TU Delft

Delft University of Technology

Measuring, Predicting and Controlling Disruption Impacts for Urban Public Transport

Yap, Menno

DOI
10.4233/uuid:b48f17ad-41c6-4996-b976-e1b01cc09fec

Publication date
2020

Document Version
Final published version

Citation (APA)
Yap, M. (2020). *Measuring, Predicting and Controlling Disruption Impacts for Urban Public Transport*. [Dissertation (TU Delft), Delft University of Technology]. TRAIL Research School. https://doi.org/10.4233/uuid:b48f17ad-41c6-4996-b976-e1b01cc09fec

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Measuring, Predicting and Controlling Disruption Impacts for Urban Public Transport

Menno Yap

Delft University of Technology, 2020

This thesis is the result of the TRANS-FORM (Smart transfers through unravelling urban form and travel flow dynamics) project funded by NWO grant agreement 438.15.404/298 as part of JPI Urban Europe ERA-NET CoFound Smart Cities and Communities initiative.

Cover design by Sascha Linguard

Measuring, Predicting and Controlling Disruption Impacts for Urban Public Transport

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus, Prof.dr.ir. T.H.J.J. van der Hagen,
chair of the Board for Doctorates
to be defended publicly on
Wednesday 26 February 2020 at 10:00 o'clock

by

Menno Damiën YAP
Master of Science in Transport, Infrastructure & Logistics, Delft University of Technology, the Netherlands
born in 's-Gravenhage, the Netherlands

This dissertation has been approved by the:
Promotors: Prof.dr.ir. S.P. Hoogendoorn and Dr. O. Cats
Copromotor: Dr.ir. N. van Oort.

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Prof.dr.ir. S.P. Hoogendoorn	Delft University of Technology, promotor
Dr. O. Cats	Delft University of Technology, promotor
Dr.ir. N. van Oort	Delft University of Technology, copromotor

Independent members:

Prof.dr.ir. C.G. Chorus	Delft University of Technology
Prof.dr. F. Corman	Eidgenössische Technische Hochschule Zürich, Switzerland
Prof.dr. M. Munizaga	Universidad de Chile, Chile
Dr. J. Törnquist Krasemann	Blekinge Tekniska Högskola, Sweden
Prof.dr.ir. J.W.C. van Lint	Delft University of Technology, reserve member

TRAIL Thesis Series no. T2020/3, the Netherlands Research School TRAIL

TRAIL
P.O. Box 5017
2600 GA Delft
The Netherlands
E-mail: info@rsTRAIL.nl

ISBN: 978-90-5584-261-2

Copyright © 2020 by Menno Yap

All rights reserved. No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without written permission from the author.

Printed in the Netherlands

Change things you can't accept
Accept things you can't change

Acknowledgements

This research was performed as part of the TRANS-FORM (Smart transfers through unravelling urban form and travel flow dynamics) project funded by NWO grant agreement 438.15.404/298 (Chapters 2, 3, 5, 7 and 8**)** and the Swedish Research Council (Formas grant agreement 942 - 2015-2034) (Chapter 8) as part of JPI Urban Europe ERA-NET CoFound Smart Cities and Communities initiative. For the research performed in Chapter 7**,** we acknowledge the SETA project funded by the European Union's Horizon 2020 research and innovation programme.

I would like to thank HTM - Peter Tros, Rien van Leeuwen, Janiek de Kruijff and Sandra Nijënstein in particular - for the valuable, long-term cooperation and data provision in relation to the research performed in Chapters 2, 3, 4, 6, 7 and 8. I thank Shuixian Yu for her inputs for the research in Chapter 3. Besides, I would like to thank Goudappel Coffeng B.V. for the cooperation related to the research performed in Chapter 4. I would also like thank the Washington Metropolitan Area Transit Authority (WMATA) - especially Jordan Holt - for the cooperation and data provision for the research in Chapter 5. For the contribution related to the research performed in Chapter 6, I would like to thank RET and ProRail. At last, thank you Sascha, for all your editorial help and the cover design.

A big thanks to Oded, Niels and Serge for your great support during my research as supervisors, but most importantly as colleagues and as a person. I would like to thank you for your huge flexibility and trust during my Ph.D. research. First, to allow me doing my research in a 50% part-time construction, instead of a much more common 80% or 100% employment. Second, to support me to chase my dreams, when I had the desire to move to London in 2017. Thank you for allowing me to perform the majority of my research abroad, working from home and discussing my work over Skype. Without your flexibility, moving to London would not have been possible. Oded, thank you for all the great discussions we had, and your methodological, systematic, detailed and fast feedback on my research, which has been extremely helpful. At last, I would like to thank Goudappel Coffeng, and Eric Pijnappels in particular, for supporting and stimulating me to start my Ph.D. research in 2016 whilst I was working for Goudappel.

Content

Summary **1**

Samenvatting (summary in Dutch) **7**

CHAPTER 1 - Introduction **13**
1.1 Importance of Reliable Public Transport 13
1.2 State-of-the-Art and Research Gaps 15
1.3 Research Objective, Questions and Scope 23
1.4 Research Contribution 26
1.5 Research Context 29
1.6 Outline 29

I Measuring Passenger Demand and Behaviour during Disruptions

CHAPTER 2 - A Robust Transfer Inference Algorithm for Public Transport Journeys During Disruptions **33**
2.1 Introduction 34
2.2 Methodology 34
2.3 Results 40
2.4 Conclusion 41

CHAPTER 3 - Crowding Valuation in Urban Tram and Bus Transportation based on Smart Card Data **43**
3.1 Introduction 44
3.2 Methodology 45
3.3 Results and Discussion 54
3.4 Conclusions 58

CHAPTER 4 - Improving Predictions of Public Transport Usage during Disturbances Based on Smart Card Data **61**
4.1 Introduction 62
4.2 Methodology 63
4.3 Case Study 70
4.4 Results 73
4.5 Conclusions 78

II Predicting Disruptions and Disruption Impacts

CHAPTER 5 - Predicting Disruptions and their Passenger Delay Impacts for Public Transport Stops **83**
5.1 Introduction 84
5.2 Methodology 87
5.3 Case Study 93
5.4 Results and Discussion 97
5.5 Conclusions 103

CHAPTER 6 - Identification and Quantification of Link Vulnerability in Multi-level Public Transport Networks: A Passenger Perspective **105**
6.1 Introduction 106
6.2 Methodology 108
6.3 Results 114
6.4 Conclusions and Further Research 120

III Towards Controlling Disruption Impacts

CHAPTER 7 - Where Shall We Sync? Clustering Passenger Flows to Identify Urban Public Transport Hubs and their Key Synchronisation Priorities **125**
7.1 Introduction 126
7.2 Methodology 128
7.3 Case Study 137
7.4 Results and Discussion 138
7.5 Conclusions 144

CHAPTER 8 - Quantification and Control of Disruption Propagation in Multi-level Public Transport Networks **147**
8.1 Introduction 148
8.2 Methodology 151
8.3 Case Study 161
8.4 Results and Discussion 165
8.5 Conclusions 168

CHAPTER 9 - Conclusions **171**
9.1 Main Findings 171
9.2 Implications for Practice 175
9.3 Recommendations for Future Research 177

References **181**

About the Author **195**

List of Publications **197**

TRAIL Thesis Series **201**

Summary

Study Relevance
Public transport systems can be subject to disruptions, which have negative impacts on passengers. Disruptions can result in additional in-vehicle time, waiting time, transfer time and extra transfers for passengers. In addition, perceived journey times might increase due to higher crowding levels on public transport services. Public transport disruptions can also result in revenue losses, rescheduling costs, reimbursement costs and fines for the public transport service provider. Although it is thus important to reduce the impact of public transport disruptions, it is particularly challenging to foresee and study disruptions due to their uncertainty and variety. They occur in an environment with complex interactions between decisions made by both passengers and public transport service provider in response to these disruptions, surrounded by various sources of uncertainty in relation to disruption type, location and duration. In this research, we propose a generic, stepwise approach to reduce the passenger impacts of disruptions:

- Step 1: *Measure* current disruption impacts.
- Step 2: *Predict* future disruptions frequencies and impacts.
- Step 3: Develop and evaluate measures aimed to *control* these disruption impacts.

Research Objective and Research Questions
The main research objective of this study is '*to improve methods to measure, predict and control disruption impacts for urban public transport'*. Based on a review of state-of-the-art research methods for measuring, predicting and controlling disruption impacts, different research gaps are identified. This results in the following three research questions:

1. How can we measure and characterise the behavioural and demand response of passengers during planned and unplanned urban public transport disruptions?
2. How can we incorporate disruption frequency and impact predictions in a public transport vulnerability analysis for urban and multi-level public transport networks?
3. How can we predict and control the direct and propagated impacts of disruptions on the urban public transport network in a multi-level network environment?

Scope

The scope of our research is as follows:

- *Focus on urban public transport, in a multi-level public transport network environment.* Our research focuses on disruption impacts for the urban public transport network, consisting of metro, light rail, tram and bus services. Whilst other network levels are not the focus of this research, the multi-level network environment is considered. This implies we consider the role the train network might play both as means to mitigate impacts of urban network disruptions, and as a source for train network disruptions propagating to the urban public transport network level.
- *Address both recurrent and non-recurrent disruptions*. Our research focuses on both the smaller, more frequent recurrent disruptions, and the larger, non-recurrent disruptions, with no, partial or full infrastructure degradation as result. We do not consider extreme events such as natural disasters or terror attacks in this research.
- *Consider both unplanned and planned disruptions*. We study the impact of unplanned disruptions, as well as the impact of planned disruptions, such as planned track maintenance works.

Research Contribution

The following scientific contributions are made in this research (see **Figure I.1**):

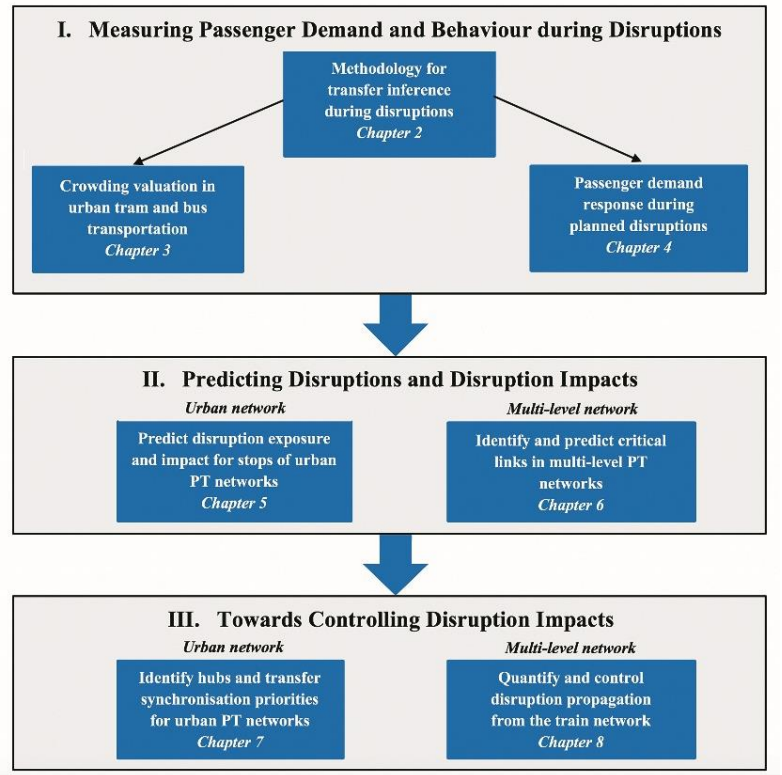


Figure I.1. Research structure and contribution

- Development of an improved transfer inference algorithm for urban public transport journeys during disruptions.
- Estimation of crowding perception multipliers for urban tram and bus journeys based on Revealed Preference.
- Estimation of mode and route choice coefficients for passengers during planned public transport disruptions based on empirical data.

- Development of a methodology to predict disruption frequencies and disruption impacts for urban networks.
- Development of a methodology to identify the links which contribute most to vulnerability of multi-level public transport networks.
- Identification of urban public transport hubs and the key routes serving these hubs to prioritise for public transport synchronisation.
- Development of a methodology to predict the disruption impact propagation from train network disruptions to the urban public transport network.
- Evaluation of the impact of different train rescheduling strategies on the integrated multi-level public transport network.

Main Findings

Based on our research results, we provide answers to the three formulated research questions.

1. How can we measure and characterise the behavioural and demand response of passengers during planned and unplanned urban public transport disruptions?

To measure the passenger impacts of a disruption, passengers' generalised journey costs need to be inferred from empirical data and compared between a disrupted and an undisrupted journey. As a first step, we develop a robust transfer inference algorithm with the ability to infer passenger journeys from individual smart card transactions during disrupted and undisrupted circumstances (**Chapter 2**). It relaxes existing state-of-the-art transfer inference algorithms to incorporate the atypical passenger behaviour that can be observed during a disruption, and considers an alighting a transfer if it satisfies the following temporal, spatial and binary criteria:

- The temporal criterion states that an alighting is a transfer if a passenger boards the first reasonable vehicle arriving at a transfer location, thereby incorporating required transfer walking time, crowding levels and potential denied boarding for this vehicle.
- The spatial criterion indicates that the maximum transfer walking distance should not exceed a certain threshold - for our case study calibrated to 400 Euclidean metres - unless a passenger uses public transport services at another network level or from another service provider as intermediate journey stage during a disruption.
- The binary criterion states that a transfer to the same line is only possible when made to the next vehicle of this same line, to incorporate the effect of operational measures as short-turning or deadheading possibly being applied during a disruption.

A partial validation shows that our algorithm improves inference during disruptions, without compromising inference results during undisrupted circumstances.

A second step when measuring disruption impacts is to infer how passengers perceive the different journey components, especially in relation to crowding (**Chapter 3**). The following results from our estimated discrete choice model with panel effects are entirely based on Revealed Preference route choice observations:

- The average in-vehicle time crowding multiplier for urban trams and buses equals 1.16 when all seats are occupied and no passengers are standing. In case occupancies increase to an average standing density of 3 passengers per m^2, this in-vehicle time multiplier equals 1.34.
- For frequent passengers, these two values equal 1.31 and 1.75, respectively.
- Infrequent passengers do not incorporate crowding in their route choice, due to the lack of prior knowledge about crowding levels.

- Our estimated crowding multipliers are lower than values found in previous Stated Preference experiments. This illustrates the tendency of Stated Preference experiments to overestimate values of coefficients, compared to Revealed Preference based studies.

A third step is to infer passengers' demand response in the event of planned disruptions (**Chapter 4**). For this study purpose, we calibrate route and mode choice parameters of a public transport ridership prediction model based on empirical data from two planned disruptions, which we validate using two different planned disruptions. A three-step rule-based procedure is developed for this. Our results suggest the following:

- Passengers perceive in-vehicle time in replacement buses about 11% more negatively compared to the tram line being replaced.
- Waiting time perception for rail-replacement buses is ≈30% higher than for regular trams or buses, potentially caused by limited facilities at temporal bus stops and by uncertainty about service headways and reliability.
- The new parameter set improves prediction accuracy up to 13% compared to the default parameter set used to predict impacts of structural network changes.

2. How can we incorporate disruption frequency and impact predictions in a public transport vulnerability analysis for urban and multi-level public transport networks?

Predictions of the frequencies and impacts of public transport disruptions are necessary to identify the most critical components of a public transport network in a vulnerability analysis. We develop an improved pre-selection method and an improved full scan method to perform this analysis, which explicitly account for disruption frequencies next to disruption impacts. In **Chapter 6**, we propose a pre-selection method for multi-level public transport networks, which uses expected direct and indirect disruption exposure, as well as the expected number of affected passengers as indicators. Case study results indicate that busy links of the metro / light rail network generally have the largest contribution to vulnerability of the multi-level network, as both exposure and the number of affected passengers are relatively high. Our study results show the relevance of incorporating disruption frequencies in vulnerability analyses, as the list of most critical links differs substantially from a list based only on expected disruption impacts.

We also develop a data-driven full scan methodology to identify the most critical stations of an urban public transport network within reasonable computation times (**Chapter 5**). A supervised learning approach is developed to predict the probability of each disruption type, and to predict passenger delay impacts of each disruption type for each individual station, based on demand predictors, temporal predictors and network topology predictors. In a last step, stations are clustered using unsupervised learning based on their expected contribution to network vulnerability. This improves the transferability of our case study results, as this indicates which types of stations contribute most to network vulnerability. Case study results from the Washington, D.C. metro network show that stations with high train frequencies and high passenger volumes on central trunk sections are most critical, together with transfer stations and terminals.

3. How can we predict and control the direct and propagated impact of disruptions on the urban public transport network in a multi-level network environment?

To control disruption impacts for the urban public transport network, one can apply control to urban public transport services, or apply control to train services to mitigate disruption propagation to the urban network level or to mitigate the impact of urban network disruptions.

Real-time synchronisation of urban public transport services is one type of control measure which can be applied to urban networks. Synchronisation of urban services is currently only optimised for relatively small case study networks, as the optimisation problem becomes difficult to solve within reasonable times for larger networks. Our contribution is the development of a generic, preparatory method to reduce dimensionality of this problem by identifying key locations and routes to prioritise for optimal synchronisation (**Chapter 7**).

- First, using passengers' transfer patterns as input, we apply a density based clustering technique to determine the subset of public transport hubs where synchronisation needs to be prioritised.
- Second, we represent the transfer patterns within each hub using a C-space inspired topological network representation. By using a community detection algorithm, groups of lines are identified for which it is recommended to synchronise them simultaneously.

When applied to the urban public transport network of The Hague, the Netherlands, as case study, results show that 70% of all transfers occurring within identified transfer locations would be captured by considering less than 1% of all transfer locations for synchronisation, thus substantially reducing the complexity of solving the optimal transfer synchronisation problem.

To control disruption propagation from the train to the urban network level, we develop a method which combines a train rescheduling optimisation model and a dynamic public transport assignment model in an iterative procedure (**Chapter 8**). We incorporate the number of transferring passengers from the train to the urban network level in the objective function of the train rescheduling model. Then, we test the impact of this using the dynamic assignment model based on updated train timetables from the optimisation process. The train rescheduling model is then iteratively updated based on train passenger volumes resulting from the assignment model. In our case study, the propagation of passenger delays to the urban network could be reduced by up to 14-27%, without increasing delays for passengers on the train network.

We also illustrate how the train network level can be used as means to reduce the impact of disruptions occurring on the urban level. To this end, a societal cost-benefit analysis framework is established (**Chapter 6**). A case study application to the southern part of the Randstad (the Netherlands) illustrates that mitigation measures applied to the train network can reduce the total disruption impacts resulting from an urban network disruption by 8%. Hence, this illustrates the potential of the multi-level network to mitigate disruption impacts.

Implications for Practice

We formulate several implications of our research for public transport service providers and for public transport authorities in relation to policy-making.

- This research enables public transport authorities and service providers to improve the accuracy of their passenger predictions during planned and unplanned disruptions.
- Methods developed in this research can result in an improved and easier quantification of disruption impacts based on empirical data.
- Our study supports transport policy makers in prioritising the locations and disruption types for which to develop and implement robustness measures.
- Results of our study can improve the real-time control decisions taken by controllers to mitigate disruption impacts, hence reducing the passenger impact of disruptions.

These implications have the potential to result in better project assessments, better decision-making during disruptions, and a higher level of service provided to passengers. This can improve passenger satisfaction and therefore potentially increase public transport ridership.

Recommendations for Future Research

Based on our work, we formulate several recommendations for future research directions:

- To set up a detailed study towards passengers' demand response during planned disruptions, particularly focusing on the extent that passengers use alternative modes of transport, such as ride-hailing services or bicycle-sharing schemes.
- To study passengers' dynamic en-route choice behaviour during unplanned disruptions in detail, thereby incorporating factors such as real-time information provision or flexible working arrangements.
- To investigate the influence of information provision to passengers before and during disruptions on the impact of disruptions.
- To develop a more advanced method to attribute observed passenger delays from Automated Fare Collection (AFC) systems to individual disruptions.
- To compare the performance and computation times of our proposed two-step approach to prioritise locations and routes for public transport synchronisation with new approaches for network-wide synchronisation.

In summary, the objective of the developed methods in this research is to strengthen the passenger perspective when measuring, predicting or controlling disruption impacts. The application of our methods to different case study networks worldwide confirms our methods can be applied in practice. Although results might differ from case to case, our empirical evaluations and model application results suggest that passenger benefits can be realised when applying our approaches. Our research provides generic methods and tools for the public transport industry to apply to their specific public transport network. We recommend a close cooperation between science and the public transport industry, to implement methods and results from this research in the daily business of the public transport sector. This has the potential to further improve the public transport product delivered to passengers.

Samenvatting

Relevantie van betrouwbaar openbaar vervoer
Verstoringen in het openbaar vervoer, zoals een defect voertuig of een aanrijding, kunnen leiden tot extra in-voertuigtijd, wachttijd, overstaptijd en tot extra overstappen voor reizigers. Daarnaast kan ook de ervaren reistijd toenemen als gevolg van toegenomen drukte. Verstoringen in het openbaar vervoer kunnen ook resulteren in kosten voor de vervoerder of vervoersautoriteit, bijvoorbeeld door inkomstenderving, bijsturingskosten, boeteclausules en restitutie van reiskosten. Het is daarom belangrijk om de impact van verstoringen te verminderen. Het is echter moeilijk om grip te krijgen op verstoringen als gevolg van onzekerheid en variatie waar en wanneer deze plaats vinden. Verstoringen vinden plaats in een omgeving met complexe interacties tussen beslissingen van zowel de reiziger als de vervoerder als reactie op verstoringen, omgeven door onzekerheid wat betreft locatie en duur van verschillende soorten verstoringen. In dit onderzoek ontwikkelen we een generieke, stapsgewijze benadering om de impact van verstoringen voor reizigers te verminderen:

- Stap 1: Het *meten* van de huidige verstoringsimpact.
- Stap 2: Het *voorspellen* van de frequentie en impact van toekomstige verstoringen.
- Stap 3: Het ontwikkelen en evalueren van maatregelen om deze verstoringsimpact te *beheersen*.

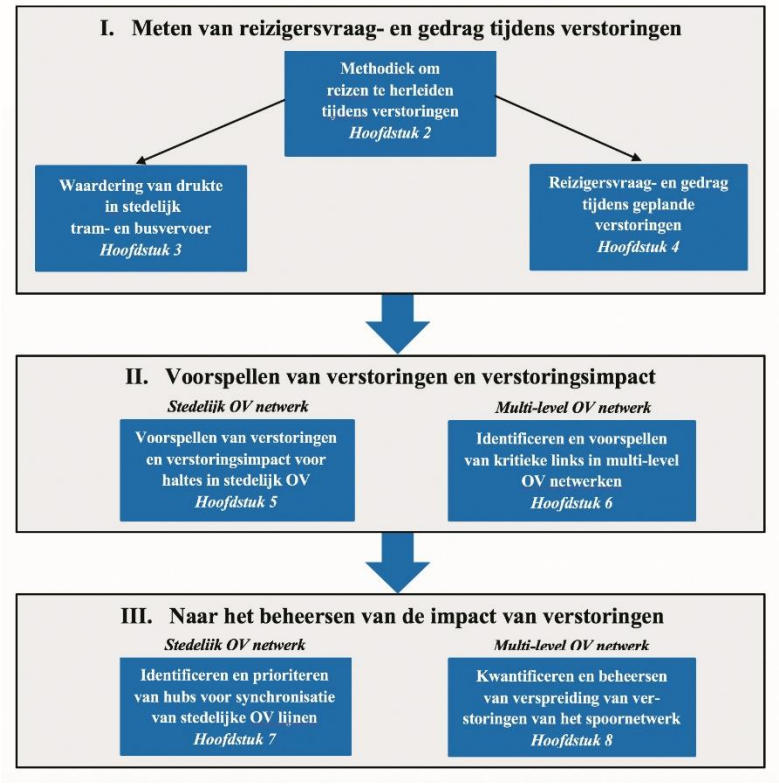
Onderzoeksdoel en onderzoeksvragen
Het primaire onderzoeksdoel van deze studie is *'het verbeteren van methoden om de impact van verstoringen voor het stedelijk openbaar vervoer te meten, voorspellen en beheersen'*. Op basis van literatuuronderzoek identificeren we hiaten qua methoden om verstoringsimpact te meten, voorspellen en beheersen. Dit resulteert in de volgende drie onderzoeksvragen:

1. Hoe kunnen we route- en vervoerwijze-keuze van OV reizigers als reactie op geplande en ongeplande verstoringen in stedelijk openbaar vervoer meten en kenmerken?
2. Hoe kunnen voorspellingen van de frequentie en impact van verstoringen worden gebruikt in een kwetsbaarheidsanalyse voor stedelijke en multi-level OV netwerken?
3. Hoe kunnen we de directe en indirecte impact van verstoringen voor het stedelijk openbaar vervoernetwerk voorspellen en beheersen in een multi-level netwerkcontext?

Scope
Dit onderzoek is als volgt afgebakend:

- *Focus op stedelijk openbaar vervoer, in een multi-level netwerkcontext.* Ons onderzoek richt zich op de impact van verstoringen voor het stedelijk openbaar vervoernetwerk (metro, lightrail, tram en bus). Hoewel andere netwerkniveaus van het openbaar vervoer netwerk, zoals het regionale spoornetwerk, niet de focus zijn van dit onderzoek, nemen we deze integrale, multi-level netwerkcontext wel in beschouwing. Dit betekent dat wordt meegenomen wat de potentiële rol van het spoornetwerk kan zijn als middel om de impact van verstoringen op het stedelijk netwerk te verminderen, maar ook als bron van verstoringen op het treinnetwerk die naar het stedelijk netwerk kunnen doorwerken.
- *Gericht op terugkerende en niet-terugkerende verstoringen*. Onze studie richt zich zowel op de kleinere, vaker voorkomende, terugkerende verstoringen, als op de grotere, minder frequente verstoringen, die kunnen leiden tot het geheel, gedeeltelijk of niet beschikbaar zijn van infrastructuur. Extreme gebeurtenissen zoals natuurrampen of aanslagen vallen buiten de scope van dit onderzoek.
- *Onderzoeken van geplande en ongeplande verstoringen*. We analyseren de impact van zowel ongeplande als geplande verstoringen, zoals geplande werkzaamheden.

Studiebijdrage
Dit onderzoek resulteert in de volgende wetenschappelijke bijdragen (zie **Figuur II.1**):



Figuur II.1. Onderzoekstructuur en bijdrage

- Het ontwikkelen van een verbeterd algoritme om reizen van passagiers in het stedelijk openbaar vervoernetwerk af te leiden tijdens verstoringen.
- Vaststellen hoe reizigers drukte meewegen in hun routekeuze voor reizen met stedelijk tram- en busvervoer op basis van *Revealed Preference*.

- Het kalibreren van parameters voor vervoerwijze- en routekeuze voor passagiers tijdens geplande verstoringen in het openbaar vervoer op basis van empirische data.
- Het ontwikkelen van een methode om de frequentie en de impact van verstoringen te voorspellen voor stedelijke openbaar vervoernetwerken.
- De ontwikkeling van een methodologie om trajecten te identificeren die het meest bijdragen aan de kwetsbaarheid van multi-level openbaar vervoernetwerken.
- Identificatie van hubs en clusters van OV lijnen die deze hubs bedienen in stedelijke openbaar vervoernetwerken om te prioriteren tijdens OV synchronisatie.
- Het ontwikkelen van een methode om te voorspellen hoe de impact van verstoringen op het spoornetwerk doorwerkt naar het stedelijk openbaar vervoernetwerk.
- Evaluatie van de impact van verschillende bijsturingsmaatregelen toegepast op het spoornetwerk voor het multi-level openbaar vervoernetwerk.

Resultaten
Op basis van de onderzoeksresultaten kunnen de drie onderzoeksvragen worden beantwoord.

1. Hoe kunnen we route- en vervoerwijze-keuze van OV reizigers als reactie op geplande en ongeplande verstoringen in stedelijk openbaar vervoer meten en kenmerken?

Om de verstoringsimpact voor reizigers te meten, moeten de gegeneraliseerde reiskosten van passagiers afgeleid worden uit empirische data en vergeleken worden tussen een verstoorde en onverstoorde reis. De eerste stap hiervoor is het ontwikkelen van een robuust algoritme, wat reizen van passagiers afleidt van individuele chipkaart transacties tijdens zowel verstoorde als onverstoorde situaties (**Hoofdstuk 2**). Op basis van verschillende criteria worden bij elkaar horende chipkaart transacties gekoppeld tot een reis. Dit algoritme is een relaxatie van eerder ontwikkelde algoritmes hiervoor, en neemt het atypische reizigersgedrag tijdens verstoringen in beschouwing. In dit algoritme wordt een uitstappende reiziger alleen als overstapper geclassificeerd, indien voldaan wordt aan een temporeel, ruimtelijk en binair criterium:

- Het temporele criterium houdt in dat een uitstapbeweging alleen als overstap wordt beschouwd, indien een passagier vervolgens instapt in het eerstvolgende logische voertuig, waarbij rekening wordt gehouden met de benodigde overstaplooptijd, drukte en de mogelijkheid dat reizigers door drukte op de halte moeten achterblijven.
- Het ruimtelijke criterium stelt een maximale overstaploopafstand vast - bij onze case study 400 meter hemelsbreed - tenzij een passagier tussentijds gebruik maakt van OV op een ander netwerkniveau of van een andere vervoerder tijdens een verstoring.
- Het binaire criterium houdt in dat een overstap naar dezelfde lijn alleen mogelijk is, indien wordt overgestapt naar het direct achteropkomende voertuig van diezelfde lijn. Hiermee wordt rekening gehouden met de mogelijkheid dat tijdens verstoringen bijsturingsmaatregelen, zoals kort-keren, worden toegepast die hiertoe leiden.

Een gedeeltelijke validatie laat zien dat ons algoritme het afleiden van OV reizen tijdens verstoringen verbetert, zonder dat resultaten verminderen tijdens onverstoorde situaties. Dit betekent dat het zowel tijdens verstoorde als onverstoorde situaties kan worden toegepast.

De tweede stap bij het meten van verstoringsimpact is om af te leiden hoe passagiers de verschillende componenten van de reis ervaren, met name wat betreft het ervaren van drukte (**Hoofdstuk 3**). Hiervoor is een discreet keuzemodel met paneleffecten geschat. Dit model is geheel gebaseerd op geobserveerde routekeuze van reizigers op basis van chipkaart transacties, en op geobserveerde kenmerken van de verschillende routes afgeleid van voertuiglocatie data, chipkaart data en de fusie van deze databronnen om de ervaren drukte in het voertuig af te leiden. Dit model laat de volgende resultaten zien:

- Gemiddeld ervaren reizigers één minuut in-voertuigtijd in trams en bussen als 1,16 minuut, wanneer alle zitplaatsen bezet zijn en er geen staande reizigers zijn. Wanneer de bezetting toeneemt tot gemiddeld 3 staande reizigers per m^2, wordt één minuut in-voertuigtijd gemiddeld als 1,34 minuut ervaren.
- Voor frequente passagiers zijn deze factoren respectievelijk 1,31 en 1,75.
- Niet-frequente reizigers nemen drukte niet mee tijdens hun routekeuze, doordat zij vooraf geen kennis hebben van de verwachte drukte op de verschillende routes.
- De geschatte druktebeleving is minder negatief dan geschat in eerdere studies op basis van *Stated Preference* experimenten. Dit illustreert de neiging van *Stated Preference* onderzoeken om coëfficiënten te overschatten, ten opzichte van schattingen op basis van daadwerkelijk vertoond gedrag.

De derde stap is het meten van de impact van geplande verstoringen op het gebruik van openbaar vervoer (**Hoofdstuk 4**). In dit onderzoek kalibreren we coëfficiënten voor route- en vervoerwijze-keuze voor een model om reizigers te voorspellen op basis van empirische data van twee geplande verstoringen, gevalideerd met data van twee andere geplande verstoringen. De hiervoor ontwikkelde *rule-based* methode geeft de volgende resultaten:

- Passagiers ervaren in-voertuigtijd in rail vervangend busvervoer ongeveer 11% negatiever ten opzichte van de tramlijn die vervangen wordt.
- Wachttijd voor rail vervangend busvervoer wordt ≈30% negatiever ervaren dan voor reguliere trams en bussen. Mogelijk liggen de vaak beperkte faciliteiten bij tijdelijke bushaltes hier aan ten grondslag, of speelt onbekendheid aangaande de frequentie en betrouwbaarheid van vervangend busvervoer een rol.
- De nieuwe set parameters verbetert de nauwkeurigheid van de voorspellingen tijdens geplande verstoringen tot 13% ten opzichte van de oorspronkelijke set parameters, die gebruikt wordt om het effect van structurele netwerkwijzigingen te voorspellen.

2. Hoe kunnen voorspellingen van de frequentie en impact van verstoringen worden gebruikt in een kwetsbaarheidsanalyse voor stedelijke en multi-level openbaar vervoernetwerken?

Het voorspellen van de frequentie en impact van verstoringen is noodzakelijk om in een kwetsbaarheidsanalyse de meest kritische elementen van een openbaar vervoernetwerk te identificeren. We ontwikkelen zowel een preselectie methode als een *full scan* methode hiervoor, die beide behalve de verstoringsimpact expliciet de frequentie van verstoringen in beschouwing nemen. In **Hoofdstuk 6** is een preselectie methode voor multi-level openbaar vervoernetwerken ontwikkeld, welke zowel de verwachte directe en indirecte blootstelling aan verstoringen, als het verwachte aantal getroffen passagiers gebruikt als indicatoren. Case study resultaten laten zien dat met name drukke metro / lightrail trajecten het meest kritisch zijn in een multi-level netwerk, aangezien daar zowel het aantal verstoringen als het aantal getroffen reizigers relatief hoog is. Daarnaast tonen onze studieresultaten het belang aan om de frequenties van verstoringen een plaats te geven in kwetsbaarheidsanalyses, aangezien de lijst van meest kritische links substantieel verschilt van de lijst die alleen gebaseerd is op de verwachte verstoringsimpact.

