same boarding stop after the line configuration is changed as they did before. This is a plausible assumption for most railway networks, since stations are located far from each other, which means that passengers often only consider boarding stops within walking distance. Additionally, due to the high transfer penalty, it is assumed that passengers will wait at their boarding stop until a train providing direct service (when one such exists) is available. If no direct service is available for the desired O/D-pair, passengers are assumed to board the first vehicle at their desired boarding station *bs*. The initial waiting time is thus derived as follows:

$$\sum_{l}\sum_{c} ½ * (h_{l(direct),bs} * y_{ij} + h_{l,bs} * (1 - y_{ij})) * d_{ijc}, \forall i = bs, j \qquad (10)$$

Here $h_{l(direct),bs}$ is the headway of direct lines *l* serving the boarding stop *bs*, while $h_{l,bs}$ is the headway of all lines *l* serving boarding stop *bs*. Binary variable y_{ij} is 1 if there is a direct line *l* between the station pair *i*=*bs* and *j*, zero otherwise. The assumption that the initial headway is equal to half the headway of the attractive set of lines applies for high frequent transit services where passengers are assumed to arrive to the boarding station at random.

$$y_{ij} = \begin{cases} 1, & t_{ij} = 0 \\ 0, & t_{ij} > 0 \end{cases} \qquad (11)$$

The binary variables y_{ij} are closely related to t_{ij} as outlined in equation (11). Again, t_{ij}, the minimum number of transfers between stations *i* and *j*, changes according to the x^l variables from the analytical problem formulation or, equivalently, to the swaps imposed in the heuristic solution algorithm.

Frequency

The frequency cost which is the second part of the objective function is not an integrated part of the heuristic solution algorithm. This is because an inherent part of the algorithm is maintaining the frequency on each track segment, which means that the total vehicle kilometres, and thus also the operating costs are maintained throughout the optimisation. However, as a result of the changes made to the line plan configuration, the need for rolling stock may change. To assess this change, a simplified calculation of the requirements for rolling stock is made for the existing line configuration and the optimised line configuration, respectively. Due to the simplicity of the calculations, the resulting requirements for rolling stock only provide a rough estimate. On the other hand, since the calculations are done similarly before and after optimising the line configuration, it can be expected that the results provide a realistic picture of the change, i.e. whether the new line configuration can be operated by more or less trains than before, although the exact change in the number of vehicles required might differ in reality.

The assessment is derived as follows. Between all pairs of terminal stations, where railway lines are operated, the total travel time and the total layover time are added. The travel time and the layover time together make up the total round trip time. The total round trip time is divided by the desired frequency and rounded up to the nearest integer, which then is the minimum required number of rolling stock.

4.3 Upper level – tabu search algorithm

In the tabu search algorithm, we determine which swaps to impose. The idea is to select the combination of swaps that has the largest total optimisation potential. Every line can only be involved in maximum one swap. The reason for this is that when a line is involved in more than one swap, the uncertainty in the prediction of passengers' travel behaviour becomes too large. Consequently, it is more difficult to predict whether a swap has an improving impact on the objective value or not.

When each line is only allowed to be involved in one swap, it is not straightforward which swaps to select. Therefore, we apply a metaheuristic, a tabu search, which has the ability to find a good solution fast. The improving tabu search algorithm needs an initial solution before the actual tabu search algorithm is initiated. Therefore, an initial solution needs to be constructed.

Construction of initial solution

I. Swaps, $[l_1,l_2,s]$ are selected greedily based on their optimisation potential until all lines are, at most, involved in one swap in the solution *Sol*.
II. From the optimisation potential, the initial solution has a value *SolVal*, which is denoted as the best solution value *SolVal**.

Improving heuristic

1. For each swap, $[l_1,l_2,s]$ in *Sol*, a new *Sol* is formed by swapping l_1 and l_2, with other lines.
2. Assess the new *SolVal*.
3. If *SolVal>SolVal**, then *Sol*=Sol* and *SolVal* = SolVal*.
4. After adding new swaps to *Sol*, the reverse swaps are labelled tabu until being released from the tabu list.
5. If iteration limit is met, go to 6. Otherwise, go to 1.
6. Terminate improving heuristic. Return *Sol**.

When the improving heuristic is terminated, the swaps in *Sol** are imposed, and the upper level is terminated which means that a new passenger assignment has to be run.

4.4 Lower level – public assignment model

Passengers' adapted route choice behaviour is derived from a passenger assignment model. The model applies a utility-based approach describing passengers' perceived generalised travel costs, *GenCost*. The utility function reflects passengers' generalised travel cost from origin *i* to destination *j* for trip purpose *c* as follows.

$$\begin{aligned} C_{ijc} &= \beta_c * WaitTime_c + \beta_c * FirstWait_c + \beta_c * WalkTime_c \\ &+ \beta_c * ConnTime_c + \beta_c * \#Changes_c + \beta_c * TotalIVT_c, \forall i, j, c \end{aligned} \tag{12}$$

C_{ijc} is the utility, *WaitTime* is the transfer waiting time, *FirstWait* is the waiting time at the boarding stop, *WalkTime* is the walking time used when transferring, *ConnTime* is the time used getting from home to the boarding stop, *#Changes* is the number of transfers in a trip and *TotalIVT* is the time spent in transit vehicles. For each trip purpose, beta values (value-of-time) are outlined (table 2) (Nielsen & Frederiksen, 2006). All beta values except for *#Changes* (Danish kroner per transfer) are in Danish kroner[4] per minute.

Table 6 – Beta values

	Walk Time	*Wait Time*	*ConnTime*	*FirstWait*	*#Changes*	*TotalIVT*
Commuter	0.633	0.633	0.75	0.28	8.8	0.45
Business	4.50	4.50	4.50	1.217	64	3.783
Leisure	0.467	0.467	0.33	0.117	4	0.15

5 Case study

To test the bi-level heuristic approach, the suburban railway network of the Greater Copenhagen area (outlined in figure 2) in Denmark is used. All lines in figure 2 apart from the F-line were subject to changes in line configuration. The reason for disregarding the F-line was that it does not share infrastructure with the remaining lines. We consider passenger demand for the morning peak hours from 7 am to 9 am. Demand is asymmetric in the peak hours, however, close to symmetric over the day. Consequently, the selected lines are operated in both directions with equal frequencies.

[4] 1 euro is approximately 7.5 Danish kroner

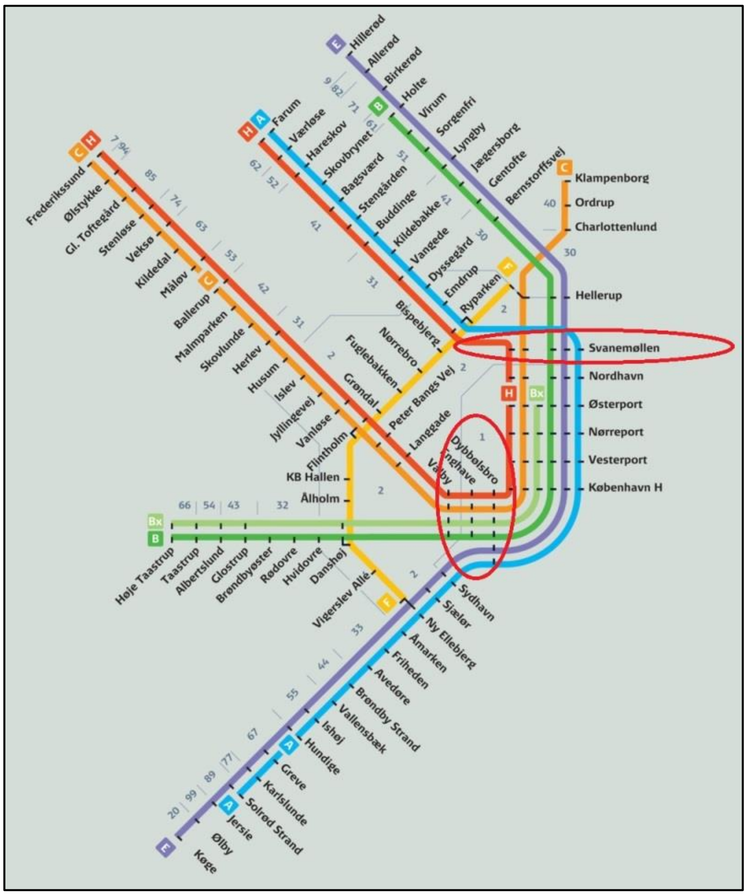


Figure 28 – Suburban railway network (Greater Copenhagen area)

The two intersection stations outlined with a red circle in figure 2 are the two stations where swaps can take place, namely *Dybbølsbro* and *Svanemøllen*. As can be seen, all lines stop at these two stations, consequently, several swapping options are feasible. On top of this, every line has a number of line variants, namely 61 different line variants exist for the suburban railway system. Line variants are defined as groups of trains sharing characteristics (e.g. short turned at the same station, skipping the same stops, running in the same direction etc.) belonging to the same railway. All of these line variants are eligible for changes in line plan configurations.

5.1 Results

The results obtained when applying the improving passenger-oriented line planning approach to the suburban railway network in the Greater Copenhagen area are exhibited in table 3. All results are weighted by the number of passengers. Therefore, the values for the commuters are much larger than the two other passenger groups since 84.8% of all trips in the morning peak hours are commuter trips. Business and leisure trips account for 2.4% and 12.8% of all trips, respectively.

Table 7 - Results

	Trains only		Entire transit system				
Trip type	#Changes	Waiting time	#Changes	FirstWait	Waiting time	IVT	GenCost
Benchmark							
Commute	12,229.3	34,557.7	75,589.5	10,038.6	183,175.2	3,438,482.6	3,368,389.2
Business	334.1	916.0	2369.9	1.4	5,512.8	102,650.0	756,221.3
Leisure	1,304.5	2,823.9	12,217.4	559.7	28,477.7	585,793.6	227,325.5
Total	13,867.8	38,297.5	90,176.8	10,599.8	217,165.6	4,126,926.2	4,351,936.0
Iteration final							
Commute	11,729.7	33,092.6	74,970.2	9,465.5	178,166.4	3,439,621.0	3,355,579.0
Business	322.5	872.6	2353.7	0.8	5,379.6	102,542.1	753,825.1
Leisure	1,285.1	2,845.5	12,200.8	508.9	28,222.3	586,228.1	227,466.5
Total	13,337.3	36,810.7	89,524.7	9,975.2	211,768.3	4,128,391.1	4,336,870.6
Percentage change							
Commute	-4.09	-4.24	-0.82	-5.71	-2.73	0.03	-0.38
Business	-3.46	-4.73	-0.68	-43.20	-2.42	-0.11	-0.32
Leisure	-1.49	0.77	-0.14	-9.09	-0.90	0.07	0.06
Total	-3.83	-3.88	-0.72	-5.89	-2.49	0.04	-0.35

The table presents the results at three levels. The upper level outlines the benchmark results (existing line plan) in Danish kroner (DKK), specifically the initial line plan with the corresponding timetable and frequencies. The middle level, named iteration final, presents the results (in DKK) for the improving line plan configuration. The lower level illustrates the percentage change between the results for the initial and the final line plan configuration for all relevant parameters impacting passengers' route choice. Apart from the horizontal tri-partition, a vertical bi-partition outlines relevant performance indicators for the entire transit system and for the suburban train system, respectively.

The purpose of the table partition is to show that, in particular, the suburban railway passengers are better off in terms of reduction in the number of transfers and waiting time after the optimisation. In fact, passengers on the suburban railway lines experience a reduction close to four percent in both the number of transfers and the transfer waiting time as a result of the improved line plan. The new line plan provides more direct connections where these are needed in terms of accommodating passenger demand.

When examining the reduction in the number of transfers and the transfer waiting time, for the average transit user, it is 0.72% and 2.49%, respectively. This is expected since the improvement in line plan configuration is applied only to the suburban railway network. The reason for exhibiting the results for the entire system is that we want to emphasise that the improved conditions for the suburban railway passengers are not obtained at the expense of the remaining transit passengers.

The reduction in transit passengers' boarding time, *FirstWait*, stems from the fact that more direct connections are now provided for the suburban railway passengers when each line is weighed by the passenger demand. Before the change in line configuration, passengers had to wait longer at their boarding stop for a direct line compared to after the improvement.

The change in passengers’ in-vehicle travel time is negligible, which indicates that passengers’ in general tend to use the same path between origin and destination after the optimisation as they did before. Consequently, in this case, only out-of-vehicle times are changed when improving the line plan configuration.

In the final column, passengers’ generalised travel cost (explained in section 4) is outlined. As seen in the table, transit passengers are on average slightly better off after the optimisation. This is clearly a result of the improvement for the railway passengers. Remaining transit users do not experience any changes.

Edge frequencies are maintained throughout the optimisation since swapping two lines does not affect which track segments are visited and how frequent. On the other hand, the rolling stock requirements increase from 95 to 99 when changing the line plan configuration from the initial to the optimised one. These numbers are derived in a post-optimisation assessment in the simplified way as outlined in section 4.2 *Frequency*.

6 Conclusions and future work

This paper developed a bi-level improving heuristic algorithm to solve the passenger-oriented line planning problem. The developed solution framework was applied to the suburban railway network in the Greater Copenhagen area. The improving algorithm yielded a significant reduction in passengers’ weighted number of transfers and the waiting time they experience when boarding and transferring.

Our study contributes to the literature by proposing a new optimisation tool, which has its strength in the line plan configuration optimisation and in the accuracy of the results since passengers’ adapted travel behaviour is considered explicitly. Another contribution is the application to a real-sized network which is proved by the case study presented in the paper.

A topic for future research is to include the rolling stock circulation plan in the objective of the heuristic solution algorithm. Also, testing different weight settings in the objective function of the heuristic solution algorithm would allow the operator to prioritise between the reduction in passengers’ travel impedance and the fleet size requirements. At the same time, the operator would be able to create several outputs from various weight settings in order to test different line plan configurations. It is expected that integrating rolling stock circulation explicitly in the heuristic algorithm will reduce the solution space. Therefore, passengers’ benefit will most likely be reduced compared to the results outlined in table 3. On the other hand, it would be expected that the requirements for rolling stock would be reduced or at least would not increase. Likewise, the applicability of the model would also be improved by incorporating crew scheduling in the model. Both extensions of the model are left for future research.

References

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

Bussieck, M. R., Lindner, T., & Lübbecke, M. E. (2004). A fast algorithm for near cost optimal line plans. Mathematical Methods of Operations Research,59 (2), 205-220.

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

Ceder, A. (2007). Public transit planning and operation: theory, modeling and practice. Elsevier, Butterworth-Heinemann.

Fu, H., Nie, L., Meng, L., Sperry, B. R., & He, Z. (2015). A hierarchical line planning approach for a large-scale high speed rail network: The China case. Transportation Research Part A: Policy and Practice, 75, 61-83.

Glover, F. (1990). Tabu search: A tutorial. Interfaces, 20(4), 74-94.

Goossens, J. W., Van Hoesel, S., & Kroon, L. (2004). A branch-and-cut approach for solving railway line-planning problems. Transportation Science, 38(3), 379-393.

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

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

Kepaptsoglou, K, & Karlaftis, M. (2009). Transit route network design problem: review. Journal of transportation engineering, 135(8), 491-505.

Lee, Y. J., & Vuchic, V. R. (2005). Transit network design with variable demand. Journal of Transportation Engineering, 131(1), 1-10.

Mandl, C. E. (1980). Evaluation and optimization of urban public transportation networks. European Journal of Operational Research, 5(6), 396-404.

Meng, L. and Zhou, X. (2011). Robust single-track train dispatching model under a dynamic and stochastic environment: a scenario-based rolling horizon solution approach. Transportation Research Part B: Methodological, 45 (7), 1080–1102.

Nachtigall, K., & Jeros, K. (2008). Simultaneous network line planning and traffic assignment. In OASIcs-Open Access Series in Informatics (Vol. 9). Schloss Dagstuhl-Leibniz-Zentrum für Informatik.

Nielsen, O. A., & Frederiksen, R. D. (2006). Optimisation of timetable-based, stochastic transit assignment models based on MSA. Annals of Operations Research, 144(1), 263-285.

Szeto, W. Y., & Jiang, Y. (2014) Transit route and frequency design: Bi-level modeling and hybrid artificial bee colony algorithm approach. Transportation Research Part B: Methodological, 67, 235-263.

Schmidt, M, & Schöbel, A. (2010). The Complexity of Integrating Routing Decisions in Public Transportation Models. In Proceedings OASIcs, 2757.

Schöbel, A., & Scholl, S. (2005). Line planning with minimal transfers. In 5th Workshop on Algorithmic methods and Models for Optimization of Railways (No. 06901).

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

United Nations. Department of Economic and Social Affairs. Population Division. (2014). World urbanization prospects: The 2014 revision. UN.

Wan, Q. K., & Lo, H. K. (2003). A mixed integer formulation for multiple-route transit network design. Journal of Mathematical Modelling and Algorithms, 2(4), 299-308.

Wang, L., Jia, L. M., Qin, Y., Xu, J., & Mo, W. T. (2011). A two-layer optimization model for high-speed railway line planning. Journal of Zhejiang University SCIENCE A, 12(12), 902-912.