In ons onderzoek is daarnaast ook een data-driven *full scan* methodologie ontwikkeld om binnen een acceptabele rekentijd de meest kritische stations in een stedelijk openbaar vervoernetwerk te identificeren (**Hoofdstuk 5**). Een *supervised learning* benadering is toegepast om zowel de kans op elk verstoringstype, als de reizigersvertraging als impact van elke verstoring te voorspellen voor elk station. Hiervoor gebruiken we vraag-gerelateerde, netwerk-topologische en temporele voorspellers. Ten slotte wordt een *unsupervised learning*

methode toegepast om de verschillende stations te clusteren op basis van de mate waarin ze bijdragen aan de kwetsbaarheid van het openbaar vervoernetwerk. Hierdoor kunnen case study resultaten gegeneraliseerd worden, aangezien dit aangeeft welk type station het meest kritisch is. De case study resultaten toegepast op het metronetwerk van Washington, D.C. laten zien dat stations met de hoogste frequenties en de grootste passagiersaantallen op centrale secties van het netwerk het meest kritisch zijn, samen met overstaplocaties en begin- en eindpunten.

3. Hoe kunnen we de directe en indirecte impact van verstoringen voor het stedelijk openbaar vervoernetwerk voorspellen en beheersen in een multi-level netwerkcontext?

Om de verstoringsimpact voor het stedelijk openbaar vervoer te verminderen kan men beheersmaatregelen toepassen op het stedelijke netwerk zelf, of bijsturing toepassen op het treinnetwerk om het doorwerken van verstoringen van het spoornetwerk naar het stedelijk netwerk te verminderen of om de impact van verstoringen op het stedelijk netwerk zelf te mitigeren. Het real-time synchroniseren van stedelijke OV ritten is een van de mogelijke beheersmaatregelen voor het stedelijke netwerk. Synchronisatie wordt momenteel alleen geoptimaliseerd voor relatief kleine netwerken, aangezien dit optimalisatieprobleem moeilijk oplosbaar is voor grotere netwerken. De bijdrage van deze studie is het ontwikkelen van een generieke, voorbereidende methode om dimensionaliteit van dit probleem te verminderen, door een selectie van locaties en lijnen te prioriteren voor synchronisatie (**Hoofdstuk 7**).

- Ten eerste gebruiken we een op dichtheid gebaseerde clusteringtechniek die, op basis van overstappatronen van reizigers, een subset van hubs vaststelt waar synchronisatie geprioriteerd moet worden.
- Ten tweede representeren we overstappatronen binnen elke hub met een op *C-space* geïnspireerde topologische netwerkrepresentatie. We gebruiken een *community detection* algoritme om groepen van lijnen te identificeren die binnen elke hub simultaan gesynchroniseerd zouden moeten worden.

De resultaten - na toepassing op het stedelijk openbaar vervoernetwerk van Den Haag - laten zien dat 70% van alle overstappen binnen de geïdentificeerde overstaplocaties in beschouwing wordt genomen wanneer minder dan 1% van alle overstaplocaties geselecteerd en geprioriteerd wordt voor synchronisatie. Dit laat zien dat de complexiteit van dit optimalisatieprobleem voor synchronisatie aanzienlijk kan worden verminderd met deze methode.

Om het doorwerken van een verstoring op het treinnetwerk naar het stedelijk OV netwerk te verminderen, combineren we een *train rescheduling* optimalisatiemodel met een dynamisch toedelingsmodel in een iteratief proces (**Hoofdstuk 8**). Het aantal reizigers dat overstapt van het spoornetwerk naar het stedelijk openbaar vervoernetwerk wordt als input in het optimalisatiemodel gebruikt. Vervolgens wordt de impact hiervan getest met het dynamische toedelingsmodel, op basis van een geüpdatete spoordienstregeling als resultaat van het optimalisatieproces. Het optimalisatiemodel wordt vervolgens iteratief geüpdatet op basis van het aantal treinreizigers wat resulteert uit het toedelingsmodel. In onze case study kan de verstoringsimpact die doorwerkt naar het stedelijke netwerk tot 14-27% worden verminderd, zonder dat dit resulteert in meer vertraging voor reizigers op het spoornetwerk.

Ons onderzoek laat ook zien hoe het spoornetwerk gebruikt kan worden als middel om de impact van verstoringen die op het stedelijk netwerk plaats vinden te verminderen. Hiervoor is een maatschappelijk kosten-batenanalyse framework opgesteld (**Hoofdstuk 6**). Een toepassing voor het multi-level openbaar vervoernetwerk van het zuidelijke deel van de Randstad illustreert dat bijsturingsmaatregelen op het spoornetwerk de totale verstoringsimpact - resulterend van een verstoring op het stedelijk netwerk - met 8% kunnen verminderen. Dit laat de potentie van het multi-level netwerk zien om de impact van verstoringen te mitigeren.

Implicaties voor de OV sector

Resultaten van ons onderzoek hebben de volgende implicaties voor vervoerders en vervoersautoriteiten:

- Dit onderzoek helpt vervoerders en vervoersautoriteiten om de nauwkeurigheid van reizigersvoorspellingen tijdens geplande en ongeplande verstoringen te vergroten.
- Methoden ontwikkeld in dit onderzoek leiden tot een verbeterde en snellere kwantificering van de impact van verstoringen op basis van empirische data.
- Ons onderzoek ondersteunt beleidsmakers om robuustheidsmaatregelen te prioriteren voor locaties en verstoringstypen die de grootste invloed op de robuustheid van het openbaar vervoernetwerk hebben.
- Dit onderzoek kan de kwaliteit van real-time bijsturingsmaatregelen verbeteren, waardoor de reizigersimpact van verstoringen verminderd kan worden.

Deze implicaties hebben de potentie om de nauwkeurigheid van haalbaarheidsstudies te verbeteren, besluitvorming tijdens verstoringen te verbeteren, en reizigers een beter product te verstrekken. Dit kan leiden tot een hogere klanttevredenheid en in potentie tot een reizigerstoename in het openbaar vervoer.

Aanbevelingen voor toekomstig onderzoek

Op basis van dit onderzoek formuleren we de volgende aanbevelingen voor toekomstige onderzoeksrichtingen:

- Het opzetten van een gedetailleerde studie naar de vervoerwijze-keuze van passagiers tijdens geplande verstoringen, met name gericht op de mate waarin passagiers gebruik maken van alternatieve vervoerwijzen, zoals deelfietsen of aanbieders van vervoerdiensten als *Uber* of *Lyft*.
- Het uitvoeren van een vervolgstudie naar het dynamische routekeuzegedrag van reizigers tijdens hun reis in het geval van ongeplande verstoringen, en de invloed van factoren zoals real-time informatievoorziening en flexibel werken hierop.
- Onderzoeken in welke mate het verstrekken van informatie aan reizigers voor en tijdens de reis de (gepercipieerde) impact van verstoringen beïnvloedt.
- Het ontwikkelen van een geavanceerdere methode om geobserveerde reizigers-vertragingen van *Automated Fare Collection* (AFC) systemen te kunnen toewijzen aan individuele verstoringen.
- Het vergelijken van de prestatie en rekentijd van de door ons voorgestelde tweetrapsmethode om een selectie van locaties en lijnen te prioriteren voor synchronisatie, met recent ontwikkelde methoden voor netwerk-brede synchronisatie.

Samenvattend is het doel van de ontwikkelde methoden in deze studie om het reizigersperspectief te versterken bij het meten, voorspellen en beheersen van de impact van verstoringen. De toepassing van onze methoden voor verschillende case study's wereldwijd laat zien dat onze methoden toepasbaar zijn in de praktijk. Hoewel resultaten per case study zullen verschillen, laten onze empirische studies en modeltoepassingen zien dat het mogelijk is om reizigersbaten te realiseren. Dit onderzoek resulteert in generieke methoden en tools voor de openbaar vervoersector. We adviseren daarom een nauwe samenwerking tussen wetenschap en de openbaar vervoersector om methoden en resultaten van dit onderzoek te implementeren in de praktijk. Dit heeft de potentie om het openbaar vervoer voor de reiziger verder te verbeteren.

1. Introduction

1.1 Importance of Reliable Public Transport

Disruptions in public transport (PT) have negative impacts on passengers. Disruptions can result in additional in-vehicle time, waiting time, transfer time and extra transfers for passengers. Besides, perceived journey times might increase due to higher crowding levels on alternative PT services. For example, Cats and Jenelius (2014) found that a 30-minute closure of a link on the Stockholm metro network during the morning peak increases the nominal and perceived passenger journey time on the total PT network by up to 11%, depending on the location of the disruption and the information provided to passengers. In Cats et al. (2016b), we calculated that yearly passenger disruption costs resulting from disruptions on one single light rail link in the metropolitan PT network of The Hague and Rotterdam, the Netherlands, can exceed €900,000. Meanwhile, in London all disruptions on Transport for London's underground network during a four-week period from 28 April to 25 May 2019 have resulted in 2.2 million lost customer hours (Transport for London, 2019b). This number expresses the total perceived journey time increase for all passengers affected by the disruptions and illustrates the severity of the impact PT disruptions can have on passengers.

PT disruptions can also result in revenue losses, rescheduling costs, reimbursement costs and fines for the PT service provider. Several service providers refund the fare to passengers if a delay exceeds a certain threshold. For example, the PT agency in Washington D.C., WMATA, fully reimburses passengers whose journey is delayed by more than 10 minutes during rush hours (WMATA, 2019). Transport for London automatically refunds passengers in case selected disruption types result in a delay of 15 minutes or more (Transport for London, 2019a). In the Netherlands, passengers receive a partial or full reimbursement of their fare from the Dutch Railways (NS) when a delay exceeds 30 minutes (Nederlandse Spoorwegen, 2019). Besides, PT service providers can be required to pay a fine to the PT authority in the event of delays or disruptions, depending on the contractual agreements between authority and service provider. For example, PT service providers in the Amsterdam area in the Netherlands are fined if the number of delayed PT trips exceeds the contractually agreed threshold (GVB Holding NV, 2018). As another illustration, in 2016 MTR - service provider of the Hong Kong metro -

was required to pay HK$14.5 million (€1.7 million) due to delays (Straits Times, 2017). Additionally, service providers can suffer from temporary or systematic revenue losses when passengers decide not to travel by public transport in response to a disruption. Saberi et al. (2018) found an 85% increase in usage of bicycle sharing schemes during a strike on the London Underground network, whilst Shires et al. (2018) studied the impacts of planned rail closures on passengers' mode choice, destination choice and trip frequency choice. Depending on the level of awareness and quality of the alternative service provision, they found a temporary rail demand reduction ranging from 5% up to 32% during planned rail closures. These examples illustrate the financial impact PT disruptions might have for the PT service provider involved.

Passengers consider public transport reliability an important quality aspect. Based on Maslow's hierarchy of human needs (Maslow, 1948), Van Hagen (2011) determined that safety and reliability are the most fundamental needs in the hierarchy of customer needs when using public transport. The perceived importance of PT reliability by passengers is also shown in studies by for example Bates et al. (2001) and Rietveld et al. (2001). Susilo and Cats (2014) and Abenoza et al. (2017) state that reliability is an important determinant of PT customer satisfaction, whilst Cats et al. (2015a) and Abenoza et al. (2019) conclude that passengers are systematically dissatisfied with information provided during planned and unplanned service disruptions, based on a study conducted in Stockholm. Olsson et al. (2012) illustrate that negative events, such as PT disruptions, leave a longer lasting mark on customer satisfaction. Van Oort (2016) states that unreliability is an important determinant for passenger route choice (e.g. Schmöcker and Bell, 2002; Liu and Sinha, 2007) and mode choice (Turnquist and Bowman, 1980). Hence, this stipulates the importance of reliable public transport, as (perceived) unreliability can result in dissatisfaction and PT ridership reductions.

Although it is important to reduce the impact PT disruptions have on passengers, PT service providers and customer satisfaction, it is particularly challenging to foresee and study disruptions due to the uncertainty and variety with which they occur. As disruptions occur relatively infrequently, it is difficult to predict when and at which location a certain disruption will occur, and what the disruption duration will be. Besides, there is a wide range of disruptions varying from relatively small, recurrent disruptions (such as a train delay or cancellation), non-recurrent disruptions (such as a train breakdown) to extreme events (such as strikes or natural disasters). There can be differences in susceptibility to these different disruption types for different locations on the PT network, during different seasons or for different periods of the day. Predicting the impact once a disruption happens is also a far from trivial task. This depends on the disruption type, location and duration, as well as the number and type of passengers affected (such as the mixture between commuting vs. leisure passengers). This also depends on the response of the PT service provider in terms of service adjustments and information provision to passengers, and passengers' behavioural response based on actions of the PT service provider, prior knowledge and previous experiences. We therefore conclude that PT disruptions occur in an environment with complex interactions between decisions made on the demand and supply side of the PT system, surrounded by various sources of uncertainty.

For a systematic approach to mitigate PT disruption impacts, we need to understand the current disruption impacts, predict future disruption frequencies and impacts, and then develop and evaluate measures aimed at controlling these disruption impacts. Our general framework for how to approach PT disruptions is presented in **Figure 1.1**. The first step is to measure disruptions and quantify the impacts of PT disruptions empirically for past disruptions. This provides a better understanding of the spatial and temporal distribution of disruptions and the magnitude of their impacts. Given the rare nature of many disruptions, in a second step it is necessary to predict how often, and at which locations different disruption types will occur in the future, together with predicting their impacts on passengers and PT service providers. This

enables performing a systematic PT vulnerability analysis to identify and quantify the vulnerability of different parts of the PT network to different disruptions. Once predictions of disruption frequencies and impacts are available, controlling disruption impacts takes place in the third step. In this step, measures aimed at reducing the disruption frequency and/or mitigating disruption impacts are developed. Potential measures can range from strategic (infrastructure or service network related), tactical (planning related) to real-time control (for example retiming, reordering or cancelling PT trips). Predicting the costs and benefits of potential measures is important to support policy makers in their decision with which measures to proceed towards the implementation phase. This research covers measuring, predicting and controlling disruptions impacts, as well as predicting disruptions (**Figure 1.1**). Measuring and controlling disruptions (rather than their impacts) both fall outside the scope of our research.

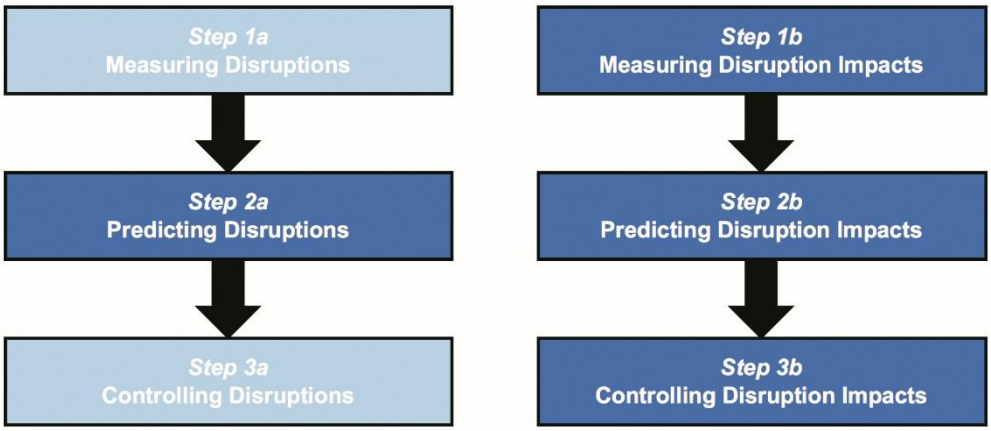


Figure 1.1. Framework to reduce PT disruptions and disruption impacts for passengers

1.2 State-of-the-Art and Research Gaps

1.2.1 Measuring disruption impacts

Over the last decades, metrics to measure PT disruption impacts have shifted from supply-oriented to passenger-oriented indicators. These metrics aim at measuring PT (un)reliability: the extent to which the realised passenger journey deviates from the scheduled or expected passenger journey. This should reflect the impact on the total passenger journey, including in-vehicle time, waiting time, walking time, crowding and the number of transfers. Traditionally, punctuality-based metrics measure the percentage of PT trips which depart or arrive with a delay smaller than a certain maximum number of minutes from/at a set of predefined stations. For example, for the Dutch railway network this threshold is set to 5 minutes (Vromans, 2005). Additionally, average punctuality can be calculated for each PT line (e.g. Van Oort, 2011). For high-frequent urban PT networks, vehicle regularity is often more important than punctuality. Hence, many studies focusing on urban PT networks use the Coefficient of Variation (CoV) of actual headways as a metric instead of punctuality (see for example Engelson and Fosgerau, 2011). Based on the CoV, the additional waiting time and variation in waiting time for PT passengers caused by irregularity can be computed (see for example Turnquist and Bowman, 1980; Van Oort, 2011). Under the assumption of random passenger arrivals at the PT stop, irregularity results in a larger passenger segment experiencing longer waiting times, whilst a smaller passenger segment experiences shorter waiting times. This results in an increase in both the average waiting time and variance in waiting time as consequence.

Recent years, passive data from Automated Vehicle Location (AVL), Automated Fare Collection (AFC) and Automated Passenger Count (APC) systems has become widely available in the PT sector, as well as data from GPS and mobile phone (e.g. Trépanier and Yamamoto, 2015). These data sources provide opportunities to quantify metrics in a fast and automated manner. The availability of AVL data with scheduled and realised vehicle departure and arrival times enables automated quantification of the abovementioned metrics for all PT trips. Nevertheless, a clear disadvantage of these metrics is that all trips are weighted equally, regardless of the number of passengers affected. Passenger-weighted train punctuality aims to correct for this, by weighting arrival punctuality based on the expected train load (Vromans, 2005). This metric is however still problematic due to its focus on separate train trips. None of the abovementioned metrics incorporates how a single PT vehicle delay affects the complete passenger journey, including the possibility of missed connections to other trains, or to trams and buses at the urban PT network level (as for example considered by Lee et al., 2014).

Excess Journey Time (EJT) compares the realised passenger journey time with the scheduled journey time (Zhao et al., 2013; Hendren et al., 2015). Based on tap in and tap out data resulting from AFC systems, realised and scheduled times can be compared for the total passenger journey per origin-destination (OD) pair. However, this metric does not incorporate the difference in *perceived* journey times, for example caused by higher crowding levels. To capture the passenger disruption impact more completely, the total realised generalised journey time (GJT) needs to be compared with the planned or expected GJT. When expressed in monetary terms, passenger disruption impacts can then be expressed as welfare change between realised and scheduled journey (Cats and Jenelius, 2014).

To measure the realised GJT empirically, all journey time components need to be obtained (typically using data from AFC, AVL and APC systems) and multiplied with their respective coefficients, which reflect how passengers perceive the different components (see our framework in **Figure 1.2**). As a first step, passenger journeys need to be inferred from the individual AFC transactions. Several studies have developed destination inference algorithms to infer the journey leg destination in case of AFC systems where passengers only need to tap in, or in case passengers (un)deliberately do not tap out in an entry-exit AFC system (e.g. Trépanier et al., 2007; Munizaga and Palma, 2012). Once destinations are inferred for individual AFC transactions, transfer inference algorithms are required to determine which transactions form one passenger journey. These transfer inference algorithms vary from relatively simple (such as applying a maximum transfer time threshold between a tap out and consecutive tap in by the same smart card) (e.g. Seaborn et al., 2009), to more complex (such as Gordon et al., 2013). These rule-based algorithms assume a certain behavioural logic in passenger route choice when distinguishing transfers from final destinations. Therefore, these can be applied during regular, undisrupted cases. However, during PT disruptions passengers can be confronted with service adjustments, imperfect knowledge about alternative routes and lack of information. This implies that the assumed logic for journey inference during undisrupted scenarios does not necessarily apply during disruptions, as passenger route choice behaviour during disruptions - such as making additional transfers, detours, or not boarding the first vehicle due to excessive crowding - would be considered illogical if there would be no disruption. Therefore, existing transfer inference algorithms are currently not suitable to capture passengers' specific route choice behaviour accurately during disruptions. When this results in inadequate journey inference, this can potentially lead to an incorrect comparison of GJT between scheduled and observed journeys and an incorrect measurement of disruption costs.

Once passenger journeys are established, the second step to measure disruption impacts is obtaining the values of the journey time components. If the considered PT system has on-board devices for passengers to tap in, the characteristics of each journey leg can be directly

obtained from AFC and AVL data. Examples can be found for urban tram and bus networks in the Netherlands (Van Oort et al., 2016) or in Brisbane, Australia (Alsger et al., 2016). In case of gate lines at the station, a passenger-to-train assignment needs to be performed first based on AFC and AVL data, as for example proposed by Hörcher et al. (2017) and Zhu et al. (2017). Once passenger itineraries are observed or inferred, the realised in-vehicle time for each journey leg can directly be obtained from AVL data. To measure the perceived in-vehicle time caused by crowding, load profiles for each trip and each line segment are necessary. Spatiotemporal load profiles and thus crowding levels can be obtained from APC systems, or by fusion of passenger itineraries (observed or inferred from AFC data) with AVL data (e.g. Luo et al., 2018). Transfer time can be calculated by taking the difference between the AVL arrival time or AFC tap out time of one journey leg, and the AVL departure time or AFC tap in time of the consecutive journey leg. This transfer time can be divided into transfer walking time and waiting time, based on an assumed walking speed distribution (Hänseler et al., 2016; Zhu et al., 2017). We can conclude that inference of the values of the different journey time components based on empirical data is a well-studied topic where important research gaps have already been addressed.

The third step to measure disruption impacts is to determine how passengers perceive the different journey time components. Many studies have been performed on how waiting time or walking time coefficients relate compared to in-vehicle time coefficients, for example in the UK (Wardman, 2004) and in the Netherlands (Bovy and Hoogendoorn-Lanser, 2005). In addition, during the last decades many studies have been performed on how in-vehicle time is perceived as a function of on-board crowding levels. Extensive meta-analyses of in-vehicle time crowding multipliers as function of load factor or standing density were performed by Wardman and Whelan (2011) and Li and Hensher (2011). However, these crowding multipliers are typically based on stated preference (SP) research, rather than using observed choice behaviour. SP research has the inherent limitation that there might be a discrepancy between the behaviour stated by respondents and their realised behaviour, the latter being used in revealed preference (RP) research. Therefore, there is a risk of potential bias when estimating crowding multipliers based on SP research, as suggested by studies where selected RP data was used to validate SP results (Kroes et al., 2014; Batarce et al., 2015). The availability of large amounts of AFC and AVL data nowadays does provide an opportunity to re-estimate models on how passengers value on-board crowding. This is especially relevant during PT disruptions, as train delays or cancellations typically result in increased crowding levels on alternative services. Using SP based crowding multipliers can result in incorrect disruption impact measurements. In addition, it can result in incorrect route choice predictions when predicting the impact of future disruptions. Only recently two studies have been performed to estimate crowding multipliers for metro passengers in Singapore (Tirachini et al., 2016) and Hong Kong (Hörcher et al., 2017) purely based on RP data. No studies have been performed in a European context, nor for other PT modes such as light rail, trams and buses.

After measuring the journey time components and how these are perceived, a fourth step is to consider the behavioural and demand response of passengers when a PT disruption occurs. Most studies to unplanned disruptions focus on en-route choice effects for passengers, and assume no PT demand suppression (e.g. Cats and Jenelius, 2014; Cats and Jenelius, 2015). Passengers are assumed to redistribute over the PT network, as there is no awareness assumed when starting the PT journey. Generally, in PT vulnerability analyses there is a strong focus on unplanned disruptions. The topic of planned disruptions, for example related to maintenance works, is relatively understudied (Shires et al., 2018). During planned disruptions a fixed PT demand assumption does however not apply. Due to awareness, passengers might change their mode, destination or trip frequency choice. There are however very few empirical studies which study passengers' demand response in the event of planned disruptions. Van Exel and Rietveld

(2001) reviewed passenger behaviour specifically in response to public transport strikes based on 13 studies. The recent work of Shires et al. (2018) is one of the limited studies towards this demand response during planned rail closures in general, thereby focusing on long-distance trains in the UK. So far, no studies have been performed focusing on passengers' behavioural and demand response during planned disruptions for urban PT networks.

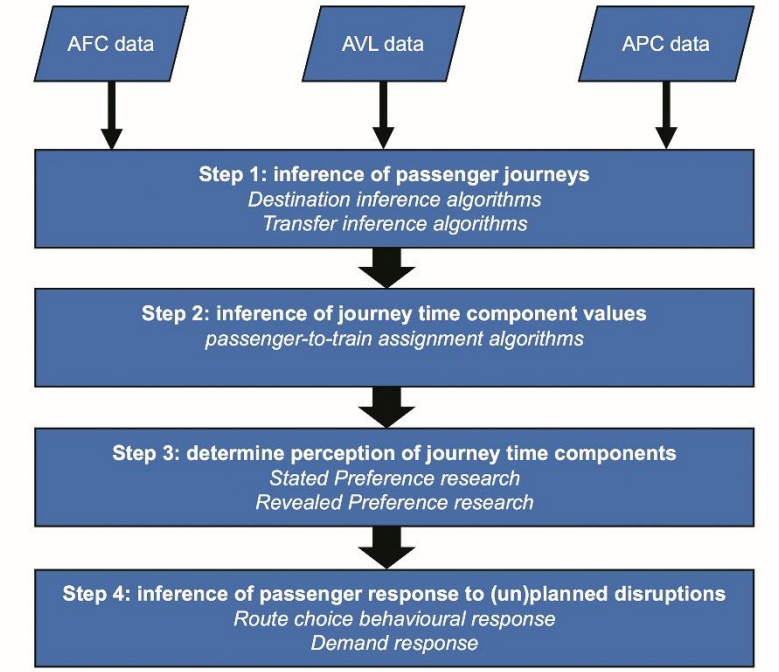


Figure 1.2. Framework to measure passenger impacts of PT disruptions from empirical data

Based on the review of state-of-the-art research towards measuring PT disruption impacts, we can identify the research gaps as defined below. Addressing these research gaps is important to obtain a more accurate and more comprehensive understanding of PT disruption impacts based on empirical data.

- **1.1** There is no transfer inference algorithm which is able to infer public transport journeys under disrupted circumstances, as existing inference algorithms rely on behavioural route choice logic which does not necessarily apply during public transport disruptions.
- **1.2** It is unknown how passengers perceive public transport crowding during their journeys based on realised, empirically observed route choice behaviour, particularly for urban trams and buses and in a European context.
- **1.3** The demand response of public transport passengers in the event of planned disruptions on the urban public transport network is unknown.

1.2.2 Predicting disruption frequencies and disruption impacts

As PT disruptions occur relatively infrequently, measuring disruption impacts can only be performed for instances for which empirical data is available. This means that typically passenger impacts can be measured for a selection of locations, time periods and disruption types. Once historical data for a given time period (e.g. one or two years) would be considered, it is unlikely that empirical disruption observations will be available for all disruption types, for each location of the PT network, during each time period of the day due to the large number of

possible combinations of disruption types, locations and time periods. Prediction of PT disruption impacts is therefore necessary for future disruptions or disruptions for which empirical data is lacking, as well as for the evaluation of possible interventions to mitigate disruption impacts. This section discusses state-of-the-art work to predict PT disruption frequencies and disruption impacts. First, we address methods to predict disruption frequencies and disruption impacts. Second, methods used to perform PT vulnerability analyses to identify the locations in a PT network which contribute most to PT vulnerability are discussed.

Traditionally static, frequency-based or schedule-based, PT assignment models are used to predict PT disruption impacts (Gentile et al., 2016), in some cases combined with variable demand models to capture mode choice impacts in the event of planned disruptions in particular. The disadvantage of using static assignment models for this purpose is their assumption that passenger route choice is determined before the journey starts, based on knowledge on how PT services are amended in response to a disruption. This means these models might be used to predict the impact of planned disruptions, where passengers are aware of the disruption when commencing their journey. These models are however unable to incorporate the dynamics of especially unplanned PT disruptions. Typically, passengers become aware of unplanned disruptions during their journey, requiring them to adjust their route during their journey, often based on limited information of the service adjustments or disruption duration. Static models also assume a stable PT service network during the disruption, whilst neglecting the transition from undisrupted to disrupted network and the recovery time the PT system needs once the disruption is resolved. In addition, the dynamic interaction between PT demand and supply during disruptions or delays, which can result in vehicle queuing or bunching, cannot be captured in static assignment models.

To predict impacts of unplanned disruptions - assuming a fixed PT demand - there is a need for more advanced, dynamic PT assignment models, which are able to capture the demand and supply dynamics and their interactions during disruptions. In recent years, there have been several developments to use this type of models in transportation, instead of the aforementioned traditional static assignment models. For this purpose, mesoscopic, agent-based assignment models are developed for road networks (e.g. De Souza et al., 2019) and for PT networks. For example, Cats et al. (2016a) use BusMezzo, a dynamic, mesoscopic PT assignment model, for urban and metropolitan PT networks. As individual PT vehicles and passengers are simulated, it is possible to account for dynamic, en-route passenger route choice and test the impact of real-time information provision or day-to-day learning effects (Cats and Jenelius, 2014) for complete and partial service degradations (Cats and Jenelius, 2018). This model type however assumes operations without explicitly considering a railway signalling system. To predict impacts of heavy rail disruptions, often microscopic or mesoscopic simulation models such as Open Track (Nash and Huerlimann, 2004) are used, which specifically incorporate railway characteristics such as a signalling system, acceleration and braking characteristics of different rolling stock types and block lengths. These simulation models focus primarily on simulating trains, whilst passengers and their route choices are often incorporated in a simplified way. Not incorporating the full, dynamic interactions between PT demand and supply is typically less problematic for heavy rail networks, compared to high-frequent urban PT systems, due to railway characteristics. Heavy rail systems generally have lower service frequencies and early departures from stations are often prohibited due to the signalling system in place. Effectively, this implies that bunching between subsequent train services is less likely to occur, as trains are subject to holding until their scheduled departure time at the majority of the stations. Additionally, heavy rail networks have a lower network density than urban PT networks, which means that the number of feasible route alternatives available to passengers will be more limited than for urban networks. Hence, train simulations models are often used to predict disruption

impacts for heavy rail networks, in contrast to macroscopic or mesoscopic passenger assignment models being used for urban PT systems.

Because of the different types of models used to predict disruption impacts for heavy rail and urban PT networks, the impact of disruptions for the integrated multi-level PT network is hardly calculated. The total PT network consists of different functional network levels - such as the (inter)regional train network level and the urban tram and bus network level - which are hierarchically connected to each other. We use the term *multi-level network* to refer to the total PT network consisting of these different network levels. As an illustration, a train network disruption can result in missed transfers to the urban tram or bus network, thereby affecting the journey time and crowding levels on the urban PT network level. Models generally only consider the transport modes which are subject to similar traffic and control regimes (such as light rail, trams or buses on the one hand, and heavy rail systems on the other hand). Hence, existing models only predict the impact of a disruption on the same network level as where the disruption originated, for those transport modes sharing the same traffic regime. The propagation of disruption impacts from one PT network level to another network level, operated by transport modes with different traffic regimes, is not considered.

A PT vulnerability analysis is used to identify the network elements which contribute most to PT network vulnerability, and to quantify this contribution. This enables prioritising mitigation measures for the most critical network parts. Identification of the nodes or links which contribute most to the vulnerability of a transport network is a well-studied research topic. Two different approaches are broadly distinguished in literature: pre-selection methods and full computation methods. The first approach uses criteria to pre-select a smaller number of nodes or links which are expected to contribute most to network vulnerability. In a second step, full disruption impacts are only calculated for these selected nodes or links, thereby reducing the required number of model runs. The disadvantage however is that there is no guarantee the most critical nodes or links are selected. In the latter approach, disruption impacts are predicted for all nodes or links in the transport network, which enables listing the contribution of each node / link to network vulnerability (e.g. Knoop et al., 2012). The disadvantage is that these methods are very time consuming when considering large, real-world transport networks, as a large number of predictions (often using model runs) would be required. This method is therefore only applicable to smaller or test networks within reasonable computation times. This illustrates that methods to predict disruption impacts for all locations of a large PT network in a systematic manner within reasonable computation times are currently lacking.

There is a variety of indicators developed in scientific literature for pre-selection methods to identify nodes or links with the largest contribution to network vulnerability. Pre-selection methods for road networks use for example the volume/capacity ratio of links, or the Incident Impact Factor as possible criteria (e.g. Tampère et al., 2007). For PT networks, potential criteria found in literature are amongst other node degree, betweenness centrality, or PT link volume (e.g. Bell, 2003; Cats and Jenelius, 2014). An important shortcoming is that existing pre-selection criteria only identify nodes or links for one network level. For example, Derrible and Kennedy (2010) only consider metro networks, whilst Cats and Jenelius (2014) only focus on urban PT networks. Existing studies do not identify the most critical links of the total multi-level PT network. To prioritise mitigation measures to reduce disruption frequency and impact, it is however important to understand how links of different PT network levels contribute to the vulnerability of the total multi-level PT network.