Zhao, F., Ubaka, L. & Gan, A. (2005). Transit network optimization: minimizing transfers and maximizing service coverage with an integrated simulated annealing and tabu search method. Transportation Research Record: Journal of the Transportation Research Board, 1923(1), 180-188.

Zhao, F. & Zeng, X. (2008). Optimization of transit route network, vehicle headways and timetables for large-scale transit networks. European Journal of Operational Research, 186(2), 841-855.

Appendix 5: Parbo & Lam (2015)

Modelling degradable capacity of a transit network

Jens Parbo [a], William H.K. Lam [b]

[a] DTU Transport, Technical University of Denmark Bygningstorvet 116B, 2800 Kgs. Lyngby, DK-Denmark

[b] Department of Civil and Structural Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China

Submitted to *The journal of Public Transportation*

Abstract
In this paper we propose a model to assess the degradable capacity of a transit network. The aim is to cancel individual train runs without violating vehicle capacity constraints. To take the interaction between supply and demand into account, a bi-level modelling framework is developed. The upper level determines the maximum degradable capacity while the lower level is a schedule-based transit assignment model for capturing the responses of passengers to the change of the transit services. For complexity reasons, a heuristic solution algorithm is proposed for solving the bi-level problem. We show in a simple numerical example that a Braess-like paradox might occur when individual transit runs are cancelled leading to a reduction in passengers' generalised travel cost. The proposed model and solution algorithm are applied to the suburban railway network in the Greater Copenhagen area in Denmark.

Keywords
Railway timetabling, passenger behaviour, degradable capacity, large-scale application.

Introduction

In many countries road traffic congestion is becoming an increasingly big problem. People are delayed while waiting in queues (European Commission, 2012), and fuel is wasted. Consequently, attractive alternatives are required to satisfy travel demand, e.g. in the form of public transport.

For public transport to offer a competitive alternative to car traffic, several characteristics need to be fulfilled, e.g. short travel time, high comfort and satisfactory reliability (Ceder, 2007). This paper develops a model to assess the degradable capacity of a transit network. The aim is to assess how many trains can be cancelled without passengers failing to board due to overcrowded vehicles. From the operator's perspective this analysis serves as a tool e.g. for contingency planning. For policy-makers, the assessment can safeguard the network to the future growth in demand by e.g. increasing frequency during certain time periods or by building new infrastructure.

In transportation research, passengers' travel behaviour in transit networks is modelled either by frequency-based or schedule-based transit assignment models. Frequency-based models reveal average line loads, while schedule-based models reveal exact run loads. By applying a schedule-based transit model, the well-known common lines problem no longer occurs. At the same time, frequency-based models limit all transit lines to be operated with even headway, while there is no problem in transit lines having uneven headway when the scheduled-based approach is applied (Nuzzolo et al., 2012). Furthermore, the schedule-based approach allows deriving individual run loads, thereby, enabling the creation of a more demand responsive timetable.

In railway transportation, disruptions occur regularly due to both internal and external factors. A typical approach used to restore the planned schedule is cancelling trains, thus releasing capacity and reducing delay propagation (Hofman et al., 2006). This paper develops a model that could be used in the case of disruptions to assist the operators in selecting which train runs to cancel in order to mitigate the negative impact it has on passengers' travel cost.

Literature review

Assessing degradable capacity is a topic under network vulnerability. According to Wang et al. (2014), who conducted a review of vulnerability studies from between 2001 and 2013, to some extent all network vulnerability studies addressed the following two questions:

- What are the weak (vulnerable) components (links or stations) in the network?
- How vulnerable is the network to failures on these components?

In the present paper, network vulnerability is narrowed down to considering the degradable capacity of a transit network. Capacity degradability has been studied by e.g. Nagurney & Qiang (2007), Chen et al. (2007), Matisziw et al. (2009), Nagurney & Qiang (2009), von Ferber et al. (2012), Chen et al. (2012), Rodríguez-Núñez & García-Palomares (2014) and Cats & Jenelius (2014). While the first vulnerability studies in general were concerned with road networks (e.g. Wong & Yang, 1997; Chen et al., 1999; Yang et al., 2000; Nagurney & Qiang, 2007; Chen et al., 2007; Nagurney & Qiang, 2009; Sumalee et al., 2009; Miandoabchi & Farahani, 2011; Chen et al., 2012), within the last ten years applications for transit networks have been developed as well (e.g. Criado et al., 2006; von Ferber et al., 2012; Cats & Jenelius, 2014; Rodríguez-Núñez & García-Palomares, 2014; Cats & Jenelius, 2015). In the literature, simulation and mathematical modelling approaches have received most attention. Simulation-based approaches were adopted by Chen et al. (2007), Nagurney & Qiang, (2007), Matisziw et al. (2009), Nagurney & Qiang (2009) and von Ferber et al. (2012). Simulation-based approaches in combination with graph theoretical network measures were adopted by Chen et al. (2012), Cats & Jenelius (2014), Rodríguez-Núñez & García-Palomares (2014) and Cats & Jenelius (2015). Combinations (often bi-level programming) were adopted by Wong & Yang (1997), Chen et al. (1999), Yang et al. (2000), Sumalee et al. (2009) and Miandoabchi & Farahani (2011).

Methodologies

In network vulnerability studies, the objective is often to identify bottlenecks or weaker parts of the network. The identification could be enabled by e.g. increasing demand, reducing supply or a combination of these. Given the broad spectrum of different scenarios, planners need to identify a representative subset of these to perform a satisfactory vulnerability test (Matisziw et al., 2009).

Nagurney & Qiang (2007) assessed the ability to cope with disruptions in a road network. They incrementally reduced the capacity limits on all links until capacity constraints were violated. Heydecker et al. (2007) developed a new measure to assess how much of the non-commuter demand that could be accommodated in a road network. Chen et al. (2007) applied network accessibility measures to evaluate the impact of one or more link failures. Travellers' behaviour was subject to changes in route choice, mode choice and destination choice. Matisziw et al. (2009) considered a similar problem instance and applied simulation to assess how degraded capacity impacted car flow and network connectivity. Nagurney & Qiang (2009) tested differences in travel behaviour by applying system-optimal and user-equilibrium behaviour, respectively, when capacity was degraded on the roads in a transport network.

Criado et al. (2006) applied several graph theoretical performance indicators to compare the vulnerability of subway networks around the world. They concluded that star-like network structures were more vulnerable to disruptions than complete graph structures. This was supported by von Ferber et al. (2012) who analysed the impact on network connectivity when removing links or stations in the transit networks of Paris and London, respectively. Higher average node degree implied that networks were less vulnerable to disruptions. The more widespread network in Paris was thus less vulnerable to disruptions than the star-like transit network in London (von Ferber et al., 2012).

Cats & Jenelius (2014) explored the value of providing real-time information to passengers in the case of transit network disruptions. Typically, passengers' travel behaviour is assumed to be based on complete knowledge of the timetable. However, this is an optimistic assumption. Cats & Jenelius (2014) tried to cope with that by letting passengers adapt their travel behaviour en-route based on the disturbances they faced. Considering forced degradation of capacity in the case of reduced economic resources, D'Acierno et al. (2014) adopted two strategies, *Change the Least Possible* and *Change the Framework*. The former is applicable when the initial services are able to satisfy users' needs, while the latter is applicable when services are insufficient. Rodríguez-Núñez & García-Palomares (2014) measured the vulnerability of a transit network by assessing how link/node disruptions impacted travel time and network connectivity. Cats & Jenelius (2015) assessed how enhancing capacity on alternative links in a transit network reduced the impact on passengers' travel time when disruptions occurred. A two-stage approach was applied. First, important links were identified. Second, for each of these links, the increase in capacity required to mitigate the impact of the disruptions was determined.

Objective and contributions

The objective of the present paper is to develop a method to explore the degradable capacity of a transit network. This is done by assessing how many individual transit runs can be cancelled before passengers fail to board the vehicles due to capacity limitations. The practical contribution of the present study is a tool for planners to be used in the case of rolling stock shortage. In that case, the current model will assist planners in selecting which runs to cancel. The academic contribution lies in the level of detail of the transit network vulnerability assessment. To derive passengers' route choice, a schedule-based transit assignment model is applied, thereby explicitly accounting for variations in demand and uneven headways. Based on the O/D-demand, exact vehicle loads can be computed for all transit runs on all arcs. Also, changes in passengers' mode choice and route choice are considered explicitly.

Later, we show the existence of a Braess-like paradox occurring when individual transit runs are cancelled. Finally, the model is verified on the suburban railway network in the Greater Copenhagen area in Denmark. Supporting these contributions, a large literature review on vulnerability studies from 2001 to 2013 showed that only few vulnerability studies considered transit network applications and that among these studies, dynamics in demand were rarely captured (Wang et al., 2014). A limitation of the current study is that vehicle capacity and on-board discomfort are not taken into account in passengers' travel cost. However, by penalising the travel time for heavily used vehicles, we try to approximate how travellers' route choice would actually be if capacity constraints were considered.

Methodology

Analytical formulation – capacity degradability

The mathematical formulation of the upper level problem of the bi-level model is

$$Minimise \sum_{r \in K} Y_r \tag{13}$$

subject to

$$X_{ra} * Y_r \leq VehCap, \forall r \in K, a \in A \tag{14}$$

$$Y_r \in \{0,1\} \tag{15}$$

The objective is to cancel as many runs as possible without violating the in-vehicle capacity constraints for any runs *r* at any track segments (arcs *a*). The variable Y_r is equal to one if run *r* is operated and zero if not. $X_{r,a}$ is the accumulated passenger load on run *r* on arc *a*. Since we assume a uniform fleet of vehicles, the vehicle capacity *VehCap* has no index. *K* is the set of runs and *A* is the set of track segments (arcs).

Analytical formulation – transit assignment

The formulation of the lower level problem is based on the schedule-based path choice model proposed by Nuzzolo et al. (2012).

Path choice

For a certain O/D-pair *od* and target (departure) time τ_{TTi}, the probability $P_\tau^{od,\tau_{TTi}}[\tau_{Di}, s, r]$ of choosing a given path *p* identified by departure time τ_{Di} access stop *s* and run *r* is derived as:

$$P_{\tau}^{od,\tau_{TTi}}[\tau_{Di}, s, r] = P_{\tau}^{od,\tau_{TTi}}[\tau_{Di}] * P_{\tau}^{od,\tau_{TTi}}[s|\tau_{Di}] * P_{\tau}^{od,\tau_{TTi}}[r|s, \tau_{Di}] \quad (16)$$

Where, $P_{\tau}^{od,\tau_{TTi}}[\tau_{Di}]$ is the probability of departing at time τ_{Di} and $P_{\tau}^{od,\tau_{TTi}}[s|\tau_{Di}]$ is the probability of choosing boarding stop s. $P_{\tau}^{od,\tau_{TTi}}[r|s, \tau_{Di}]$ is the probability of selecting run *r*. All three parts are based on a given O/D-pair *od* and target (departure) time τ_{TTi}.

Path loads

The path loads are found by multiplying path choice probabilities by the demand for each O/D-pair $d^{od,\tau}{}_{TTi}$.

$$h_{\tau}^{od,\tau_{TTi}}[\tau_{Di}, s, r] = d^{od,\tau_{TTi}} * P_{\tau}^{od,\tau_{TTi}}[\tau_{Di}, s, r] \quad (17)$$

This yields path loads relative to departure time τ_{Di}, run *r* at arriving stop *s* at time τ.

Run loads at arcs

Run loads on each arc X_{ra} are derived by multiplying the path loads by the incidence matrix $\delta_{r,a}^{od,\tau_{TTi},p}$, which denotes the relationship between paths, runs and arcs.

$$X_{ra} = \sum_{od} \sum_{\tau_{TTi}} \sum_{p} \delta_{r,a}^{od,\tau_{TTi},p} * h_{\tau}^{od,\tau_{TTi}}[\tau_{Di}, s, r], \forall r \in K^{s}, a \in A \quad (18)$$

Solution framework – capacity degradability

The capacity degradability problem in this paper is formulated as a bi-level model. The upper level is a capacity degradability model, where specific runs are cancelled, while the lower level is a transit assignment model revealing passengers' route choice, thus also the occupancy of each individual train run. Since transit assignment calculations are computationally expensive when considering several disruption scenarios for large-scale networks (Matisziw et al., 2009; Chen et al., 2012), a heuristic surrogate model framework is applied. The purpose of the surrogate model is to approximate the transit assignment (lower level) in the upper level problem to speed up the convergence, i.e. limit the number of required transit assignment calculations (Koziel et al., 2011).

Cancelling trains at the run level, allows a selective cancellation scheme, which means that the transit services can be downscaled in order just to meet the demand even in the cases where demand is varying significantly. When cancelling single runs, the headways might end up uneven. However, since demand is rarely uniform over time, a timetable with uneven headways could be the optimal solution for the passengers (Li et al, 2010).

Surrogate model - stepwise approach

The surrogate model works according to the following step-wise approach which is explained in the following sub sections.

I. Run transit assignment.
II. Assess the potential for cancelling runs.
III. Impose run cancellations.
IV. If stopping criteria is met, stop.
V. Otherwise, go to I.

I. Transit assignment

The transit assignment calculation yields passengers' travel behaviour. The output enables the construction of the calculation graph sketched in figure 1. The calculation graph is applied in the subsequent steps II and III of the surrogate model.

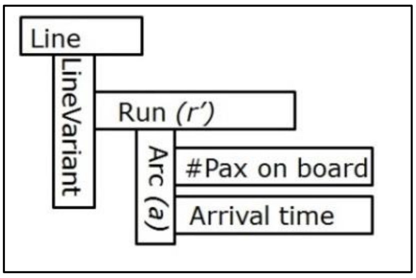


Figure 29 - Calculation graph

II. Potential of cancelling runs

When assessing the degradable capacity of a transit network, it is important to carefully select which runs to cancel. The remaining runs should be able to absorb the extra passenger demand. Examining whether or not passengers from the cancelled runs can be accommodated by the remaining operating runs would ideally require a transit assignment calculation. Since this is too cumbersome in terms of computation time, the following cancellation potential is developed.

Runs subject to cancellation are the runs having the minimum maximum passenger arc load. When fewer passengers are rescheduled, the probability of being able to board the neighbouring attractive runs is higher. The runs are identified in the following way:

1. For each run *r*, store the maximum vehicle arc load ([*RunID; MaxArcLoad*]).
2. Sort runs in ascending order and load into a list *cancellable runs*, abbreviated *cr*. In *cr*, run *1* hence represents the run with the minimum maximum load.

III. Impose run cancellations - algorithmic framework

Based on the *cr*, run cancellations are imposed according to the following procedure:

1. While *cr* is not empty, establish the neighbourhood of run *1*.
2. Derive the absorbing potential of run *1*'s neighbourhood of runs.
3. Cancel run *1* if the absorbing potential allows it. Otherwise, remove run *1* from *cr* and go to 1.
4. Remove runs neighbouring run *r'* from list *cr* and go to 1.

Establish neighbourhood of run r'

The neighbouring runs are the runs passengers on the cancelled run find attractive, thus those they are assumed to board instead of the cancelled run. The neighbourhood is defined as the subsequent runs available between the same stopping pairs. The number of subsequent runs considered in the neighbourhood of attractive alternative runs is limited by a maximum waiting time threshold ρ. When passengers have to wait longer than the threshold ρ, runs are considered unattractive. Figure 2 exhibits the neighbourhood (red runs) of run *r'* on arc *a*. To identify the neighbourhood of a run *r'*, the following stepwise approach is developed.

Identifying neighbouring runs

1. For each arc a traversed by run *r'*, identify all other runs operating and add them to the list of parallel operating runs, called *L*.
2. Sort runs in *L* ascendingly according to their arrival time at arc a.
3. Calculate the difference in arrival time at arc a between run *r'* and run *1* (i.e. the earliest departing run) in *L* as follows
   $t^{a}_{r',1} = arr^{a}_{1} - arr^{a}_{r'}$
   where arr^{a}_{1} arr is the arrival time of run *1* on arc a and $arr^{a}_{r'}$ is the arrival time of run *r'* at arc a. If $t^{a}_{r',1} < 0$, then select the second earliest arriving run in *L* and calculate $t^{a}_{r',2}$.
4. The runs where $0 < t^{a}_{r',r} \leq \rho$ are added to the set of neighbouring runs $N_{r'}$ belonging to run *r'*.

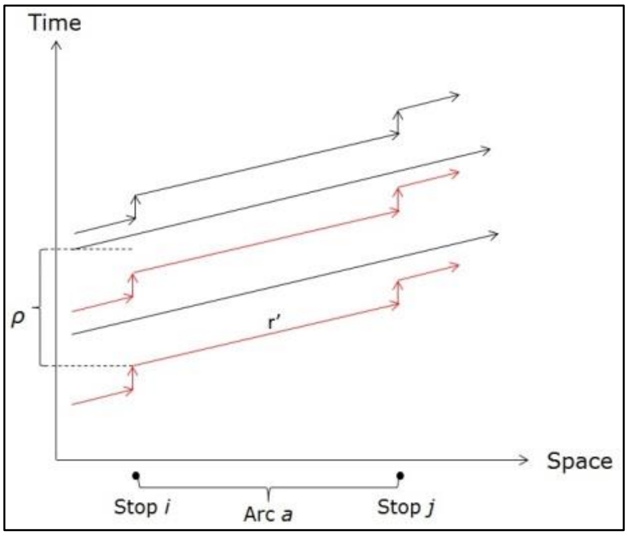


Figure 30 - Neighbourhood of run *r'* on arc *a*

Only trains serving the same arc are included in the neighbourhood, because through- going trains are assumed to be unavailable for the passengers travelling only on that particular arc. The neighbourhood of run *r'* is the set of runs satisfying the time requirement (ρ) on at least one arc where stopping patterns are similar to the run *r'*.