To make PT vulnerability analyses useful for decision-making, predictions of both *disruption impact* and *disruption frequency* for each network location are necessary. When only disruption impacts are predicted, network elements with the most severe, yet rare, disruptions might be prioritised incorrectly. For a correct prioritisation, the joint contribution in terms of

both disruption frequency and disruption impact should be predicted for each network element. The methods discussed above however only focus on the prediction of disruption impacts, without explicitly considering disruption frequencies or disruption locations. The same applies for existing pre-selection criteria, which are only based on expected disruption impacts, without considering how often different nodes or links are exposed to different disruption types. This can often be explained by the lack of usable disruption log data, as log data of sufficient quality from a longer period of time would be needed to properly assess or predict the frequency of different disruption types at different locations, given their relatively infrequent occurrence. Notwithstanding, without assessing how often each disruption type occurs, together with its duration and location, it becomes difficult to correctly predict vulnerability and to incorporate robustness benefits of potential mitigation measures in appraisal studies or cost-benefit analyses.

Based on this review of state-of-the-art research regarding the prediction of disruption frequencies and disruption impacts, the following research gaps are defined:

- **2.1** Methods to predict disruption impacts do not incorporate the propagation of disruption impacts between different network levels of the integrated, multi-level public transport network, but focus on one single network level instead.
- **2.2** There is no methodology to predict disruption frequencies and disruption impacts for all network locations and time periods in a systematic manner for large, real-world public transport networks within acceptable computation times.
- **2.3** Methods for public transport vulnerability analysis do not incorporate the frequency and location of different disruptions in the multi-level public transport network when identifying the nodes or links which contribute most to network vulnerability and when predicting this contribution.

1.2.3 Controlling disruption impacts

There is a wide variety of measures on a strategic, tactical and operational level aimed at controlling PT disruption impacts. The reader is referred to Van Oort (2011) for a comprehensive overview of potential measures. Strategic measures are for example related to the realisation of additional switches to enable short-turning during disruptions, or focused on a robust configuration of a rail terminal (Van Oort and Van Nes, 2010). On a tactical level measures can relate to the design of robust timetables for railways (e.g. Andersson et al., 2015) or urban bus networks (e.g. Gkiotsalitis et al., 2019), or to introduce holding stop points on a PT route to prevent early departures and enable delayed vehicles to adhere to the timetable again (e.g. Van Oort et al., 2012). Tactical measures can also focus on reducing driver schedule complexity, to limit disruption impact propagation and recovery time (Yap and Van Oort, 2018). In the research area of railways, many studies are performed on real-time rescheduling of rolling stock and crew, typically using optimisation-based approaches (see for example Corman et al., 2010; Cacchiani et al., 2014; Van der Hurk et al., 2018). For urban PT networks studies focus on comparing rescheduling strategies (for example short-turning or diverting a trip) (Roelofsen et al., 2018), or on real-time synchronisation between services. Gavriilidou and Cats (2019) provide an overview of the extensive amount of research performed on PT synchronisation. They develop a rule-based controller for synchronisation of two urban rail lines, thereby incorporating expected crowding levels and passengers potentially being denied boarding. Daganzo and Anderson (2016) use simulation to evaluate transfer synchronisation between metro and bus, whilst Laskaris et al. (2018) use BusMezzo as simulation-based assignment model to evaluate different multiline holding control strategies. Nesheli and Ceder (2015) optimise real-time transfer synchronisation between three bus lines at two different

transfer locations, whereas Hadas and Ceder (2010) develop an optimal transfer synchronisation strategy for a case study network consisting of one train line and three bus routes. In addition to mitigating disruption impacts, measures can also aim at reducing the disruption frequencies, for example by maintenance optimisation (using techniques as described by De Jonge and Scarf, 2019) or by the introduction of predictive maintenance for a train or bus fleet (see for example Killeen et al., 2019).

We identify two limitations regarding existing approaches to control PT disruption impacts. The first limitation is related to transfer synchronisation, being one of the possible real-time control measures to control disruption impacts. Optimal transfer synchronisation becomes computationally challenging for larger, real-world urban PT networks, due to NP-hardness of the transfer synchronisation problem (Desaulniers and Hickman, 2007). Existing studies to optimal PT synchronisation generally focus on heavy rail networks or smaller urban PT case study networks. Railway studies seldom fully account for the stochasticity in PT demand and supply: stochastic passenger route choice, stochastic demand or stochastic vehicle running times, and interactions between demand and supply due to bunching are often not considered or approximated in a simplified way (e.g. D'Ariano, 2007; Törnquist Krasemann, 2012; Binder et al., 2017). This may be justified for lower-frequent railway services, where train separation is arranged by a signalling system and where early departures are often not possible. For high-frequent urban PT networks these stochastic and dynamic aspects are however considerably more relevant, but at the same time this induces extra challenges for solving the transfer synchronisation problem within reasonable computation times. Hence, abovementioned applications of optimal synchronisation for urban PT networks are limited to smaller networks. When considering larger urban PT networks, there can be many transfer locations with many different PT lines, for which synchronisation can be potentially relevant. Network managers need to determine in a systematic way at which transfer locations and between which routes PT synchronisation needs to be prioritised, once network-wide optimisation is not feasible. Transfer synchronisation can then be applied for selected locations and between selected routes (see for example Lee et al., 2014). At this moment, guidance to support controllers to prioritise locations and routes for synchronisation is lacking.

The second limitation relates to the incorporation of disruption propagation impacts when devising and evaluating measures to control disruption impacts. Existing real-time control measures are focusing on one PT network level only, without considering the impacts on the other PT network levels. For example, optimal train rescheduling strategies only consider the effect on the train network, whilst neglecting the impact of these control decisions on the urban tram and bus network, for example via missed connections or increased crowding levels. The methodological difficulty here is related to incorporating the urban PT dynamics in the train rescheduling optimisation problem. Nevertheless, ignoring the disruption impact propagation to the lower PT network level potentially results in suboptimal rescheduling from a total passenger perspective.

Based on the abovementioned review of state-of-the-art research related to controlling and mitigating PT disruption impacts, we can define the following research gaps:

- **3.1** When solving the network-wide transfer synchronisation problem is computationally not feasible for large, urban public transport networks, a systematic approach to support planners and controllers to prioritise transfer locations and routes for synchronisation is missing.
- **3.2** In the process of determining an optimal real-time control strategy for trains in response to a train network disruption, the impact of control decisions on the lower-level urban public transport network is not quantitatively incorporated, potentially resulting in the application of suboptimal control strategies.

1.3 Research Objective, Questions and Scope

1.3.1 Research objective

As PT disruptions can have major implications for passengers, PT service provider and authority, it is important to quantify the impacts of these disruptions and to evaluate strategies aimed at mitigating disruptions and their impacts. This can result in a better PT product delivered to passengers and increase the attractiveness of public transportation. The review of state-of-the-art research shows there are several research gaps related to this topic which need to be addressed. The main research objective of this study is therefore formulated as follows:

'To improve methods to measure, predict and control disruption impacts for urban public transport'

1.3.2 Research questions

We formulate three research questions which contribute to the main research objective. The first research question focuses on the first level of the disruption framework (**Figure 1.1**) and aims to improve measuring disruption impacts. This question addresses research gaps **1.1**, **1.2** and **1.3** and focuses on both unplanned and planned disruptions. It considers how measuring the behavioural response of passengers during disruptions based on empirical data sources can be improved, particularly in relation to transfer inference and crowding perception during disruptions. It also aims to better understand passengers' demand response specifically for planned disruptions.

1. How can we measure and characterise the behavioural and demand response of passengers during planned and unplanned urban public transport disruptions?

The second research questions aims to improve the predictions of the frequency and impacts of PT disruptions (the middle level of **Figure 1.1**) and addresses research gaps **2.2** and **2.3**. This research question aims to improve PT vulnerability analyses by developing new criteria to identify critical links based on disruption frequency and impact in multi-level PT networks. It also focuses on the development of an improved method to predict disruption frequency and impact for all locations in a larger urban PT network within acceptable computation times.

2. How can we incorporate disruption frequency and impact predictions in a public transport vulnerability analysis for urban and multi-level public transport networks?

The third research question aims to improve methods for prediction and control of PT disruption impacts (middle and lower level in **Figure 1.1**) and relates to research gaps **2.1**, **3.1** and **3.2**. The research question focuses on the prediction of disruptions impacts for the urban PT network in a multi-level network context. This entails disruptions occurring on the urban PT network, thereby acknowledging the availability of the multi-level PT network for potential mitigation, but also predicting and mitigating the propagation of disruption impacts from train network disruptions to the urban network. Besides, it considers how real-time control synchronisation measures can be prioritised for specific locations and routes to mitigate disruption impacts for larger, real-world urban PT networks.

3. How can we predict and control the direct and propagated impacts of disruptions on the urban public transport network in a multi-level network environment?

1.3.3 Scope, definitions and conditions

In line with the formulated research objective and research questions, our research focuses on disruption impacts for the urban public transport network, consisting of metro, light rail, tram and bus services. Whilst other PT network levels are not the focus of this research, the multi-level PT network environment is considered. For example, this means we compare the contribution of urban PT links and train network links to PT network vulnerability. Besides, we consider the role the train network might play both as means to mitigate impacts of urban PT disruptions, and as a source for train disruptions propagating to the urban PT network level.

In this research, we focus on recurrent and non-recurrent disruptions. Recurrent PT disruptions, such as a train door malfunctioning or a delayed departure from the terminal, occur relatively frequently whilst the impact is generally small. To the contrary, non-recurrent PT disruptions are relatively rare, but typically have larger impacts once they occur. One can think of examples as a faulty train, signal failure or vehicle derailment. Recurrent and non-recurrent disruptions are conceptualised in the framework shown in **Figure 1.3**. It should be noted that there is no explicit demarcation between recurrent and non-recurrent disruptions. Instead, they can be considered as two ends of the same scale. We do not consider extreme events such as natural disasters or terror attacks in this research. These events differ substantially from typical PT disruptions in terms of magnitude and behavioural response by passengers, PT service providers and authorities, for which a bespoke research approach is necessary (see for example Markolf et al., 2019).

Unplanned disruptions as well as planned disruptions are incorporated in our research scope. The impact of planned disruptions - such as planned track maintenance works - is generally smaller than the impact this same disruption would have if unplanned. This is due to awareness and route and mode choice adjustments by passengers, as well as due to planned resource allocation by the service provider in anticipation of this disruption. Hence, the unplanned disruption impact can be considered an upper bound for the disruption impact of the same planned disruption.

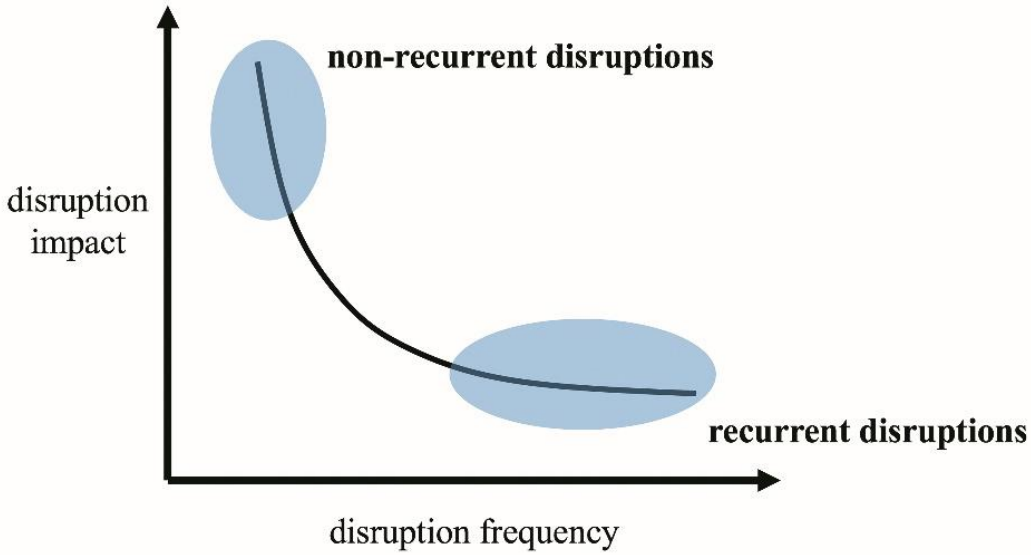


Figure 1.3. Conceptual framework of recurrent and non-recurrent disruptions

We define *disruptions* (interchangeably referred to as *disturbances*) in this study as distinctive incidents which result in deviations from normal operations. This encompasses discrete events of which an underlying cause can be traced, in contrast to normal stochasticity in system input,

such as variability in passenger volumes or travel times, which is not considered a disruption. Within our research scope we consider a wide spectrum of disruptions, ranging from recurrent to non-recurrent disruption types. Some parts of this research focus only on specific types of disruptions (see **Section 1.6**), such as non-recurrent disruptions or planned disruptions which last for multiple days or weeks, while other parts consider the full range of disruptions.

In this research, we define *vulnerability* as the degree of susceptibility of a PT network to disruptions and the ability of a PT network to cope with these disruptions. This definition is obtained by combining definitions as suggested by Rodriguez-Nunes and Garcia-Palomares (2014) and Oliveira et al. (2016). Vulnerability thus refers to both *exposure*, the degree to which a PT system is exposed to disruptions, and to the *impact* once a disruption occurs. *Robustness* of a PT network is the inverse of vulnerability (Snelder, 2010). Recurrent and non-recurrent disruptions result in degradation of the performance of a PT system, with unreliability for passengers as result. *Resilience* is defined as the ability of a PT system to maintain its function when exposed to disruptions (Rose, 2007) and the ability to return to its regular performance once the disruption has been resolved (i.e. recovery time) (Pimm, 1984). The relation between the different concepts is shown in **Figure 1.4**, obtained from McDaniels et al. (2008). Resilience is reflected by the blue coloured area. The conceptual impact of *ex ante* mitigation measures and *ex post* measures to reduce recovery time are visualised in this figure as well.

In this study we refer to *predictions* as estimating the outcomes for unseen cases based on information or based on observations. For example, we use empirical data about disruption frequencies and disruption impacts to predict the future number of disruptions or the passenger impact of future disruptions for which no empirical data is available. In addition, we use PT models to predict passenger flows for non-observed PT disruptions based on empirically derived information about passengers' behavioural and demand response during disruptions. In our study we perform relatively aggregate predictions (i.e. predicting future disruption frequencies or impacts), rather than real-time, short-term predictions of - for example - the exact time a disruption will occur, or the passenger flows for time t+15 minutes given demand at t.

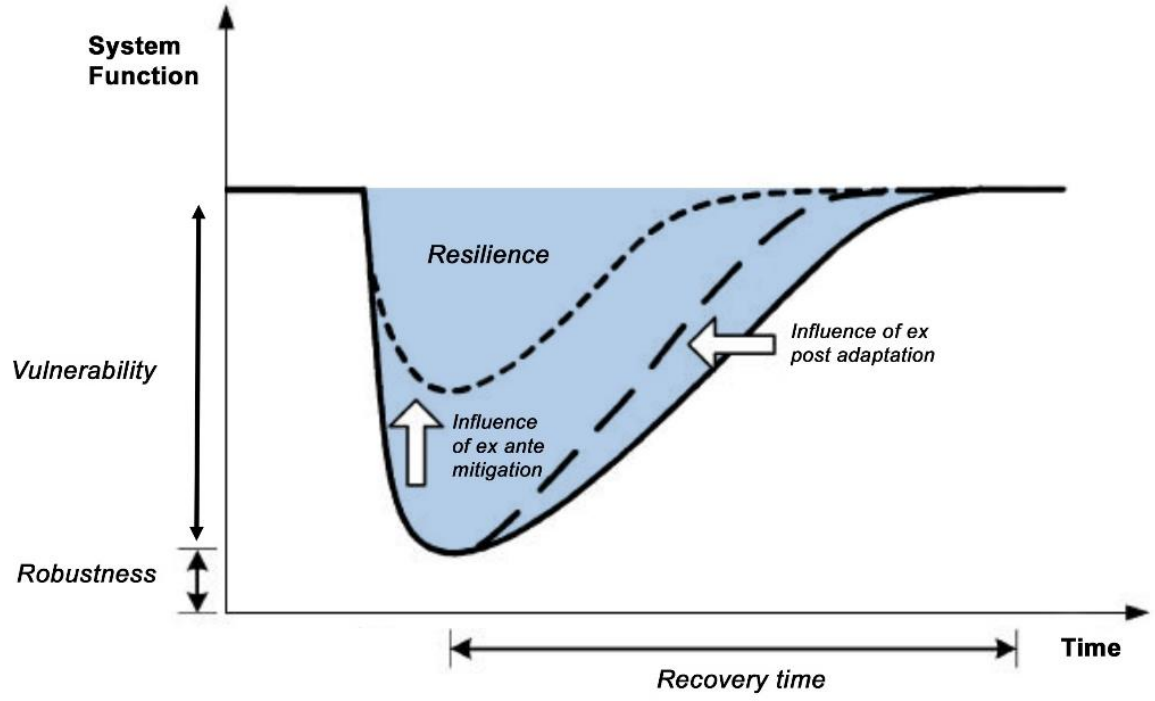


Figure 1.4. Conceptual relations between vulnerability, robustness and resilience
Source: McDaniels et al. (2008, p.312) ('vulnerability', 'resilience' and 'recovery time' added by author)

Our research assumes the availability of passenger demand data from AFC systems and vehicle position data from AVL systems of the considered public transport network as input for our proposed methods. Our methods can be applied to both open and closed AFC systems, with entry-exit or entry-only tap in requirements. Depending on the AFC system in place for the PT network of interest, destination inference (for entry-only systems) or passenger-to-train assignment (for AFC systems with gates at the stations) might be required as a preparatory stage. In this research, developed methods are illustrated in a case study using the urban or multi-level PT network of The Hague, the Netherlands, or using the metro network of Washington, D.C.

1.4 Research Contribution

1.4.1 Scientific contribution

The main scientific contributions of this research are the following.

1. Development of an improved transfer inference algorithm for urban public transport journeys during disruptions (Chapter 2)
In this research we adjust existing transfer inference algorithms - which are based on logic for passenger route choice during undisrupted circumstances - such that it is possible to infer PT journeys from individual AFC transactions during both disrupted and undisrupted circumstances. This is based on an empirical study to passenger route choice during disruptions in The Hague, the Netherlands based on AFC and AVL data. The work results in a robust transfer inference algorithm where no *a priori* demarcation between disrupted and undisrupted network state is required.

2. Estimation of crowding perception multipliers for urban tram and bus journeys based on Revealed Preference (Chapter 3)
This study estimates crowding multipliers for in-vehicle time during urban tram and bus journeys based on observed route choices. It contributes to the variety of Stated Preference based studies to crowding valuation, by using observed route choice from AFC data and observed attribute values from AVL and AFC data as input. This Revealed Preference approach corrects for potential inconsistencies between stated and realised behaviour in Stated Preference studies. The crowding multipliers are estimated using a discrete choice modelling approach.

3. Estimation of mode and route choice coefficients for passengers during planned public transport disruptions based on empirical data (Chapter 4)
This study contributes to a better understanding how PT passengers adjust their mode and route choice in the event of planned PT disruptions. Whilst the majority of PT vulnerability studies typically focuses on unplanned disruptions, this study puts the emphasis on planned disruptions such as rail construction works. Based on empirical data obtained from AFC and AVL systems for multiple planned disruptions, this study calibrates route choice parameters for the use and perception of rail-replacement buses, as well as mode choice elasticities which reflect passengers' demand response to planned disruptions.

4. Development of a methodology to predict disruption frequencies and disruption impacts for urban networks (Chapter 5)
This study proposes a methodology to predict how often different stops of a metro network are exposed to different disruption types and to predict the passenger delay impact of these

disruptions. The contribution of this work lies in the development of a systematic approach to predict disruption frequency and impact for each individual station, time period and disruption type, even though empirical data is not available for all possible combinations and disruption scenarios. We develop supervised learning models to predict disruption frequency and impact using incident log data, AFC and AVL data as input, and apply unsupervised learning to cluster stations according to their contribution to PT vulnerability. This approach enables performing a full scan vulnerability analysis within reasonable computation times.

5. Development of a methodology to identify the links which contribute most to vulnerability of multi-level public transport networks (Chapter 6)
In this research, an improved methodology is developed to identify the links which contribute most to vulnerability of multi-level PT networks based on pre-selection criteria. While existing studies use pre-selection criteria reflecting disruption impacts for a specific PT network level, our new methodology explicitly incorporates both disruption frequency and disruption impact in the pre-selection criteria. Our method also compares links of the train, metro / light rail and tram network in terms of their disruption exposure and impact, to identify the most critical links for the total multi-level PT network. Incident log data for different PT modes combined with outputs from PT assignment models are used as input to develop this methodology.

6. Identification of urban public transport hubs and their key routes to prioritise for public transport synchronisation (Chapter 7)
This research supports PT synchronisation for large urban PT networks by identifying locations and routes to prioritise for synchronisation. Optimal transfer synchronisation is in current studies applied to train networks or to smaller urban networks, whilst rule-based synchronisation approaches are used for larger urban networks due to computational challenges. We use clustering to identify the most important urban PT hubs and to find groups of lines for each hub to prioritise for synchronisation. This results in a selection of locations and routes which can be incorporated in an optimal transfer synchronisation algorithm.

7. Development of a methodology to predict the disruption impact propagation from train network disruptions to urban public transport network (Chapter 8)
The contribution of this study is the quantification of disruption impact propagation from train network disruptions to the urban PT network. Existing studies predict disruption impacts for each PT network separately, without considering how this impact might propagate to another network level. The proposed methodology combines a mesoscopic optimisation-based train rescheduling model and an agent-based dynamic PT assignment model for this purpose.

8. Evaluation of the impact of different train rescheduling strategies on the integrated multi-level public transport network (Chapter 8)
Current research towards strategies to control disruption impacts of train network disruptions only focus on the train network impacts. Train rescheduling strategies are optimised based on expected impacts on this network level, without incorporating the impact of these control decisions for passengers on the urban PT network level. The novelty of this study is that it incorporates predictions for disruption propagation to the urban network in the train rescheduling based on simulation-based optimisation. This enables testing different train rescheduling strategies to predict impacts on both the train and urban PT network.

1.4.2 Societal contribution

The societal contributions of this research are divided into contributions for public transport service providers and for public transport authorities in relation to policy-making.

Contribution for public transport service provider
Our work helps public transport service providers to better understand characteristics of current and future passenger demand. The developed transfer inference algorithm can be used to improve OD matrix estimation during undisrupted and especially disrupted circumstances. The insights gained from the behavioural and demand response of passengers during disruptions can be used as input for transport planning to improve route choice and ridership forecasts. Updated crowding multipliers, demand elasticities and perception coefficients for rail-replacement buses can be incorporated in variable demand and assignment models to improve their prediction accuracy. This can support PT service providers in better aligning their PT supply during disruptions with predicted demand. This enables the provision of sufficient capacity for the remaining passenger demand during planned disruptions, which is potentially beneficial for customer satisfaction levels. In addition, it reduces provision of too much residual capacity on alternative services during planned disruptions, thereby reducing operational costs.

Our research supports network managers and controllers in improving the quality of the control decisions taken in the event of disruptions. Identifying key synchronisation locations and routes for real-time control helps them to prioritise their work. Incorporating disruption propagation impacts provides insights to controllers of the impact of their decisions on the total PT network and therefore supports them in their decision-making process.

Contribution for public transport authority
The identification of stations and links which contribute most to PT network vulnerability can be used by the public transport authority in policy-making to prioritise the most important locations to develop and implement mitigation measures for within a limited budget. By predicting both the frequency and impacts of future disruptions, robustness benefits of potential mitigation measures can be monetised and incorporated in appraisal studies or cost-benefit analysis frameworks. Incorporating the impact of disruptions and mitigation measures for the integrated PT network, rather than for one network level only, results in a more complete and therefore more accurate quantification of robustness benefits of different measures.

Additionally, our work can support the PT authority in the development of passenger-oriented reliability metrics. Our methods allow for the prediction of the impact of real-time control measures taken by the PT service provider for the integrated multi-level PT network, including the propagation to different network levels. This provides insights to the authority about the effectiveness of different reliability metrics to incorporate in contractual agreements with the PT service provider, so that the PT service provider undertakes the appropriate real-time control measures consistent with these metrics, thereby putting passengers at the centre of their decision-making process.

1.4.3 Highlights

The aim of this research is to improve methods to measure, predict and control disruption impacts for urban public transport networks. Methods are developed to measure passengers' behavioural and demand response during (un)planned disruptions. Additionally, we propose methods to predict the frequency and impact of disruptions for different locations of a PT network, and methods to identify those locations contributing most to network vulnerability based on expected disruption exposure and impact. At last, methods are developed to prioritise,

devise and evaluate control measures aimed at reducing impacts of disruptions - which originate either on the urban level, or on another network level - for the urban PT network.

1.5 Research Context

This research results from the TRANS-FORM (Smart transfers through unravelling urban form and travel flow dynamics) project, as part of JPI Urban Europe ERA-NET CoFound Smart Cities and Communities initiative. The overarching objective of the TRANS-FORM project is *'to better understand transferring dynamics in multi-modal public transport systems and to develop insights, strategies and methods to support decision-makers in transforming public transport usage to a seamless travel experience by using smart data'* (TRANS-FORM, 2019). The TRANS-FORM project develops, implements and tests real-time traffic management strategies to support proactive and adaptive public transport operations. Concepts and methods of behavioural modelling, passenger flow forecasting and network state predictions are integrated into real-time operations. New empirical knowledge and modelling foundations are developed by undertaking a multi-level approach for monitoring, mapping, analysing and managing dynamics of interchanging passenger flows. Analysis of passenger flows on the hub, urban and regional networks is facilitated by data secured from case studies in Switzerland, the Netherlands and Sweden, respectively. The outcomes contribute to improving coordination between different public transport modes, in particular in cases of public transport disruptions.

The project consortium consists of five partners from four different countries, of which four are academic partners (Delft University of Technology, the Netherlands (main applicant and project coordinator); Blekinge Institute of Technology, Sweden; Linköping University, Sweden; École Polytechnique Fédérale de Lausanne, Switzerland) and one an industrial partner (ETRA I+D, Spain). This research is conducted as one of the main contributions of the Delft University of Technology to the TRANS-FORM project. This research focuses on analysing passenger flows on the urban public transport network during public transport disruptions, thereby incorporating the interactions with pedestrian flows within public transport hubs and passenger flows on regional train networks.

1.6 Outline

The outline of this research is shown in **Figure 1.5**. Black arrows in this figure indicate that research output from one chapter is directly used as input for the considered chapter. The thicker blue arrows reflect the general disruption framework as introduced in **Figure 1.1**, moving from measuring disruption impacts (step 1), via predicting disruption frequencies and impacts (step 2), towards controlling disruption impacts (step 3).

This research is divided into three different phases. Part I is related to measuring PT disruption impacts and addresses Research Question 1. Based on the identified research gaps, phase I focuses on *Measuring passenger demand and behaviour during disruptions*. It consists of three chapters. In **Chapter 2**, a transfer inference algorithm is developed to infer PT journeys during disruptions. This improved inference algorithm is used as one of the inputs for **Chapter 3**, where passengers' crowding valuation in urban tram and bus journeys is estimated. The improved transfer inference can also be used as input for **Chapter 4**, which studies the behavioural and demand response of passengers particularly during planned disruptions.

Part II of this research focuses on *predicting disruptions and disruption impacts* as the second step in the disruption framework. This phase answers Research Question 2 and partially Research Question 3. The aim of **Chapter 5** is the development of a methodology to predict

disruption frequencies and disruption impacts for each location of an urban metro network. In **Chapter 6**, the urban PT network is considered together with the multi-level network environment: a method is proposed to identify PT links from the train, metro / light rail and tram networks with the largest contribution to total network vulnerability. In addition, this chapter predicts the impact of an urban PT network disruption on these identified critical links, given the availability of the total multi-level PT network for passengers.

In Part III of this research, we move *towards controlling disruption impacts*, which relates to the third step of the disruption framework in **Figure 1.1**. Research Question 3 is addressed in this phase. When mitigating disruption impacts for the urban PT network, **Chapter 7** focuses on synchronisation measures to be applied at the urban network level itself. The focus here is to enable optimal synchronisation for large urban PT networks, by proposing a preparatory method to prioritise key locations and routes for synchronisation. In **Chapter 8**, we consider the propagation of impacts of disruptions occurring on the multi-level network to the urban network level. We develop a method to quantify and control the propagation of train network disruptions to the urban PT network level. Finally, in **Chapter 9** we formulate the main conclusions from our research, together with implications for the PT industry and recommendations for future research.

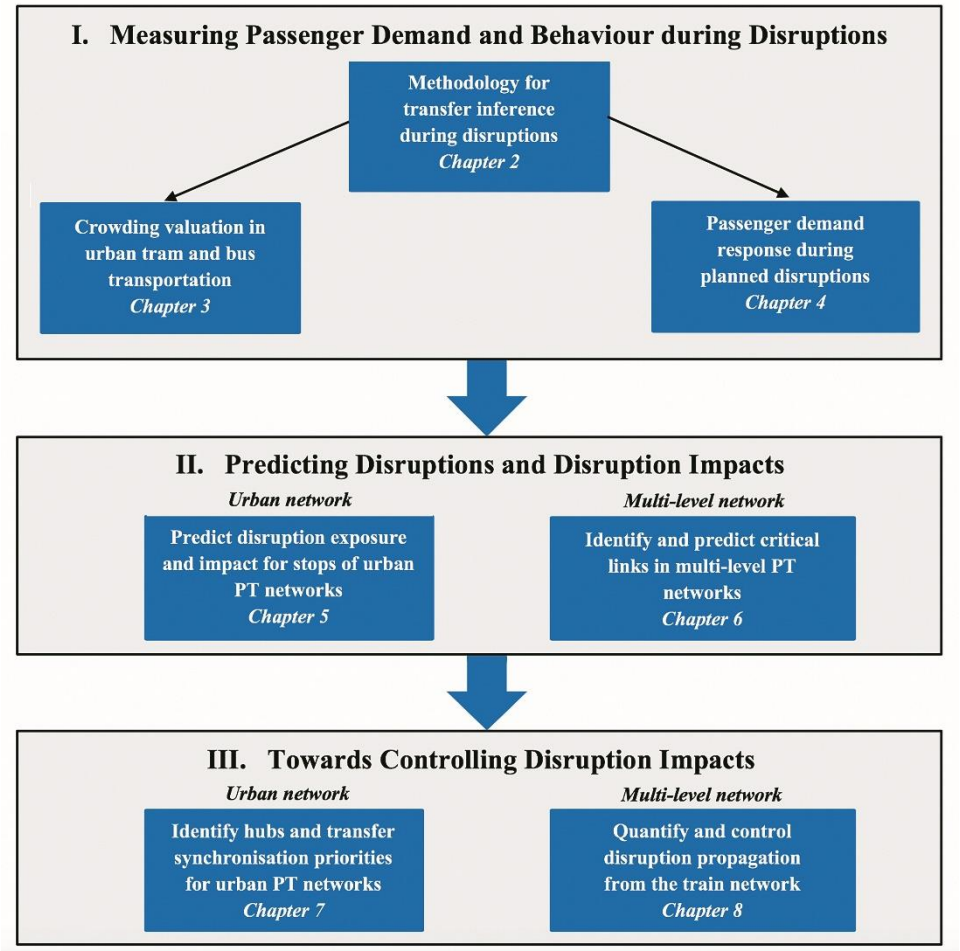


Figure 1.5. Research structure

Part I

Measuring Passenger Demand and Behaviour during Disruptions

2. A Robust Transfer Inference Algorithm for Public Transport Journeys during Disruptions

This chapter is the first component of Part I of this research, which contributes to answering the first research question (as defined in **Section 1.3**): how can we measure and characterise the behavioural and demand response of passengers during planned and unplanned urban public transport disruptions? The scientific contribution of this chapter is the development of a transfer inference algorithm to infer passenger journeys during both disrupted and undisrupted circumstances. This is based on an empirical study to passenger route choice during disruptions in The Hague, the Netherlands, based on data from Automated Fare Collection and Automated Vehicle Location systems. This can be considered a first step to correctly measure passenger disruption impacts. When comparing the realised and scheduled passenger journey times for each passenger journey, comparing the correct journeys in each scenario is essential. An incorrect inference of passenger journeys during disruptions - where passenger behaviour typically differs from behaviour during regular, undisrupted circumstances - can result in an incorrect comparison and thus result in an incorrect measurement of passenger disruption impacts.

This chapter is based on an edited version of the following article:

Yap, M.D., Cats, O., Van Oort, N., Hoogendoorn, S.P. (2017). A robust transfer inference algorithm for public transport journeys during disruptions. *Transportation Research Procedia, 27*, 1042-1049.
© 2017 The Authors. Published in Elsevier B.V.

2.1 Introduction

Disruptions in public transport can have a major impact on passengers' nominal and perceived journey time. The operation of public transport services without disruptions is considered a key quality aspect of public transport by passengers (Golob et al., 1972; Van Oort, 2011). Therefore, it is important to get insight in passenger behaviour during disruptions. Passive data availability the last decades provides an opportunity to get more insight in this. Automated fare collection (AFC) data, automated vehicle location (AVL) data, and data from automated passenger count (APC) systems are used for many purposes by scientists and practitioners on a strategic, tactical and operational level (Pelletier et al., 2011). Passive data availability allows for a comparison between the realised journey time during a disruption and the undisrupted travel time on an individual level, and therefore enables quantification of disruption costs.