Absorbing potential

Run *r'* is cancelled when the following inequality is satisfied.

$$\sum_{r \in N_{r'}} \left(VehCap - X_{r,a}\right) \geq X_{r',a}, \forall a \in A_{r'} \qquad (19)$$

VehCap is the passenger capacity of a train. X_{ra} is the total number of passengers on run *r* and arc a. $N_{r'}$ is the neighbourhood of runs belonging to run *r'*. $A_{r'}$ is the set of arcs traversed by run *r'*.

Imposing run cancellations
When a run *r'* is cancelled, runs from the neighbourhood of run *r'* are prohibited from being cancelled, i.e. the runs are removed from the list of cancellable runs *cr*. In order for these runs to be subject to cancellation again, another transit assignment needs to be run.

Stopping criterion
The process of cancelling runs continues until *cr* is empty. Then the upper level is terminated and another transit assignment calculation is required before all remaining runs are loaded into the list *cr* again.

IV. Stopping criterion
If no remaining runs are cancellable according to the absorption criteria after a new transit assignment has made all runs eligible to cancellation, the entire model is terminated.

Exploring a Braess-like paradox

The transit network outlined in figure 3 is used as a test network to show how degrading capacity, counterintuitively, may lead to performance improvement. The idea is an extension of the study by Szeto & Jiang (2014), who proved the occurrence of a Braess-like paradox when reducing average line frequency. The Braess paradox was first proven on a road network by Braess (1968). In this paper we explore whether a Braess-like paradox can occur when individual runs are cancelled.

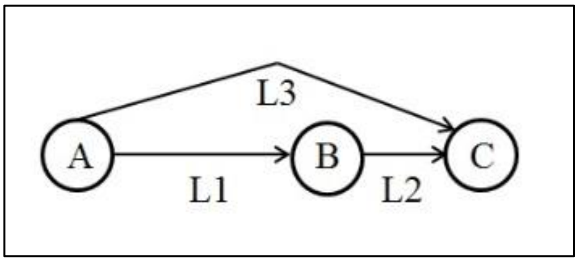


Figure 31 - Test network

In the test network, three lines are operated. Table 1 outlines the paths, headways and travel times for lines *L1*, *L2* and *L3,* respectively. The passenger capacity of each transit run is 120. Between stations *A* and *C*, two paths are available, either *L3* or a combination of *L1* and *L2*. Between stations *B* and *C*, *L2* is the only option.

Table 8 - Line data

Line number	Path	Travel time	Transit runs per hour	Headway (min.)	St. dev.
L1	A->B	9	5	10, 14, 12, 10, 11	1.5
L2	B->C	3	6	28, 17	5.5
L3	A->C	14	3	9, 12, 7, 8, 8, 13	2.2

To reflect the impedance when travelling, passengers' generalised travel cost is used. The generalised travel cost consists of in-vehicle travel time, transfer time, boarding waiting time and an in-vehicle congestion cost.

While the former three depends on the timetable, the latter is a function of the number of passengers on-board a specific transit run. Each component, *Att*, is multiplied by a weight reflecting passengers' perception of the component (table 2) and the number of passengers.

$$GTC_{O,D} = \sum Att * Pax_a * \beta_{att} \qquad (20)$$

Table 9 – Attribute weights

In-vehicle time	1
Waiting & transfer time	2
In-vehicle congestion cost	2

In-vehicle congestion cost reflects passengers' discomfort, and is derived as follows.

$$Cong_r = (\max_r(load_r, 0.75 * VehCap) - (0.75 * VehCap)) * \frac{1}{15} \qquad (21)$$

Where *load* is the number of passengers on run *r* and *VehCap* is the vehicle capacity, set to 120. Passengers only experience discomfort when the occupation rate is above 75%.

Timetable and demand

To show that a Braess-like paradox can be observed when individual transit runs are cancelled from a transit network, a timetable is built for the 3-node test network. The timetable for one hour is exhibited in a time-space diagram in figure 4, where the runs (*L1(3)* and *L1(5)*) to be deleted are highlighted (dashed lines). Total passenger demand is 360 passengers per hour for both O-D pairs (*A->B* and *A->C*). Passengers are launched uniformly every minute. The idea is to explore how passengers' generalised travel cost changes when individual runs are cancelled.

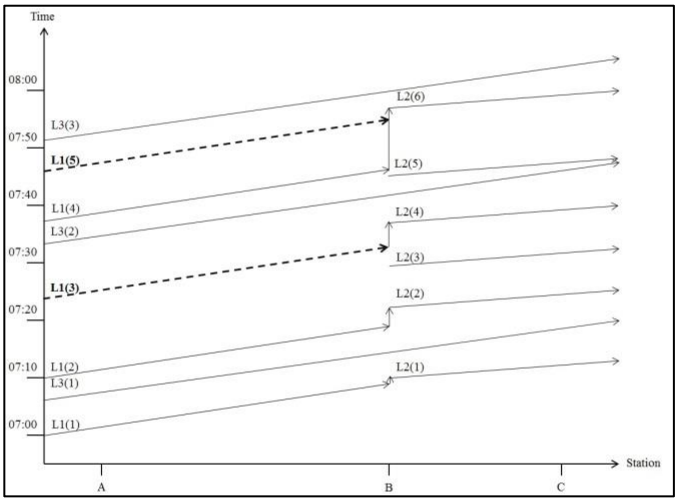


Figure 32 - Timetable (deleted runs highlighted)

Results

Table 3 exhibits the results in three similar tables where generalised travel cost is separated into four different components according to equation 8. The extra waiting time experienced by passengers who fail to board is also added to the table.

Table 10 - Passengers' generalised travel cost

Original				
	L1	L2	L3	Total
In-vehicle time	1782	1512	**1764**	5058
Boarding time	1866	3012	**966**	5844
Transfer time	**1608**	0	0	1608
In-vehicle congestion cost	0	**1036.8**	0	1036.8
Failed-to-board passengers	0	840	0	840
Total	5256	6400.8	2730	**14 386.8**
L1(3) cancelled				
In-vehicle time	1242	1296	**2604**	**5142**
Boarding time	930	3012	**2766**	**6708**
Transfer time	**936**	0	0	936
In-vehicle congestion cost	0	**556.8**	480	1036.8
Failed-to-board passengers	0	600	480	1080
Total	**3378**	**5464.8**	6330	**14 902.8**
L1(3) & L1(5) cancelled				
In-vehicle time	756	1224	**3360**	**5340**
Boarding time	444	3012	**3792**	**7248**
Transfer time	**612**	0	0	612
In-vehicle congestion cost	0	**153.6**	480	633.6
Failed-to-board passengers	0	360	480	840
Total	**1812**	**4749.6**	8112	**14 673.6**

When *L1(3)* and *L1(5)* are cancelled, *L2* experiences a reduction in passenger load, which is reflected in the total cost. The decrease in patronage on *L2* is caused by fewer runs from *L1* feeding *L2*. Furthermore, the new timetable structure encourages more passengers to choose *L3*. When more passengers choose *L3*, fewer passengers have to transfer. Since in-vehicle travel time is higher on *L3* compared to *L1->L2*, a total increase in in-vehicle travel time is observed. Due to the reduced frequency of vehicles departing from station *A*, passengers' boarding time increases.

In-vehicle congestion cost on *L2* reduces as transit runs on *L1* are cancelled, since *L1* now feeds fewer passengers to *L2*. For the same reason, a reduction in the number of passengers failing to board is seen when *L1* feeds fewer passengers to *L2*. Due to the uneven headway and the long waiting time between *L1(2)* and *L3(2)*, several passengers board *L3(2)* leading to in-vehicle congestion on *L3(2)*.

Occurrence of Braess-like paradox

As expected, passengers' overall travel cost increases when *L1(3)* is cancelled. Counterintuitively, it decreases when *L1(5)* is cancelled. The paradox occurs when cancelling runs leads to a change in passengers' route choice, yielding a decrease in in-vehicle congestion and extra boarding waiting time that is larger than the increase in in-vehicle time and boarding waiting time. For this particular case, the runs feeding the line where the in-vehicle congestion is most pronounced are cancelled. Consequently, a decrease in in-vehicle congestion cost and the number of passengers who fail to board is observed. This decrease is larger than the total increase in boarding time, transfer time and in-vehicle time, hence leading to the Braess-like paradox.

In reality, it is hard to outline specific guidelines on the occurrence of the Braess-like paradox as a result of run cancellations. The most certain way to detect the occurrence of the paradox would be to assess run cancellations with a transit assignment calculation.

Case study

The transit network of the Greater Copenhagen area (figure 5) is used as test network; black lines represent the suburban railway network (1679 suburban trains), while the red lines are the remaining transit system (5065 buses, 361 regional trains and 12 intercity trains). Only suburban trains are subject to cancellation. A utility-based approach for modelling the travel choice behaviour is applied. C_{ijpc}, reflects the generalised travel cost on a path *p* from origin station *i* to destination station *j* for passenger group *c* as follows:

$$C_{ijpc} = \sum \beta_c * Att_{ijp} \qquad (22)$$

Here, β_c is the relative weight assigned to attribute *Att*. The attributes considered are in-vehicle time, transfer time, boarding waiting time and walking time.

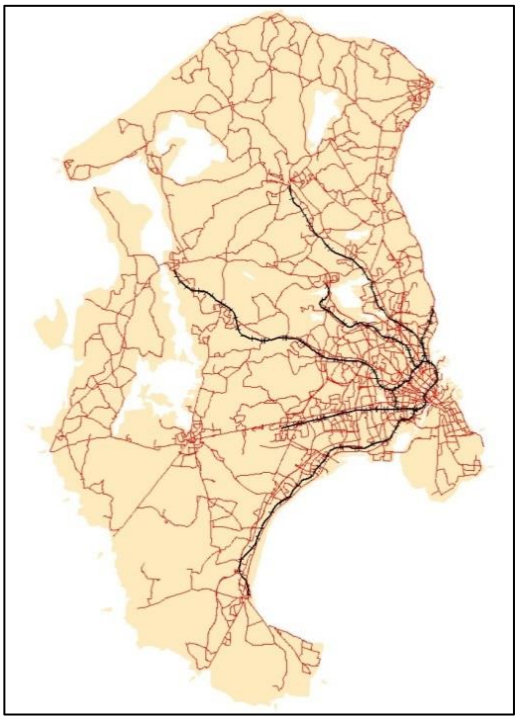


Figure 33 - Transit network (Greater Copenhagen area)

Since in-vehicle congestion cost is not an explicit part of the generalised travel cost, an alternative strategy is applied. For each track segment, where the on-board passenger load exceeds the capacity limit, the travel time on the particular track segment is increased slightly for the particular run. Consequently, some passengers are assumed to find a more attractive path, thus reducing the passenger load on the train run.

Degrading capacity reduces the number of operated suburban trains to 1021 (-39.2 %). This is a severe degradation of the capacity, but among the cancelled runs, some runs barely have any passenger load. Figure 6 outlines, for each line (1: forward, 2: backward), the average passenger load before (grey) and after degrading capacity (black).

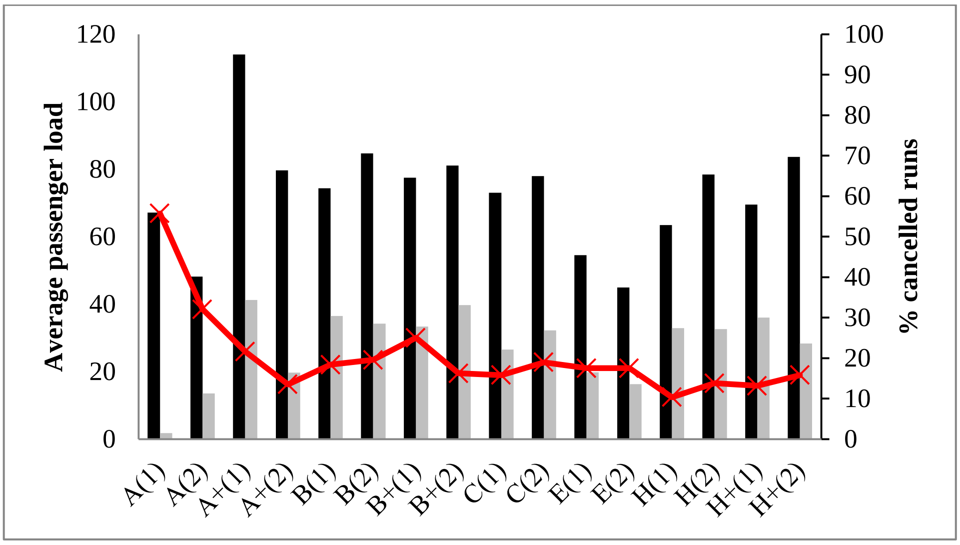


Figure 34 – In-vehicle load (before and after) and percentage of cancelled runs

All lines experience a significant increase in average vehicle load. Although, the model ensures that no passengers are prevented from boarding, the on-board discomfort experienced by the passengers is assumed to increase significantly. The red line indicates the percentage of cancelled runs on each line. There is no clear relation between the share of cancelled runs and the increase in passenger load. The lack of correlation is a result of passengers' adapted mode and route choice. Initially, suburban trains accounted for 71% of all passenger kilometres, while buses accounted for 20%. After degrading capacity, the share of suburban trains fell to 65%, while buses' share increased to 25 %. Intercity trains and regional trains accounted for the remaining part.

Figure 7 exhibits the percentage change in travel cost as a result of the run cancellations. Among the five corridors, passengers along the northernmost corridor and the second southernmost corridor (served by lines B and H, respectively) barely experience any change in travel cost. On the other hand, passengers in the second northernmost corridor (served by line A) experience a large increase in travel cost, which is expected since service on these lines are degraded the most. Additionally, it should be noted that the southernmost corridor experiences a significant increase in travel cost, especially on the northern part of the corridor. This is because some of the A-lines only operate on this part.

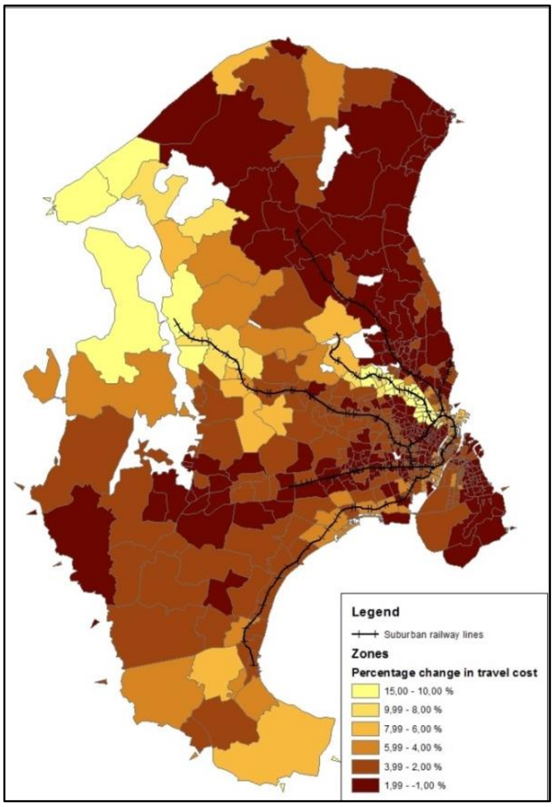


Figure 35 - Change in travel cost at the zone level

Figure 8 displays the percentage change in boarding waiting time. The second northernmost corridor experiences the largest increase in boarding waiting time. Increases are also seen on the northernmost and southernmost corridors. The two remaining corridors do not experience a significant increase. This is explained by fewer runs being cancelled, but also the remaining transit network which is very dense in this area. Examining the northernmost corridor, shows that that while boarding waiting time increases significantly, overall travel cost is barely affected. The cancelled runs are a part of the explanation. However, since several express buses are operated in the corridor, passengers have a competitive alternative, which is also reflected in the negligible change in travel cost. The middle corridor is mostly affected on the western part in terms of increase in travel cost. Here, only few travel alternatives exist. Therefore, the cancelled runs (reflected through increased boarding waiting time) have a larger impact on the travellers compared to the eastern part where several travel alternatives exist. For the southernmost corridor, a large increase in boarding waiting time is experienced, while travel cost only increases slightly. This is because trips from these zones on average are longer. Consequently, boarding waiting time comprises a relatively smaller part of the total travel time, thus not reflected as clearly in the overall travel cost.