A first important requirement for this comparison is that journeys can be inferred in a valid way. When no valid distinction is made between transfers and destinations, this can result in a biased journey identification and thus a biased journey level quantification of disruption impacts. Last decade several studies are performed to estimate origin-destination (OD) matrices based on individual AFC transaction data (see for example Trépanier et al., 2007; Zhao et al., 2007; Seaborn et al., 2009; Wang et al., 2011; Munizaga and Palma, 2012; Gordon et al., 2013; Nunes et al., 2016). These studies propose advanced algorithms to infer journeys from passive data for regular circumstances. However, these algorithms are based on a certain logic in passenger route choice. For example, when the next transaction is made in the same public transport line as the previous transaction, current algorithms infer an activity since there is no other reason why passengers would alight from a vehicle and then board a next vehicle of the same line again (Gordon et al., 2013). However, during disruptions passengers might have to adjust their route choice due to limited service availability, which can result in routes which would be illogical in case there were no disruptions. For example, due to operator rescheduling measures as deadheading or short-turning during disruptions, passengers might have to make an additional transfer to the subsequent vehicle of the same line. This means that the logic on which current transfer inference algorithms are based is not suitable to infer transfers during disruptions, given the illogical route and transfer choice passengers might be forced to during disruptions. Applying existing algorithms leads to biased transfer inference and thus to a biased OD matrix estimation. As a consequence, quantifying disruption costs on an OD level will be biased as well.

To be able to infer transfers during disruptions therefore places additional challenges to transfer inference algorithms, since these algorithms must be robust to infer transfers during disruptions, while still providing valid results for undisrupted situations as well. This is necessary, since it is often difficult to infer the exact time demarcation between disrupted and undisrupted circumstances from disaggregated AVL and AFC data sources. This research develops such transfer inference algorithm. **Section 2.2** discusses the methodology. The developed algorithm is applied to a case study, of which results are presented in **Section 2.3**. Conclusions and recommendations for further research are formulated in **Section 2.4**.

2.2 Methodology

2.2.1 The Hague case study network

In our study we use passive data from HTM, the urban public transport operator of The Hague, the Netherlands. The urban network in The Hague consists of light rail, tram and bus lines. The set of public transport lines is denoted by L. Each public transport line $l \in L$ is defined as an

ordered sequence of stops $S_l = (s_{l,1}, s_{l,2}..s_{l,|l|})$. Each line $l \in L$ is operated by an ordered set of runs (run sequence), denoted by R_l. A run $r \in R_l$ is performed by one vehicle serving the ordered stop sequence S_l in one direction. For each run $r \in R_l$ there exists a schedule with scheduled arrival times $\widetilde{t_a}$ and scheduled departure times $\widetilde{t_d}$ for each stop $s_{l,j} \in S_l$.

When travelling in light rail, trams or buses in the Netherlands by smart card, passengers are required to tap in and tap out at devices which are located within the vehicle. This means that the passenger fare is based on the exact distance travelled in a specific public transport vehicle. Especially for buses, this is different from many other cities in the world where often an open, entry-only system with flat fare structure is applied, for example in London (Gordon et al., 2013) and Santiago, Chile (Munizaga and Palma, 2012). This means that for each individual transaction the boarding time and location, and the alighting time and location of each journey stage are known. Also, it is known in which public transport line, vehicle number and trip number (a unique number assigned to each one-directional run $r \in R_l$) each passenger boarded with their unique smart card number. The AVL data provides the scheduled times $\widetilde{t_a}$ and $\widetilde{t_d}$, and the realised times $t_a$and t_d for each run at each stop, where each run is indicated by the same trip number as appears in the AFC data. By integrating AFC and AVL data based on the corresponding trip number, vehicle occupancy can be inferred for each run between each stop $s_{l,1}, s_{l,2}$.

2.2.2 Full validation of destination inference algorithm

Before starting the analyses, data cleaning and data processing is required. First, transactions where a system error occurred are removed from the dataset. In these cases there occurred an error in the AFC devices, leading to unrealistic alighting times or alighting locations, or to missing or unrealistic trip numbers. For The Hague, this percentage varies between 0.05% and 0.50% of the daily transactions. The within-vehicle AFC systems implies that in general destinations of journey stages are directly available from the data, so no destination inference is needed. Therefore, destination inference needs to be performed only for transactions where there was a missing tap out. This occurs when passengers unintendedly forget to tap out when alighting from the vehicle, or deliberately do not tap out if the distance based travel costs are higher than the deposit deduced from the card when boarding for relatively long trips. The daily percentage of transactions with a missing tap out in The Hague varies between 1% and 2% on average. For destination inference we apply the well-known trip chaining algorithm as applied by Trépanier et al. (2007), Zhao et al. (2007) and Wang et al. (2011). The aim is to estimate the alighting stop $\hat{s}^a_{pjk}$ of the jth journey stage of the total number of journey stages m made by passenger p on day k. The indices s^a and s^b reflect the alighting and boarding stop, respectively. The following basic assumptions are applied in this algorithm:

- If $m > 1$ and $j \neq m$: the most likely alighting location of j is the stop which is closest to $s^b_{p(j+1)k}$.
- If $m > 1$ and $j = m$: the most likely alighting location of j is the stop which is closest to $s^b_{p(j=1)k}$. Assumed is that passengers return to the location where the first journey stage started (e.g. home) at the end of the day.
- If $m = 1$: trip chaining is not possible and no destination can be inferred. In that case, the transaction is removed from the dataset. Contrary to Trépanier et al. (2007), we did not incorporate travel behaviour made by the same card number on previous days in the algorithm. Since destination inference is not the main research goal of this study, we aimed to prevent too much noise in the dataset from complex destination inference algorithms.

The set of candidate stops A_{pjk} for $\hat{s}_j^a$ in case $m > 1$ and $j = m$ is shown by **Eq.1** and contains all stops from the registered boarding stop s_l^b at line l downstream to $s_{l,|l|}$. In case $m > 1$ and $j \neq m$ an additional constraint is added, which guarantees that the realised arrival time t_{as} of run r at stop s should be earlier than the boarding time at s_l^b of the next journey stage $j + 1$. This is expressed by **Eq.2**.

$$A_{pjk} = \{s_{lj}^b .. s_{lj}^+\},\ j = m_{pk} \tag{1}$$

$$A_{pjk} = \{s_{lj}^b .. s_{lj}^+\},\ j < m_{pk} \quad st.\, t_{asrj} < t_{dsr(j+1)} \tag{2}$$

The selection of $\hat{s}_{pjk}^a$ from A_{pjk} is based on minimising the Euclidean distance d between the candidate alighting stop and $s_{p(j+1)k}^b$ or $s_{p(j=1)k}^b$. We minimise the Euclidean distance, instead of the generalised travel time as proposed by Munizaga and Palma (2012) and Sánchez-Martinez (2017). Using generalised travel time is mostly beneficial if the set of candidate stops contains stops of a public transport line in both directions. Minimising the Euclidean distance could then infer a stop of the line in the opposite direction which is just slightly closer to the next boarding location, while neglecting the substantially longer in-vehicle time to reach that stop. Since our candidate set is one-directional and only contains stops downstream the boarding location, we can minimise the Euclidean distance without problems. A maximum walking distance threshold d_{walk} is applied. If no candidate stops can be found within a reasonable walking distance, it is likely that this passenger used another mode as intermediate journey stage. In that case no destination can be inferred. **Eq.3** shows the applied destination inference algorithm.

$$\hat{s}_{pjk}^a = argmin\{d(s_{p(j+1)k}^b, \hat{s}_{pjk}^a)\} \quad \forall\, \hat{s}_{pjk}^a \in A_{pjk}, m_{pk} > 1 \quad st.\, d \leq d_{walk} \tag{3}$$

Validation of the applied destination inference algorithms shows to be difficult in other studies. Inferred destinations can be validated with passenger counts in vehicles or at stops, or by using surveys to a small sample of the population. Besides, a variety of walking distance thresholds is applied, varying between 400m (Zhao et al., 2007), 750m (Gordon et al., 2013), 1000m (Wang et al., 2011; Munizaga and Palma, 2012) and 2000m (Trépanier et al., 2007). The fact that in the Dutch urban public transport network both tapping in and tapping out are required, however enables a full validation of the algorithm and allows for the selection of an optimal value for d_{walk} resulting in the most accurate destination inference. We selected all complete transactions made on the HTM network on one working day (≈286,000 transactions) and removed the alighting location. We applied the destination inference algorithm with varying values for d_{walk} to predict back these alighting locations and considered the percentage of destinations what was correctly, incorrectly and not inferred, respectively. **Table 2.1** and **Figure 2.1** provide the results. From **Table 2.1** can be seen that, depending on d_{walk}, in total between 70% and 87% of all destinations could be inferred. This is higher than percentages found by Trépanier et al. (2007) and Zhao et al. (2007) ranging between 66% and 71%. The higher d_{walk}, the more destinations could logically be inferred. However, with an increasing number of inferred destinations the number of incorrectly inferred destinations increases faster than the number of correctly inferred destinations. From all inferred destinations, the percentage correctly inferred drops from 71% for d_{walk}=200 to 65% for d_{walk}=1600. This shows there is a trade-off between the quantity and accuracy of inferred destinations.

To find the optimal d_{walk}, we maximise the number of correctly inferred destinations $\hat{s}_{pjk}^{a,c}$ corrected for incorrectly inferred destination $\hat{s}_{pjk}^{a,w}$, as shown by **Eq.4**. We increased d_{walk}

stepwise by 200m starting from 200 to 1600 Euclidean metres. **Figure 2.1** shows that this value is maximised when applying a maximum walking threshold of 400 Euclidean metres (on average ≈550 real metres). For d_{walk}=400 we investigated the error margins for a subset of 100 transactions. For 72% of incorrectly inferred destinations, the chosen destination was only one stop further upstream or downstream. This probably reflects passengers performing an activity between two stops and selecting the stop on the other side of the activity for boarding again.

$$d_{walk} = argmax\left(\hat{s}_{pjk}^{a,c} - \hat{s}_{pjk}^{a,w}\right),\ d_{walk}\{d_{200}, d_{400}..d_{1600}\} \tag{4}$$

Table 2.1. Destination inference results for varying values of d_{walk}

Variable	d_{200}	d_{400}	d_{600}	d_{800}	d_{1000}	d_{1200}	d_{1400}	d_{1600}
% inferred destinations	69.6%	76.4%	80.6%	83.2%	84.5%	85.6%	86.1%	86.6%
% correctly inferred from all inferred destinations	70.6%	70.1%	68.3%	66.9%	66.1%	65.7%	65.4%	65.1%
% correctly inferred from total transactions	49.1%	53.6%	55.0%	55.6%	55.9%	56.2%	56.3%	56.4%
% incorrectly inferred from total transactions	20.5%	22.8%	25.5%	27.5%	28.6%	29.4%	29.8%	30.2%
% not inferred from total transactions	30.4%	23.6%	19.4%	16.9%	15.5%	14.4%	13.9%	13.4%

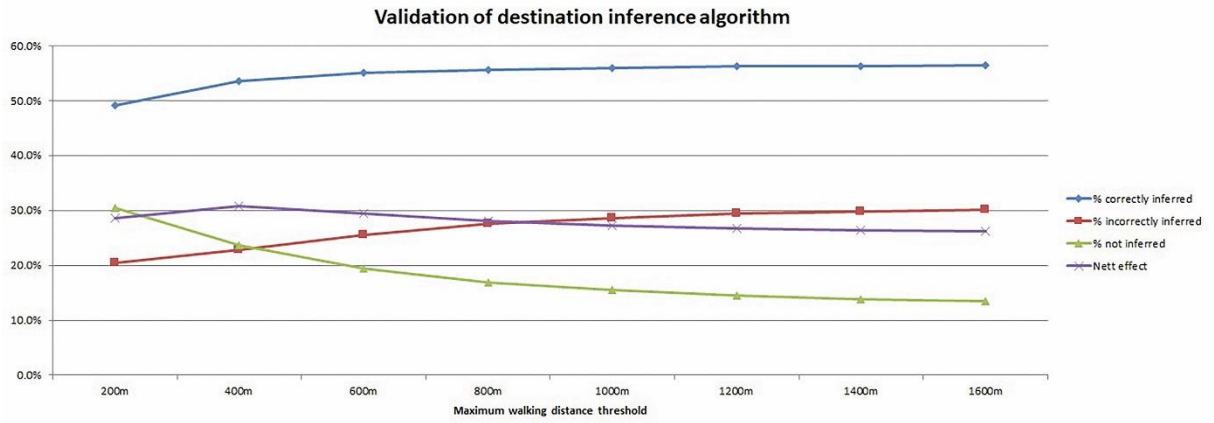


Figure 2.1. Performance of destination inference algorithm for different maximum Euclidean walking distance values

2.2.3 Robust transfer inference algorithm

We show the state-of-the-practice and state-of-the-art transfer inference algorithms and then illustrate limitations of these algorithms based on a theoretical network. The state-of-the-practice criterion to identify an alighting as transfer as applied in the Netherlands is based on a maximum time threshold between the previous tap out and next tap in with the same smart card ID. If the time between t_{asrj} and $t_{dsr(j+1)}$ is larger than a transfer threshold time $t_{t,max}$, the alighting is classified as activity. In the Netherlands, $t_{t,max}$=35 minutes. This criterion can lead to biased transfer inference, mainly because it tends to underestimate short journeys. If activities are performed which last shorter than 35 minutes, two separate journeys are incorrectly considered as one journey. If this is a back-and-forth trip of which s_{pjk}^{b}= $s_{p(j+1)k}^{a}$, this journey is not included in the OD matrix at all. Given the high frequent services in urban public transport, a transfer time of 35 minutes will not often be exceeded. During disruptions longer transfer times are however also possible, which could also overestimate the number of journeys.

When considering state-of-the-art transfer inference algorithms, three criteria are formulated which should all be satisfied to define an alighting as transfer.

- *Temporal criterion.* The temporal constraint as applied in practice is replaced by a criterion which expresses whether a passenger took the first passing vehicle at a transfer, by integrating AFC and AVL data (Gordon et al., 2013). Based on t_{apjk} and the stop coordinates of the alighting stop and next boarding stop, the first realistic passenger arrival time at the next boarding time can be calculated. In this study, we correct the Euclidean distance by $\sqrt{2}$ to obtain realistic transfer distances. We use the 2.5th percentile of the walking speed distribution derived by Hänseler et al. (2016) instead of the average walking speed, to prevent that on average 50% of transferring passengers might not be considered as transferring. The first realistic passenger arrival time at the next boarding stop is compared with realised vehicle departure times of the chosen line. If the first passing vehicle after the first realistic passenger arrival time is boarded, the alighting is considered a transfer. A minimum transfer allowance of 5 minutes is applied.
- *Spatial criterion.* The spatial criterion constrains the maximum transfer distance to d_{walk}. In this study we set d_{walk} to 400 Euclidean metres, in line with results from **Section 2.2.2**.
- *Binary criterion.* The binary criterion indicates whether the next boarding line is equal to the previous boarding line. In this case, the alighting is considered to be related to an activity. Successive services in opposite direction indicate a return trip from an activity, whereas successive services in the same direction also indicate a performed activity (Gordon et al., 2013).

We adjust this state-of-the-art algorithm to make the algorithm robust to transfer inference during disruptions. To illustrate the relevance of this algorithm, we use a selection of the HTM case study network as shown in **Figure 2.2**. It shows a light rail connection between the centre of the satellite city Zoetermeer (nodes A-B on the left) via intermediate stop D to the centre of a main city (nodes F-G on the right). There is also a separate transit line between D-E. The public transport lines A-B-D-F-G (line 4), C-B-D-F-G (line 3), and D-E (line 19) are operated by the same operator (HTM). This means that both AFC and AVL data from these lines are available. Stops C and F provide transfer connections to the train network, which stations are indicated by C1 and F1. The train service C1-F1 is operated by another operator, which means that only AVL data (open data in the Netherlands) is available.

We assume a disruption (e.g. a signal failure) occurs on the light rail track between D and F, leading to reduced capacity between D and F. In line with HTM disruption management, 50% of the light rail services eastbound short-turn at D, whereas 50% of the westbound light rail services short-turn at F. We consider three different passenger journeys over this network, shown by **Table 2.2**.

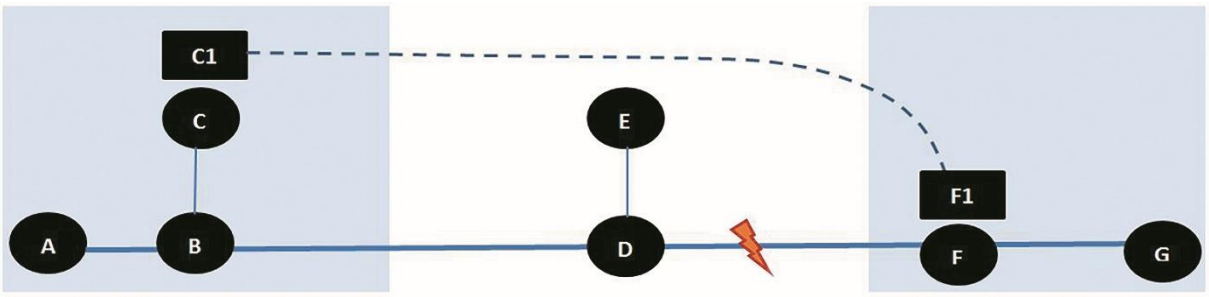


Figure 2.2. Selection of The Hague case study public transport network

Table 2.2. Illustration new transfer inference algorithm for three OD journeys E-G, B-G, A-G

Boarding stop	Alighting stop	Smart card ID	Transit line	Transfer? *Current algorithm*	Transfer? *New algorithm*
E	D	1233	19	No	Yes
D	G	1233	3		
B	C	1234	3	No	Yes
F	G	1234	4		
A	D	1235	3	No	Yes
D	G	1235	3		

- *Temporal criterion.* We consider a journey from E to G. Due to reduced services between D and F, remaining vehicles can get very crowded. This means that transferring passengers at D (from E) can experience denied boarding in busy urban networks and might not be able to board the first passing vehicle. It is also possible that some passenger decide to wait for a next service, if they notice the very crowded vehicle arriving at the station. Applying the original temporal criterion would therefore incorrectly classify the alighting of these passengers at D as activity, since they did not take the first passing vehicle. Under regular circumstances, denied boarding in the Netherlands is very exceptional. However, in disrupted situations the frequency of denied boarding and very crowded vehicles increases substantially. To account for this we adjust this criterion such that an alighting is considered a transfer, if a passenger takes the first *reasonable* passing vehicle at a transfer location. Boarding a reasonable vehicle is quantified by adding an extra constraint to the temporal criterion. An alighting is considered a transfer if a passenger boards the first vehicle of a service after the first realistic passenger arrival time at the stop, of which the occupancy is lower than the norm capacity. By integrating AFC and AVL data, vehicle occupancies are derived. When occupancies are higher than the norm capacity, it can be expected that passengers decide to skip this vehicle or are even denied boarding.
- *Spatial criterion.* We consider a journey from B to G. These passengers adjust their route choice, by using the train network at the side of the city centres as alternative. By transferring from C to C1, and back from F1 to F, the disruption is avoided. Especially in dense urban networks, passengers can use the total multi-level public transport network which remains available after a disruption (see e.g. Cats et al., 2016b). Since train services are operated by another operator, no AFC data of this journey stage is available. Since the distance between C and F is substantially larger than common values for d_{walk} (in this study 400m), applying the original spatial criterion would incorrectly classify the alighting at C as journey destination and categorise the trip F-G as new journey. Since urban public transport is fully covered by AFC data, there is only the higher-level train network as multi-level alternative which is not covered in the data. Therefore, we add a binary indicator to each stop which equals 1 if a train station is located within the maximum transfer distance d_{walk}. If both $\boldsymbol{s}^{a}_{pjk}$ and $s^{b}_{p(j+1)k}$ are equal to 1, we apply the temporal condition as explained above. We determine, given the train AVL data, whether the boarding time of a passenger in the urban network in F shows that this passenger took the first reasonable train alternative from C1 to F1. If the realised boarding time at F does not exceed the expected travel time given the realised train departure and arrival times, it is likely that another public transport mode is taken as intermediate journey stage. The alighting at C is then considered a transfer. A further relaxation of the original spatial criterion is applied, by considering a tap out and consecutive tap in to the same vehicle and trip number as transfer, even if the transfer distance exceeds d_{walk}. This indicates that a passenger did not really alight the

vehicle, but (un)intendedly tapped out and in during the ride. This can be explained by passengers in doubt if the tap in was successful, holding their card to the device again and then tapping out. Another explanation relates to passengers who deliberately tap out during a part of the trip to save travel costs, given the fully distance based fare.

- *Binary condition.* We consider a journey from A to G. A part of the passengers who keep using the light rail service have to alight from their short-turning vehicle at D and wait for a next service of this line headed for G. This means that the disruption forces these passengers to make a transfer to the same line 4. The original algorithm would therefore incorrectly infer the alighting at D as journey destination. We therefore adjust the algorithm such that when a transfer to the same line is made, this is considered a transfer if and only if a passenger boards the first run of this same line after the alighted run. This allows transfer inference in case of rescheduling measures as short-turning, stop-skipping or deadheading. By only allowing the first run after the alighted run as transfer, in high frequency urban networks the headway will be short. Measures as short-turning or deadheading are in practice especially performed if the next run already bunches behind the previous one. This means that false positive transfer inferences are very unlikely, since the time to perform an activity will be very short. This adjustment is also relevant in case part of the disrupted track is replaced by bus services operating under the same line number. Besides, this adjustment can be of relevance during undisturbed situations in case of non-typical route topologies, in which transfers to the same line number occur. For example, in case of lines with loops or in case of lines where short-services are operated under the same line number, it can occur that passengers have to transfer to a vehicle of the same line under planned circumstances.

2.3 Results

We compare the performance of the state-of-the-practice, state-of-the-art and newly developed transfer inference algorithm. For this aim, we use a dataset consisting of individual AFC transactions during two different non-recurrent disruptions which occurred on the HTM network in November 2015. The dataset contains transactions from passengers who specifically travelled over one of the disrupted lines and disruption location during one of these disruptions (in total ≈23,300 transactions). We applied the three different transfer inference algorithms to this dataset. For the state-of-the-practice algorithm we used a maximum transfer time of 35 minutes. For the state-of-the-art and new developed algorithm we set d_{walk} equal to 400 Euclidean metres. Based on a normal distributed walking speed $N(1.34, 0.34)$ we applied a 2.5th percentile walking speed of 0.66 m/s (Hänseler et al., 2016).

Table 2.3. Performance comparison between three different transfer inference algorithms

	State-of-the-practice	State-of-the-art	New developed algorithm
Average # trips per journey	1.44	1.18	1.21
% journeys with 0 transfers	68%	85%	82%
% journeys with 1 transfer	23%	14%	15%
% journeys > 3 transfers	0.8%	0.0%	0.1%

Table 2.3 shows that the algorithm as currently applied in practice results in a relatively high number of trips per journey (1.44), indicating that alighting is relatively often classified as transfer. This is because of the applied transfer threshold time without further behavioural rules. In the case study network, journeys with more than 3 transfers are highly unlikely. Only additional transfer behaviour during disruptions could require more than 3 transfers in total in rare occasions. Given that 0.8% of all journeys have more than 3 transfers, it shows that this state-of-the-practice algorithm overestimates transfer behaviour. The state-of-the-art algorithm

on the other hand shows less transfers and no journeys with more than 3 transfers. However, this algorithm is too strict in classifying an alighting as transfer, especially in case of disruptions. In our developed algorithm there is some relaxation of the transfer criteria of the state-of-the-art algorithm. This results in a slightly higher average number of trips per journey (1.21 instead of 1.18). As shown in **Section 2.2.3**, this new developed algorithm results in an improved transfer inference during disruptions. Besides, it also shows that still hardly journeys with >3 transfers are identified when applying this new algorithm, despite these relaxations.

Figure 2.3 expresses the ratio between travelled distance by public transport and the Euclidean distance per identified journey. Journeys with a very high ratio are symptomatic for two separate journeys which are incorrectly identified as one (i.e. a back-and-forth travel identified as one journey). Thus, this ratio can be used to validate the performance of transfer inference algorithms. As can be seen, both the state-of-the-art and new algorithm prevent journeys with unrealistic high ratios, which are found using the state-of-the-practice algorithm. Our new proposed algorithm thus improves transfer inference during disruptions, without compromising on general inference quality.

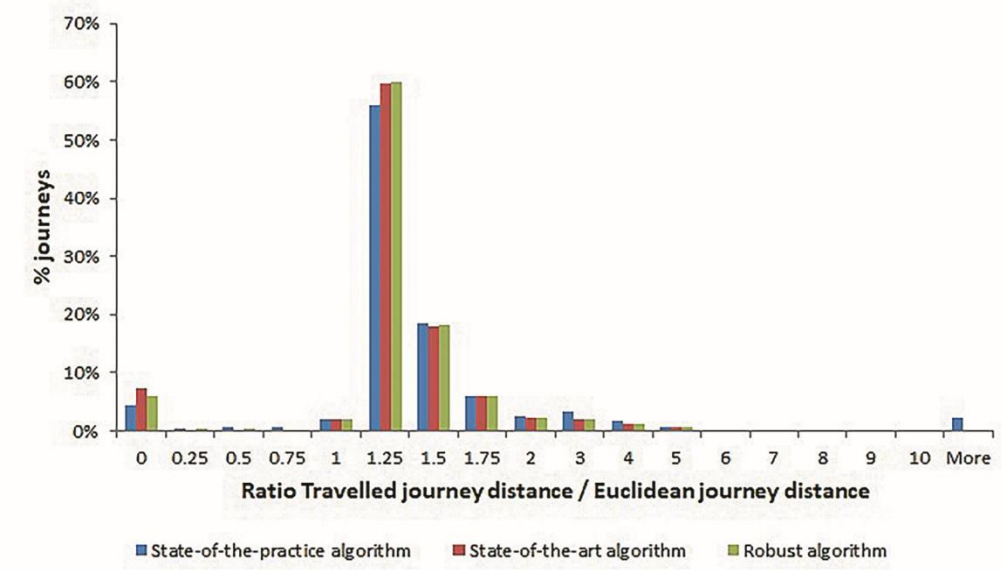


Figure 2.3. Distribution of ratio travelled/Euclidean distance for state-of-the-practice, state-of-the-art and new robust transfer inference algorithm

2.4 Conclusion

Several rule-based algorithms exist to infer whether a passenger alighting and subsequent boarding is categorised as transfer or final destination where an activity is performed. Although this logic can infer transfers during undisrupted public transport operations, these algorithms have limitations during disruptions: disruptions and subsequent operational rescheduling measures can force passengers to travel via routes which would be non-optimal or illogical during undisrupted operations. We develop a new transfer inference algorithm which infers journeys from raw smart card transactions in an accurate way during both disrupted and undisrupted operations. In this algorithm we incorporate the effects of denied boarding, transferring to a vehicle of the same line, and the use of public transport services of another operator on another network level as intermediate journey stage during disruptions. A further validation of the proposed transfer inference algorithm is recommended for future research.

3. Crowding Valuation in Urban Tram and Bus Transportation based on Smart Card Data

The purpose of this chapter is to determine how crowding is valued by passengers during urban tram and bus journeys entirely based on observed route choices and observed attribute values for each route alternative. This chapter belongs to Part I of this research and contributes to answering how passengers' behavioural response can be characterised during disruptions (Research Question 1 as defined in **Section 1.3**). This chapter uses the results from the transfer inference algorithm as proposed in **Chapter 2** as input. Applying this transfer inference algorithm results in individual passenger journeys between different origin-destination pairs inferred from individual transactions from Automated Fare Collection systems. For selected origin-destination pairs, these inferred passenger journeys together with their observed route choice are used in this chapter to estimate a discrete choice model to determine how crowding is valued and incorporated in passenger route choice. A better understanding of crowding valuation based on realised passenger route choice as performed in this study - instead of using stated choices from Stated Preference experiments - is particularly relevant to measure public transport disruption impacts accurately, as disruptions typically result in increased crowding levels on public transport services.

This chapter is based on an edited version of the following article:

Yap, M.D., Cats, O., Van Arem, B. (2018). Crowding valuation in urban tram and bus transportation based on smart card data. *Transportmetrica A.* DOI: 10.1080/23249935.2018.1537319
© 2018 The Authors. Published by Informa UK Limited, trading as Taylor & Francis Group

3.1 Introduction

Crowding in public transport can have major influence on passengers' travel experience and therefore affect route and mode choice. Because of the expected increasing concentration of activities within urban agglomerations in many countries worldwide, crowding is expected to become an even more dominant factor in urban public transportation in the future. Therefore, it is important to understand and quantify how crowding in urban public transport is perceived by passengers. This can contribute to the quantification of societal benefits of measures aiming to alleviate passenger congestion, and thus potentially support policy makers in the decision-making process regarding the implementation of such measures (see for example Prud'Homme et al., 2012; Haywood and Koning, 2015; Cats et al., 2016a).

Congestion and crowding are not always incorporated in public transport modelling, contrary to highway modelling. In highway modelling, link travel times are a direct function of (amongst others) the link congestion level. In public transport modelling, congestion and crowding levels influence passengers' perceived in-vehicle times. Public transport crowding only affects the nominal travel times if dwell times would increase or in more extreme cases of denied boarding, what can result in costs for passengers and operators. Thus, in public transport there is a behavioural relation instead of a traffic flow relation between crowding and (perceived) travel times.

The majority of studies which do incorporate public transport crowding use stated preference (SP) experiments for crowding valuation. For example, MVA Consultancy (2008), Whelan and Crockett (2009), Batarce et al. (2016) and Tirachini et al. (2017) use SP surveys where crowding is represented by pictures of crowding levels in public transport carriages, and crowding valuation is expressed using the average number of standing passengers per square metre. Lu et al. (2008) express crowding as the probability on occurrence (e.g. 2 out of 5 times) in a SP experiment, whereas Li et al. (2017a) describe and show different crowding levels using colours. Douglas and Karpouzis (2005) use pictures of crowding in their SP experiment, and estimate crowding in-vehicle time multipliers for different levels of crowding (e.g. seated no crowding, seated crowding, standing, crush standing) for different durations (e.g. during 10 or 20 minutes). In some studies where crowding valuation is estimated based on SP experiments, results are validated against revealed preference (RP) data, for example by using surveys, passenger observations or cameras. For example, Kroes et al. (2014) validate SP based crowding estimates for Ile-de-France based on observed behaviour of passengers on platforms skipping a very crowded service and waiting for a next less crowded service. Batarce et al. (2015) estimate a mixed SP/RP model for the case of Santiago, Chile, combining a SP survey with pictured crowding levels and revealed passenger route choice based on smart card data. These mainly SP based results are applied in public transport models aiming to improve passenger assignment and predictions, see for example Hamdouch et al. (2011), Schmöcker et al. (2011), Nuzzolo et al. (2012), Pel et al. (2014), Cats et al. (2016a) and Van Oort et al. (2016). Extensive literature reviews regarding crowding valuation studies can be found in Wardman and Whelan (2011) and in Li and Hensher (2011).

We conclude that most crowding valuation studies are based on SP experiments. Studies which use RP data generally only use a relatively small RP dataset to validate SP estimates. It is however known that there can be a discrepancy between stated choices of respondents in SP experiments, compared to revealed choice behaviour in reality. This is a systematic discrepancy in which SP experiments tend to overestimate values compared to observed valuation in reality. This discrepancy can occur if respondents have difficulties imagining the stated, hypothetical choice situation, or if respondents lack sufficient experience with similar circumstances in reality to fully understand the trade-offs between the attributes in the choice set. The shortage of RP data usage in studies concerned with passenger perception of on-board crowding arguably

stems from the sparsity and difficulty to obtain passenger-related data in many public transport systems. However, the increasing availability of automated fare collection (AFC), automated vehicle location (AVL) and automated passenger count (APC) systems in public transport enables the estimation of crowding valuation fully based on large scale RP data. In particular, the availability of individual smart card transactions allows gaining insights into revealed trade-offs between travel time, transfers, waiting time and crowding in public transport route choice. Only a limited number of studies exploited smart card data for this purpose, namely the works by Hörcher et al. (2017) and Tirachini et al. (2016). Hörcher et al. (2017) estimated discrete choice models incorporating crowding valuation in metro systems by fusion of AFC and AVL data. In their study, 32 origin-destination (OD) pairs of the MTR metro network of Hong Kong are used to estimate the valuation of the standing probability and crowding density (expressed as the number of standing passengers per square metre). Since the AFC system in Hong Kong is a closed, station-based system where passengers have to tap in and tap out at the metro station, the exact route and trip choice had to be inferred using a passenger-to-train assignment method. Tirachini et al. (2016) estimated the passenger valuation of crowding for metro travelling on the Singapore network, fully based on observed behaviour of passengers willing to travel in the opposite direction first to secure a seat in the trip in their preferred direction.

The contribution of our study is that we estimate crowding valuation associated with urban tram and bus journeys, in a European context, entirely based on revealed route choice behaviour obtained from AFC and AVL data of the urban public transport network of The Hague, the Netherlands. We show additional evidence for the existing discrepancy between using RP and SP data for crowding valuation. Since passengers are required to tap in and tap out on-board each tram or bus vehicle, no inference of the exact route or the exact vehicle choice is required. By fusing AFC and AVL data, we directly determine the exact route and vehicle each passenger used, and deduce the stop-to-stop vehicle occupancy for each individual vehicle trip. This means we can eliminate one inference step of the methodology applied by Hörcher et al. (2017), thereby reducing potential uncertainty. We apply our methodology to a high-density public transport case study network, in which there is a large number of OD pairs between which different route choices can be observed.

This chapter is structured as follows. **Section 3.2** describes the methodology, including data processing (**Section 3.2.1**), transfer inference (**Section 3.2.2**), selection of OD pairs (**Section 3.2.3**), determination of attributes and attribute levels (**Section 3.2.4**) and choice model formulation (**Section 3.2.5**). **Section 3.3** provides the estimation results and discusses their implications. In **Section 3.4**, conclusions and recommendations for further research are formulated.

3.2 Methodology

3.2.1 Raw data semantics and processing

When travelling by light rail, tram or bus in the Netherlands, passengers are required to tap in and tap out at devices located within each vehicle. This means that there is an entry-exit, distance based fare system applied in The Hague. This is different from most urban public transport systems in the world, in which especially for buses often an open, entry-only system with flat fare structure is applied. This can for example be seen in London (Gordon et al., 2013) or in Santiago, Chile (Munizaga and Palma, 2012). The fare system as applied in the Netherlands means that for each journey leg made by each individual travelling by the abovementioned modes the boarding time and boarding location, as well as the alighting time and location, are directly available from the raw dataset. Besides, for each smart card transaction

the line number, vehicle number, trip number and smart card number are known. This means that each journey leg with its corresponding transaction information is registered as a separate row in the AFC dataset. In the Netherlands, the AFC data is closed data owned by the public transport operator. The AVL data, on the other hand, is open data and publicly available. The AVL dataset contains the scheduled and realised arrival time and departure time of each vehicle trip at each stop. In case spatiotemporal passenger information would not be directly available from the AFC system, boarding time and location, alighting time and location, and transfer times can be inferred. For example, Tu et al. (2018) propose a methodology to infer the boarding location for bus travels by fusion of smart card data and GPS trajectories, whereas Zhang et al. (2016) developed an approach to extract passengers' spatiotemporal information by inference of boarding times and transfer times.

We use the urban public transport network of The Hague, the Netherlands, to estimate public transport crowding valuation. This network consists of 12 tram lines and 8 bus lines, operated by the urban public transport operator HTM (**Figure 3.1**). Two of these 12 tram lines function as light rail service at the urban agglomeration network level, connecting The Hague with the satellite city of Zoetermeer. The other 10 tram lines and all bus lines function as urban lines within The Hague and neighbouring towns. During an average working day, more than 300,000 AFC transactions are made on the urban public transport network in The Hague.

AFC and AVL data of 28 days from November 2 – November 29, 2015 was made available by the incumbent operator for this study. This amounts to a dataset that consists of approximately 7.4 million AFC transactions and about 3.1 million AVL registrations. In the data processing phase, we apply the following steps:

- Selection of morning peak data (07:00 – 09:00).
- Removal of morning peaks with disruptions.
- Removal of incomplete AFC transactions.
- Inference and increase of occupancy data.

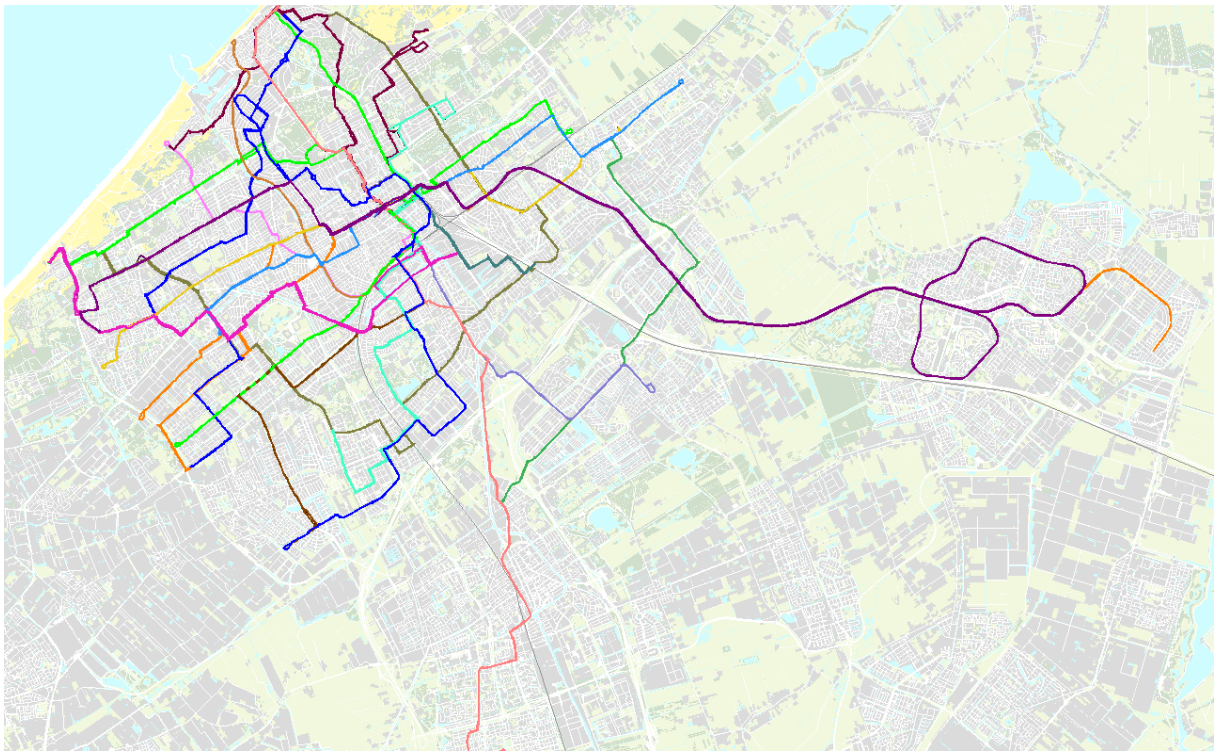


Figure 3.1. Overview of urban tram and bus lines of case study network The Hague

Since the aim of our study is to explore how passengers incorporate crowding in their route choice, it is essential that we select a time period in which crowding occurs. Compared to crowding levels reached in metro systems in cities like London, Santiago, Beijing or Tokyo, the level of crowding in The Hague can be considered quite moderate. Outside peak periods, in general no crowding occurs. In peak periods, crowding however does occur on several lines. Since public transport demand in the morning peak in the Netherlands is more concentrated within a relatively small period, compared to a more uniformly distributed demand in the evening peak, we only focus on AFC transactions during the morning peak. This means that only journeys of which the tap in record time is between 07:00 and 09:00 on working days (Monday – Friday) are considered.

In our study we focus on explaining route choice based on expected attribute values for travel time, waiting time and crowding. Therefore, it is important that only regular, undisrupted periods are incorporated in the dataset. Since disruptions can force passengers to adjust their route choice, this might cause bias in the analysis. Based on the operator log file, containing all registered disruptions with corresponding time, duration and location, we removed the AFC data from morning peaks of days where a disruption occurred. Given the possibility of second-order effects, in which a disruption on a certain public transport line might increase occupancies on other parallel lines, crowding levels on directly and indirectly affected lines can then deviate from expected crowding levels passengers have for undisrupted days (Malandri et al., 2018). Therefore, we adopted this conservative approach of excluding working day data if any disruption occurred anywhere on the considered case study network. From the 20 working days remaining in the dataset, data from 6 working days has been removed.

For the AFC data of the remaining 14 morning peaks, we removed incomplete transactions. Transactions can be incomplete due to a system error or due to a human cause (human error by forgetting to tap out, or deliberately not tapping out), in both cases leading to a missing tap out time and/or location. Although there exist many destination inference algorithms in scientific literature (e.g. the well-known trip chaining algorithm as applied by Trépanier et al., 2007, Zhao et al., 2007 and Wang et al., 2011), we removed all incomplete transactions, since the percentage of incomplete AFC transactions in The Hague is with 1.9% rather low due to the entry-exit AFC fare system. From studies validating destination inference algorithms (e.g. Munizaga et al., 2014; Yap et al., 2017), we know that between 65% and 85% of the destinations for urban tram or bus journeys are correctly inferred. By removing incomplete transactions, we avoid introducing inaccuracies resulting from possibly incorrect destination inference. Besides, given the entry-exit AFC system, the percentage of incomplete AFC transactions in The Hague is with 1.9% rather low and not expected to systematically influence results. In the Dutch public transport system a deposit is written off from each smart card when tapping in as incentive for passengers to tap out due to its distance based fare system, which is transferred back to the card when tapping out again. Hence, we expect no correlation between (higher) crowding levels and (lower willingness of) passenger tap out behaviour.

Vehicle occupancies are derived by fusion of AFC and AVL data. Since both the AFC and AVL data contain the trip number, both data sources can be coupled. This results in the occupancy of each vehicle trip between each pair of stops. These smart card data based occupancies are corrected for the percentage travellers not using a smart card, and the percentage of incomplete AFC transactions. In the Netherlands, most passengers travel using their smart card. Only passengers who buy a ticket in the vehicle at the driver or vending machine, and passengers who (un)deliberately do not tap in during their trips are not captured. Since our study focuses on experienced crowding levels, these passenger groups are however relevant. Based on passenger counts performed by the urban PT operator, for each public transport line a correction factor is determined which can be used to increase the smart card based occupancies. This factor varies between 5% and 14% for different lines. This correction

factor is applied uniformly on a line-level, which implicitly assumes that the destination choice behaviour of non-smart card users is indistinguishable from the one characterising smart card users. Since our dataset contains transactions of morning peak periods in November, the share of very infrequent passengers (e.g. tourists) for whom the abovementioned assumption might not apply is arguably very low.

3.2.2 Transfer inference

For the individual AFC transactions for each passenger trip, it is determined whether an alighting is a transfer or a final destination. To this end, we applied the transfer inference algorithm as described by Yap et al. (2017). This algorithm is an extension on the algorithm described by Gordon et al. (2013), and uses AFC, AVL and inferred occupancy data. Below, this algorithm is shortly discussed. For a more detailed explanation, the reader is referred to Yap et al. (2017). The algorithm consists of temporal, spatial and line-based criteria whether to classify an alighting activity as a transfer.

- Temporal criterion: an alighting activity is considered a transfer if the passenger boarded the first feasible vehicle of the next tap in line serving the next boarding location. A feasible vehicle is defined as the first vehicle serving the next boarding location - given the alighting time of the previous journey leg and required walking time - of which the occupancy does not exceed the norm capacity.
- Spatial criterion: an alighting is considered a transfer if the distance between the alighting location and the next boarding location does not exceed a maximum transfer walking distance of 400 Euclidean metres. An exception to this threshold is made in case passengers use an intermediate public transport service offered by another operator, for which no AFC data is available.
- Line-based criterion: an alighting is not considered a transfer if the next boarding is on the same line as the previous journey stage, since this suggests that an activity was performed in between. An exception is made in case of boarding the first passing vehicle of the same line directly following the alighted vehicle, since this can indicate (for example) a transfer from a short-service to the long-service vehicle of the same line or a transfer to the same line in case of loops. Given the relatively high frequencies of urban public transport lines, it is highly unlikely that such alighting and boarding of the next vehicle will be an activity.

In total, the dataset contains 628,839 journeys resulting from 14 working days. **Figure 3.2** (left) shows the resulting distribution of the number of transfers, whereas **Figure 3.2** (right) shows the distribution of journeys over each 30-minute period of the morning peak (using the journey tap in time for aggregation). As can be seen, almost all journeys consist of 0 or 1 transfer. The busiest part of the morning peak on a network wide level is between 08:00 and 08:30, containing 31% of all morning peak journeys.

The boarding stop, alighting stop, and the travelled route and line of each passenger journey leg are directly observed from the AFC data, because of the entry-exit smart card regime with on-board devices for tap in and tap out. Applying our study to the The Hague case study network has therefore the advantage that no inference is required to determine passenger route choice, so that our study fully relies on directly observed route choice. Vehicle occupancies are also the direct result of fusing empirical AFC and AVL transactions, without relying on any further inference. The transfer inference algorithm, which is the only inference algorithm applied to the dataset, only relates to the interpretation whether an alighting is considered a transfer or final destination, but is not consequential in determining route choice or occupancies.

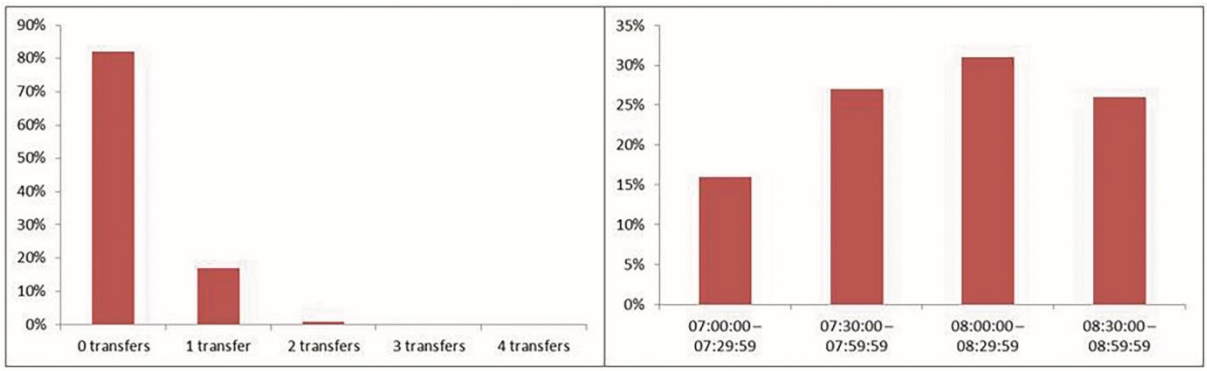


Figure 3.2. Journey distribution by number of transfers (left) and distribution per half-hour of the morning peak (right)

3.2.3 Selection of origin-destination pairs

In total, the database consists of 49,231 different chosen routes. This value can be considered as the sum-product of the number of chosen OD pairs and the number of chosen route alternatives for each OD pair. Different routes are considered to belong to one and the same OD pair, if the boarding stops are located in each other's vicinity, as well as the alighting stops. The following criteria are used to select OD pairs to be included in the estimation of the discrete choice model with crowding effects:

- Minimum choice set size of 2 for each OD pair.
- Minimum number of 100 observed choices for each OD pair.
- Minimum observed choice probability of 0.1 for each route alternative in the choice set.
- Minimum expected seat occupancy of 50% on minimal one link of one route alternative.
- Attribute variation over all OD pairs.

In our study we only consider the observed route choice set. This prevents making assumptions regarding the incorporation of non-observed, possibly feasible route alternatives in the choice set and allows us to infer attribute levels for all alternatives entirely based on observed AFC and AVL data. The first criterion thus means that there should be an observed choice between at least two route alternatives for a given OD pair. A route is defined as a unique sequence of boarding locations, alighting locations and intermediate (combination of) lines. In case of routes with the same origin stop and the same destination stop, the routes should physically differ from each other, to be considered as separate route alternative in the choice set. In case of a bundle of different lines sharing the same infrastructure (e.g. for journeys within the city centre), passengers might take the first arriving vehicle suitable for their destination. Given our aim to explain route choice, we consider route alternatives as different from each other, only if these do not share the exact same geographical path (i.e. 100% overlap).

A minimum number of 100 observed choices for each OD pair is deemed necessary to incorporate sufficient choices of individual passengers to infer generic coefficients from. By setting a minimum value for the observed choice probability of 10% per chosen route alternative, we ignore route alternatives which are chosen only in a very limited number of cases. Such route alternatives might be chosen in case of (non-registered) delays, disruptions or a-typical passenger behaviour (e.g. passengers travelling for fun). Another requirement is that there should be at least a certain amount of crowding expected on (at least) one of the route alternatives of an OD pair. Since we want to estimate crowding valuation, we did not want to set a crowding constraint *a priori*. However, if no crowding occurs on all route alternatives at all, it is not possible to examine the influence of crowding on route choice. Therefore, we opted

for a light requirement here. If the expected seat occupancy exceeds 50% during some period of the morning peak on some part of one of the observed route alternatives, this requirement is fulfilled. A seat occupancy threshold of 50% is used, since passengers start having to sit next to each other from a seat occupancy of 50% or higher. It is expected that negative crowding experiences might arise from this value onwards. The robustness of estimation results towards the latter assumption has been investigated in a sensitivity analysis. Reducing this threshold value by 20% (to 40%) showed not to influence the number of OD pairs and observations in the dataset. An increase of this threshold by 20% (to 60%) resulted in only one OD pair not satisfying this criterion anymore, thereby reducing the number of observations in the dataset by 2%. Since the total number of observations in the dataset is hardly influenced by this parameter value, this sensitivity test attests to the robustness of the estimation results to different values of this threshold. At last, we checked over all remaining OD pairs together whether they contain all attributes of the discrete choice model to be estimated. This means that at least some OD pairs should consist of a transfer, tram or bus, in order to be able to estimate the valuation of a transfer or to estimate the in-vehicle time perception in a tram compared to a bus.

Applying the abovementioned criteria results in 58 remaining OD pairs, with a total of 17,994 journeys (= 17,994 observed choices). The route set of 16% of these OD pairs consists of at least one route that involves a transfer. These 17,994 journeys are made by 7,083 different smart card numbers. Under the assumption that each passenger uses one, unique smart card, this means that there are on average ≈2.5 observations per passenger in the total dataset.

3.2.4 Attributes and attribute levels of route choice alternatives

In this section, we discuss the different attributes and attribute levels for the estimated models.

In-vehicle time

The expected in-vehicle time t^{ivt} is determined for each journey leg of each route alternative separately. Based on the AVL data, we calculate the expected in-vehicle time by taking the average realised in-vehicle time over all observations using this route alternative for this OD pair. This means we use the expected in-vehicle time rather than the scheduled in-vehicle time as value for t^{ivt}. Since the scheduled travel times are constant during the whole morning peak, t^{ivt} is calculated for the whole morning peak as well. For each journey leg, index m indicates whether the journey leg is made by tram or bus.

Waiting time

The expected waiting time t^{wait} expresses the initial waiting time before boarding the first journey leg. Since the AFC system in urban public transport in the Netherlands only has tap in / tap out devices on-board the vehicle, it is not possible to empirically derive the passenger arrival time at the initial boarding stop from the smart card data. In order to quantify the initial waiting time, we assume a random passenger arrival pattern. Given the high frequency of the urban public transport services in the considered case study network, this assumption is considered to be reasonable. Therefore, we set t^{wait} as equal to half of the scheduled headway of the boarding line in the corresponding time period.

Transfer time

The expected transfer time t^{trans} expresses the time between the alighting time from the first journey leg and the boarding time of the next journey leg of a certain route alternative. This means that t^{trans} equals the sum of the transfer walking time and transfer waiting time, and equals zero for a route alternative without transfers. The expected value t^{trans} for a certain

route alternative is calculated by taking the average realised transfer time over all observed morning peak journeys using this route alternative.

Number of transfers

n^{trans} is an integer variable which reflects the number of transfers of a certain route alternative, and equals 0 in case of a route alternative without a transfer. This variable is used to determine the perceived transfer penalty, which expresses the additional penalty associated with the inconvenience of performing a transfer beyond the additional (perceived) travel time it induces.

Path size

We calculate the path size factor to determine and correct for overlap between route alternatives of a certain OD pair. The natural logarithm of the path size factor, denoted by r, is used and incorporated into a standard MNL model. We quantify this commonality factor using the distance-based amount of overlap between route alternatives, as shown in **Eq.1**. We consider route alternative i from all route alternatives j of the observed choice set for a certain OD pair. Each route i consists of a sequence of links $a_i \in A_i$ with length l_a. The number of route alternatives of the choice set using link a is indicated by $|j|_a$. In case of two route alternatives without any overlap, r_i equals $\ln(1)$, whereas r_i equals $\ln(0.5)$ is case of two fully overlapping route alternatives.

$$r_i = \ln\left(\sum_{a_i \in A_i}\left(\left(\frac{l_a}{\sum_{a_i \in A_i} l_a}\right) \cdot \left(\frac{1}{|j|_a}\right)\right)\right) \quad (1)$$

Crowding: seat occupancy and standing density

To quantify the valuation of public transport crowding, we use two different attributes: the seat occupancy q and standing density d. Given the non-uniformly distributed demand pattern over the morning peak (**Figure 3.2**, right), it is not sufficient to calculate the average values for q and d over the whole morning peak. The attribute values for both attributes are therefore calculated per line, per link (stop-to-stop line segment), per 30 minutes time period. Since crowding levels vary across trips, using a too large time period can result in the use of average crowding levels which do not match with the expected and experienced crowding levels as they evolve during the morning peak. Notwithstanding, passengers will usually not have knowledge of the expected crowding levels for each individual trip departure. By dividing the morning peak into four periods of 30 minutes, we aim to balance between excluding non-uniformly distributed demand on the one hand, and applying a time period for which passengers can have realistic crowding expectations on the other hand.

The seat occupancy q is calculated using **Eq.2**, and expresses the ratio between the expected passenger load l and the vehicle seat capacity κ. If the expected load exceeds the seat capacity, q remains equal to 1. The expected passenger load is calculated based on the average realised occupancy for each link $a_i \in A_i$ (stop-to-stop line segment) for each 30-minutes time period $t \in T$. κ is determined using data provided by the operator, based on the vehicle type used for each line. To calculate q_{it} for each journey leg for alternative i in time period t, the weighted average value is calculated based on the expected seat occupancy and expected travel time $t_{a_i}^{ivt}$ per link $a_i \in A_i$ of the journey leg.

$$q_{it} = \min\left(\frac{\sum_{a_i \in A_i} \frac{l_{at}}{\kappa_{at}} \cdot t_a^{ivt}}{\sum_{a_i \in A_i} t_a^{ivt}}, 1\right) \quad (2)$$

The standing density d is calculated using **Eq.3**, and reflects the expected number of standing passengers per m^2. In line with Wardman and Whelan (2011), we use the standing density per

m^2 instead of the occupancy rate if $l > \kappa$, in order to account for different vehicle layouts. If the expected passenger load l does not exceed κ, this value equals zero. This expresses the assumption that passengers will stand only when all seats are occupied. When $l > \kappa$, d is calculated by dividing the number of standing passengers by the total surface available in each vehicle type for standing θ, thus assuming an equal distribution of standing passengers over the available standing surface. The expected value of d_{it} for alternative i in time period t is calculated for each link $a_i \in A_i$ (stop-to-stop line segment) for each 30-minutes time period $t \in T$ over all observed choices for route alternative i for a certain OD pair. The expected value per journey leg is computed by using the (by expected travel time $t_{a_i}^{ivt}$) weighted average over all links.

$$d_{it} = \max\left(\frac{\sum_{a_i \in A_i} \frac{l_{at} - \kappa_{at}}{\theta_{at}} \cdot t_a^{ivt}}{\sum_{a_i \in A_i} t_a^{ivt}}, 0\right) \quad (3)$$

Table 3.1 depicts the seat capacity κ and surface available for standing passengers θ per vehicle type for our case study network. **Table 3.2** shows the minimum and maximum values observed for q_t and d_t per half hour time period when calculated for the total dataset.

Table 3.1. Seat capacity and standing surface per vehicle type (HTM data)

Vehicle type	Mode	Lines (November 2015)	Seat capacity	Standing surface
GTL-8	Tram	1,6,9,11,12,15,16,17	73	25.1 m^2
Citadis	Light rail	3,4,19	86	32.0 m^2
Avenio	Tram	2	70	33.8 m^2
MAN	Bus	18,21,22,23,24,25,26,28	31	8.9 m^2

Table 3.2. Min/max seat occupancy and standing density in dataset

Time period	Seat occupancy		Standing density	
	Min	*Max*	*Min*	*Max*
07:00 – 07:30	0.10	1.0	0	0.81
07:30 – 08:00	0.08	1.0	0	2.48
08:00 – 08:30	0.13	1.0	0	2.04
08:30 – 09:00	0.11	1.0	0	2.31

3.2.5 Model formulation

In this study we estimate four discrete choice models in total. Model 1 reflects a model without crowding, whereas model 2 is an extension of model 1 in which the two crowding attributes are incorporated. Model 1 and 2 estimate coefficients averaged over all passenger segments. Since our study focuses on incorporating expected crowding levels in the route choice, it can be hypothesised that a segmentation between frequent and infrequent travellers is relevant. Passengers travelling frequently over a certain OD pair have a better expectation of crowding levels on the route choice alternatives based on their prior experiences, whereas infrequent passengers are expected to have limited or no prior crowding expectations on the specific route. Since the smart card number is known for all AFC transactions, it is possible to explicitly distinguish between frequent and infrequent travellers for each OD pair. We apply a binary classification, in which passengers who travel on average at least once per week on a certain route (in our case study: a minimum of four observations for a certain OD pair) are considered frequent travellers. Model 3 is a segment model without crowding, whereas model 4 is a segment model which incorporates crowding. We use a traditional utility maximisation framework, in which it is assumed that each respondent chooses the route alternative with the

largest utility (smallest disutility). The expected utility of a certain route choice alternative $U(V, \vartheta, \varepsilon)$ consists of the structural utility component V, which is a vector of observable attributes with their corresponding weights, and a random utility component ε. The logarithm of the path size factor r in incorporated into all four models to account for overlapping between route alternatives. This allows us to estimate standard MNL models as basis. Since there are multiple route choice observations in our dataset made by the same smart card number (= the same individual), we extend the standard MNL model to a mixed logit model with panel effects to correct for possible correlations between choices made by the same respondent. Therefore, U also consists of an individual specific utility component ϑ.

We use Biogeme as software package for performing the maximum likelihood estimations (Bierlaire, 2003). In order to reduce the number of draws, we perform Halton draws from a normal distribution to incorporate the panel structure of the model. In order to determine the number of required Halton draws, we started with an initial number of five Halton draws and then doubled the number of draws and checked whether the model outcome can be considered stable. All four model showed to be very stable directly after doubling the number of Halton draws to 10.

Model 1: no crowding, no segmentation

Eq.4 shows the calculation of V, the structural deterministic part of the utility function, for model 1. The attributes corresponding to the first journey leg are denoted by index 1; attributes corresponding to the second journey leg are denoted by index 2. We experienced with all combinations between estimating only generic coefficients for $t^{wait}, t^{ivt}, t^{trans}, n^{trans}$ and estimating all mode-specific coefficients. A model with mode-specific in-vehicle time coefficients, and generic waiting+transfer time and transfer penalty coefficients showed to give most reasonable results and the highest value for McFadden's adjusted R^2. Hence, as can be seen in **Eq.4**, generic coefficients are estimated for the initial waiting time and transfer time simultaneously, and for the transfer penalty. Mode-specific coefficients are estimated for in-vehicle time. A 'tram bonus' indicating a lower perceived in-vehicle time for tram/rail travelling compared to bus travelling has been previously reported in the literature (Bunschoten, 2013). The selected model specification allows us to quantify this 'tram bonus' based on RP data as well. The estimated coefficients for all attributes are denoted by β.

$$V = \beta^{wait} \cdot t^{wait} + \beta_m^{ivt} \cdot t_{m,1}^{ivt} + \beta^{wait} \cdot t^{trans} + \beta^{trans} \cdot n^{trans} + \beta_m^{ivt} \cdot t_{m,2}^{ivt} + \beta_r \cdot r \quad (4)$$

Model 2: crowding, no segmentation

Model 2, being an extension of model 1, estimates the same mode-specific in-vehicle time coefficients and generic waiting+transfer time and transfer penalty coefficients. **Eq.5** shows the structural part of the utility function when the seat occupancy q_t and standing density d_t for each 30-minutes time period t with their corresponding coefficients β^q and β^d are incorporated. As can be seen, the total in-vehicle time coefficient is now equal to the original in-vehicle time coefficient β^{ivt}, multiplied by a crowding multiplier which is equal to $(1 + (\beta^q \cdot q_t) + (\beta^d \cdot d_t))$

$$V = \beta^{wait} \cdot t^{wait} + (\beta_m^{ivt} \cdot t_{m,1}^{ivt} \cdot (1 + (\beta^q \cdot q_{t1}) + (\beta^d \cdot d_{t1}))) + \beta^{wait} \cdot t^{trans} + \beta^{trans} \cdot n^{trans} + (\beta_m^{ivt} \cdot t_{m,2}^{ivt} * (1 + (\beta^q \cdot q_{t2}) + (\beta^d \cdot d_{t2}))) + \beta^r \cdot r \quad (5)$$

Model 3: segmentation, no crowding

Model 3 can be considered an extension of model 1, to which initially an interaction-term is added for each estimated generic coefficient. The interaction-term is the product of an estimated

interaction-coefficient and a dummy-coded indicator-variable indicating whether an observed choice made by a certain respondent is classified as frequent or infrequent traveller. The indicator-variable equals 1 in case of a frequent traveller, and equals 0 in case of an infrequent traveller. This means that each total estimated coefficient is equal to the estimated generic coefficient which applies to both segments, plus the estimated interaction-coefficient multiplied by the indicator-variable which only applies to the frequent traveller segment. Initially, interaction-coefficients are estimated for all coefficients. Presented results in the next section show the estimation results for the final estimated model in which only significant interaction-coefficients are incorporated in the model specification. The interaction-coefficient β^{int} is denoted by superscript int.

Model 4: segmentation, crowding
Model 4 can be considered an extension of model 2, where also interaction-terms are added to the estimated crowding model. We adopt a similar approach here as described for model 3. In the final model only significant interaction-coefficients are incorporated in the model specification.

3.3 Results and Discussion

This section first shows the estimation results of the four models in **Section 3.3.1**. In **Section 3.3.2**, implications of these results are discussed.

3.3.1 Results

Table 3.3 shows the values of the estimated coefficients with corresponding t-values for all four models, the number of estimated coefficients, the final log-likelihood and McFadden's adjusted Rho-square.

Table 3.3. Estimation results

	Model 1 (no crowding, no segments)	**Model 2 (crowding, no segments)**	**Model 3 (segments, no crowding)**	**Model 4 (segments, crowding)**
β_{tram}^{ivt} (in-vehicle time tram)	-0.158** (-20.6)	-0.151** (-13.7)	-0.156** (-16.9)	-0.154** (-16.8)
β_{bus}^{ivt} (in-vehicle time bus)	-0.262** (-15.4)	-0.250** (-18.0)	-0.261** (-22.3)	-0.255** (-21.5)
β^{wait} (waiting+transfer time)	-0.398** (-24.9)	-0.395** (-24.7)	-0.397** (-24.9)	-0.387** (-24.4)
β^{trans} (transfer penalty)	-0.994** (-9.06)	-1.20** (-10.3)	-1.42** (-9.49)	-1.33** (-11.6)
$\beta^{trans,int}$ (transfer penalty interaction)	-	-	0.681** (3.95)	-
β^{r} (log-path size factor)	2.65** (3.01)	2.37* (2.65)	-	-
$\beta^{r,int}$ (log-path size factor interaction)	-	-	4.02** (3.89)	3.44** (3.54)
β^{q} (seat occupancy)	-	0.158** (4.97)	-	-
$\beta^{q,int}$ (seat occupancy interaction)	-	-	-	0.305** (7.94)
β^{d} (standing density)	-	0.0611* (2.15)	-	-
$\beta^{d,int}$ (standing density interaction)	-	-	-	0.149** (3.98)
σ^{panel} (sigma constant for panel effects)	-0.000 (-0.00)	0.000 (0.00)	0.000 (0.00)	0.000 (0.00)
Number of observations	17,994	17,994	17,994	17,994
Number of individuals	7,083	7,083	7,083	7,083
Number of Halton draws	10	10	10	10
Number of estimated coefficients	6	8	7	8
Final log-likelihood	-11,404	-11,384	-11,398	-11,356
Adjusted Rho-square	0.085	0.087	0.086	0.089

*t-values in parentheses. * p < 0.05; ** p < 0.01*

From **Table 3.3** we can conclude that all estimated coefficients are significant, except for the random parameter σ^{panel} introduced to capture panel effects. We can also see that the direction of all estimations of time and transfer related coefficients is negative, which is plausible. When extending model 1 with crowding (model 2), the adjusted Rho-square increases by 2.4%. Although the explanatory power of model 2 is only slightly higher than model 1, the LRS-test indicates that the improvement in goodness-of-fit is significant. The LRS-value of 40.6 is larger than the critical χ^2 value of 5.99 (with 8-6 = 2 degrees of freedom for α=0.05). Extending segment model 3 with crowding (model 4) results in 3.5% increase in the adjusted Rho-square. Also in this case the improvement in explanatory power is significant, since the LRS-value of 84.0 is larger than the critical χ^2 value of 3.84 (with 8-7 = 1 degree of freedom for α=0.05).

Table 3.4. Scaled estimation results

	Model 1 (no crowding, no segments)	**Model 2 (crowding, no segments)**	**Model 3 (segments, no crowding)**		**Model 4 (segments, crowding)**	
			Frequent	***Infrequent***	***Frequent***	***Infrequent***
in-vehicle time tram	0.6	0.6	0.6	0.6	0.6	0.6
in-vehicle time bus	1.0	1.0	1.0	1.0	1.0	1.0
waiting+transfer time	1.5	1.6	1.5	1.5	1.5	1.5
transfer penalty	3.8	4.8	2.8	5.4	5.2	5.2
seat occupancy	-	1.16	-	-	1.31	1.00
standing density	-	1.06	-	-	1.15	1.00

In order to expose the trade-offs, utility function coefficients are expressed in relation to travel time on-board a bus. **Table 3.4** shows the scaled estimation results for in-vehicle time, waiting+transfer time, transfer penalty and crowding, in which the in-vehicle time coefficient for bus t_{bus}^{ivt} is set equal to 1. The trade-offs presented in **Table 3.4** based on the estimation results for model 2 and model 4 (with crowding) exhibit plausible relations. A clear 'tram bonus' can be observed, since 1 minute in-vehicle time by bus is perceived as 0.6 minute in-vehicle time by tram. Earlier research based on SP experiments indicated values ranging between 0.67 and 0.80 (Bunschoten, 2013), which means that our research suggests an even more substantial 'tram bonus'. One minute waiting time is perceived 1.5 to 1.6 times more negatively, compared to one minute in-vehicle time. This is in line with values found in other studies (e.g. Balcombe et al., 2004), and also shows evidence that this multiplier has a lower value than assumed in earlier studies based on SP experiments. In addition, the transfer penalty of 5 minutes for each (urban) transfer is also plausible. In general, we see that RP estimates for the transfer penalty are somewhat lower than values reported in SP studies (e.g. Schakenbos et al., 2016).