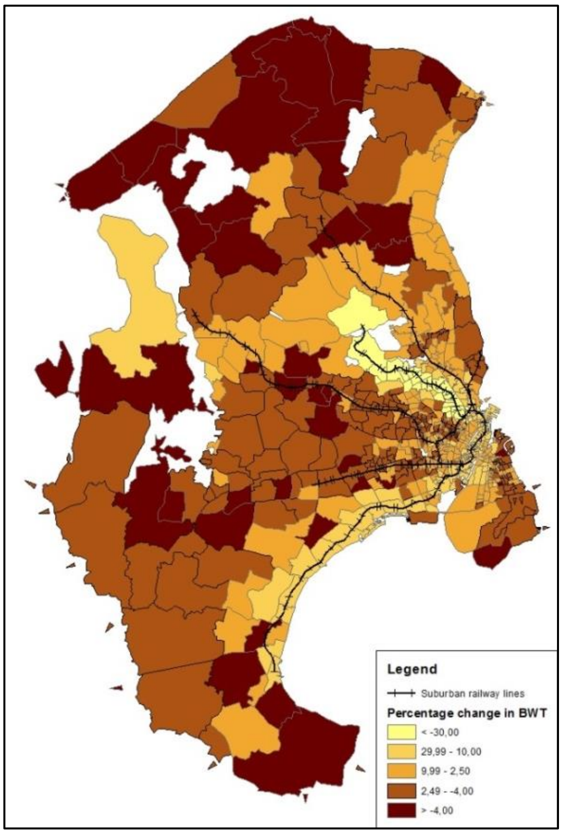


Figure 36 – Percentage change in boarding waiting time at the zone level

Table 4 shows aggregate values for passengers' travel cost when runs are cancelled. Despite the increased vehicle load, passengers' travel cost is barely affected. Transfer waiting time and walking time increase, which is expected since degrading capacity is equivalent to reducing frequency. The increase may be a result of changed mode choice, changed route choice to paths with extra transfers and/or longer transfer walks or maybe the occurrence of a Braess-like paradox. Boarding waiting time improves a bit when aggregating the numbers for all zones (table 4). However, when examining the change locally (figure 8), the picture is less clear. In some areas, boarding waiting time decreases, probably as a result of route/mode choice changes. This change could also help explaining the small increase in in-vehicle travel time.

Table 11 – Percentage change in KPIs

	IVT	BWT	Transfer time	#Transfers	WalkT	Travel cost
Home-Work	0.39	0.14	3.17	0.07	1.07	0.64
Work-Work	0.52	-1.06	4.44	0.18	1.22	0.53
Leisure	0.44	-0.05	4.18	0.08	0.50	0.77
Total	0.45	-0.36	**3.89**	0.11	**0.96**	**0.55**

Conclusions and future work

In this paper, a bi-level model was proposed to investigate the capacity degradability of a transit network, where both supply and demand were modelled at a disaggregate level. A heuristic solution algorithm was developed for solving the capacity degradability problems in practice. The proposed model was applied to the public transport network in the Greater Copenhagen area in Denmark for demonstration purposes. It was that this particular transit network was very resilient towards run cancellations on the suburban train network since a cancellation of 4 out of 10 runs on average only had a minor impact on the overall travel cost. However, when examining the consequences in a local area at a disaggregate level, it is seen that along some of the railway corridors, passengers' travel cost increases quite remarkably, particularly for those stations with lower passenger demand.

In addition, it was shown in a numerical example that the capacity degradability problem may lead to a Braess-like paradox. When individual transit runs were cancelled, passengers' travel cost was counterintuitively reduced when accounting for the discomfort related to in-vehicle congestion. A topic for future research would be to identify the causality between degrading capacity and reductions in passengers' travel cost, and thereby create an exhaustive set of guidelines on how to avoid the occurrence of the Braess-like paradox when transit runs are cancelled.

Another topic for further study would be an extension of the proposed model which explicitly takes into consideration passengers' inconvenience related to on-board congestion or in-vehicle crowding discomfort in their travel cost. Finally, further applications of the proposed solution algorithm to urban transit networks could be carried out, e.g. by comparing the results to the results obtained when solving an analytical model to optimality.

Acknowledgements

This work is jointly supported by research grants from the Research Grant Council of the Hong Kong Special Administrative Region (Project No. PolyU 152074/14E), and the RISUD of Hong Kong Polytechnic University (Project Nos. 1-ZVBX and 1-ZVBY).

References

D'Acierno, L., Gallo, M., Biggiero, L., & Montella, B. (2014). Replanning public transport services in the case of budget reductions. Urban Transport XX, 138, 77.

Braess, P. D. D. D. (1968). Über ein Paradoxon aus der Verkehrsplanung. *Unternehmensforschung*, *12*, 258-268.

Cats, O., & Jenelius, E. (2014). Dynamic vulnerability analysis of public transport networks: mitigation effects of real-time information. *Networks and Spatial Economics*, *14*(3-4), 435-463.

Cats, O., & Jenelius, E. (2015). Planning for the unexpected: The value of reserve capacity for public transport network robustness. *Transportation Research Part A: Policy and Practice*.

Ceder, A. (2007). *Public transit planning and operation: theory, modeling and practice*. Elsevier, Butterworth-Heinemann.

Chen, B. Y., Lam, W. H., Sumalee, A., Li, Q., & Li, Z. C. (2012). Vulnerability analysis for large-scale and congested road networks with demand uncertainty. *Transportation Research Part A: Policy and Practice*, *46*(3), 501-516.

Chen, A., Yang, C., Kongsomsaksakul, S., & Lee, M. (2007). Network-based accessibility measures for vulnerability analysis of degradable transportation networks. *Networks and Spatial Economics*, *7*(3), 241-256.

Chen, A., Yang, H., Lo, H. K., & Tang, W. H. (1999). A capacity related reliability for transportation networks. *Journal of advanced transportation*, *33*(2), 183-200.

Criado, R., Hernandez-Bermejo, B., & Romance, M. (2007). Efficiency, vulnerability and cost: An overview with applications to subway networks worldwide. *International Journal of Bifurcation and Chaos*, *17*(07), 2289-2301.

European Commission. (2012). Measuring road congestion. *JRC Scientific and policy reports.*

von Ferber, C., Berche, B., Holovatch, T., & Holovatch, Y. (2012). A tale of two cities. *Journal of Transportation Security*, *5*(3), 199-216.

Heydecker, B. G., Lam, W. H., & Zhang, N. (2007). Use of travel demand satisfaction to assess road network reliability. *Transportmetrica*, *3*(2), 139-171.

Hofman, M., Madsen, L., Jespersen Groth, J., Clausen, J., & Larsen, J. (2006). Robustness and recovery in train scheduling-a case study from DSB S-tog a/s. *OASIcs-OpenAccess Series in Informatics* (Vol. 5). Schloss Dagstuhl-Leibniz-Zentrum für Informatik.

Koziel, S., Ciaurri, D. E., & Leifsson, L. (2011). Surrogate-based methods. In *Computational Optimization, Methods and Algorithms* (pp. 33-59). Springer Berlin Heidelberg.

Li, Z. C., Lam, W. H., Wong, S. C., & Sumalee, A. (2010). An activity-based approach for scheduling multimodal transit services. *Transportation*, *37*(5), 751-774.

Matisziw, T. C., Murray, A. T., & Grubesic, T. H. (2009). Exploring the vulnerability of network infrastructure to disruption. *The Annals of Regional Science*, *43*(2), 307-321.

Miandoabchi, E., & Farahani, R. Z. (2011). Optimizing reserve capacity of urban road networks in a discrete network design problem. *Advances in Engineering Software*, *42*(12), 1041-1050.

Nagurney, A., & Qiang, Q. (2009). A relative total cost index for the evaluation of transportation network robustness in the presence of degradable links and alternative travel behavior. *International Transactions in Operational Research*,*16*(1), 49-67.

Nagurney, A., & Qiang, Q. (2007). Robustness of transportation networks subject to degradable links. *EPL (Europhysics Letters)*, *80*(6), 68001.

Nielsen & Frederiksen (2006) "Optimisation of timetable-based, stochastic transit assignment models based on MSA.

Nielsen & Frederiksen (2008) "Large-scale Schedule-based transit assignment- further optimization of the solution algorithms".

Niu, Yi-Feng, Lam, H.K. William & Gao, Ziyou (201x). "Modeling degradable capacity of road networks".

Nuzzolo, A., Crisalli, U., & Rosati, L. (2012). A schedule-based assignment model with explicit capacity constraints for congested transit networks. *Transportation Research Part C: Emerging Technologies*, *20*(1), 16-33.

Rodríguez-Núñez, E., & García-Palomares, J. C. (2014). Measuring the vulnerability of public transport networks. *Journal of transport geography*, *35*, 50-63.

Szeto, W. Y., & Jiang, Y. (2014). Transit route and frequency design: Bi-level modeling and hybrid artificial bee colony algorithm approach. *Transportation Research Part B: Methodological*, *67*, 235-263.

Sumalee, A., Luathep, P., Lam, W. H., & Connors, R. D. (2009). Evaluation and design of transport network capacity under demand uncertainty. *Transportation Research Record: Journal of the Transportation Research Board*, *2090*(1), 17-28.

Wang, Z., Chan, A. P., Yuan, J., Xia, B., Skitmore, M., & Li, Q. (2014). Recent Advances in Modeling the Vulnerability of Transportation Networks. *Journal of Infrastructure Systems*.

Wong, S. C., & Yang, H. (1997). Reserve capacity of a signal-controlled road network. *Transportation Research Part B: Methodological*, *31*(5), 397-402.

Yang, H., Bell, M. G., & Meng, Q. (2000). Modeling the capacity and level of service of urban transportation networks. *Transportation Research Part B: Methodological*, *34*(4), 255-275.

Appendix 6: Sels et al. (2015)

Towards a better timetable for Denmark reducing total expected passenger time

Peter Sels[a], Katrine Meisch[b], Tove Møller[b], Jens Parbo[c], Thijs Dewilde[a], Dirk Cattrysse[a] & Pieter Vansteenwegen[a]

[a] KU Leuven, Leuven Mobility Research Centre, CIB Celestijnenlaan 300, 3001 Leuven, Belgium

[b] Banedanmark, Amerika Plads 15, 2100 København, Denmark

[c] DTU Transport, Technical University of Denmark Bygningstorvet 116B, 2800 Kgs. Lyngby, DK-Denmark

Submitted to *Public Transport*

Abstract
With our Periodic Event Scheduling Problem (PESP) based timetabling method we are able to produce a passenger robust timetable for all 88 hourly passenger trains running on tracks managed by the Danish Infrastructure Manager Banedanmark. The objective function of our model is the total expected passenger journey time in practice and is minimised. The result of this is that the produced timetable reduces the expected journey time of all corresponding train passengers together by 2.9% compared to the original timetable defined by Banedanmark. Our simulations show that the average probability of missing a transfer is also reduced from 11.34% to 2.45%. The computation of this timetable takes only 65 minutes. The major innovations of our approach are the addition of a complete objective function to the PESP model and the addition of a particular cycle constraint set that reduces computation times. In this paper, we demonstrate that these combined innovations result in a method that quickly generates cyclic timetables for a train network spanning an entire country and that these timetables also reduce the expected passenger travel time in practice.

Keywords Expected Passenger Time, Integer Linear Programming, Optimal Cyclic Railway Timetabling, Periodic Event Scheduling Problem.

1 Introduction

This paper's topic is the automatic construction of a cyclic, macroscopic railway timetable. The word cyclic means that there is a timetable period, here 1hour, by which every train repeats itself. The word macroscopic means that a standard value for the minimum headway times of 3 minutes is assumed and inside stations, the microscopic headway constraints that arise from the block sections staircase model are not enforced. We also assume that line planning is fixed including the halting pattern for each line. This means that for each train, for each station, only the arrival and departure time are to be determined. In other words, only ride and dwell supplements are to be chosen. Of course, many solutions exist, but these supplements have to be chosen so that the resulting timetable possesses some desirable properties. We previously constructed a Periodic Event Scheduling Problem (PESP) based model which has as objective function: the total expected passenger journey time in practice over all passengers (Sels et al, 2015b). In Dewilde et al 2013), the authors conclude that, unlike to what is the case for some alternative definitions of robustness, this objective function is a practical method to obtain robustness and that the obtained robustness is ideal for passengers. Our objective function integrates and makes a trade-off between efficiency and robustness. It penalises supplements that are so big that they would lower efficiency too much but also penalise supplements that are so small that robustness would be compromised. In Sels et al (2015b), this MILP model is generated for the set of all 196 hourly trains in Belgium. The main results were that a timetable, automatically generated in about 2 hours, saves about 3.8% of total expected passenger journey time. This timetable also significantly reduced the percentage of missed transfers from 13.9% to 2.6%. To study how generally applicable this model is to practice, we now also test it on the set of all 88 hourly trains using Banedanmark's infrastructure.

2 Timetabling Methodology and Assumptions

Our timetabling approach consists of the basic constraints of the popular PESP model (Seraffini and Ukovich, 1989; Schrijver and Steenbeek, 1993; Nachtigall, 1996; Goverde, 1998a,b; Peeters, 2003; Kroon et al, 2007; Liebchen, 2007; Kroon et al, 2009; Caprara et al, 2011; Sparing et al, 2013) using a standard event activity network. We impose its classic constraints enforcing minimal ride times and minimal dwell times. As described in detail in Sels et al (2011), we automatically construct all potential transfers. By this, we mean that if two trains stop in the same station, a transfer edge will be added between the arrival time of the feeder train and the departure time of the target train. Currently, a minimum of 3 minutes is assumed for each transfer. Headway edges and the respective minimum headway time constraints are also automatically constructed between entry times of each pair of trains that enter the same infrastructure resource and similarly also between all pairs of exit times. For single track sections, between each leaving and each entering train, a similar headway time constraint is imposed. The headway minimum time assumed on this macroscopic level is 3 minutes. This summarises all hard constraints in our model. For more details, we refer to Sels et al (2015b), where all these mandatory constraints and some supplementary ones that are merely intended to speed up computation are discussed. We will also only give a qualitative description of our objective function here, as the main focus of this paper is the application of our timetabling model on the Danish train network. As derived formally in detail in Sels et al (2013b) and Sels et al (2013a), our objective function consists of the sum of the expected passenger time for each edge (action) in the event activity graph *G(V;E)* that corresponds to a passenger activity. So, for each ride, dwell and transfer edge we model an expected passenger time. We express this expected passenger time of an edge as a function of its minimum time and its added supplement time. The shape of this function mainly depends on the expected primary delay distribution and consequently, so does the value of the supplement that should be ideally added. The scale of this function depends on the number of passengers involved. This indicates the relative importance of the expected passenger time of one edge compared to that of another and these are balanced by the objective function.

For the primary delays, as do Schwanhässer (1974); Meng (1991); Ferreira and Higgins (1996); Goverde (1998a); Vansteenwegen and Van Oudheusden (2006); Kroon et al (2006) and Yuan (2006), we assume negative exponential distributions. These distributions have an average (=expected value) that can be set to a certain fixed percentage `a' of the minimum time for that action. This average can in theory be determined by inspecting logs of trains as they are running in the current timetable. This has been described by Goverde and Hansen (2000) and Daamen et al (2009) for the Dutch and by Labermeier (2013) for the Swiss infrastructure. So, for example, if the minimum time of a ride action is 5 minutes from one stop to the next, if one sets `*a*' to 10%, the average primary delay on that ride action is assumed to be 0.5 minutes. By this one parameter, the negative exponential distribution $p(d)$ of the primary delay d is unambiguously defined, as $p(d) = exp(-d/a)/a$. For now, we assume the same value of *`a'* for all ride, dwell and transfer edges, for all trains and for all tracks. The value of `*a*' is typically chosen in the range of 1% to 5% (Goverde, 1998a). Depending on the action type that passengers participate in, the expected passenger time is another type of function of the supplements added to these actions. We now discuss these types of passengers and associate cost functions. For through passengers, experiencing a ride and subsequent dwell action, the expected time, as a function of the added ride and dwell supplements s, as can be seen in the example in figure 1, is almost the function $f(s) = P*s$, with P equal to the number of participating passengers. This is logical, since for whatever supplement is added to a ride or dwell action, the through passengers just have to sit it through. So high values of s are not beneficial to these passengers.

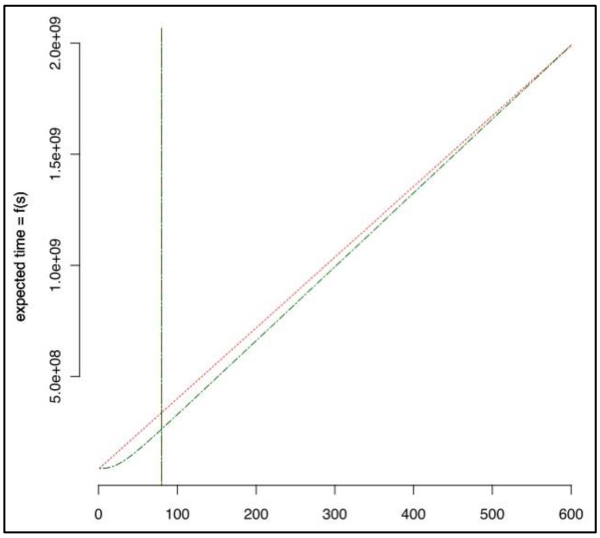


Figure 37 – Through and arriving passenger expected time as a function of the chosen supplement. All time is given in 6 second multiples

At low values of s, the slope of $f(s)$ is a little flatter because small delays occur more often than large delays and so, waiting for the end of s takes a smaller fraction of time s on average than for larger supplements. The larger the supplement, the smaller the fraction that common delay sizes form compared to it. So for larger supplements this secondary 'curving effect' diminishes. The situation is entirely similar for arriving passengers, experiencing a ride plus sink action, and so the cost function for arriving passengers is also similar to the one shown in figure 1. Note that the green vertical line shows that an 8 minute supplement was chosen by the solver. A supplement equal to 0 minutes would be locally optimal, but other hard constraints like headway constraints may forbid this here.