Table 3.5. Crowding multiplier as function of seat occupancy and standing density

Seat occupancy q (% seats occupied)	**Standing density d (standing pass / m^2)**	**Crowding multiplier Model 2**	**Crowding multiplier Model 4**	
			Frequent	***Infrequent***
0	0	1.00	1.00	1.00
1	0	1.16	1.31	1.00
1	1	1.22	1.45	1.00
1	2	1.28	1.60	1.00
1	3	1.34	1.75	1.00

Next, we analyse the crowding effects on route choice. **Table 3.5** and **Figure 3.3** show the estimated crowding multiplier as function of the seat occupancy and standing density. As can be seen, in the generic model (model 2) the crowding multiplier equals 1.16 when all seats are occupied. In case the occupancy level increases further, the crowding multiplier increases with 0.06 for each increase in the integer number of standing passengers per m^2, additional to the crowding multiplier of 1.16 at seat capacity. For instance, in case of on average three standing passengers per m^2, the crowding multiplier thus equals 1.16 + (3 * 0.06) = 1.34. When segmentation is applied, a clear distinction can be observed between frequent and infrequent travellers. For frequent travellers, the crowding multiplier equals 1.31 when all seats are occupied. The multiplier further increases with 0.15 for each increase in passengers per square metre, additional to this value of 1.31. In case of three standing passengers per square metre on average, this results in a crowding multiplier of 1.75. This shows that frequent travellers incorporate crowding significantly more in their route choice than the average passenger, as estimated in model 2 without segmentation. On the other hand, the insignificant generic seat occupancy and standing density coefficients clearly indicate that infrequent travellers do not incorporate anticipated crowding levels in their route choice. This is plausible, given their lack of prior knowledge and experience regarding expected crowding levels. From **Table 3.4** it can be seen that the sum of coefficients estimated for the seat occupancy and standing density (generic coefficient plus interaction-term) remains equal to 1 - the nominal in-vehicle time - for infrequent travellers.

We also tested the estimation of a non-linear crowding function both related to the seat occupancy and the standing density. No plausible and significant results could however be found, thereby indicating a linear relationship between crowding and perceived in-vehicle time. Testing the estimation of a model with mode-specific crowding coefficients, next to the already incorporated mode-specific in-vehicle time coefficient, also resulted in implausible results.

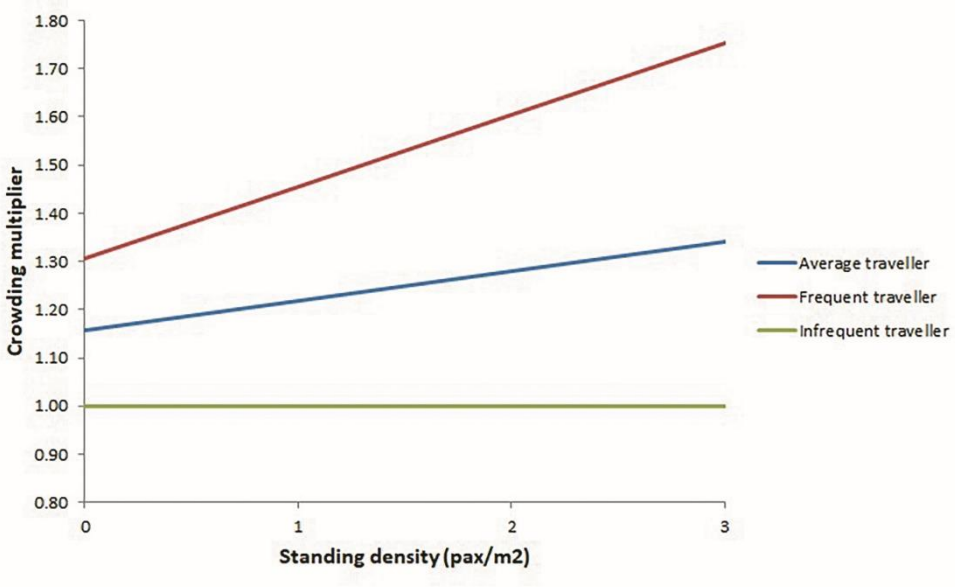


Figure 3.3. Crowding multiplier as function of the standing density for different traveller segments

Figure 3.4 shows the estimated crowding multiplier as function of the load factor, which equals the passenger load divided by the seat capacity. As can be seen, there are different crowding functions for the different vehicle types operating on the considered case study network. This shows the relevance of using the standing density instead of the load factor, when

estimating crowding if passenger loads exceed seat capacity. It can be seen that the crowding multiplier increases steeper for vehicle types which have a relatively high number of seats, compared to the total capacity (e.g. vehicle type 'GTL' and 'bus' in **Figure 3.4**). For vehicle types with relatively few seats compared to the total capacity, the crowding function increases at a slower pace (e.g. the light rail vehicle types 'Citadis' and 'Avenio' in **Figure 3.4**). Besides, clear differences in crowding multiplier can be seen between frequent and infrequent travellers, where the crowding multiplier for infrequent travellers remains equal to 1 for all vehicle types.

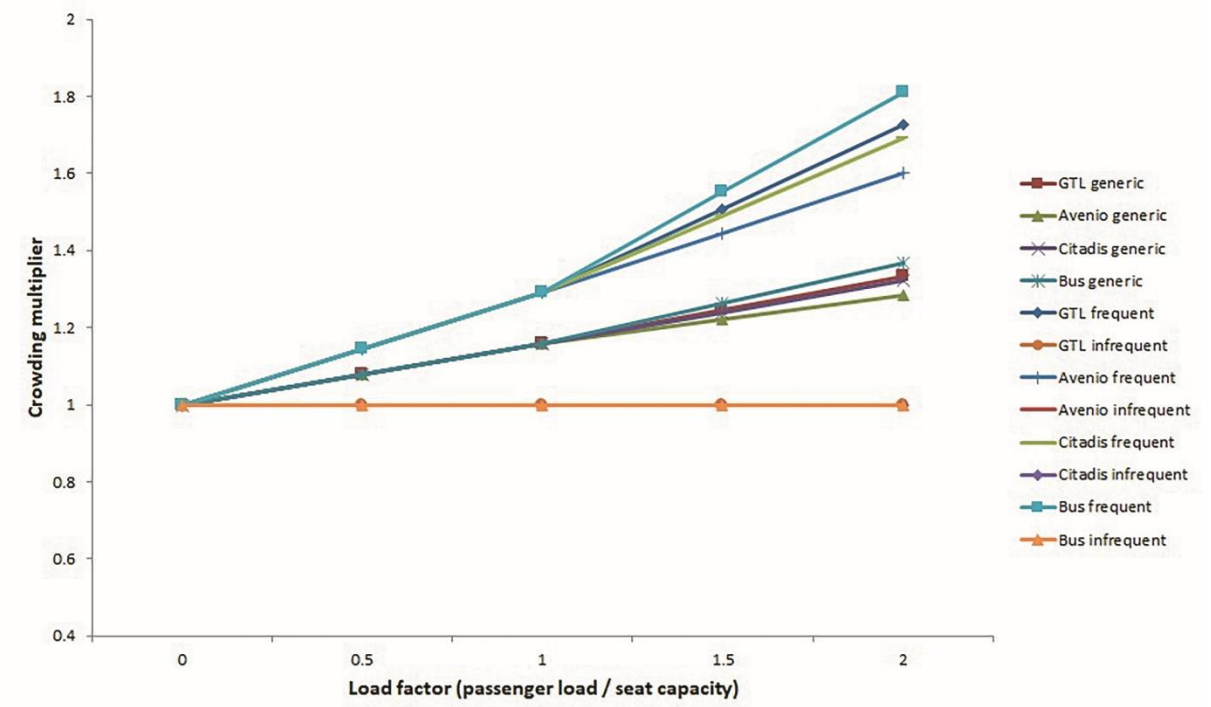


Figure 3.4. Crowding multiplier as function of the load factor per vehicle type

3.3.2 Discussion

The estimated coefficients exhibit plausible relations which are in line with directions found in earlier studies. When comparing model 3 (no crowding, segmentation) to model 1 (no crowding, no segmentation), a significant interaction coefficient is found for the transfer penalty and path size factor. These results suggest that frequent travellers have a less negative perceived transfer penalty compared to infrequent travellers. This might be explained by the higher level of knowledge about the network, transfer location and likelihood of reliability of the connecting service, compared to infrequent travellers. However, it can be seen that when crowding is incorporated in the segmented model (Model 4), the suggested difference in perceived transfer penalty between frequent and infrequent passengers in the segment model without crowding (Model 3) disappears. Besides, frequent travellers have a preference for non-overlapping route alternatives, whereas for infrequent travellers this does not significantly add explanatory power to their route choice. A possible explanation here is again the higher knowledge level of frequent travellers of other, non-overlapping route alternatives available in the network.

Regarding crowding we see that estimated crowding multipliers are lower than values found in SP experiments. For example, MVA Consultancy reports crowding multipliers up to 1.6-1.9 when seat capacity has been reached, compared to the highest value of 1.3 found in our study for frequent travellers. This gives evidence for the tendency of SP experiments to overestimate values of coefficients, compared to RP based studies. In the context of choice

experiments in a survey, respondents are arguably more inclined to attach greater importance to crowding in their stated route choice, compared to their decision-making in real-world settings. When we compare the results of this study with the RP results found for the Hong Kong MTR metro by Hörcher et al. (2017), we note that for the Hong Kong case the estimated crowding multiplier equals 1.98 at a density of six standing passengers per square metre. This is the sum of a standing multiplier of 1.265 and an increase of the crowding multiplier by 0.119 for each additional standing passenger per square metre. If we would linearly interpolate these results, the crowding multiplier for three standing passengers per m^2 is estimated to be 1.62. In the RP-based study by Tirachini et al. (2016) the estimated crowding multiplier is 1.55 in case of a standing density of three passengers per m^2. Our results suggest that the crowding multiplier equals 1.34 and 1.75 for average and frequent travellers, respectively, for a standing density of three passengers per m^2. Our estimation for the average crowding coefficient thus yields somewhat lower values than those found for the Hong Kong and Singapore metro case studies. However, for frequent travellers the estimated crowding multiplier of 1.75 is slightly higher than the average values found for Hong Kong and Singapore.

From **Table 3.4** it can be observed that 1 minute waiting time is perceived as 1.5 minutes in-vehicle time in a non-crowded vehicle by both frequent and non-frequent travellers. For frequent travellers, 1 minute travelling in a PT service with an average of 1.33 standing passengers per square metre is also perceived as equivalent to 1.5 minutes travelling in a non-crowded vehicle. This shows the trade-offs passengers perceive between waiting time and crowding, and can contribute to ridership forecasts for different types of measures. For example, increasing the frequency of a non-crowded service from 6 to 8 services per hour will reduce the perceived travel time by passengers by 1.88 minutes under the assumption of a random arrival pattern at stops. Alternatively, reducing the crowding level for a 10-minute trip from an average of 2.0 to 1.0 standing passengers per square metre, reduces the perceived in-vehicle time by 1.5 minutes for frequent travellers. Since infrequent travellers do not have knowledge about the expected crowding levels, measures purely aimed at reducing crowding without increasing the service frequency (e.g. use of longer vehicles) are not expected to increase the ridership levels of infrequent travellers, contrary to measures which target service headway.

We note that we only estimated coefficients for expected crowding levels, which can be different from *a posteriori* experienced crowding levels. Our study results also show the potential of crowding information provision to infrequent or less frequent travellers. Given the estimated values for frequent travellers, we might expect that infrequent travellers will incorporate crowding in a similar way in their route choice as frequent travellers if information is provided to them. This can affect route choice and occupancy levels on routes.

3.4 Conclusions

Crowding in public transport can have major influence on passengers' travel experience and therefore affect route and mode choice. Due to the increasing concentration of activities within urban agglomerations in many countries worldwide, crowding is expected to become an even more dominant factor in urban public transportation in the future. Besides, it is important to incorporate valuation of crowding in a correct way within a cost-benefit analysis. Therefore, it is important to understand how crowding in urban public transport is perceived by passengers. In this study, the availability of individual smart card transactions was used to gain insights into revealed trade-offs between travel time, transfers, waiting time and crowding in public transport route choice entirely based on revealed preference data.

Model estimations confirm that crowding plays a significant role in passengers' route choice in public transport. The average crowding multiplier of in-vehicle time equals 1.16 when all seats are occupied. For frequent passengers, this value further increases to 1.31 when all

seats are occupied. In case of standing passengers, the average crowding multiplier further increases with 0.06 for each increase in the integer number of standing passengers per m^2. For frequent travellers, this increase per square metre is estimated to be equal to 0.15. Our results show that infrequent travellers do not incorporate expected crowding in their route choice, probably due to the lack of prior experience. Our results suggest that crowding valuation studies using stated preference experiments can have a tendency to overestimate crowding values.

Our study is the first in which crowding valuation is determined entirely based on revealed preference data particularly for urban tram and bus services, in a European context, thereby adding to the knowledge gained from crowding valuation studies for metro networks. The insights gained from our study can support the decision-making process by quantifying benefits of measures aiming to reduce crowding levels. For further research, it is recommended to consider how passengers are distributed throughout the vehicle. In case of unequal distributions, the experienced crowding can deviate from expected crowding levels based on equal passenger distributions. At last, an important limitation for this study is that no information is available regarding the realised passenger arrival time at the stop. Our study assumes that passenger route choice is fully based on expected crowding levels. Due to the lack of this information, we cannot determine whether a passenger boarded the first vehicle arriving at the stop, or deliberately skipped a crowded vehicle for a less crowded alternative. Information about this would enable us to investigate to which extent the real-time crowding level of the arriving vehicles affects route choice, compared to expected crowding levels based on prior experiences.

4. Improving Predictions of Public Transport Usage during Disturbances based on Smart Card Data

The main contribution of this chapter is to infer passengers' route and mode choice behaviour during planned disruptions, such as track maintenance works, based on empirical data from multiple planned disruptions. This chapter contributes to the first research question as formulated in **Section 1.3** to measure and characterise passengers' behavioural and demand response, in this case tailored to planned disruptions. This completes Part I of this research. When passenger journeys are inferred using a transfer inference algorithm (**Chapter 2**), the journey time components of each journey can be obtained or inferred from passive data sources and multiplied by a corresponding coefficient which reflects how passengers perceive each component of the journey (see **Chapter 3** for crowding perception in particular). The last step to measure passenger disruption costs is to measure how passengers adjust their route and mode choice in the event of a disruption. Particularly for planned disruptions which can be communicated to passengers before commencing their journey, it is important to understand how passengers change their mode choice in response to this, and how passengers perceive temporary service adjustments made by the public transport service provider, such as the provision of rail-replacement bus services. The number of passenger journeys for each origin-destination pair, which results from applying the proposed transfer inference algorithm in **Chapter 2**, is used as direct input in this study. This enables a comparison between the number of passenger journeys during different planned disruptions and during regular, undisrupted circumstances, based on which route and mode choice parameters are calibrated to capture the passenger response.

This chapter is based on an edited version of the following article:

Yap, M.D., Nijënstein, S., Van Oort, N. (2018). Improving predictions of public transport usage during disturbances based on smart card data. *Transport Policy, 61,* 84-95.
© 2017 Elsevier Ltd.

4.1 Introduction

The last decade, in several cities worldwide automated fare collection (AFC) systems are introduced for the public transport system by public transport operators and authorities. For these AFC systems, passengers need to use a smart card for public transport travelling. Open systems in which passengers only need to tap in, as well as closed systems in which both a tap in and tap out are required, are applied in practice. Although the main purpose of the introduction of AFC systems was to enable an easier way of revenue collection, additionally large amounts of data are generated which can be used to get more insight in passengers' travel behaviour. Over the last years, data from AFC systems is used for many purposes by scientists and practitioners on a strategic, tactical and operational level (Pelletier et al., 2011). Data from AFC systems is for example used for destination inference in case of systems with tap in only (e.g. Trépanier et al., 2007; Nunes et al., 2016), transfer inference (e.g. Hofmann and O'Mahony, 2005; Jang, 2010) and journey inference to estimate origin-destination (OD) matrices (e.g. Seaborn et al., 2009; Wang et al., 2011; Munizaga and Palma, 2012; Zhao et al., 2012; Gordon et al., 2013). Other studies focus on fusion of smart card data of different operators (e.g. Nijënstein and Bussink, 2015) or clustering public transport stops in order to identify and classify public transport activity centres based on smart card data (Cats et al., 2015b).

In addition to the aforementioned studies which use smart card data to describe, analyse, cluster and visualise current travel patterns, there are also studies focusing on public transport ridership prediction based on smart card inferred travel patterns. Idris et al. (2015) developed several mode choice models based on revealed preference, contrary to traditional mode choice models having the tendency to overestimate public transport ridership. Wei and Chen (2012) developed a forecasting approach for short-term ridership predictions in metros using a combination of empirical mode decomposition and neural networks, whereas Li et al. (2017b) predict metro ridership under special events using a multiscale radial basis function (MSRBF) network. Ding et al. (2016) predict metro ridership using gradient boosting decision trees, thereby incorporating temporal features and bus transfer activities. In Van Oort et al. (2015a) a smart card based prediction model is developed which allows the prediction of effects of changes in public transport supply, such as increasing the frequency or rerouting public transport services. This model considers the total urban public transport network and uses an elasticity approach, where parameter values are obtained based on revealed preference studies. Effects of crowding can also be incorporated in this short-term ridership prediction model (e.g. Van Oort et al., 2015b). This type of prediction model is of added value to improve prediction accuracy of the impact of structural network changes, which are usually implemented by operators on one or on a few fixed dates in the year. However, in practice many public transport operators are confronted with temporary closures of infrastructure many more times per year. These temporary infrastructure closures are for example caused by maintenance work, track renewal or redesign of public space. These closures usually result in longer travel time, more transfers, lower ridership, lower passenger satisfaction, and less revenues. In the Netherlands, a tendency can be observed of more, larger and more long-lasting rail infrastructure closures. For example, HTM, the urban public transport operator in The Hague, the Netherlands, was confronted with more than 20 temporary track closures in 2015. It therefore becomes more urgent for operators to predict the impact of these (planned) disturbances on their passengers, ridership and revenues. This impact of temporary track closures on demand and supply is different compared to the impact of structural network changes. Passengers might be willing to postpone a single trip, change their mode choice or route choice, or accept the use of rail-replacement bus services for temporary situations. Operators on the other hand have to accept the temporary reduction in level of service - because of rail-replacement bus services, additional

travel time and transfers - and might accept the temporary additional operational costs for these bus services and communication. It can be concluded that the responses of passengers and operators differ in case of temporary network changes, compared to structural network changes. In order to predict passenger impacts of temporary network changes with sufficient accuracy, other/additional parameters and/or different parameter values in the public transport ridership prediction models are therefore required.

This study aims to improve the prediction accuracy of the impact of planned, temporary disturbances on public transport usage. To this end, in this study a new parameter set is calibrated and validated to predict public transport ridership in case of planned disturbances. This parameter set is based on smart card data derived from AFC systems during several planned disturbances which occurred in The Hague in 2015. The study results in a new set of parameter values allowing to better predict passenger impacts of planned disturbances in urban public transportation. With this result, more insight is gained in passenger behaviour during disturbances. It also supports operators to predict the impact on their revenues, and to better align supply of rail-replacement services on alternative routes to the remaining demand, in order to efficiently use their scarce resources. This chapter is structured as follows. **Section 4.2** describes the methodology to calibrate and validate the parameter set of the ridership prediction model. **Section 4.3** describes the case study network to which the methodology is applied. **Section 4.4** discusses the results of this study. At last, in **Section 4.5** conclusions and recommendations for further research are formulated.

4.2 Methodology

4.2.1 Origin-destination matrix estimation

When travelling in trams or buses in the Netherlands by smart card, passengers are required to tap in and tap out at devices which are located within the vehicle. This means that in the Netherlands the passenger fare is based on the exact distance travelled in a specific public transport vehicle. Especially for buses, this is different from many other cities in the world where often an open, entry-only system with flat fare structure is applied, for example in London (Gordon et al., 2013) and Santiago, Chile (Munizaga and Palma, 2012). This means that for each individual transaction the boarding time and location, and the alighting time and location of each trip leg are known. Also, it is known in which public transport line and vehicle each passenger boarded and alighted with its unique smart card number. This within-vehicle system therefore eases the destination and journey inference, compared to open entry-only systems. When merging this within-vehicle AFC system with Automated Vehicle Location (AVL) data, vehicle occupancies can be inferred directly from the transaction data for each line segment and vehicle.

For an urban public transportation network with tram and bus lines, journeys can be inferred by combining registered trip legs made with the same smart card ID. In this study we used a simple temporal criterion to determine whether a passenger alighting is considered a final destination or a transfer. When the boarding time in a vehicle follows within a certain time window after the alighting time of the previous trip leg made with that same card, two AFC transactions are considered as one journey. This approach is also used, for example, by Hofmann and O'Mahony (2005) and Seaborn et al. (2009). We are aware that in scientific literature more advanced transfer inference algorithms have been developed (e.g. Zhao et al., 2007; Munizaga and Palma, 2012; Gordon et al., 2013; Yap et al., 2017). In Dutch practice however, operators apply only a time window threshold between the previous alighting and next boarding as transfer inference criterion. In order to compare the prediction accuracy of the

new proposed parameter set with the earlier operator predictions with the default parameter set, we decided not to adjust the transfer inference algorithm in this study. In this way, we can evaluate purely the effects of our new parameter set on the prediction accuracy, while not also changing the transfer inference algorithm simultaneously. In the Netherlands, a maximum threshold transfer time of 35 minutes is applied to classify trip legs made by the same smart card ID as one journey. By aggregating all journeys, a stop-to-stop smart card based OD matrix can be inferred. In the ridership prediction model, zones are located at the stop locations. Only stop codes which belong to the same stop from a passenger perspective, are aggregated to one zone. This means that stop codes of platforms of the same stop in opposite directions, or stops located at the same junction, are represented by one zone. This is done to prevent passenger travel patterns to be relying too strong on the exact current stop codes of boarding and alighting in the undisturbed scenario. Under the assumption that the distribution of destinations j from each origin i for non-card users is similar to the distribution of smart card users, which is in line with the assumption applied by Munizaga and Palma (2012) to correct for missing tap outs, the zone-based OD matrix can be scaled based on the small percentage of non-card users in the Netherlands. Determination of the share of non-card users is based on passenger counts.

When travelling by train or metro in the Netherlands, there is an entry-exit system where transactions are required during boarding and alighting. For train and metro, devices are however located at the station gates. This means that train-train or metro-metro transfers, as well as exact chosen routes cannot be determined directly from the data and that trip and transfer inference algorithms are necessary to obtain these insights.

4.2.2 Public transport ridership prediction model

For the prediction of public transport usage in case of planned disturbances, in this study the public transport ridership prediction model as described in Van Oort et al. (2015a) is used as basis. For an urban public transportation network, let the set of public transport stops and lines be denoted by S and L respectively. Each line $l \in L$ is defined by an ordered sequence of stops $l = (s_{l,1}, s_{l,2\ldots}, s_{l,|l|})$. $L^t \in L$ and $L^b \in L$ represent the subset of tram lines and bus lines of the considered network, respectively. $|L|$ expresses the number of public transport lines in the total set L. Trip schedules are imported in the model, based on which the frequency and stop-to-stop travel times are inferred for each line $l \in L$ in each distinguished time period t. Public transport demand is connected to this network by an OD matrix between all stops $s \in S$ for each distinguished time period t. The OD matrix of the undisturbed base scenario δ_0 is based on smart card data and estimated as explained in **Section 4.2.1.** We use a conversion table between the stop ID of the boarding and alighting location in the smart card transaction data and the modelled zones in the prediction model, in order to connect travel demand to the modelled urban public transportation network.

For public transport ridership predictions, this model is based on a demand elasticity. For each OD pair i, j the generalised travel costs - being the sum of costs for in-vehicle time, transfer walking time, waiting time, transfers and travel fares with their corresponding weights - are calculated for the base scenario δ_0 and for each scenario δ. **Eq.1** shows the calculation of the generalised costs, expressed in monetary terms. Applying a demand elasticity parameter to the relative change in generalised travel costs between δ_0 and δ for each OD pair allows for the calculation and assignment of a new public transport OD matrix for each scenario δ. **Eq.2** shows the calculation of new public transport demand.

The default parameter values for a_1, a_2, a_3, a_4, a_5 used in this prediction model for structural network changes are obtained based on a combination of model calibration and literature review (Balcombe et al., 2014). In this calibration process, model assignment results (number of passengers and passenger-distance on the network, per line $l \in L$ and per link) for

the undisturbed base scenario δ_0 were compared with the raw smart card transaction data. The parameter set resulting in the highest fit between assignment results and raw smart card data, with parameter values within bounds found in literature, is applied to this model. The weight of in-vehicle time a_1 equals 1.0, whereas one minute walking time a_2 or waiting time a_3 are valued 1.5 times more negatively compared to one minute in-vehicle time. This is also in line with values found in literature (e.g. Balcombe et al., 2004; Arentze and Molin, 2013). Given the focus on an urban public transport network with usually relatively short trips, a relatively small transfer penalty of 3 minutes is applied for a_4. In this prediction model we only consider the marginal travel costs per travelled kilometer, without incorporating the base fare of €0.88 which applies for all passengers and all trips in urban public transport in the Netherlands. This is justified since this fixed cost component, which is the same for each public transport route, does not add explanatory power to passenger route choice in the model. The marginal travel costs per travelled kilometre in the model are reflected by a_5 and are equal to €0.05/km. Compared to the marginal travel costs of €0.15/km currently in the Netherlands (MRDH, 2016), this value shows a limited price sensitivity. This can be explained due to the fact that also passengers which are price-inelastic are incorporated in the data. These passengers do not have to pay for their tickets themselves (e.g. business trips paid by the company, or student trips paid by the Dutch government), have monthly or yearly travel passes (where the marginal travel costs are usually lower), or travel with discount (e.g. elderly, children). The Value-of-Time for the Dutch situation is determined based on Significance et al. (2013).

In the model used in this study no segmentation in parameter values is used for passengers with different trip purposes or for different time periods. Differences in sensitivity to generalised cost components can for example be related to trip purpose or socio-economic status of passengers. In the model used for this study, no explicit distinction is made between the home-end and activity-end of a journey. Therefore, it is not possible to determine the home-end of each journey explicitly, and to relate this for example to the socio-economic status of a specific area of the city. Also regarding trip purpose an explicit segmentation cannot directly be inferred from the AFC data. For example, in The Hague there are 58 different product types available for travelling by smart card (e.g. business card, student cards, monthly cards). For some card types it is relatively easy to match card type (e.g. student card) and trip purpose (e.g. education). However, for many other card types this match is less trivial. Besides, for this study no information was available regarding the product type used for each AFC transaction. In addition, our AFC dataset did not contain the ID of the smart card used for travelling. Therefore, it was not possible to infer trip purpose from long-term travel patterns. Different sensitivities during different periods of the day are mostly related to different mixtures of passenger segments (e.g. a high percentage commuters during peak periods, or a high percentage leisure / shopping passengers during off-peak periods), as for example already shown by Nazem et al. (2011). Instead of aiming to infer these different travel segments from available travel patterns, we decided to come up with a robust parameter set which is - on average - able to improve prediction accuracy over these different passenger segments. For further research, there is definitely potential to further improve the parameter set specific for different passenger segments and areas of the city.

$$C_{ij} = (\alpha_1 \cdot IVT_{ij} + \alpha_2 \cdot WKT_{ij} + \alpha_3 \cdot WTT_{ij} + \alpha_4 \cdot NT_{ij}) * VoT + \alpha_5 \cdot d_{ij} \qquad (1)$$

With:

C_{ij}	Generalised costs for OD pair *i,j*
α_1,α_2,α_3,α_4, α_5	Weight coefficients in generalised costs calculation
IVT_{ij}	In-vehicle travel time for OD pair *i,j*
WKT_{ij}	Walking time for OD pair *i,j*

WTT_{ij}	Waiting time for OD pair *i,j*
NT_{ij}	Number of transfers for OD pair *i,j*
VoT	Value-of-Time (€/hour)
d_{ij}	Distance travelled in public transport for OD pair *i,j*

$$D_{ij}^{\delta} = \left(E \cdot \left(\frac{C_{ij}^{\delta}}{C_{ij}^{\delta_0}} - 1 \right) + 1 \right) \cdot D_{ij}^{\delta_0} \tag{2}$$

With:

D_{ij}^{δ}	Demand for OD pair i,j in scenario δ
E	Elasticity
C_{ij}^{δ}	Generalised costs in scenario δ
$C_{ij}^{\delta_0}$	Generalised costs in base scenario δ_0
$D_{ij}^{\delta_0}$	Demand for OD pair i,j in base scenario δ_0

We can thus conclude that there is already a calibrated parameter set which is used to predict public transport ridership for undisturbed situations. In this study, we specifically investigate to what extent this parameter set needs to be adjusted to perform accurate passenger predictions in case of planned disturbances.

4.2.3 Evaluation framework

An evaluation framework is developed to evaluate the accuracy of different parameter sets for ridership predictions in case of (planned) disturbances. First, the subset $L^a \in L$ is determined, where L^a contains all public transport lines which are directly or indirectly affected by a certain disturbance δ. Public transport lines of which the route, mode and/or frequency is adjusted due to a disturbance, are considered directly affected lines. Public transport lines parallel to the directly affected lines, which can function as alternative for passengers to reach their destination, are considered indirectly affected lines. Although route and frequency of these lines are not changed, the number of passengers and passenger-kilometres travelled over these lines can be affected due to the directly disturbed public transport lines. For all lines $l \in L^a$ we consider the impact of a disturbance for the different distinguished time periods t. Since we do not estimate different parameter sets for different time periods, it is important to explicitly evaluate and compare the prediction accuracy of different parameter sets over the different time periods in order to develop a robust parameter set.

We use two indicators to express the impact of a disturbance: the relative impact of a disturbance on the number of passengers P and on the number of passenger-kilometres PK. This means we consider the relative increase or decrease in P and PK for all affected public transport lines in all distinguished time periods during a specific disturbance δ, compared to the undisturbed base scenario δ_0. For each disturbance, this leads to a total number of $|L^a| * |T| * 2$ elements for which the empirical and predicted relative effect of a disturbance are compared. **Eq.3** and **Eq.4** formalise the calculation of the relative impact of a disturbance on the number of passengers P and passenger-kilometres PK, respectively.

$$\Delta P = \left(\frac{(P_{\delta,lt} - P_{\delta_0,lt})}{P_{\delta_0,lt}} \right) * 100 \qquad \forall\, l \in L^a, t \in T \tag{3}$$

$$\Delta PK = \left(\frac{(PK_{\delta,lt} - PK_{\delta_0,lt})}{PK_{\delta_0,lt}} \right) * 100 \quad \forall\, l \in L^a, t \in T \tag{4}$$

With:

$P_{\delta,lt}$	Number of passengers in disturbed scenario δ
$P_{\delta_0,lt}$	Number of passengers in undisturbed base scenario δ_0
$PK_{\delta,lt}$	Number of passenger-kilometres in disturbed scenario δ
$PK_{\delta_0,lt}$	Number of passenger-kilometres in undisturbed base scenario δ_0

In this study we use the well-known R^2 measure to quantify the prediction accuracy of different parameter sets for all $|L^a| * |T| * 2$ elements, which compares empirically derived values $\Delta\widetilde{P_{lt}}$ and $\Delta\widetilde{PK_{lt}}$ with predicted values $\Delta\widetilde{P_{lt}}$ and $\Delta\widetilde{PK_{lt}}$. The used prediction model consists of an undisturbed base scenario δ_0, of which the number of passengers and passenger-kilometres are calibrated based on imported smart card data corresponding to this undisturbed base network (**Section 4.2.2**). The passenger impact of a disturbed scenario δ is predicted using the described elasticity approach after modelling the network corresponding to each scenario δ. Based on the model output for $P_{\delta_0}, PK_{\delta_0}$ and P_{δ}, PK_{δ}, predicted values $\Delta\widetilde{P_{lt}}$ and $\Delta\widetilde{PK_{lt}}$ are calculated. The values $\Delta\widetilde{P_{lt}}$ and $\Delta\widetilde{PK_{lt}}$ can be inferred from the AFC data. For both the undisturbed base scenario δ_0 and each disturbed scenario $\delta \in \Delta$, the raw smart card data is scaled for non-card users, thereby applying the same scaling factor as applied in the prediction model. Since the period of the year which is used to infer P_{δ_0} and PK_{δ_0} differs from the period which is used to infer P_{δ} and PK_{δ}, a correction for possible ridership variation which cannot be attributed to the adjusted public transport network needs to be applied. There can by systematic differences in travel patterns between the days used for δ_0 and δ, for example due to holiday periods. Also seasonal differences can occur (e.g. different public transport ridership in January compared to June), as well as daily differences due to weather (rain or sunny weather), day of the week and random fluctuations. To prevent a biased comparison between $\Delta\widetilde{P_{lt}}$ and $\Delta\widetilde{P_{lt}}$, and between $\Delta\widetilde{PK_{lt}}$ and $\Delta\widetilde{PK_{lt}}$, we made sure that the generic travel patterns are similar for δ_0 and δ. This means that if δ occurs within the summer holiday, we also inferred data from the (undisturbed part of this) summer holiday for δ_0. Similarly, in case δ occurs during regular working periods, δ_0 is also based on days during a working period. We also made sure that the share of the different days of the week was similar for δ_0 and δ - by using AFC data from an equal number of Mondays, Tuesdays etc. for δ_0 and δ - to prevent bias due to differences in ridership over the different days of the week. This however does not exclude seasonal and daily random variation from the comparison yet. This is especially relevant, given the large number of disturbances occurring on the urban tram network of operators in the Netherlands. As illustration: for the tram network of The Hague, only 6 months out of the 36 months between July 2014 and July 2017 were totally free from planned disturbances. For example, if δ lasted during the whole month of November, there is no other option than using AFC data from one of these undisrupted months (e.g. March) for δ_0. In order to correct for possible seasonal and random variation between these periods, we applied a correction to the inferred values P_{δ_0} and PK_{δ_0} based on the ridership differences found on all public transport lines $l \notin L^a$ between δ_0 and δ. We calculated the average difference for P_t and PK_t over all public transport lines which are not directly, nor indirectly, affected by a disturbance δ, for each distinguished time period. Under the assumption that the seasonal and random effect for all lines $l \in L^a$ is similar to the average effect found for all lines $l \notin L^a$, we corrected the originally directly inferred values $\widetilde{P_{lt,\,\delta_0}}'$ and $\widetilde{PK_{lt,\,\delta_0}}'$ using **Eq.5**. This allowed us to compute $\Delta\widetilde{P_{lt}}$ and $\Delta\widetilde{PK_{lt}}$ without bias using **Eq.3** and **Eq.4**, based on which the R^2 measure could be computed using **Eq.6.** We illustrate the calculations in **Eq.5** and **Eq.6** for P, whilst these apply for calculating PK as well.

$$\widetilde{P_{lt,\,\delta_0}} = \widetilde{P_{lt,\,\delta_0}}' \cdot \left(\frac{1}{|L|\notin L^a}\right) \cdot \sum_{l\notin L^a} \frac{\widetilde{P_{lt,\delta}}}{\widetilde{P_{lt,\,\delta_0}}} \qquad \forall\, l \in L^a, t \in T \tag{5}$$

$$R^2 = 1 - \frac{\sum_{t\in T}\sum_{l\in L^a}\left(\widetilde{\Delta P_{lt}} - \widetilde{\overline{\Delta P_{lt}}}\right)^2}{\sum_{t\in T}\sum_{l\in L^a}(\widetilde{\Delta P_{lt}} - \widetilde{\overline{\Delta P}})^2} \qquad \forall\, l \in L^a, t \in T \tag{6}$$

4.2.4 Experimental design

In this study four planned disturbances which occurred on the HTM network in 2015 are considered, denoted by the total set Δ. Subsets $\Delta_A \in \Delta\,\{\delta_1, \delta_2\}$ and $\Delta_B \in \Delta\,\{\delta_3, \delta_4\}$ are defined, which both contain 50% of the investigated disturbances for calibration and validation purposes, respectively. For the calibration phase, a rule-based three-step procedure is used to determine an improved parameter set which leads to a better fit between empirical AFC data and predicted public transport ridership during planned temporary disturbances.