Note that all cost functions in figures 1, 2, 3 and 4 show an actual expected time cost function in green that is used in evaluation and a piecewise linear approximation of it in red which is used in linear optimisation. The green vertical line indicates an example of an actual chosen supplement. Its associated expected passenger time cost can then also easily be read from the graph. In each case, we see that the linearisation error is relatively small.

To departing passengers, experiencing a source plus ride action, it is beneficial when the train they get onto departs as scheduled. This is ensured by providing enough time buffers against primary delay on this train on the sections this train traverses before these departing passengers embark on it. The curve in figure 2 shows indeed that the selection of a larger buffer on the previous sections for this train statistically leads to lower expected delay for departing passengers than a lower buffer. However, it also demonstrates that a supplement larger than 10 minutes does not significantly increase the buffering effect compared to a 10 minute supplement. The green vertical line shows that the MIP solver decided to set the supplement to 8 minutes. This is not the local minimum, 60 minutes, but due to competition with other terms in the objective function this could be a reasonable choice. The value 8 minutes is the basis of the crossing of the two red segments which meet on the green curve so the linearisation error is 0 here.

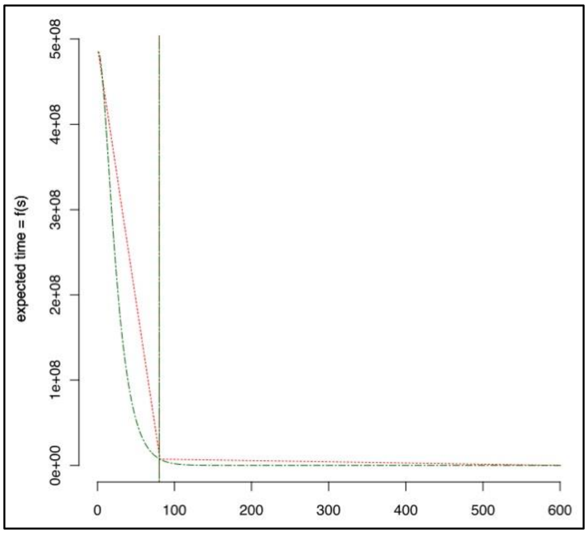


Figure 38 - Departing passenger expected time as a function of the chosen supplement. All time is given in 6 second multiples

For passengers who are changing between trains, experiencing a ride plus transfer action, we model an expected transfer time that depends on the chosen supplement for this transfer, on top of the minimum of 3 minutes. If the supplement is low, the probability that the transfer is missed is high. If the transfer is missed, we conservatively assume a penalty waiting time of the timetable period, here 1 hour. If the supplement is high, the probability of missing the transfer is low, but the transfer passenger will always have to wait until the supplement time has elapsed. The above means that the expected passenger time for a transfer is a U-shaped function of the supplement. An example of a transfer cost curve is given in figure 3. So there is a trade-off and a locally optimal value for the transfer supplement somewhere between 0 and 60 minutes. This supplement range is very broad and naturally very large supplements will rarely be added. Exceptionally, like when a transfer is only taken by very few people, and a small supplement on this transfer would mean a large supplement on an action with more people, a very large supplement on this less important transfer can occur though. The allowed range for supplements is defined as 0 to 60 minutes to avoid infeasibility problems. Note that a transferring passenger can be seen as the combination of both an arriving and a departing passenger and this is rejected in the cost function in figure 3 being the addition of the cost functions of figures 1 and 2. The vertical green line in figure 3 shows that the MIP solver was able to select a supplement equal to 4.5 minutes which minimises the local linearised expected transfer time. This also coincides with the minimum of the green curve.

As for secondary delays, or knock-on delays, our model already contains the graph edges associated to these. Indeed, they are the same edges as the headway edges, temporally separating pairs of trains that use the same infrastructure resource. So for each headway edge, we also add a term in the objective function that represents the knock-on time or secondary delay that passengers on the second train may experience in case the first train is delayed. In our model, as derived in Sels et al (2013a), this time depends on the delay distributions of both trains and on the number of passengers on the second train. Obviously, the total knock-on time is proportional to the number of passengers on the second train. Also, the expected knock-on passenger time forms a decreasing function of the train separating supplement s_{ij}, since the higher the time separation between two trains i and j, the lower the expected knock-on delay.

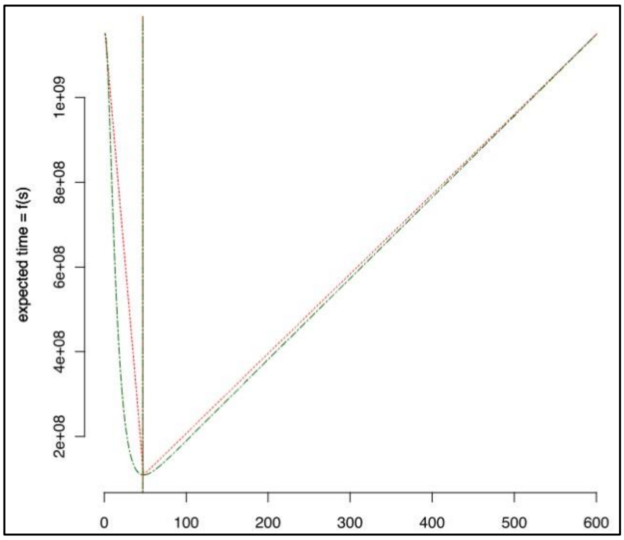


Figure 39 - Transfer passenger expected time as a function of the chosen supplement. All time is given in 6 second multiples

Figure 4 shows an example of a knock-on delay cost function. The horizontal axis shows the supplement between 0 and 60 minutes and on the vertical axis the expected knock-on time is given. Note that our MIP model optimises over all possible train orders. This means that when N trains use a common resource, for all train pairs, cyclically, $N(N - 1)$ knock-on terms are added to the objective function. Knock-on costs are a major determinant for the optimal train orders, but major transfers will also play a role in this.

We could also consider the expected waiting time that passengers experience at their station of departure. This depends on the spreading between alternative trains in the timetable. In this paper we did not add these terms to the objective function since our model developed to estimate this expected time does not scale well yet to networks with many trains (Sels et al, 2015a).

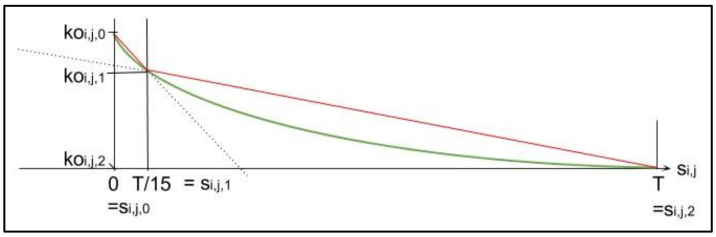


Figure 40 - Shape of expected knock-on delay as a function of the chosen supplement. T is the timetable period, which is 60 minutes here. The vertical axis has no specific scale.

All types of objective function time terms described are seen as objective time. No subjective weights are added. This concludes our qualitative discussion of the objective function of our PESP MILP model representing the timetabling problem. In the next section, we apply our model to the train network of all passenger trains in Denmark and show the results.

3 Application to the Danish Railway System

Our complete method first constructs an event activity graph representing the train service network. Then, we route passengers over this graph to derive local passenger ow numbers for every ride, dwell and transfer action in this graph. We subsequently reschedule trains, deriving ideal arrival and departure times for all trains in all stations. We report results for each of these three phases.

3.1 Constructing the Event Activity Graph

For this project, Banedanmark started from the infrastructure they manage. This is 1956km or 79.5% of the total of 2636km of railway track in Denmark. These tracks are visualised in figure 5. Subsequently, for an 'average' Wednesday in 2013, all trains running on this infrastructure were collected and slightly adapted, so that the timetable became exactly periodic with one hour. One representative hour for this network contains 84 passenger trains and 4 freight trains. Note that we do not schedule the suburban trains on the infrastructure of S-bane. The S-bane operates in the Copenhagen area and is completely independent of the rest of the network, so it has no effect on our case. Some private operators run trains that briefly also use the Banedanmark infrastructure in just three places. These trains have not been modelled but are expected to have little influence on our main results. Freight trains were defined in the input only on sections where Banedanmark knows that there is a capacity bottleneck. For other sections, no freight trains were defined. It is assumed that they can be fitted between the scheduled passenger trains later. We then generated the event activity network that corresponds to this service. This graph contains 88 trains, 264 stations, 3346 vertices and 9918 edges. The number of ride edges is 1541. Table 1 shows more problem instance statistics for this Danish event activity network.

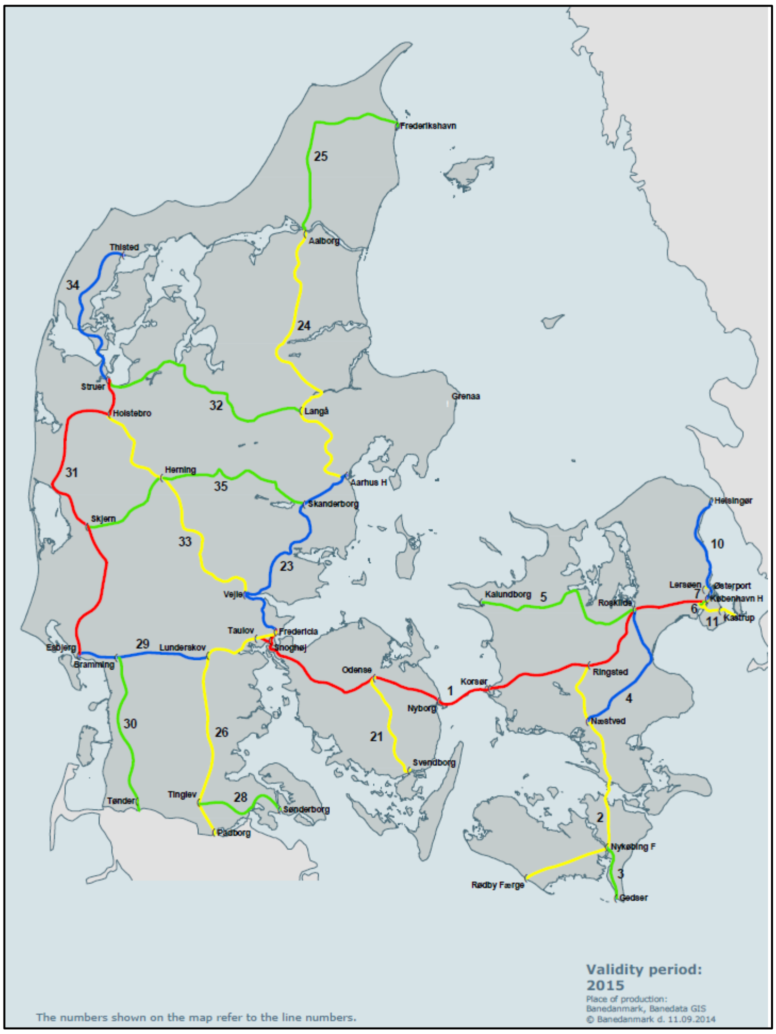


Figure 41 - Danish train infrastructure lines managed by Banedanmark

3.2 Routing: Reflowing

Now that the basic service graph is constructed, we mimic the process where passengers decide what train to take if they go from an origin station (O) to a destination station (D). The number of commuters per day is 394377.

Table 12- Graph and timetable MIP problem instance statistics

# ride edges	**1533**
# dwell edges	**1445**
# turn-around edges	**0**
# knock-on (headway) edges	**13596**
# major transfer edges	**4908**
# model rows	**47335**
# model columns	**32057**
# model non-zero elements	**140516**
# objective function terms for major flows	**16652**
# objective function terms in post-optimisation evaluation	**21522**

The morning peak OD matrix of these commuters is used to route passengers over this train service network, according to the routing algorithm described in Sels et al (2011). This is a modified Dijkstra algorithm implemented in C++. For efficiency, the modified Dijkstra algorithm was parallelised both on the core-level (using open MP, 2013) and the machine-level (using open MPI, 2014). For every OD-pair in the OD matrix, the best routings from O to D are calculated independently. First the modified Dijkstra algorithm is run to find the route with the lowest planned time, based only on the sum of minima for its ride and dwell actions. To avoid too many transfers in a route we penalise the choice of a transfer with 15 minutes. Note that the actual duration of a transfer is not known yet at this point. Next, all edges forming this route are eliminated from the graph and a new route search is performed. This route finding process is repeated until the new found route takes more than 20% more time than the first route found. At this point, it is assumed that no passengers will still opt for such a slower route. Passengers for a specific OD-pair are then distributed over the different OD-routes found, where more are assigned to the shorter routes than to the longer routes. Note that in our method, routing passengers comes before timetabling. This means that arrival and departure times are still unknown and so is their spreading out across one timetable hour. We simplify by assuming that these factors play no role in the passenger distribution over different routes for a given OD-pair (Jolliffe and Hutchingson, 1975). This assumption will be more realistic with good temporal spreading than with bad temporal spreading of alternative trains (Sels et al, 2015a). After the routing phase, which is parallelised for all OD-pairs, a non-parallelised merging phase, for each action (ride, dwell, transfer) on each link of the network is performed. Passenger numbers from the different OD-streams passing along an action are accumulated. We obtain the passenger number for every action (edge) in the event activity graph. Note that the freight trains in our system start in a technical station that passengers do not have access to. The freight trains also do not halt nor stop in passenger stations and so, in our routing algorithm, no passengers can get on or off these trains, as is the case in practice. This means that in our timetabling model, a freight train is treated like a passenger train with no passengers on, so it will be of lower priority during scheduling. If one wants a higher importance, one could assign a virtual number of passengers to each freight train.

The results from the full passenger routing phase, accumulated per track section, are given graphically in figure 6. In this figure, the area of each circle incident to a track section is proportional to the number of people traveling on trains that travel along that track section. It is clear that the set of trains in the area around Copenhagen transport the most passengers. All trains together going from Høje-Taastrup to Hedehusene carry 29215 passengers in the morning peak. This is the maximum ow present in the graph. The second highest passenger flows occur on the tracks from Copenhagen westwards to Odense and back and also from Fredericia North to Aarhus and back. It can be seen that the collected trains for other track sections in the rest of Denmark each transport a lot less passengers.

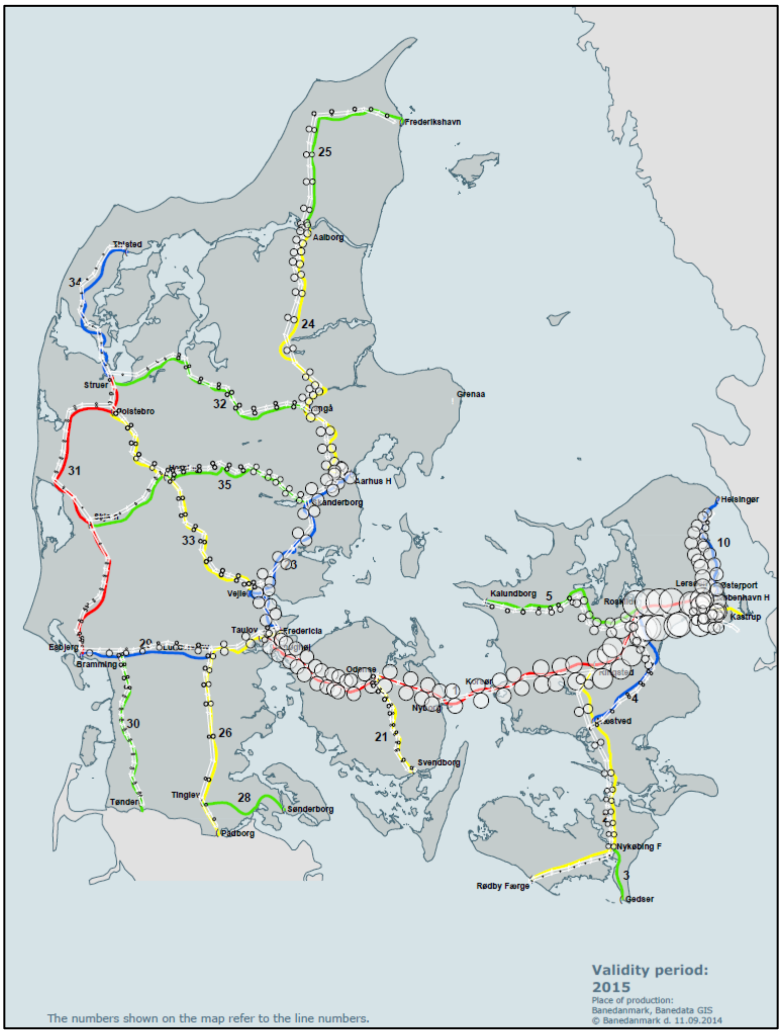


Figure 42 - Passenger flows in Denmark for a typical Wednesday morning peak

3.3 Scheduling: Retiming

Now that we know the number of passengers for each ride, dwell, transfer and knock-on action, we perform timetabling, according to the methodology described in section 2. We use the obtained local passenger numbers as fixed weights in the objective function.

4 Results

With different parameter settings, different MILP timetabling models were constructed. With each model, we construct a different timetable. Our software has a solver independent architecture, using the open source library MILP-logic (Sels, 2012). This way, a simple solver setting and recompilation allows the software to call any solver supported by MILP-logic. Currently, these are CPLEX, Gurobi, XPRESS. In this paper, we restrict ourselves to reporting of results with Gurobi. Each of our timetabling models was tackled by the MILP solver Gurobi version 6.0.0 on an Intel Xeon E31240 3.3GHz processor with 16GB of RAM. When constructing and optimising a MIP model, we noticed that computation times were sensitive to the amount of passenger flows we consider in the objective function. When all streams are considered, computation time becomes excessive so we defined a threshold of number of passengers. Streams with fewer passengers than this threshold are not considered in the objective function. The threshold of 210 passengers per morning peak gave manageable computation times. A further parameter is the required MIP gap. Setting this to 74% resulted in schedules with a lower total expected passenger time than the original schedule. Gap values lower than 74% result in better schedules but computation time also rises. For these parameter values 210 and 74% we get an optimised timetable. This is the timetable we report results for in sections 4.1 and 4.2. Section 4.1 describes that for this optimised timetable, there are no minimum headway time violations. Section 4.2 shows that large time supplements can be and are here assigned to train actions where no passengers are expected. Section 4.3 shows that for various parameter settings, the total passenger time in practice that is expected for the resulting optimised timetables, is always reduced compared to the current timetable.

4.1 No Collisions or Headway Violations

The current and optimised timetables were verified by Banedanmark by visual inspection of space-time graphs per infrastructure line. Some examples of these graphs are given as figures 7, 8, 9 and 10.

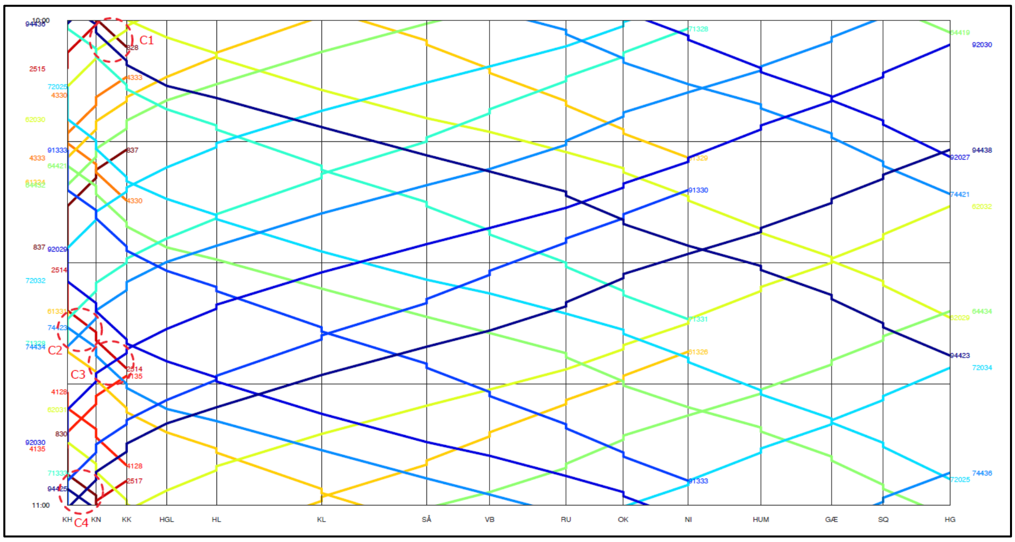


Figure 43 - Space time graph for the original timetable for line 10

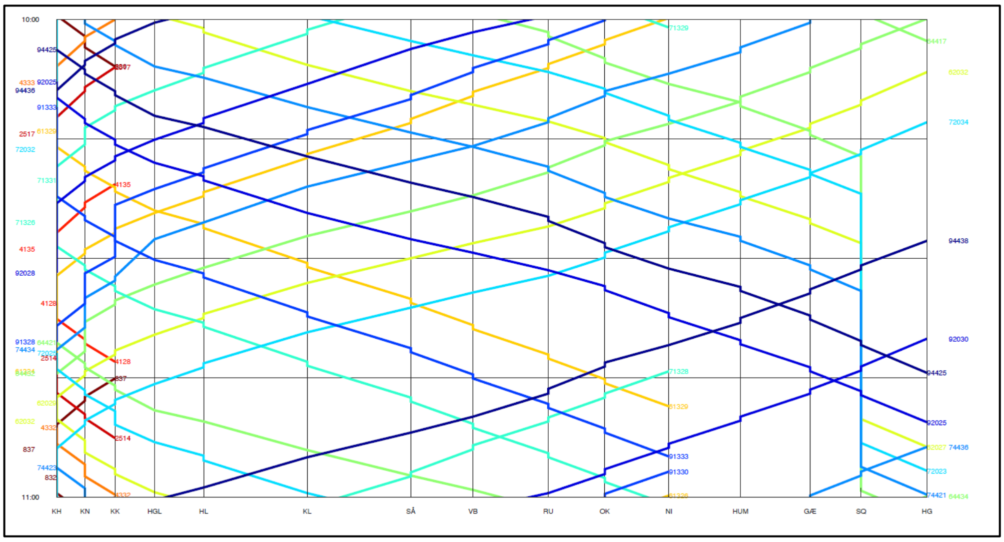


Figure 44 - Space time graph for the optimised timetable for line 10

Figure 7 shows the space-time diagram of trains running on the train infrastructure line 10 between Copenhagen (KH) and Helsingoer (HG) and back for the original timetable. Figure 8 shows the same trains but now for the optimised timetable. In figure 7 it can be seen that the original table generally leaves the required 3 or more minutes between each couple of subsequent trains except for 4 cases between Copenhagen (KH) and Oesterport (KK) and back as indicated by the red dashed circles C1 to C4. Indeed, in circles C1 and C4, train 828 (brown) and train 94423 (dark blue) only have a headway time of 2 instead of 3 minutes between them. The same happens between train 2514 (dark red) and train 74423 (semi-light blue) in circles C2 and C3. In the optimised timetable in figure 8, it can be seen that no such violations of the minimal headway time constraints of 3 minutes occur.

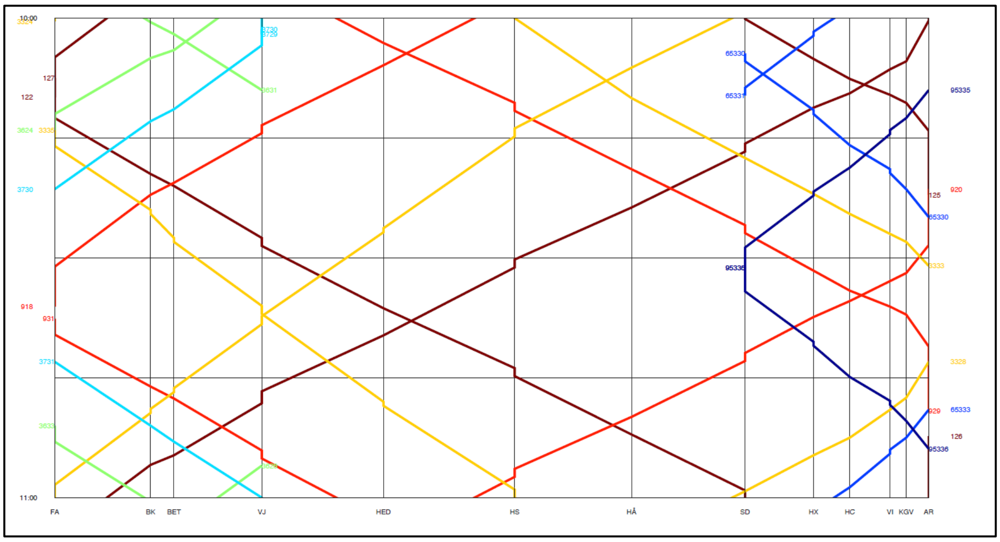


Figure 45 - Space time graph for original timetable for line 23

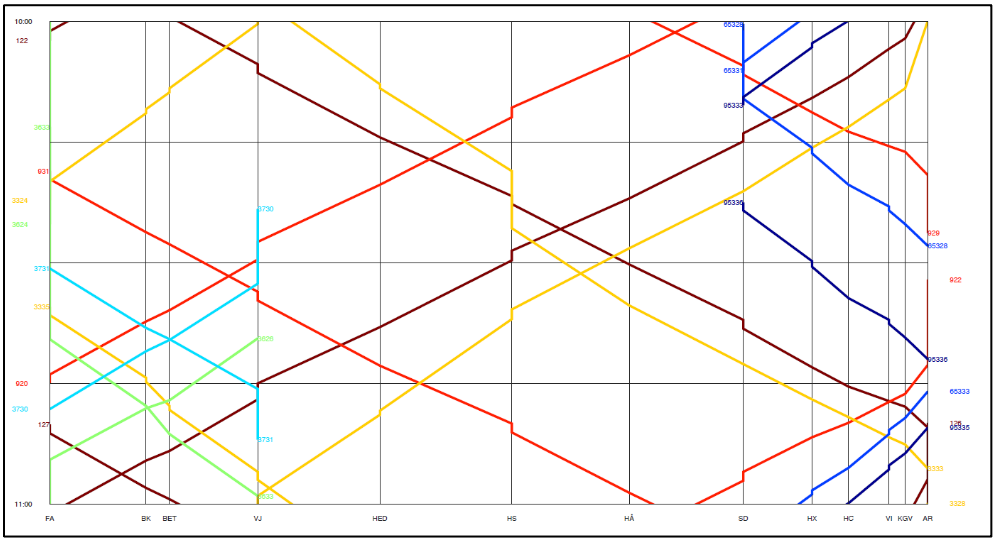


Figure 46 - Space time graph for the optimised timetable for line 23

Figures 9 and 10 show the space-time diagram of trains running on the train infrastructure line 23 between Fredericia (FA) and Aarhus (AR) and back, respectively for the original and the optimised timetable. One can verify that on this line, for both timetables, no single train collision or violation of minimal headway time constraints occurs. For all other infrastructure lines, similar graphs were generated and verified as well and as such Banedanmark declared the optimised timetable as free of headway conflicts.

4.2 Large Dwell Times on Line 10 Explained

Figure 8 shows that, at the station Snekkersten (SQ), 4 trains heading for Helsingoer (HG) are assigned large dwell times. These trains are (64421 (medium green), 72025 (light blue), 62029 (yellow-green) and 74423 (semi-light blue)). This is caused by the fact that our routing phase resulted in no passengers between Snekkersten (SQ) and Helsingoer (HS). This can be seen in figure 6, where no white circle occurs between Snekkersten and Helsingoer. This also means that these dwell times are not penalised in our objective function of our timetabling model. They can become arbitrarily large without having an effect on any passengers indeed.

Furthermore, it should be noted that our current timetable is only ideal for passengers traveling in the morning. Since one usually wants a timetable that is the same for morning and evening, one can express that by supplying an OD matrix that contains both morning and evening OD-pairs together. If then, all ride and dwell actions of all trains will have at least some passengers on them, in both directions, none of these dwell or ride times will stay unaccounted for in the objective function of our timetabling model. As such, all these actions will also have sensible supplements assigned to them.

Also note that in Snekkersten (SQ), in practice, there is not enough platform tracks in the station to allow simultaneous dwelling of 4 trains. Our timetabling model does indeed not take microscopic issues like this into account. Again, when some passengers would be assigned to these dwell actions, shorter dwell times will result and with that the number of simultaneously dwelling trains will most likely be significantly reduced.

4.3 Reduced Expected Passenger Time

By construction, our optimised timetables contain no single violation of hard (minimum run time, minimum dwell time, minimum headway time) constraints. For headway times this was illustrated in the previous sections graphically. In this section, we show that the optimised timetable also results in lower expected passenger time in practice than the original timetable. The relevant results are shown in table 2. Each of our timetabling models was tackled by the MILP solver Gurobi version 6.0.0 on an Intel Xeon E31240 3.3GHz processor with 16GB of RAM. Results for the different optimisations and their respective input parameter values are ordered from less to more demanding from top to bottom. By more demanding, we mean that either the required MILP gap (column 3) is lower or the number of transfers considered in the optimisation is higher or a combination of both. The transfer threshold (column 2) is the number of people that are required as minimum for a transfer to be considered in the optimisation. Column 6 shows the reduction in percent from original to optimised timetable of the expected time as evaluated over all streams, also the ones with fewer people than the threshold value. Column 7 shows the missed transfer probability in the original timetable as simulated over all streams and column 8 shows the same for the optimised timetable. Column 9 shows the reduction in percent of the planned ride and dwell supplements from the current to the optimised timetable. We see that setting the transfer threshold to 420 makes that the solver spends a lot of time (19421 and 62417 seconds) before it finds a solution with an optimality gap below the required one. When the transfer threshold is lowered to 210 transfer passengers, resulting in more transfers considered in the optimisation, the model seems to become easier for Gurobi. When subsequently also lowering the required gap from 79% to 74% (column 3), timetable solutions are found within 1534 to 3922 seconds (column 5) and corresponding savings of total expected passenger time increase from 0:82% to 2:90% (column 6). Lowering the required gap further to 73% still improves the solution with a total reduction of expected passenger time of 3:16%, however, the computation time then increases significantly to 20726 seconds, being 5.76 hours. To test if lowering the transfer threshold further below 210 reduces computation time, we investigate whether a threshold of 195 combined with a not so demanding required gap of 76% gives us a good timetable quickly. The last line of table 2 shows that after 101000 seconds, no acceptable timetable solution was found yet, since the solver is still at a gap of 76.8%. So the value 210 as a transfer threshold somehow seems a good trade-off between giving Gurobi enough information about a good timetable and not too many terms in the objective function.

Table 13 – Results for different timetable optimisations of all 88 hourly Danish trains. Req- = required, obt. = obtained, exp. time = expected passenger time, red. = reduction, eval. = evaluation, orig.tt = original timetable, rd. + dw. t = ride + dwell train time.

1	2	3	4	5	6	7	8	9
a	Transfer threshold	Gap req.	Gap obt.	Solver time	Exp. time red. eval.	Missed orig. tt	Transfers: opt. tt	Planned rd. +dw. t red.
(%)		(%)	(%)	(%)	(%)	(%)	(%)	(%)
2	420	75	74.92	19421	1.67	11.34	2.07	-4.88
2	420	74	73.63	62417	1.96	11.34	5.21	-3.95
2	210	79	78.07	1534	0.82	11.34	2.83	-7.77
2	210	77	76.62	2436	1.59	11.34	3.20	-4.89
2	210	75	74.96	2924	2.45	11.34	1.12	-3.08
2	210	74	73.83	**3922**	**2.90**	11.34	**2.45**	-2.53
2	210	73	72.96	20726	3.16	11.34	2.07	-2.05
2	195	76	≥76.8	≥101000				

For the timetable that reduces the passenger time by 2.90% compared to the original one, we show the expected passenger time and its components graphically in figure 11. This figure stacks expected time components on top of each other to reveal the total expected time for all passenger streams, large and small, for this optimised timetable. Expected time components can indeed be added together since all of them are expressed in the same units: (tenths of) passenger minutes. In figure 11, the left bar indicates the original timetable (orig) and the right bar indicates the optimised timetable (opt). The vertical dimension represents expected passenger time, also for its constituent components: ride (blue), dwell (yellow), transfer (orange), knock-on (purple). For dwell and transfer time, all ride time of the ride action preceding it, is convoluted with it, which is what the blue shading refers to. On the left of each bar, the percentages (orig. *m* and opt. *m*) indicate the ratio of the total expected passenger time part that can be seen as the consequence of the planned minima (m), to its total bar height. Note that this part is equivalent to the planned passenger minimum time. On the right, the percentages (orig. *s* and opt. *s*) indicate the ratio of the total expected passenger time part, that can be seen as the consequence of the planned supplements (s), to its total bar height.

This part is equivalent to the difference of the total expected time minus the total planned passenger minimum time. For each colour, the minima are shown in a darker tone of the colour and the supplements in a lighter tone of the same colour. Figure 11 shows clearly that the obtained reduction of total expected time of 2.9% is caused by the net effect of three main changes. First, the amount of time spent in supplements on ride and dwell actions are significantly lowered from 7.51% to 4.57%. Second, the expected knock-on delay time is reduced from 3.14% to 2.59% of the total expected time. Third, the expected transfer time is increased from 5.75% to 7.26% of the total expected time. In absolute terms, the transfer time increase is smaller than the sum of decreases in expected time spent in ride and dwell supplements and in knock-on events. This means the net result is a reduction in total expected passenger time. We go back to table 2. For the best timetables found, its last column mentions that these possess between 3.08% and 2.05% more train weighted planned ride and dwell time than the original timetable. Even then, the total passenger time is reduced.