Step 1. Selection of parameters with potentially different values during disturbances
In order to predict public transport usage in case of planned disturbances, it is important to determine which parameter values could be different, compared to the values used to predict regular passenger route choice and ridership for undisturbed scenarios as described in **Section 4.2.2**. In theory, values can be different for all six parameters of **Eq.1** - $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, VoT$ - and the elasticity parameter E of **Eq.2**. In a first step narrowing down the solution space, we determine which parameters can have potentially different values during disturbances based on theory. It is confirmed using a sensitivity analysis that parameters excluded from the solution space based on theory, do not or hardly affect prediction accuracy indeed.

First, the value of the elasticity parameter E_δ in case of disturbances is of relevance. As mentioned in **Section 4.1**, passengers react differently to temporary network changes compared to structural network changes. On the one hand, passengers might accept a longer travel time for a certain amount of time (indicating a less negative value of E_δ). On the other hand, passengers might decide to change their mode choice or destination choice, or to postpone their trip in case of temporary track closures, until regular operations are restored (indicating a more negative value of E_δ). Second, the modelling of rail-replacement services is of relevance. Let $L^R \in L$ be the subset of rail-replacing bus services. In many cases, operators will supply rail-replacing bus services in case of track closures. These rail-replacing services differ from regular bus lines in several ways. For example, the existence, route and stop locations of such services are often less well known by passengers. Given the temporary existence of these lines, passengers are less familiar with aspects as departure time, travel time and reliability. When these buses replace rail services, these services have to use temporary stop locations nearby the closed rail stop, which often have less visibility and equipment like dynamic arrival information or shelters. It is therefore possible that passengers experience waiting time for rail-replacement services more negatively compared to waiting time for regular tram or bus services (indicating a higher value of parameter a_3, related to waiting time WTT^R specific for rail-replacement services). Besides, these services transport passengers who are familiar with rail-bound services. From literature it is known that when a bus service is transformed into a tram line, travel time is perceived less negatively compared to bus travelling (Bunschoten et al., 2013). Therefore, it can be hypothesised that the replacement of a tram line by buses will be perceived more negatively by passengers familiar with rail-bound travelling. Therefore, the value of parameter a_1 related to in-vehicle time perception in rail-replacement busses IVT^R might be more negative compared to regular trams or buses. Rail-replacement buses usually operate with higher frequencies than the original tram line, to compensate for the lower capacity of a bus compared to a tram. However, it is unclear to what extent passengers really perceive and

incorporate this theoretical benefit in their route and mode choice. It is therefore questionable whether modelling the realised frequencies of the rail-replacement services f^R, or the original frequencies of the tram line which is being replaced f^T, leads to more accurate predictions.

The remaining parameters from **Eq.1** - a_2 (multiplier for walking time perception), a_4 (fixed transfer penalty), a_5 (marginal travel costs) and VoT (value of time) - were expected to have no or a limited effect on the prediction accuracy. There is no rationale to assume that the marginal travel costs or Value of Time differ during disturbances, since the fare system applied during disturbances for rail-replacement bus services is exactly equal to the fare system used for regular tram and bus lines. Although it could be hypothesised that the walking time or a transfer to rail-replacement bus services can be perceived more negatively compared to regular tram or bus lines, a first sensitivity analysis showed only a very limited impact on the prediction accuracy using **Eq.6** when varying parameter values for a_2 and a_4. Increasing / decreasing parameter values of a_2 and a_4 with 33% for $\Delta_A \in \Delta\{\delta_1, \delta_2\}$, did not increase/decrease the prediction accuracy with more than 3% (**Figure 4.1**). **Table 4.1** shows the remaining parameter values which are included in the search procedure, together with the expected direction of the parameter values compared to the default parameter value used for predictions for undisturbed network changes.

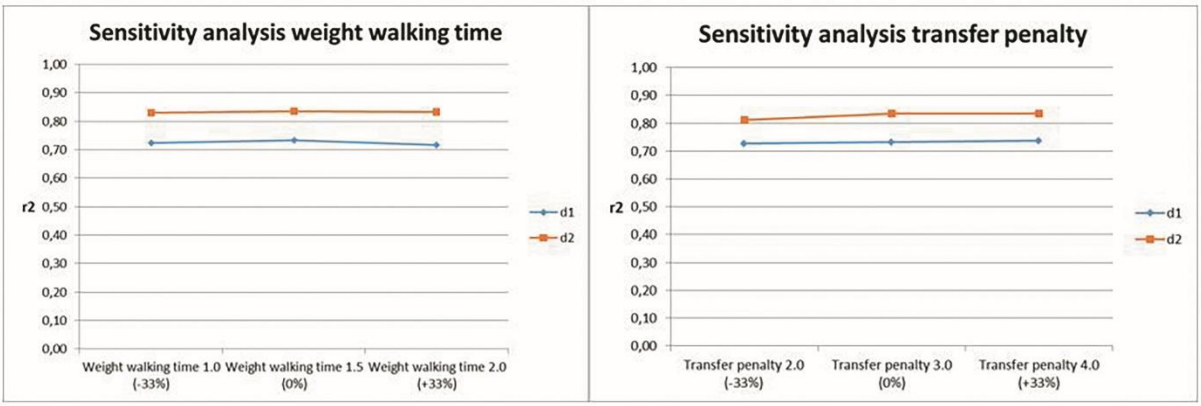


Figure 4.1. Results sensitivity analysis to walking time perception (left) and transfer penalty (right) to rail-replacement bus services

Table 4.1. Parameter selection for systematic grid-search

Parameters	Values default parameter set	Search direction parameter values
Elasticity E_δ	-1.1	More or less negative
α_3 for waiting time WTT^R	1.5	Higher
α_1 for in-vehicle time IVT^R	1.0	Higher
Frequency $\{f^R, f^T\}$	f^R	f^T

Step 2. Systematic grid-search

In the second step, a systematic grid-search is performed to scan for the best fitting parameter set(s) from predefined scenarios, using the four model parameters of **Table 4.1**. For all four parameters, plausible values are *a priori* determined. The calibrated parameter values used for passenger assignment for the undisturbed base scenario δ_0 are used as starting point (E_δ= -1.1, WTT^R=1.5, IVT^R=1.0, f^R=f^R). These values are considered as reasonable starting point, since these values are calibrated and within bounds found in literature (e.g. Balcombe et al., 2004). The direction in which each parameter value can change when predicting ridership during disturbances, compared to regular ridership predictions, is explained in the first part of this section and shown in **Table 4.1**. The upper and lower bound values for E_δ, WTT^R and

IVT^R are selected in such way, that they remain within literature bounds on one hand, but show sufficient variation to explore the solution space on the other hand. The modelling of the frequency of rail-replacement bus services is a binary variable, which can be equal to f^R or f^T. The second row of **Table 4.2** shows the resulting parameter values. All combinations between these predefined parameter values are systematically evaluated using the R^2 measure for prediction accuracy from **Eq.6**. Using the Cartesian product of the chosen plausible values, this results in 24 scenario combinations of parameter values (**Table 4.2**). The first row of **Table 4.2** summarises the four parameters which are hypothesised to have different values when modelling passenger behaviour during disturbances specifically.

Based on the results of this second step the solution space can be narrowed down further, so that a further in-depth search can be performed in the third step of the search procedure. In this second step we did not only look at the exact R^2 value resulting from the 24 different scenarios, but we also investigated which patterns and scenarios were robust over the two evaluated disturbances $\Delta_A \in \Delta \{\delta_1, \delta_2\}$. Since we aim to come up with one improved parameter set to predict impacts of different (types of) disturbances, it is important to have a robust, good performing parameter set rather than optimising too much for these specific two disturbances. We determine the most promising parameter sets from these pre-defined scenarios as result of this second step of the procedure.

Table 4.2. Experimental design

Parameters	**Elasticity E_δ**	**Waiting time WTT^R**	**In-vehicle time IVT^R**	**Frequency**
Parameter values	***{-0.7, -1.1, -1.5}***	***{1.5, 2.0}***	***{1.0, 1.25}***	***{f^R, f^T}***
Scenario 1 (default)	-1.1	1.5	1.0	f^R
Scenario 2	-1.1	1.5	1.0	$\min(f^R; f^T)$
…	…	…	…	…

Step 3. Specific in-depth search

In the third step of the calibration, the parameter set is further improved based on the boundaries set in the second step of the procedure, using a specific in-depth search procedure. In this step, parameter values are not bound to the predefined values and scenarios anymore. Within the boundaries set in step 2, we evaluated all combinations of parameter values (with a certain minimum step size applied between the candidate parameter values). Since all these candidate parameter sets are within bounds of the promising, robust parameter sets found in step 2, we investigated which parameter set resulted in the highest prediction accuracy measured by the R^2 measure and selected this parameter set as new, preferred set. Once this parameter set is determined, this set is validated by applying it to the investigated disturbances $\delta \in \Delta_B$. For this subset Δ_B it is tested whether the prediction accuracy improved compared to the default parameter set. Note that we did not apply a full optimisation in this process of improving the parameter set. We used a rule-based approach to narrow down the solution space in a three-step procedure, meaning that not all possible combinations of parameter values are evaluated. Optimality can thus not be proved in this method.

4.3 Case Study

The methodology as described in **Section 4.2** is applied in a case study. The urban public transport network of The Hague, the Netherlands, is used in this study. Public transport services on this network are operated by HTM. The network consists of 12 tram lines and 8 bus lines. No metro services are operated in the city of The Hague. Two of the tram lines function as light rail connection between The Hague and the nearby suburb of Zoetermeer. On an average

working day, more than 250,000 trips are made on the HTM network. 93% of the passengers use a smart card for travelling (HTM, 2015). The remaining 7% buys a ticket from the driver or at the vending machine, or uses a special ticket. When modelling the HTM network, four different time periods are distinguished in the frequency-based assignment and prediction model: morning peak (7am-9am), evening peak (4pm-6pm), off-peak (9am-4pm) and the evening and early morning (6pm-7am).

Table 4.3. Overview of network changes during planned disturbances in 2015

Disturbance δ	**Period δ (# days)**	**Period δ_0 (# days)**	**Directly affected lines $l \in L^a$**	**Rail-replacement line (replaced tram)**
δ_1 Closure 'Koninginnegracht'	November (20)	March (20)	Tram 1/15/16/17: rerouted Tram 9: shortened + bus-replacement	 Bus lines 69+79 (9)
δ_2 Closure 'Loosduinseweg'	August (5)	August (5)	Tram 2: shortened + bus-replacement Tram 4: shortened Tram 6: extended (to replace tram 4)	Bus line 52 (2)
δ_3 Closure 'Westvest'	October (5)	March (20)	Tram 1: shortened + bus-replacement	Bus line 71 (1)
δ_4 Closure 'Parallelweg+Ternoot'	July (5)	August (5)	Tram 3/4: cut into two separate parts Tram 9: rerouted + bus-replacement Tram 11/12: shortened + bus-replacement	 Bus line 58 (9) Bus line 82 (11+12)

In 2015 there were several track closures on the public transport network operated by HTM. Given their entry-exit AFC system, in combination with relatively many case studies available, the HTM network is an interesting case study area to investigate the impact of planned disturbances on public transport ridership. As explained, in total four different disturbances δ which occurred in 2015 on the HTM network are investigated, which are divided into two subsets $\Delta_A \in \Delta\{\delta_1, \delta_2\}$ and $\Delta_B \in \Delta\{\delta_3, \delta_4\}$ used for calibration and validation purposes, respectively. **Table 4.3** describes the impact of each disturbance on the public transport network, as well as the period of the year the AFC data is coming from for δ_0 and δ. **Figure 4.2** shows the adjusted public transport network for all four disturbances. Closure δ_1 'Koninginnegracht' resulted in detours for several tram lines in the city centre. Besides, one of the two important connections between Central Station and Scheveningen (tram line 9) was replaced by bus services of line 69 (whole day) and 79 (only peak hours). Closure δ_2 'Loosduinseweg' resulted in the shortening of two busy tram lines 2 and 4. The shortened part of the route of tram line 2 was replaced by buses (line 52). Most stops of the shortened tram line 4 were covered by tram line 6, which route was temporarily extended over the unaffected part of the shortened route of tram line 4. During closure δ_3 'Westvest', the route of tram line 1 - connecting the city of The Hague with the city of Delft - was shortened. A rail-replacement bus line 71 was provided, although it could not stop near all original tram stops due to infrastructure constraints. Closure δ_4 'Parallelweg+Ternoot' reflects two different disturbances which occurred simultaneously on the network. The Parallelweg closure affected the route of tram lines 9, 11 and 12 through an area with a relatively high public transport dependency. Lines 11 and 12 - which share the same route through this area - were shortened, whereas tram line 9 was rerouted. The closed track of line 9 was replaced by temporary bus line 58, whereas bus line 82 replaced the closed part of the route of tram lines 11 and 12. The closure Ternoot resulted in cutting both light rail services 3 and 4 in two separate parts. Passengers could walk between the short-turning terminals of both parts of these lines.

It can be concluded that the set of disturbances Δ can roughly be divided in closures in which tram lines are rerouted (δ_1, δ_4), closures in which tram lines are shortened and replaced by bus services ($\delta_1, \delta_2, \delta_3, \delta_4$), and closures in which tram lines are divided in two separate parts (δ_4). To investigate and test that the selected parameter set is robust to perform accurate predictions for different types of closures, we aimed to have a mixture of different types of

closures in both the subset used for calibration $\Delta_A \in \Delta\{\delta_1, \delta_2\}$, as well as in the subset used for validation $\Delta_B \in \Delta\{\delta_3, \delta_4\}$.

For the reference network δ_0 used for closures δ_1 and δ_3, as well as for disturbed network δ_1, 20 working days of smart card data is used in this study. Given the ≈250,000 journeys (≈300,000 AFC transactions) on the HTM network per average working day, this roughly means that about 6 million smart card transactions are used as basis for these analyses. For the shorter lasting disturbances δ_2, δ_3, δ_4 and the reference network δ_0 used for closures δ_2 and δ_4, within the summer holiday, about 1.5 million smart card transactions (5 working days) are used (**Table 4.3**). All raw transactions are anonymised by removing personal information and by aggregating the data, to guarantee confidentiality and to obey Dutch privacy regulations.

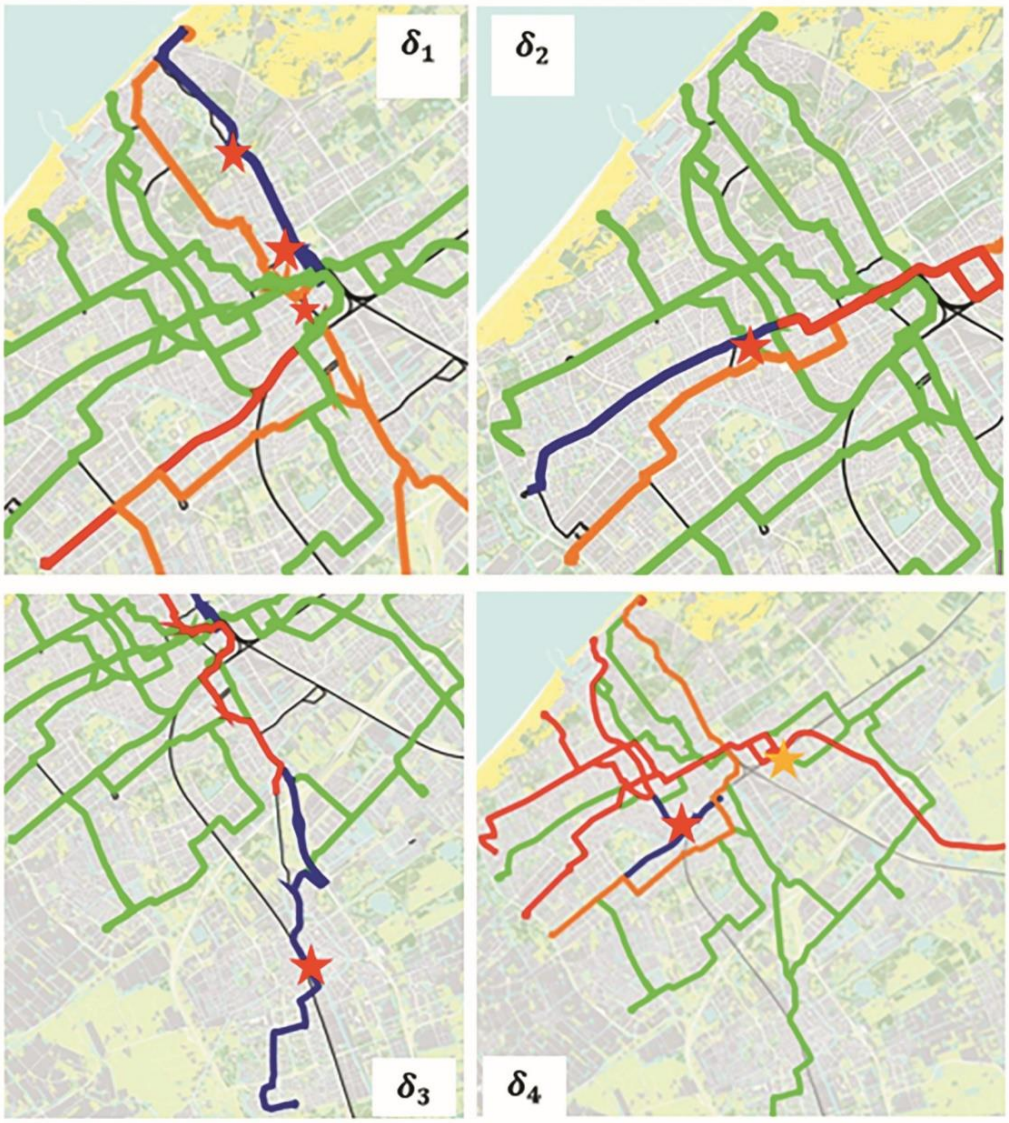


Figure 4.2. Public transport network during planned disturbances δ_1 'Koninginnegracht' (upper left), δ_2 'Loosduinseweg' (upper right), δ_3 'Westvest' (lower left), δ_4 'Parallelweg +Ternoot' (lower right)

(star: work location / orange: line rerouted / red: line shortened / blue: rail-replacement bus)

4.4 Results

4.4.1 Results rule-based search procedure

After the first selection of parameters with potentially different values during disturbances (step 1 of the three-step search procedure as described in **Section 4.2.4**), this section presents results of the second step (systematic grid-search) and the third step (specific in-depth search) of the method. **Table 4.4** and **Figure 4.3** show the resulting prediction accuracy for the two disturbances δ_1 and δ_2 used for calibration for all 24 predefined scenarios (as explained in **Section 4.2.4** and **Table 4.2**). In this step, we specifically search for patterns and robust parameter values.

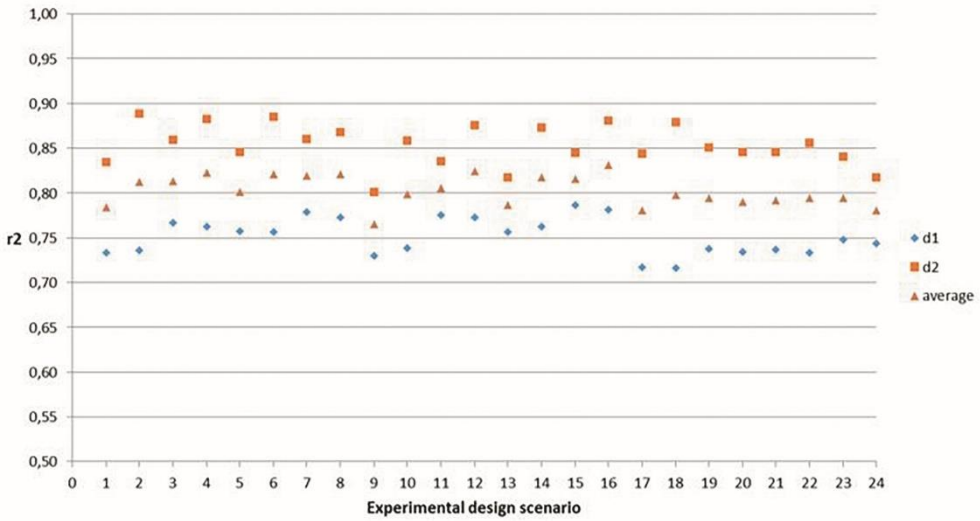


Figure 4.3. Results prediction accuracy systematic grid-search procedure

Based on **Table 4.4** and **Figure 4.3**, we conclude the following:

- Regardless the value of E_δ and the modelled frequency f, scenarios with WTT^R = 1.5 and IVT^R = 1.0 are outperformed by the other scenarios. From the resulting prediction accuracy for δ_1 we see that scenario 1+2, 9+10 and 17+18 clearly have a lower R^2 than scenarios 3-8, 11-16 and 19-24 respectively with the same value for E_δ. This indicates that this combination of values for waiting time and in-vehicle time does not put sufficient perceived disutility on usage of rail-replacement buses. This means that scenarios 1, 2, 9, 10, 17 and 18 are removed from the candidate list.
- Scenarios with E_δ = -1.5 are outperformed by scenarios using E_δ = -1.1 or E_δ = -0.7. From the resulting prediction accuracy for δ_1 we can see that for each given combination of WTT^R, IVT^R and f, scenarios 19-24 have a lower R^2 than their corresponding scenarios 3-8 and 11-16 with a different value for E_δ. This indicates that ridership loss due to temporal planned disturbances in general is not larger compared to the assumed ridership loss for structural network changes. This means that scenarios 19, 20, 21, 22, 23 and 24 are removed from the candidate list.
- Scenarios in which the realised frequency of rail-replacement buses f^R is modelled are outperformed by scenarios in which the frequency of the original tram lines which is being replaced f^T is modelled. From the resulting prediction accuracy for δ_2 we can see that for each given combination of E_δ, WTT^R and IVT^R the scenarios 4, 6, 8, 12, 14, 16 using f^T clearly outperform their corresponding scenario 3, 5, 7, 11, 13, 15 respectively

using f^R. This indicates that using the frequency of the original tram line is a better ridership predictor than using the realised frequency of rail-replacement buses, and that scenarios 3, 5, 7, 11, 13 and 15 are removed from the candidate list.

- For the remaining scenarios 4, 6, 8, 12, 14 and 16, the prediction accuracy is ranked separately for the prediction performance for δ_1 and δ_2. Scenarios 4 and 6 perform very well for δ_2, but perform mediocre for δ_1. Scenario 8 performs medium-high for both δ_1 and δ_2. Scenarios 12, 14 and 16 perform medium-high to high for δ_2, and perform medium-high to high for δ_1. Since our aim is to use a robust parameter set, scenarios 12, 14 and 16 are used as result from this second step of the search procedure, to apply the specific in-depth search to in the third step of the method.

Table 4.4. Overview of prediction accuracy for all 24 scenarios

Parameters	**Elasticity E_δ**	**Waiting time WTT^R**	**In-vehicle time IVT^R**	**Frequency**	**R^2 δ_1**	**R^2 δ_2**	**R^2 *average***
Parameter values	***{-0.7, -1.1, -1.5}***	***{1.5, 2.0}***	***{1.0, 1.25}***	***{f^R, f^T)***			
Scenario 1.1 (default)	-1.1	1.5	1.0	f^R	0.73	0.83	0.78
Scenario 1.2	-1.1	1.5	1.0	MIN(f^R; f^T)	0.74	0.89	0.81
Scenario 1.3	-1.1	1.5	1.25	f^R	0.77	0.86	0.81
Scenario 1.4	-1.1	1.5	1.25	MIN(f^R; f^T)	0.76	0.88	0.82
Scenario 1.5	-1.1	2.0	1.0	f^R	0.76	0.85	0.80
Scenario 1.6	-1.1	2.0	1.0	MIN(f^R; f^T)	0.76	0.88	0.82
Scenario 1.7	-1.1	2.0	1.25	f^R	0.78	0.86	0.82
Scenario 1.8	-1.1	2.0	1.25	MIN(f^R; f^T)	0.77	0.87	0.82
Scenario 1.9	-0.7	1.5	1.0	f^R	0.73	0.80	0.77
Scenario 1.10	-0.7	1.5	1.0	MIN(f^R; f^T)	0.74	0.86	0.80
Scenario 1.11	-0.7	1.5	1.25	f^R	0.78	0.84	0.81
Scenario 1.12	-0.7	1.5	1.25	MIN(f^R; f^T)	0.77	0.88	0.82
Scenario 1.13	-0.7	2.0	1.0	f^R	0.76	0.82	0.79
Scenario 1.14	-0.7	2.0	1.0	MIN(f^R; f^T)	0.76	0.87	0.82
Scenario 1.15	-0.7	2.0	1.25	f^R	0.79	0.85	0.82
Scenario 1.16	-0.7	2.0	1.25	MIN(f^R; f^T)	0.78	0.88	0.83
Scenario 1.17	-1.5	1.5	1.0	f^R	0.72	0.84	0.78
Scenario 1.18	-1.5	1.5	1.0	MIN(f^R; f^T)	0.72	0.88	0.80
Scenario 1.19	-1.5	1.5	1.25	f^R	0.74	0.85	0.79
Scenario 1.20	-1.5	1.5	1.25	MIN(f^R; f^T)	0.73	0.85	0.79
Scenario 1.21	-1.5	2.0	1.0	f^R	0.74	0.85	0.79
Scenario 1.22	-1.5	2.0	1.0	MIN(f^R; f^T)	0.73	0.86	0.79
Scenario 1.23	-1.5	2.0	1.25	f^R	0.75	0.84	0.79
Scenario 1.24	-1.5	2.0	1.25	MIN(f^R; f^T)	0.74	0.82	0.78

Scenarios 12, 14 and 16 have the same value for E_δ (-0.7) and for f (f^T). This means we fix these two parameter values. These scenarios differ in the parameter values used for WTT^R and IVT^R, which range between 1.5-2.0 and 1.0-1.25, respectively. This means the waiting time perception for rail-replacement buses varies between the same value as for regular tram and bus lines, and a 33% higher value. The in-vehicle time perception of 1 minute in a rail-replacement bus varies between 0.8 and 1 minute in the original tram line. In this in-depth search we explore all different combinations of these parameter values, increasing parameter values with a step size of ≈10%. This leads to values for waiting time perception $\{1.5, 1.67, 1.84, 2.0\}$ and in-vehicle time perception in the original tram line for 1 minute travelling in the rail-replacement bus $\{0.8, 0.9, 1.0\}$. Inversed this leads to in-vehicle time perception in the rail-replacement bus for 1 minute travelling with the original tram line $\{1.0, 1.11, 1.25\}$. These 3*4 combinations are evaluated, of which is known from step 2 that the combination of using a waiting time perception of 1.5 with 1.0 minute in-vehicle time perception simultaneously (scenario 2.9) does

not provide a good prediction accuracy. **Table 4.5** shows these 12 scenarios with the resulting R^2 averaged over δ_1 and δ_2.

From **Table 4.5** we can conclude that the averaged prediction accuracy is highest for scenarios 2, 3, 4 and 8. To select a final parameter set, we applied these four parameter sets to the subset of disturbances used for validation $\Delta_B \in \Delta \{\delta_3, \delta_4\}$ as well. The aim was to investigate which parameter set resulted in the highest prediction accuracy for Δ_B, in order to select the most robust parameter set. This resulted in the selection of scenario 8, in which WTT^R = 2.0 and IVT^R=1.11.

Table 4.5. Overview of prediction accuracy for in-depth search

Parameters	**Elasticity E_δ**	**Waiting time WTT^R**	**In-vehicle time IVT^R**	**Frequency**	**R^2 average**
Parameter values	***{-0.7}***	***{1.5, 1.67,1.84,2.0}***	***{1.0, 1.11,1.25}***	***{ f^T)***	
Scenario 2.1	-0.7	1.5	1.25	MIN(f^R; f^T)	0.82
Scenario 2.2	-0.7	1.67	1.25	MIN(f^R; f^T)	0.83
Scenario 2.3	-0.7	1.84	1.25	MIN(f^R; f^T)	0.83
Scenario 2.4	-0.7	2.0	1.25	MIN(f^R; f^T)	0.83
Scenario 2.5	-0.7	1.5	1.11	MIN(f^R; f^T)	0.81
Scenario 2.6	-0.7	1.67	1.11	MIN(f^R; f^T)	0.82
Scenario 2.7	-0.7	1.84	1.11	MIN(f^R; f^T)	0.82
Scenario 2.8	-0.7	2.0	1.11	MIN(f^R; f^T)	0.83
Scenario 2.9	-0.7	1.5	1.0	MIN(f^R; f^T)	0.80
Scenario 2.10	-0.7	1.67	1.0	MIN(f^R; f^T)	0.81
Scenario 2.11	-0.7	1.84	1.0	MIN(f^R; f^T)	0.81
Scenario 2.12	-0.7	2.0	1.0	MIN(f^R; f^T)	0.82

4.4.2 Resulting parameter set

Table 4.6 shows the values for the proposed new parameter set, compared to the parameter values of the default set. In case of planned disturbances, an elasticity parameter value E_δ of -0.7 resulted in the best fit with the raw smart card data. The hypothesis that passengers might perceive waiting time for a rail-replacement service more negatively compared to waiting time for regular tram and bus services was confirmed, since applying a higher waiting time coefficient did improve the prediction accuracy. Therefore, the use of a specific waiting time coefficient of 2.0 for rail-replacement services is proposed. Also, applying a more negative in-vehicle time perception in bus services replacing an existing tram line did improve the prediction accuracy. In the used prediction model in this study, this is reflected by applying a certain multiplication factor to the operational speed of a rail-replacement service. Using a speed factor of 0.9 - which equals the inverse in-vehicle time coefficient of 1.11 - resulted in the best fit with the raw smart card data. The value for the in-vehicle time coefficient derived from realisation data is somewhat lower than the speed factor found in Bunschoten et al. (2013) using a stated preference experiment, but points towards the same direction. Regarding the frequency of rail-replacement services, study results indicate that modelling the frequency of the original tram line f^T leads to a better fit than using f^R, if $f^R > f^T$. This indicates that passengers do not incorporate the benefit of the higher frequency of the rail-replacement services compared to the original tram line in their route and mode choice. From a theoretical perspective, this can be explained because vehicle capacity is not incorporated in the prediction model used in this study. The higher frequency f^R compared to f^T is often due to the lower capacity of a bus compared to a tram vehicle. Since the negative effect of a lower bus capacity is not incorporated in the model, only incorporating the positive effect of a higher bus frequency aimed to compensate for this non-incorporated capacity effect would overestimate the level of

service provided by the rail-replacement bus service. In case $f^R < f^T$, one could apply f^R in the prediction model. In this case it would even underestimate the negative effect of the bus replacement service, since only the additional waiting time (and not the lower vehicle capacity) is incorporated in the ridership prediction. Therefore, it is proposed to use the minimum of f^R and f^T as frequency of rail-replacement bus services in ridership predictions, when vehicle capacity is not incorporated in the used model.