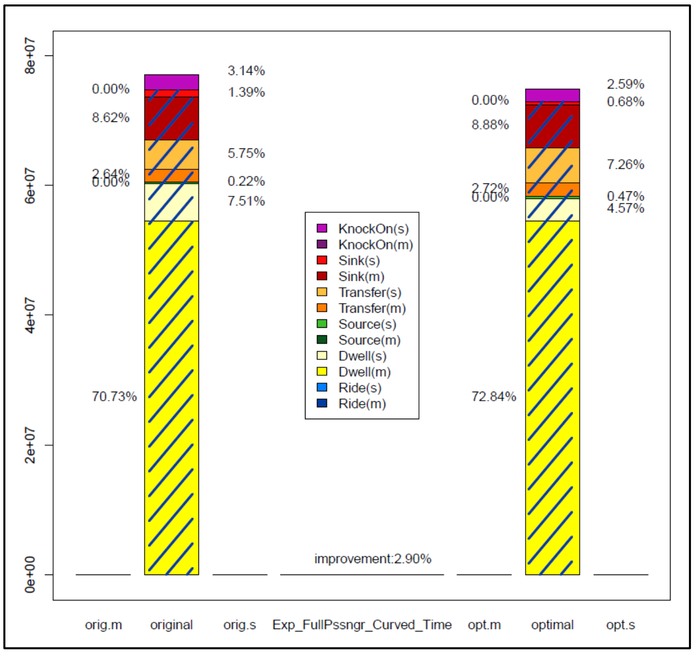


Figure 47 - Reduction of expected passenger time of 2.90% compared to the original timetable

This is possible due to a number of factors. Firstly, our method adds supplements to trains but weighs them by passengers. Secondly, supplements can cause extra robustness, so adding planned time can reduce experienced time in practice. Thirdly, classical manual timetabling uses rules of thumb like assigning a certain percentage of supplement to each train. To avoid knock-on delays, we expect these rules to perform worse than our rule of assigning supplements between each couple of trains sharing an infrastructure resource, even more so since we do this proportionally with the number of passengers on the second train and dependent on the expected delay distributions of both trains. Table 2 also mentions that the expected missed transfer probability for all passenger streams together, both large and small, is 11.34% (column 7) for the original timetable while not more than 2.45% (1.12%, 2.45% and 2.07%, column 8) for our best three timetables. This is clearly a significant improvement that will be appreciated by the railway passengers. These results were obtained by a post optimisation calculation on the obtained timetables, for all passenger streams, small and large, where expected delays are accumulated and resulting in fractions of missed and non-missed transfers. For the original timetable the percentage is always the same, 11.34%, since the value of the transfer threshold plays no role is the missed transfer calculations. Indeed all passenger streams are considered here and not only streams with more than the number of passengers indicated by the transfer threshold.

4.4 Further Verification

Further verification of realistic parameter settings like the value of *`a'* and the value of transfer minima is warranted for fair comparison with the current timetable. Also verification of other timetable quality criteria like the possible preference of some operators to avoid large inserted supplements, even for actions with very few passengers, is required and ongoing.

5 Conclusion

This paper demonstrates that our PESP based method with an objective function representing total expected passenger time in practice, improves the timetable for the whole train network of Banedanmark. Total passenger time in practice can be reduced by 2.9% and the average probability of missing a transfer is reduced from 11.34% to 2.45%. The fact that, after our successful application to the Belgian train network, the application to a second country now delivers satisfying results as well indicates that our approach is quite generally useful.

Thanks to the addition of a particular set of cycle constraints to the PESP model (Sels et al, 2015b), computation time stays limited to 65 minutes. This could lead to huge time savings in the current timetabling practice which, for the biggest part, is still carried out manually. Alternatively, the time spent on manual timetabling now, can instead be used to create more alternative line planning proposals which can be fed to our timetabling system. The line plan leading to the optimised timetable with the lowest total expected passenger time can then be selected. This would further improve passenger service.

6 Further Work

Even though the total expected passenger time of our optimised timetable is lower than the one for the original timetable, the total expected transfer time component of our optimised timetable increased. It would be interesting to see if our model could be adapted so that this expected transfer component is reduced while still also reducing the total expected passenger time. Some degree of temporal spreading of alternative trains between origin and destination is beneficial to reduce the inter-departure waiting time for passenger travelling between these points. Also considering this inter-departure waiting time at the origin and inter-arrival-time at the destination would avoid potential bunching of trains and further generalise our method.

We now produce a timetable that respects headway time minima of 3 minutes everywhere in the network, which is the most common headway minimum value for macroscopic railway models. On a microscopic level, the actually needed headways can be derived from the blocking model (Hansen and Pachl, 2014) and depend on parameters like station infrastructure, train speed and train length. Per train pair, per station, the required minimum headway between these train pairs for that station can be calculated and these values can be substituted for the 3 minute macroscopic headway minima. When our method is used with these more accurate headway minimum values as input, a microscopically feasible timetable will result.

Acknowledgements

We thank Banedanmark for supplying us with the input data and verifying the output as well as Infrabel for letting us use their software on the Danish data.

References

(2013) OpenMP Application Programmer Interface v4.0. URL http://www.openmp.org/mp-documents/OpenMP4.0.0.pdf.

(2014) The OpenMPI API specification for MPI. URL http://www.open-mpi.org/doc/.

Caprara A, Kroon L, Toth P (2011) Optimization Problems in Passenger Railway Systems. Wiley Encyclopedia of Operations Research and Management Science 6:3896-3905.

Daamen W, Goverde R, Hansen I (2009) Non-Discriminatory Automatic Registration of Knock-On Train Delays. Networks and Spatial Economics 9(1):47-61.

Dewilde T, Sels P, Cattrysse D, Vansteenwegen P (2013) Robust Railway Station Planning: An Interaction Between Routing, Timetabling and Platforming. Journal of Railway Transport Planning & Management (Young Railway Operations Research Award 2013 of IAROR) 3:68-77.

Ferreira L, Higgins A (1996) Modeling Reliability of Train Arrival Times. Journal of Transportation Engineering 22(6):414-420.

Goverde R (1998a) Optimal Scheduling of Connections in Railway Systems. Paper presented at the 8th WTCR, Antwerp, Belgium.

Goverde R (1998b) Synchronization Control of Scheduled Train Services to minimize Passenger Waiting Times. Proceedings of 4th TRAIL Annual Congress, TRAIL Research School, Delft, The Netherlands.

Goverde R, Hansen I (2000) TNV-Prepare: Analysis of Dutch Railway Operations Based on Train Detection Data. International conference on computers in railways VII:779-788.

Hansen I, Pachl J (eds) (2014) Railway Timetabling and Operations, 2nd edn. Eurailpress.

Jolliffe JK, Hutchingson TP (1975) A Behavioral Explanation of the Association between Bus and Passenger Arrivals at a Bus Stop. Transportation Science 9:248-282.

Kroon L, Dekker R, Gabor M, Helmrich MR, Vromans M (2006) Stochastic Improvement of Cyclic Railway Timetables. Tech. rep., Erasmus Research Institute of Management (ERIM).

Kroon L, Dekker R, Vromans M (2007) Cyclic Railway Timetabling: A Stochastic Optimization Approach. Algorithmic Methods for Railway Optimization Lecture Notes in Computer Science pp 41-66.

Kroon L, Huisman D, Abbink E, Fioole PJ, Fischetti M, Maroti G, Schrijver A, Ybema R (2009) The New Dutch Timetable: The OR Revolution. Interfaces 39:6-17.

Labermeier H (2013) On the Dynamic of Primary and Secondary Delay. Proceedings of 5th International Seminar on Railway Operations Modelling and Analysis (IAROR): RailCopenhagen2013, May 13-15, Copenhagen, Denmark.

Liebchen C (2007) Periodic Timetable Optimization in Public Transport. Operations Research Proceedings 2006:29-36.

Meng Y (1991) Bemessung von Pufferzeiten in Anschlüssen von Reisezügen. Veröffentlichungen des Verkehrswissenschaftlichen Institutes der Rheinischen-Westfälischen Technischen Hochschule Aachen 46:29-36.

Nachtigall K (1996) Periodic network optimization with different arc frequencies. Discrete Applied Mathematics 69:1-17.

Peeters L (2003) Cyclic Railway Timetable Optimization. PhD thesis, Erasmus Research Institute of Management (ERIM), Rotterdam.

Schrijver A, Steenbeek A (1993) Spoorwegdienstregelingontwikkeling (Timetable Construction). Technical Report, CWI Center for Mathematics and Computer Science, Amsterdam.

Schwanhäusserer W (1974) Die Bemessung von Pufferzeiten im Fahrplangefüge der Eisenbahn. Veröffentlichungen des Verkehrswissenschaftlichen Institutes der Rheinischen-Westfälischen Technischen Hochschule Aachen 20.

Sels P (2012) milp-logic: a C++ MILP Solver Abstraction Layer with a C++ Boolean Modelling Layer on Top. URL http://github.com/PeterSels/milp-logic/

Sels P, Dewilde T, Cattrysse D, Vansteenwegen P (2011) Deriving all Passenger Flows in a Railway Network from Ticket Sales Data. Proceedings of 4th International Seminar on Railway Operations Modelling and Analysis (IAROR): RailRome2011, February 16-18, Rome, Italy.

Sels P, Dewilde T, Cattrysse D, Vansteenwegen P (2013a) A Passenger Knock- On Delay Model for Timetable Optimisation. In: Proceedings of the 3rd International Conference on Models and Technologies for Intelligent Transport Systems (MT-ITS 2013), December 2-4, Dresden, Germany., pp 1-10.

Sels P, Dewilde T, Cattrysse D, Vansteenwegen P (2013b) Expected Passenger Travel Time as Objective Function for Train Schedule Optimization. Proceedings of 5th International Seminar on Railway Operations Modelling and Analysis (IAROR): RailCopenhagen2013, May 13-15, Copenhagen, Denmark.

Sels P, Dewilde T, Cattrysse D, Vansteenwegen P (2015a) Optimal Temporal Spreading of Alternative Trains in order to Minimise Passenger Travel Time in Practice. Proceedings of 6th International Seminar on Railway Operations Modelling and Analysis (IAROR): RailTokyo2015, March 23-26, Tokyo, Japan.

Sels P, Dewilde T, Cattrysse D, Vansteenwegen P (2015b) Reducing the Passenger Travel Time in Practice by the Automated Construction of a Robust Railway Timetable. Submitted to Transportation Research Part B.

Seraffni P, Ukovich W (1989) A Mathematical Model for Periodic Scheduling Problems. SIAM Journal on Discrete Mathematics 2:550-581.

Sparing D, Goverde R, Hansen I (2013) An Optimization Model for Simultaneous Periodic Timetable Generation and Stability Analysis. Proceedings of 5th International Seminar on Railway Operations Modelling and Analysis (IAROR): RailCopenhagen2013, May 13-15, Copenhagen, Denmark.

Vansteenwegen P, Van Oudheusden D (2006) Developing Railway Timetables that Guarantee a Better Service. European Journal of Operational Research 173(1)(1):337-350.

Yuan J (2006) Stochastic modelling of train delays and delay propagation in stations. PhD thesis, Technical University of Delft.

Appendix 7: Case studies

In this section a description of the public assignment model used for the case studies in the present PhD study is provided. Apart from the general introduction made in this appendix, network characteristics that are important for a specific paper (e.g. frequency, stopping patterns, passengers' path choice between selected stops and overlapping routes) can be found in the papers.

Public assignment model

The public assignment model is schedule-based, which means that every single run is described. In this model a utility-based approach is used to describe travellers' perceived travel costs. The formulation of the utility function reflects the perceived cost of travelling from zone *i* to zone *j* in time interval *t* (i.e. generalised travel cost), and is as follows.

$$\begin{aligned} c_{ijt} = {} & WaitingTime_{ij} * \mu_{WaitingTimeWeight} + WaitInZoneTime_{ij} * \mu_{WaitInZoneTimeWeight} \\ & + WalkTime_{ij} * \mu_{WalkTimeWeight} + ConnectorTime_{ij} * \mu_{ConnectorTimeWeight} \\ & + NumberOfChanges_{ij} * \mu_{ChangePenaltyWeight} + TotalInVehicleTime_{ij} \\ & * \mu_{TotalInVehicleTimeWeight}, \qquad \forall\, t \end{aligned}$$

Here, the first component in each product is the actual value of the attribute and the latter part, μ, is the specific weight factor. The weight factors used can be found in Parbo et al. (2014), Parbo et al. (2015b) and Parbo et al. (2015c). The unit of the *Time* weight factors is in Danish kroner (DKK) per minute, while the change penalty weight is in DKK per change. Travellers' route choice behaviour is intelligent and the travellers are assumed to have complete knowledge of the entire network, thus also its future states. Passengers may thus not always board the first arriving run of the attractive line. It should be noted that there are no capacity constraints on vehicles in the network. The network loading is done by an all-or-nothing assignment, where all passengers are loaded onto the routes that maximise each passenger's utility.

Danish transit network

In figure 1, all transit lines included in the public transport network in Denmark are outlined by red lines. The Danish transit network consists of the following:

- 1047 zones
- 1794 transit lines
- 8373 line variants
- 51 819 runs
- 22 008 stops

Furthermore, O/D-matrices for 10 different time intervals representing a single day are used in order to describe the demand. Within each time interval, passengers are launched 30 times per hour. Time intervals covering one day are outlined in table 1.

Table 14 – Time intervals

Start	End
05:00	06:00
06:00	07:00
07:00	08:00
08:00	09:00
09:00	15:00
15:00	16:00
16:00	17:00
17:00	18:00
18:00	21:00
21:00	05:00

The network outlined in figure 1 is used to test the applicability of the heuristic solution approach for timetable optimisation developed in Parbo et al. (2014).

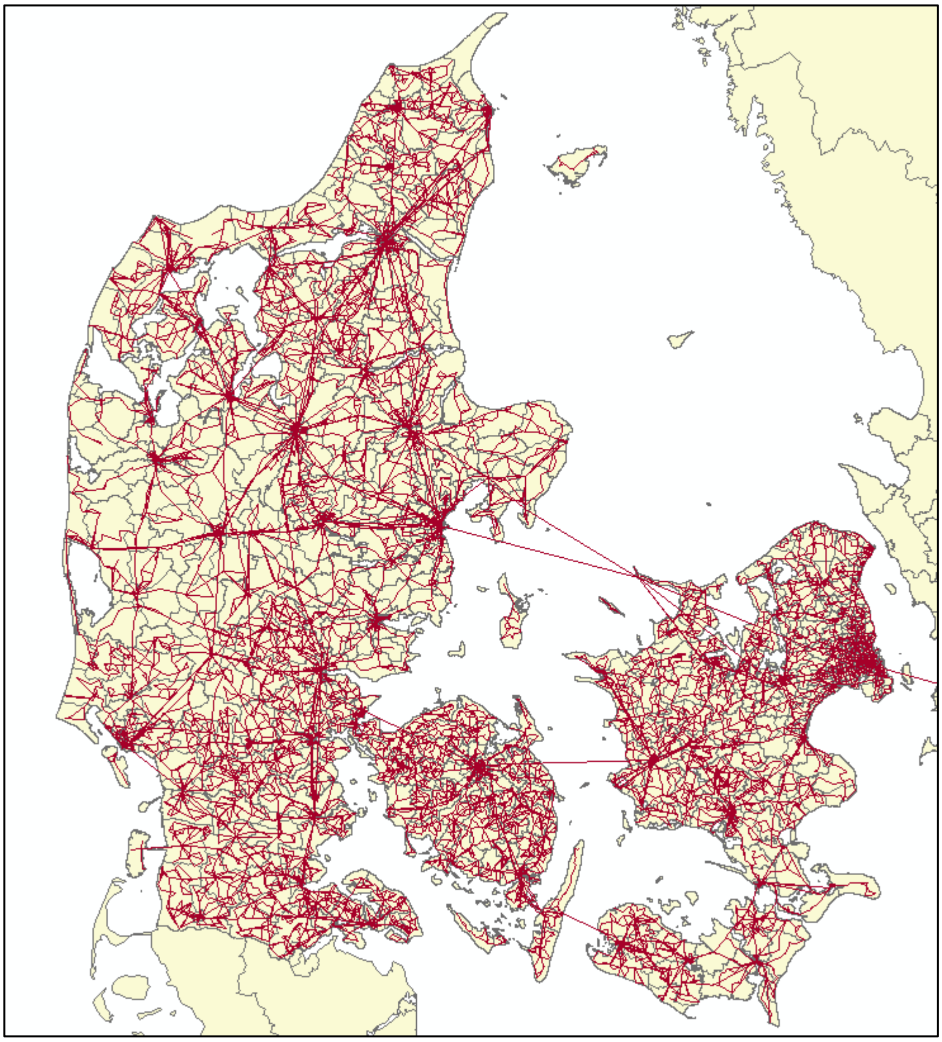


Figure 48 – Transit lines

Suburban railway network (Greater Copenhagen area)
Due to the size of the Danish transit network, the calculation time of the passenger assignment is up to 24 hours. Additionally, the output files are accordingly big, which require lots of data processing in order to build the calculation graph needed to solve the timetable optimisation algorithm developed in Parbo et al. (2014). To be able to run the algorithm faster and still be able to proof the applicability by testing on a real-size network, Parbo et al. (2015bc) and Parbo & Lam (2015) therefore apply their models to a smaller part of the Danish transit network, namely the transit network in the Greater Copenhagen area. This network is outlined in figure 2, where the black lines represent the suburban railway network and the red lines comprise the remaining transit network. This transit network consists of the following:

- 618 zones
- 275 transit lines
- 1176 line variants
- 17 927 runs
- 4704 stops

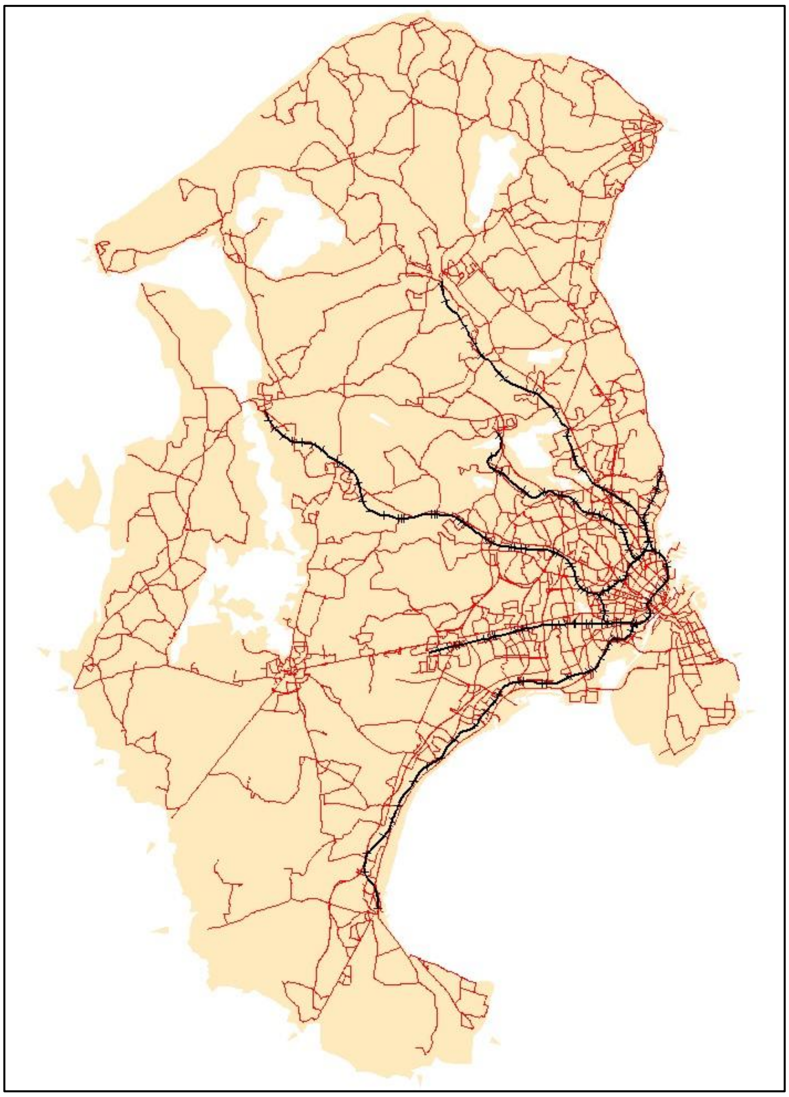


Figure 49 – Suburban railway (black) and remaining transit (red) network in the Greater Copenhagen area

The skip-stop optimisation algorithm developed for railway networks (Parbo et al., 2015b), the line plan configuration optimisation developed for railway networks (Parbo et al., 2015c) and the capacity degradability for transit networks (Parbo & Lam, 2015), are all tested on the suburban railway network in the Greater Copenhagen area. Although the alterations imposed by the three different solution algorithms only are applied to the railway network outlined in figure 3, passengers' route choice are still derived on the transit network comprising all different transit lines from the Greater Copenhagen area as outlined in figure 2. The line diagram of the suburban railway network is outlined in figure 3 and consists of the following:

- 7 lines
- 55 line variants
- 1138 runs
- 84 stops

Figure 50 – Line diagram for the suburban railway network in the Greater Copenhagen area

Appendix 8: Altering network databases

In this appendix, transit network database alterations made as part of the optimisation models are elaborated. In order to ease the understanding, the relevant input and output data (transit assignment) necessary as well as the way in which transit network data is altered are described. Relevant input data is data describing the timetable, while the relevant output data is passengers' transfer patterns, vehicle loads and zone-to-zone travel times.

Timetable data (supply)

Using a schedule-based assignment model necessitates timetable data describing every single run in the public transit network. Figure 1 exhibits all relevant tables describing the timetable. Correlation among the data tables are emphasised by black lines. The content of each data table appears from the white box.

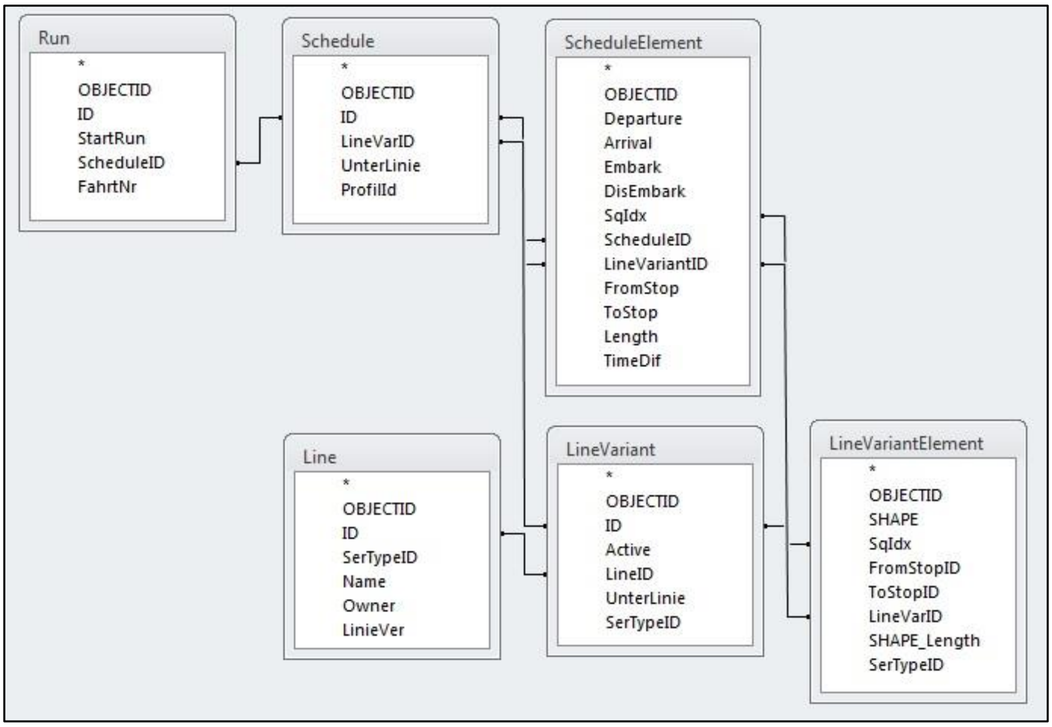


Figure 51 - Timetable data

In all the data tables, *OBJECTID* is a primary key. This key is used to uniquely identify every row of the data table. In the following, the relevant content of each data table is presented.

Line

A line in the transit context can be a bus, train, metro or a ferry line. These different types of public transport modes are identified by their *SerTypeID*.

LineVariant

A line variant is a variant of a certain line. Every line has at least one line variant. A line variant could e.g. be a bus driving from the initial station to the terminal station, while another line variant would be a bus driving the other way

LineVariantElement

For each line variant, a sequence of arcs (elements) traversed is enumerated consecutively by the *SqIdx*. Each arc is defined by the stops that are connected through that arc, *FromStopID* and *ToStopID*, respectively.

Schedule

Every line variant must have at least one schedule, e.g. one during peak hours, one outside peak hours, one during the evening and one for the night. The schedule determines the travel time for the line variant.

Run

The run table comprises information on the departure time (*StartRun*) from the initial station for each schedule belonging to a line variant.

ScheduleElement

The table provides the relevant data on when a particular line variant is supposed to arrive at/depart from a given stop.

Passenger data (demand)

Based on the structure of the timetable and the O/D-demand, passengers' route choice can be derived on the schedule-based network graph. In the following, the output tables relevant for this PhD study are presented along with an explanation of how the data is used in the particular papers (appendices 1, 3, 4 & 5).

Transfer patterns

Passengers' transfer patterns are described by the attributes outlined in figure 2. The information can be divided into three groups.

1. Feeding run data, i.e. *FromStopID*, *FromLVID*, *FLVESqIdx*, *FromRunID* and *ArrTime*.
2. Connecting run data, i.e. *ToStopID*, *ToLVID*, *TLVESqIdx*, *ToRunID* and *DepartTime*.
3. Passenger data, i.e. *CategoryID* and *TrafLoad*.

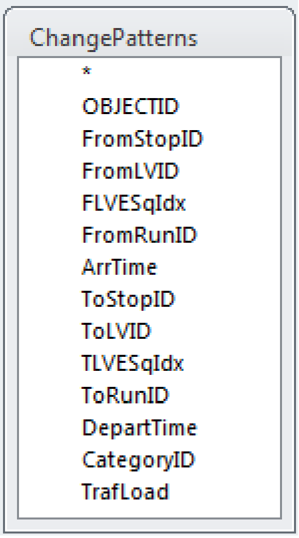


Figure 52 - Transfer patterns

Feeding data consists of the line variant (*FromLVID*), the feeding run (*FromRunID*), the arrival time (*ArrTime*) at the transfer stop (*FromStopID* and *FLVESqIdx*). Connecting data consists of the line variant (*ToLVID*), the connecting run (*ToRunID*), the departure time (*DepartTime*) from the transfer stop (*ToStopID* and *TLVESqIdx*). Passengers' transfer time is derived as the difference between the departure time of the connecting run and the arrival time of the feeding run. Passenger data consists of the trip type (*CategoryID*); commuter, business or leisure tripe, each with their own specific value-of-time parameters, and the number of transferring passengers (*TrafLoad*).

Public transport loads

Vehicle loads are described by the attributes outlined in figure 3. The table contains passenger load (*TrafLoad*) information (divided into trip types - *CategoryID*) for every single arc (*LVEleSqIdx*) traversed by all runs (*RunID*) belonging to each line variant (*LineVarID*). Furthermore, the arrival (*ArrTime*) and departure time (*DepartTime*) is outlined for each stop as well as the number of embarking (*Embarking*) and disembarking (*DisEmbark*) passengers on this particular stop.

Figure 53 - Vehicle loads

Public cost matrix

Zone-to-zone travel times are described by the attributes outlined in figure 4. Total zone-to-zone travel times (*GenCost*) are divided into boarding waiting time (*FirstWaitT*), transfer waiting time (*WaitT*), number of transfers (*NoOfChange*), waiting time at home (*ZoneWaitT*), transfer walking time (*WalkT*), walking time from home (*ZoneConT*) and in-vehicle travel time (*InVehicleT*). Furthermore, O/D-demand (*NoOfTrav*), travel distance (Length) and time interval (*TimeInteID*) is also outlined in the table. The attributes from this table are used as performance indicators when assessing the outcome of the optimisation

Figure 54 - Zone-to-zone travel time

Altering network data

The four papers, Parbo et al. (2014), Parbo et al. (2015b & c) and Parbo & Lam (2015) are all based on a bi-level optimisation method. The lower level is a schedule-based transit assignment model, while the upper level is an optimisation model changing existing network attributes. While the passenger assignment model is run from an ArcGIS-application, the upper level model is coded in C# by the authors of each paper. The algorithmic framework is comprehensively described in each of these papers. However, a description of the alterations of the transit network databases has so far been omitted. Apart from the fact that the algorithms are coded in C# and that SQL is used to alter the network databases through Object Linking and Embedding Databases (OLEDB-connection), there are quite some differences between the four papers. In the following, it is explained how the network attribute alterations are performed for each of the four papers.

Timetable optimisation

In Parbo et al. (2014) the objective is to minimise passengers' transfer waiting time by changing the departure time of selected bus lines. From figure 1, it can be seen that each line variant is linked to its runs through the schedule database. Therefore, relations between the runs of each line variant are created through the schedule database. Based on the optimisation algorithm a certain offset change is found for a subset of all line variants. The optimisation algorithm takes as input the existing timetable as well as passengers' transfer patterns (see figure 2). For each line variant and for each stop visited by runs belonging to that line variant, it is assessed how feeding and connecting passengers would be affected by an offset change. The proposed offset changes are imposed to the run database by executing an SQL-query updating the starting time (*StartRun*) from the initial station for each of the runs of the particular line variant.

Skip-stop optimisation

In Parbo et al. (2015b) the objective is to minimise passengers' travel time by skipping certain stops in an existing railway network. The relevant output data in this regard is the public transport vehicle loads (see figure 3). From this table, the number of embarking and disembarking passengers, respectively, is found for each stop served by each run belonging to a specific line from the railway network. The derivation of the potential benefit of skipping a stop on a particular railway line is explained in Parbo et al. (2015b), where passengers' route choice adaptations also are elaborated. When it is found to be beneficial to skip a stop on a certain line, the table *ScheduleElement* is altered. This is done by inactivating the stop by changing the values of *DisEmbark* and *Embark* from 1 to 0. The arrival time at (*Arrival*) and the departure time from the particular stop (*Departure*) as well as all subsequent stops have to be updated according to the travel time reduction.

Line plan optimisation

In Parbo et al. (2015c) the objective is to minimise passengers' travel time by changing the line plan configuration of a railway network. The relevant output data is passengers' transfer patterns and the passenger loads of the different vehicles. This information is crucial, when the objective is e.g. to minimise the number of required transfers. While, the optimisation potential is described in detail in Parbo et al. (2015c), the network alterations are given extra attention here. Swapping lines can be done in four different ways, called instances. The instances in this context refer to the order in which lines are swapped. Four different instances, of which three are structurally different, are considered. First instance is where the first part of a line is swapped with the first part of another line. Second instance is where the last part of a line is swapped with the last part of another line. Third and fourth instances are where either the first or last part of one line is swapped with the last or first part of another line. Network alterations are executed by adapting the two tables *LineVariantElement* and *ScheduleElement*.

For the *LineVariantElement* of the first instance, the *LineVariantID* of the swapped line variant elements are changed and the *SqIdx* is numbered ascending from 0. Those line variant elements, belonging to the two lines, which are not swapped, are only having their *SqIdx* updated according to the new first part of the line sequence. Similar alterations are made to *ScheduleID* and *SqIdx* of *ScheduleElement*, but here also the *Departure* and *Arrival* are updated.

For the *LineVariantElement* of the second instance, the *LineVariantID* of the swapped line variant elements are changed and the *SqIdx* of the swapped line variant elements are numbered ascending in accordance with the new first (swapped) part of the line variant. Similar alterations are made to *ScheduleID* and *SqIdx* of *ScheduleElement*, but here also the *Departure* and *Arrival* are updated.

For the *LineVariantElement* of the third and fourth instance, the line, which has its first part swapped with the last part of another line, needs to update its *SqIdx* according to number of inserted elements. The line which has its last part swapped needs no alterations. Both the swapped parts are updated, so that the *LineVariantID* and the *SqIdx* are now correct according to the swap of line parts. Again, similar alterations are made to *ScheduleID* and *SqIdx* of *ScheduleElement*, but here also the *Departure* and *Arrival* are updated.

All alterations outlined in this section are performed through several SQL-queries.

Capacity degradability

In Parbo & Lam (2015), the objective is to assess how much the capacity can be degraded in a network, without rejecting passengers from boarding a vehicle and without deteriorating the boarding waiting time below a certain threshold. Selecting which runs that are cancelled is based on the passenger load on each run as explained in Parbo & Lam (2015). The cancellation is performed by an SQL-query, inactivating the selected runs from the *Run* table.

What is not trivial is how we account for in-vehicle congestion in a passenger transit model that does not explicitly account for this in passengers' path choice behaviour. In-vehicle congestion is approximated by penalising arcs where the passenger load is heavy (above vehicle capacity). The penalisation is done by an SQL-query, where the travel time is increased (increasing the *Arrival, Departure* and *TimeDiff* in *ScheduleElement* for the arcs where the capacity limit is exceeded and updating the *Arrival* and *Departure* for all subsequently visited stops accordingly). After penalising all arcs where the capacity constraint is violated, a new transit passenger assignment calculation is run. The intention is to force passengers on the crowded vehicles to choose a less congested path, simply by increasing the travel time on this arc. Sequentially penalising arcs and re-calculating passengers' path choice is done until no capacity constraints are violated on any of the arcs.

DTU Transport performs research and provides education on traffic and transport planning. It advises the Danish Ministry of Transport on infrastructure, economic appraisals, transport policy and road safety and collects data on the transport habits of the population. DTU Transport collaborates with companies on such topics as logistics, public transport and intelligent transport systems.

DTU Transport
Department of Transport
Technical University of Denmark

Bygningstorvet 116B
DK-2800 Kgs. Lyngby
Tel. +45 45 25 65 00
Fax +45 45 93 65 33

www.transport.dtu.dk