Table 4.6. Comparison between default and new proposed parameter set

Parameter	Default parameter values	New parameter values
Elasticity E_δ	-1.1	-0.7
Waiting time coefficient a_3 for WTT^R	1.5	2.0
In-vehicle time coefficient a_1 for IVT^R	1.0	1.11
Frequency f^R	f^R	$min(f^R; f^T)$

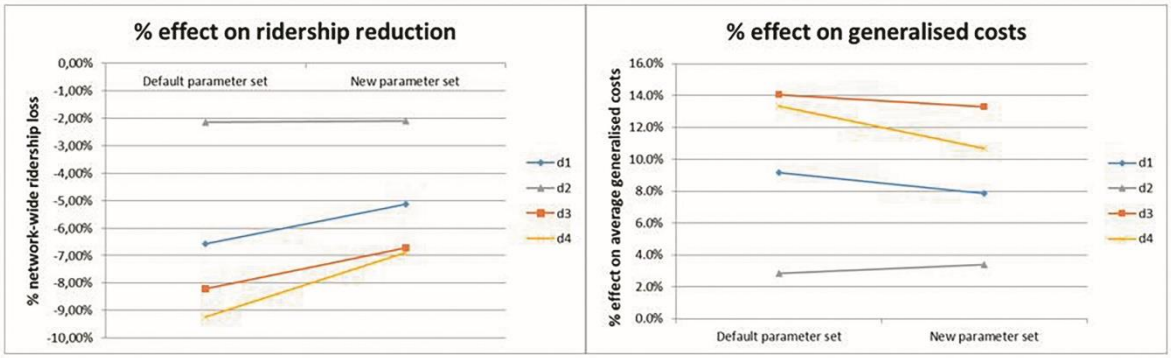


Figure 4.4. Effect of new parameter set on predicted ridership reduction (left) and on the average generalised travel costs per passenger (right)

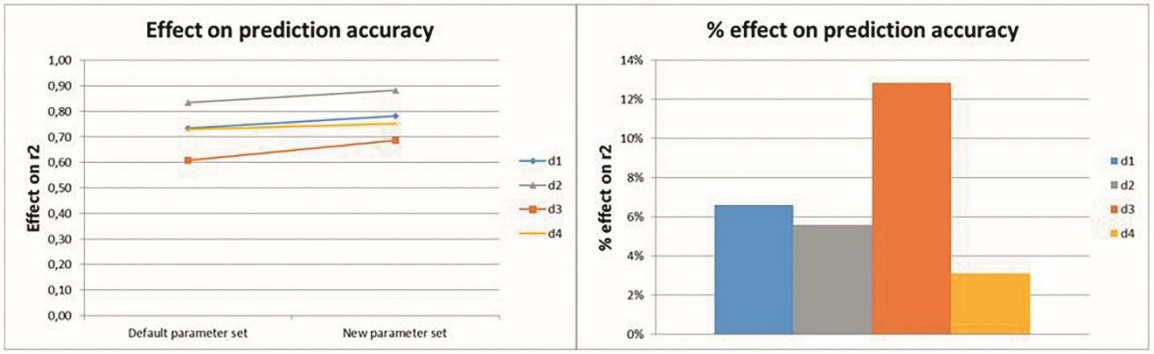


Figure 4.5. Absolute (left) and relative (right) effect of new parameter set on prediction accuracy

Figure 4.4 shows for all four disturbances investigated in this study the effect of the default and new proposed parameter set on the predicted ridership reduction (left) and on the average generalised travel costs per passenger (right), respectively. It can be seen that for all disturbances the new parameter set predicts a lower reduction in public transport ridership compared to the default parameter set. For δ_2 this effect is limited. However, for $\delta_1, \delta_3, \delta_4$ the predicted ridership reduction is 20-25% lower compared to the default parameter set. This can mainly be explained by the less negative elasticity value E_δ (-0.7 instead of -1.1). From **Figure 4.4** (right) can also be concluded that for $\delta_1, \delta_3, \delta_4$ the average generalised travel costs increase is 5-20% less when applying the new parameter set, while these increase with 20% for δ_2 compared to the default parameter set. In the new parameter set, the use of rail-replacement bus

services results in more perceived disutility compared to the default set, due to the use of a higher value for WTT^R, IVT^R and the longer average waiting time due to using $min(f^R, f^T)$ instead of f^R. This would imply an increase in average generalised costs. However, the lower value for E_δ leads to a substantial lower predicted ridership reduction for $\delta_1, \delta_3, \delta_4$, of which the net effect is a decrease in average generalised costs. Since the predicted ridership hardly decreases for δ_2 when using the new parameter set, for this disturbance this effect does not fully compensate the increased generalised costs, therefore leading to a net increase in average generalised costs.

Figure 4.5 shows that the new parameter set results in an increased predicted accuracy for all four disturbances considered in this study. This shows that the new proposed parameter set is robust to better predict ridership effects of several (types of) planned disturbances. The relative improvement in prediction accuracy ranges between 3 and 13%. In the calculation of the R^2, for $\delta_1, \delta_2, \delta_4$ in total $|L^a| * |T| * 2$ elements are computed. However, for δ_3 only $|L^a| * 2$ elements are used for the R^2 computation. This is because δ_3 occurred during the full duration of the autumn holiday. During this holiday, travel patterns differ from travel patterns during regular working days or during summer holiday, especially regarding the distribution over the day. Since there is no period of the year for which δ_0 occurred with a similar travel pattern, the effects per time period cannot be compared between δ_0 and δ_3. Therefore, we only compared these scenarios aggregated over all time periods per working day per line $l \in L^a$.

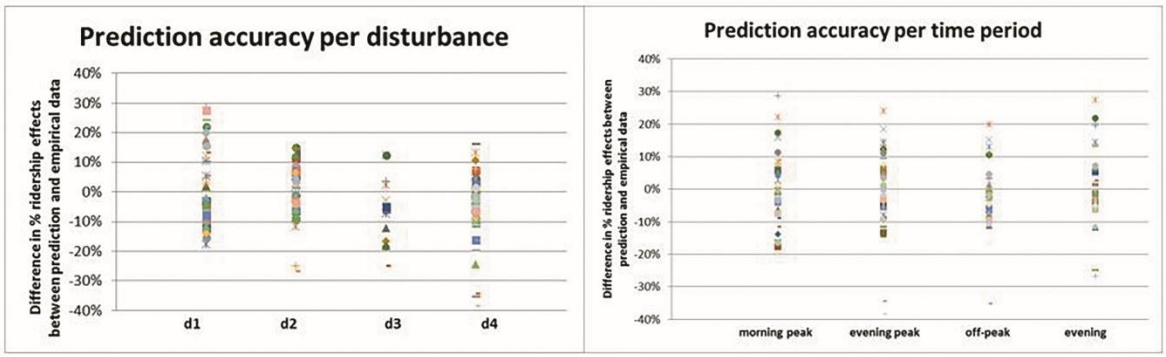


Figure 4.6. Resulting prediction accuracy for all $|L^a| * |T| * 2$ elements clustered per disturbance (left) and per time period (right)

Figure 4.6 shows the differences between the predicted and empirical relative effect of disturbances on public transport ridership for all elements, clustered per disturbance (left) and per time period (right). A positive value indicates that the predicted relative effect is less negative / more positive compared to the empirical data: ridership loss is thus *underestimated*, whereas ridership increases are *overestimated*. A negative value indicates the opposite; meaning that ridership losses are *overestimated* and ridership increases are *underestimated.* As can be seen, most predictions deviate in an acceptable range of +/- 10% from the empirical data. For δ_2, δ_4 (disturbances both occurring during summer holiday) can be observed that predictions tend to overestimate ridership losses, whereas for δ_1(a disturbance occurring during working period on a line with a high share of commuting and business passengers) predictions tend to underestimate ridership losses (**Figure 4.6**, left). We can hypothesise that frequent, commuting passengers have a higher sensitivity for network changes than average, whereas passenger segments with a relatively high share of leisure / shopping trip purposes have a lower sensitivity for network changes than average. When considering prediction accuracy for different time periods, we see similar patterns. Most predictions in the different time periods do not deviate more than +/- 15% from the empirical data. For the morning peak, a slight tendency can be

observed that predictions underestimate ridership losses, whereas for the evening period an opposite tendency can slightly be observed. Based on this we might hypothesise that passenger segments dominant during morning peak (mostly business and commuting passengers) have a relatively high sensitivity to network changes compared to passenger segments dominant during the evening period (mostly leisure passengers).

4.4.3 Reflection

In this study the accuracy of predicting the impact of temporary track closures on the number of passengers and passenger-distance is substantially improved by using the new proposed parameter set. Future research following this study can focus on four different aspects. First, in this study a rule-based three-step search procedure is applied to evaluate the prediction quality of several parameter sets. After first scanning the solution space using a systematic grid-search over different combinations of parameter values, an in-depth search around promising parameter values is performed. In order to further optimise the parameter set, it is recommended to perform a full optimisation evaluating all combinatorial possibilities, and/or to estimate a discrete choice model based on the revealed preference smart card data found for several disturbances to infer parameter values. This can lead to a set of parameter values which fit the empirical data in an optimal way and thus improving the prediction accuracy. Second, another interesting approach for future research is the application of machine learning approaches, training the prediction model based on passenger behaviour during several historical disturbances which occurred on the public transport network. Prediction accuracy results from machine learning approaches can then be compared with our developed methodology. Third, it is recommended to investigate whether location-specific, and/or purpose-specific parameter values increase the prediction accuracy. In this study only generic and mode-specific parameter values are applied. Distinguishing between areas where a disturbance occurs based on socio-economic characteristics (e.g. age, income, percentage of public transport captives) might lead to better predictions. In addition, using different parameter values for different passenger segments based on their trip purpose, or based on different time periods (peak/off-peak, weekday/weekend) might improve the prediction accuracy, since sensitivities for different parameters might be different for these segments. Fourth, a path for future research could focus on the development of supply-side indicators to identify and categorise disturbances. Our study mainly focuses on demand predictions resulting from disturbances, without making a distinction between different types of disturbances. However, developing indicators to describe disturbances and categorise the type of disturbance can be valuable, since it allows for the development of different demand prediction parameter sets for different types of disturbances. This can result in a further improvement of the prediction accuracy.

4.5 Conclusions

The introduction of automated fare collection (AFC) systems in public transport travelling the last decades leads to the availability of large quantities of smart card data, which can be used to analyse current travel patterns and to predict the effect of structural network changes on future public transport usage. Predicting the impact of planned disturbances, like temporary track closures, on public transport ridership is however still an unexplored area. In the Netherlands, this area becomes increasingly important, given the many track closures public transport operators are being confronted with the last and upcoming years. Better predicting the impact of these disturbances gives more insights in passenger behaviour during disturbances, and supports operators in aligning the supply of rail-replacement bus services with remaining demand and in predicting the impact on their revenues.

In this study, we investigated the passenger impact of planned disturbances by comparing predicted and realised public transport ridership using smart card data. Four different disturbances which occurred in 2015 on the public transport network operated by HTM in The Hague, the Netherlands, are analysed. A rule-based three-step search procedure is applied to find a parameter set resulting in higher prediction accuracy. The parameter set is calibrated using two of the four investigated disturbances, whereas the remaining two disturbances are used to validate the model.

Based on the study results we found a more negative in-vehicle time perception in rail-replacing bus services compared to in-vehicle time perception in the initial tram line. One minute tram travelling shows to be perceived as about 1.11 minute travelling in a rail-replacement bus service. Besides, when modelling rail-replacement services, it is recommended to use the frequency of the initial tram line instead of the usually higher frequency of the rail-replacement services. Passengers do not seem to perceive this theoretical benefit of higher frequencies of the rail-replacement bus, since this compensates for the lower vehicle capacity of a bus compared to a tram. If vehicle capacity is not incorporated in the prediction model, only incorporating the positive effect of a higher bus frequency aimed to compensate for this non-incorporated capacity effect would overestimate the level of service of the rail-replacement bus service. At last, also a 33% higher waiting time perception for temporary rail-replacement services could be found, compared to waiting time perception for regular tram and bus lines. The new parameter set leads up to 13% higher prediction accuracy compared to the default parameter set, and shows to be a robust and valuable tool for public transport operators. The prediction model is used by HTM in practice, in which the parameter set as recommended in this study is applied. Monitoring and further improving the prediction accuracy of the model will remain an important focus in the future.

Part II

Predicting Disruptions and Disruption Impacts

5. Predicting Disruptions and their Passenger Delay Impacts for Public Transport Stops

The objective of this chapter is the development of a methodology to predict how often different stops of an urban metro network are exposed to different disruption types and to predict the passenger delay impact of these disruptions. For this purpose, we develop supervised and unsupervised learning models. This chapter contributes to answering Research Question 2 (as defined in **Section 1.3**): how can disruption frequency and impact predictions be incorporated in a public transport vulnerability analysis for urban and multi-level public transport networks? In Part I of this research, the focus of **Chapters 2-4** was to improve methods to measure public transport impacts from empirical data. Given the relatively infrequent occurrence of public transport disruptions, empirical data is typically not available to infer disruption impacts for each location of a public transport network, during each time period, for each disruption type. Hence, it is necessary to move from *measuring* disruption impacts (Part I of this research) towards *predicting* disruptions and their impacts (Part II of this research). A systematic prediction of disruption frequencies and disruption impacts for all locations of a public transport network provides insights in the contribution of each public transport stop to total network vulnerability, thereby identifying the most critical stops in a vulnerability analysis. It supports public transport service providers and authorities in the decision-making process where and what type of mitigation measures need to be prioritised, in order to address the locations in the network which contribute most to vulnerability.

This chapter is based on an edited version of the following articles:

Yap, M.D., Cats, O. (under review). Predicting disruptions and their passenger delay impacts for public transport stops.

Yap, M.D., Cats, O. (2019). Analysis and prediction of disruptions in metro networks. *Proceedings of the 6th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS),* Krakow, Poland. (earlier version)
© 2019 IEEE

5.1 Introduction

5.1.1 Relevance

Disruptions in public transport (PT) can have major implications for passengers and service providers. Disruptions can increase passengers' nominal travel time, due to additional waiting time, in-vehicle time or transfers. Furthermore, passengers potentially experience higher crowding levels on alternative services, resulting in a more negatively perceived in-vehicle time (Hörcher et al., 2017; Tirachini et al., 2017; Yap et al., 2018a). Disruptions can also imply costs for the service provider, due to overtime payments to personnel, possible fare reimbursement for delayed passengers, and in the case of contractual agreements between service provider and authority result in fines. In the long term, disruptions can result in a loss of revenue if ridership levels decrease because of (perceived) unreliability of the PT system. It is thus in the interest of passengers and service provider to examine and assess the frequency, location and passenger delay impact of different disruption types occurring at each public transport station or link. An accurate prediction of the occurrence and impact of disruptions supports PT authorities and service providers in prioritising the locations and disruption types for which they should devise measures to reduce disruptions or their impacts.

5.1.2 Definitions and scope

For the remainder of this chapter, we first introduce several definitions used throughout this work. We apply a definition of *vulnerability*, which is obtained by combining definitions from Rodriguez-Nunes and Garcia-Palomares (2014) and Oliveira et al. (2016), with *robustness* being its antonym. Vulnerability is defined as the degree of susceptibility of a PT network to disruptions and the ability of the PT network to cope with these disruptions. This definition highlights the two components vulnerability consists of: *exposure*, the degree to which a PT system is exposed to disruptions, and the *impact* once a disruption occurs. Moving from a network level to individual elements, we define *criticality* as the degree to which an individual element of a PT system - such as a PT node or link - contributes to vulnerability. Criticality again refers to both disruption exposure and impact: it considers both *weakness*, the degree of disruption exposure for an individual stop or link, and *importance*, the impact of disruptions occurring at a stop or link (Cats et al., 2016b). The most critical nodes or links thus contribute most to PT vulnerability in terms of the product of their weakness and importance.

We consider both recurrent and non-recurrent PT disruptions in our study. Recurrent PT disruptions, such as a vehicle door malfunctioning or a delayed departure from the terminal, occur relatively frequently whilst the impact is generally limited. To the contrary, non-recurrent PT disruptions - such as a faulty train, signal failure or vehicle derailment - are relatively rare, but often have larger impacts once they occur. We do not consider extreme events as natural disasters or terror attacks in this research. These events differ substantially from typical PT disruptions in terms of frequency, location and impact, that a bespoke research approach is necessary. We focus on unplanned disruptions: planned disruptions, for example related to scheduled track maintenance works, fall outside the scope of this work.

5.1.3 State-of-the-art

Empirical data can contain information about the frequency with which different disruptions occurred, or about the disruption impact on passenger delays. However, to be able to study disruption frequencies and impacts of different disruption types for individual elements of a PT network, only using empirical data is typically insufficient. For illustration purposes, let us

consider a medium-sized PT network consisting of 100 stops, where our aim for each stop is to predict disruption frequencies and impacts for 20 different disruption types, for five different time periods of the day and week, separately for each season. This would require empirically deriving disruption frequencies and impacts for 100*20*5*4 = 40,000 instances. Consequentially, this would require sufficient empirical observations for each of these 40,000 instances to fit a probability density function for, to use empirical data reliably to predict future disruption frequencies or impacts. In practice, this means there will be insufficient empirical data available from past disruptions to use directly for future disruption occurrences and impacts for each of these instances. Therefore, some kind of prediction model becomes necessary to predict disruption frequency and impact for each individual PT network element.

In the field of transport vulnerability analysis, different approaches are applied to predict disruption impacts: *full scan* computation methods and methods using pre-selection indicators (Knoop et al., 2012). In a wider context, approaches to predict disruption impacts are broadly classified as scenario based, strategy based, simulation based or using mathematical modelling (Murray et al., 2008). For transportation networks, generally a static, dynamic or simulation based transport assignment model is used for this purpose. Full scan methods predict the disruption impact of each disruption type, at each location of a PT network. For example, in the context of highway networks, Jenelius (2007) uses a traffic simulation model where each link of the network is blocked, whilst Knoop and Hoogendoorn (2008) also incorporate dynamic spillback effects of blocked links. Full scan methods result in impact predictions for each individual network component, hence allowing all stops or links being ranked according to their contribution to network vulnerability. However, these methods are computationally prohibitive for larger networks and are typically only feasible to apply for smaller or case study networks. Instead, pre-selection methods apply indicators which result in a shortlist of locations where disruption impacts are expected to be most severe. Full disruption impacts are only modelled or simulated for this selection of locations. For example, Tampère et al. (2007) assess the expected criticality of road network links based on multiple indicators, such as the incident impact factor. Other road network vulnerability indicators used in literature are the Network Robustness Index (Scott et al., 2006) and the Modified Network Robustness Index (Sullivan et al., 2010), which are used as proxy for the impact of a full or partial link blockage on the network performance, respectively. Bell (2003) and Zhang et al. (2010) both adopt a game theoretical approach to quantify indicators for network vulnerability. To assess vulnerability of PT networks, Derrible and Kennedy (2010) propose a robustness indicator which calculates the number of available paths in the event of a disruption, based on a graph representation of 33 metro networks worldwide. Cats et al. (2016b) compare a passenger betweenness centrality measure as proposed by Cats and Jenelius (2014) and a passenger-exposure measure as PT vulnerability indicator. Aforementioned studies adopt either a node-based or link-based vulnerability approach, whilst some studies consider the vulnerability impacts of joint node and link disruptions (see for example Dhin and Thai, 2014). The disadvantage of pre-selection approaches however is that there is no guarantee that the largest impacts occur at these selected locations. This means there is no certainty whether the most critical nodes or links of a network are correctly identified. Additionally, these approaches do not allow for a comparison of disruption impacts between all individual network elements, as disruption impacts are only quantified for selected elements. The abovementioned state-of-the-art illustrates that existing methods are insufficient to predict the passenger delay impacts of disruptions for each individual PT station or link for medium- or large-sized, real-world PT networks.

Several studies focus on predicting disruption impacts, once a disruption occurs. For example, studies quantify PT disruption impacts (Cats and Jenelius, 2015), the value of spare capacity in a PT network (Cats and Jenelius, 2014), or the impact of partial rather than complete track closures (Cats and Jenelius, 2018). Corman et al. (2014) evaluate the robustness of railway

timetables once a disruption occurs. However, focusing solely on disruption impacts without considering disruption frequencies can incorrectly put the emphasis on locations where very severe yet very rare disruptions occur. Predicting how often different locations in a PT network are exposed to different disruptions is a relatively understudied topic. There has been a vast amount of work towards predicting incident frequencies for road networks in the field of traffic safety. Whereas initial road traffic research primarily used descriptive and aggregate models to predict accident probabilities (e.g. Stone and Broughton, 2003; Lord et al., 2005), more recent research has moved towards using disaggregate, predictive models (e.g. Zou and Yue, 2017). For PT networks, the use of disaggregate models remains limited. An important reason is often a lack of good quality disruption log data, as data over a longer period of time is required given the relatively infrequent occurrence of disruptions. In Cats et al. (2016b) and Yap et al. (2018c), a database consisting of logged disruptions on a PT network for a period of 2.5 year was used to fit statistical models for disruption frequencies on a network level. In these studies, relatively simple predictors such as the number of trains or train-kilometres were used to translate the network-wide number of disruptions to expected disruption exposure per station or link. This implies that location-specific characteristics - such as the type of stock serving a station, the passenger load or the geographical area where a station is located - are not considered, while these are believed to be important when predicting disruption exposure. In Tonnelier et al. (2018) a data-driven method is developed to detect individual atypical events in PT networks using anomaly detection. However, this method does not explicitly provide what type of disruption at which location initiated this anomaly, making it difficult to formulate policy recommendations concerning how to tackle PT vulnerability. This means that currently no adequate disaggregate models have been developed to predict disruption frequencies for individual PT stops or links.

5.1.4 Research approach and contribution

Our study objective can be summarised as the development of a generic methodology to predict disruptions and their passenger delay impacts accurately for different disruption types, for individual stations of a real-world PT network, thereby incorporating the specific characteristics of the different stations. This implies we develop a disaggregate modelling approach to predict disruption frequencies and to predict the passenger delay impacts of each disruption. We propose a supervised learning approach to perform these predictions, as this allows for the prediction of disruptions at individual stations for each time period, without the requirement of having sufficient empirical disruption observations available for each location and time period. This approach also enables a fast prediction of disruption impacts for a large number of disruption instances, hence addressing the computational challenges that rise when typical PT assignment or simulation models would be used for real-world PT networks.

To improve transferability of our study results, we cluster stations based on their contribution to PT vulnerability using unsupervised learning. In addition to predicting disruptions and their impacts for a specific PT network, this provides PT authorities and service providers insight in the different station types that can be distinguished based on their contribution to network vulnerability. For example, for policy purposes train stations in the Netherlands are grouped into six categories based on function and passenger volumes (Geurs et al., 2016). Our research results in a natural clustering of stations in a similar way, specifically based on vulnerability. Hence, this supports PT agencies to apply the appropriate type of measure aimed to reduce disruptions or to mitigate disruption impacts for each station type. Our research contribution is therefore defined as follows.

Scientific contributions:

- Development of a method to predict disruptions and their passenger delay impacts for individual public transport stations, incorporating the specific characteristics of each station.
- Development of prediction models which predict disruptions and their impacts based on a non-exhaustive empirical disruption dataset within acceptable computation times.

Practical contributions:

- To provide PT agencies with predicted disruption impacts for each individual station on their network, for each distinguished time period and disruption type, supporting them to prioritise locations where to put mitigation measures in place.
- Identification of different groups of public transport stations with different disruption exposure and impact characteristics, enabling PT agencies to devise appropriate measures to tackle vulnerability for different station types.

The remainder of this chapter is structured as follows. **Section 5.2** explains the methodology, whilst **Section 5.3** introduces our case study network. We discuss results in **Section 5.4**, followed by conclusions in **Section 5.5**.

5.2 Methodology

In this section we discuss the proposed methodology to predict disruption exposure and impact at different PT stations, and to cluster stations accordingly. First, we introduce the proposed modelling framework (**Section 5.2.1**). In **Section 5.2.2** we explain our supervised learning model to predict disruptions, followed by the model for disruption impact predictions (**Section 5.2.3**). At last, **Section 5.2.4** discusses our station clustering approach.

5.2.1 Modelling framework

For a given PT network, let us define each station $s \in S$, with $|S|$ being the total number of stations in the considered network. Each disruption type is defined by d, with D indicating the total set. Each distinguished time period is indicated by $t \in T$. When we define the disruption frequency f and the disruption impact w, the predicted station criticality $\check{c}$ in its simplest form is defined by **Eq.1**. To obtain station criticality, the predicted frequency of each disruption (expressed in disruptions per year) at station s is multiplied by the predicted impact, and then summed over all disruption types and time periods considered per year. In our study, we predict the total passenger delay hours $\breve{w}_{d,t,s}$ as metric for disruption impact. It should be noted that disruption impacts are generally wider than passenger delays only. Passengers' perceived travel times often increase as well, whilst PT service providers might face rescheduling costs (e.g. personnel overtime payment) or passenger reimbursement costs. In this study we however only consider the nominal travel time impact a disruption has inflicted on passengers. This means that station criticality is expressed in yearly passenger delay hours. PT network vulnerability V then equals the sum of the predicted station criticality (**Eq.2**), and expresses the predicted yearly passenger delay hours for the total PT network of interest. For the sake of simplicity, the basic impact calculation as shown in **Eq.1** does not show interdependencies between different disruptions occurring simultaneously on the considered PT network, as this can result in interaction effects affecting the disruption impact. The integrated modelling framework to calculate $\breve{c}_s$ and $\breve{V}$ is shown in **Figure 5.1**. It shows the supervised learning models used to predict disruptions and passenger delay impacts, as well as the unsupervised learning model

applied to categorise different stations. This modelling framework is explained further in the remainder of this section.

$$\breve{c_s} = \sum_{t \in T} \sum_{d \in D} \breve{f}_{d,t,s} * \breve{w}_{d,t,s} \quad (1)$$

$$\breve{V} = \sum_{s \in S} \breve{c_s} \quad (2)$$

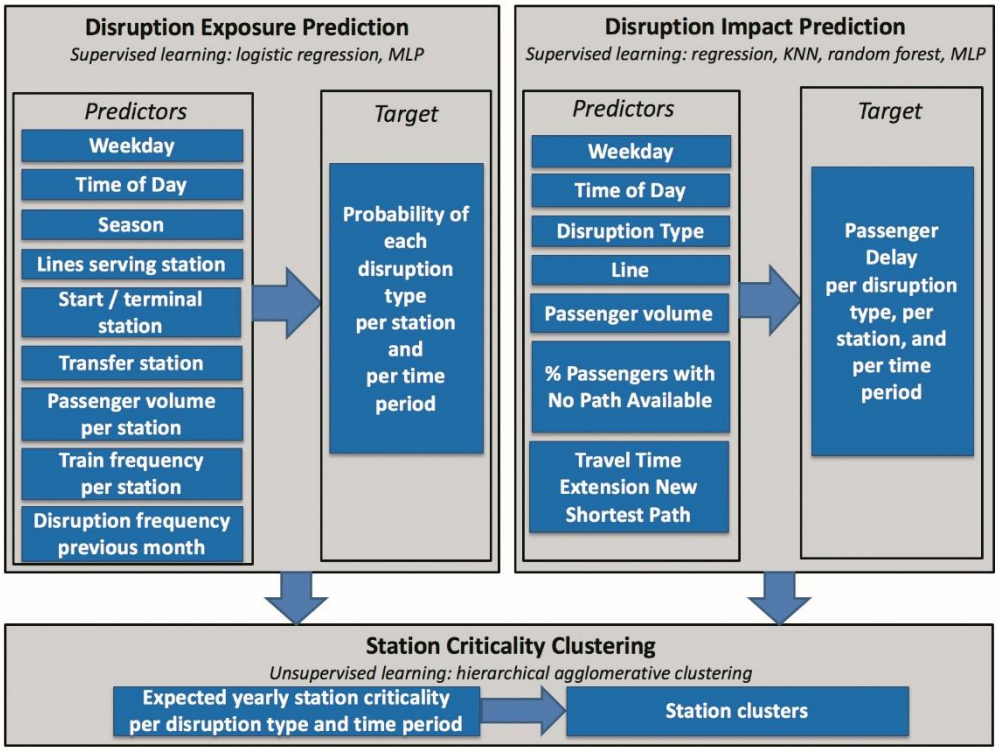


Figure 5.1. Modelling framework

For our proposed modelling framework, the following empirical data sources are required as input:

- Disruption log data, containing data for each PT disruption which occurred on the PT network within a certain time interval. As a minimum, for each disruption the start time, location and line of occurrence need to be logged, as well as the disruption type or a disruption description. Disruption end time is desirable though not mandatory for our method. This type of data is usually available at the PT authority or service provider, based on logged incident notifications from train drivers, station operators, control room staff, police and the general public.
- Individual passenger demand data from Automated Fare Collection (AFC) systems, which consists of the time and location of the first boarding and final alighting station of each individual passenger journey. This allows calculation of the realised journey time for each passenger.
- Scheduled journey times between each boarding and alighting station for each distinguished time period or day of the week, enabling a comparison between scheduled and realised passenger journey times. This data can be obtained from journey planners or can be provided by the PT service provider.

- Timetable data from GTFS or Automated Vehicle Location (AVL) systems (typically open data), which contains the planned number of PT trips for each route, during each time period and day of the week.

5.2.2 Disruption exposure prediction

In this study we adopt a supervised learning approach to predict exposure to different disruptions $d \in D$ at stations $s \in S$ during each time period $t \in T$. This allows us to find linear and non-linear relations between presumed disruption predictors and the exposure to disruptions with short computation times. As each disruption type occurs relatively infrequently at a specific station and in a specific time period, our study objective here implies predicting the occurrence of relatively rare events. For that reason we do not use $\breve{f}_{d,t,s}$ as our target, as a model always predicting zero for $\breve{f}_{d,t,s}$ would still result in a low *MSE-score* and a high average *F1-score* due to the overrepresentation of samples with zero disruptions, without providing any useful prediction for disruption exposure. Neither applying different weights for false positive and false negative predictions, nor applying a technique to correct the dataset imbalance such as a *Synthetic Minority Oversampling Technique* (SMOTE) did sufficiently improve the quality of disruption predictions. Instead, we therefore use the probability $\breve{p}_{d,t,s}$ of each disruption type occurring within each considered time period (e.g. each AM, PM, Inter Peak and Evening period for each day of the year) as target for the prediction. $\breve{f}_{d,t,s}$ is calculated by multiplying the predicted probabilities by the number of time periods $|T|$. The number of samples in our model thus equals $|S| * |T|$. To predict disruption probabilities we apply a classification algorithm, which calculates disruption probabilities for each $d \in D$ and then assigns each sample to one of the disruption categories d or to the category *no disruption* based on the highest probability. The shape of the target vector therefore equals $(|S| * |T|, 1)$, where column values can take $|D| + 1$ different values. In our case, this value equals 0 if no disruption is predicted to occur in the considered time period, and ranges between 1 and $|D|$ depending on which disruption type is predicted to occur within the time period. By dummy coding this target vector, a matrix with dimensions $(|S| * |T|, |D| + 1)$ results which contains the predicted probabilities for each disruption type.

We identify several general and location-specific station characteristics as predictors in our machine learning model (**Figure 5.1**, upper left). We first use the general predictors *Weekday, Time of Day* and *Season*. *Weekday* equals 1 if the time period is during a weekday, and 0 if during a weekend. *Time of Day* considers if the time period is during the peak (7-10AM or 3-7PM: only during weekdays), daytime off-peak (weekdays: hours outside peak until 7PM; weekend: all hours until 7PM) or evening (hours after 7PM). The aim of these predictors is particularly to capture the possible influence of differences in mixture of passenger types and travel purposes between peak, off-peak, evenings and weekends on disruption probabilities. The predictor *Seasons* aims to capture differences in disruption probabilities for different seasons. One can think of potentially more vehicle defects due to leaves in autumn, or more passenger-related incidents due to slippery surfaces in winter. Additionally, we identify several station-specific predictors. *Lines* refers to the different metro lines serving each station, as different stock types on different lines potentially influence especially railcar-related disruption probabilities. A possible difference in state and age of infrastructure between different lines can also play a role here. One-hot encoding is applied for the categorical predictors *Time of Day, Seasons* and *Lines*, resulting in separate binary predictors for each category. If a station is served by multiple lines, for example being part of a trunk section, the binary predictor equals 1 for each of these lines. Two separate binary predictors *Start station* and *Transfer station* are added, being equal to 1 if the station is a start/terminal or a metro-to-metro transfer station, respectively. It is expected that the occurrence of some disruptions is related to a station being a start/terminal, as problems such as a malfunctioning train or a late / absent train driver often arise here. It is

hypothesised that transfer stations might be more susceptible to disruptions due to more complex infrastructure (such as switches) and large passenger transfer volumes. *Passenger volume* refers to the number of boarding plus alighting passengers for each station and time period, based on Automated Fare Collection (AFC) data for an average day. This predictor is added to capture primarily passenger-related disruption probabilities. *Train frequency* equals the scheduled number of trains serving a stop during each time period and day of the week. This predictor is calculated based on timetable data, and aims to capture railcar-related disruption probabilities. *Disruption frequency previous month* is an auto regressor with respect to the number of disruptions that occurred during a certain time period, weekday/weekend at the considered train station in the previous month, for each disruption type separately. This predictor assumes disruption data of the previous month is available to predict disruption exposure in the next month. In total, $|D|$ separate predictors are used for this predictor, for each disruption type $d \in D$. Values for predictors *Passenger volume, Train frequency* and *Disruption frequency previous month* are all normalised between 0 and 1, so that all predictors use the same range.

Given our target to predict the probability of different disruption types in a certain time period at a certain station, we test two different machine learning algorithms suitable for this purpose: logistic regression and a Multilayer Perceptron (MLP) classifier, a class of feedforward artificial neural networks. The total dataset is split into an 80% training set and 20% testing set, applied in a randomised 5-fold cross validation. Applying a higher 10-fold cross validation did not significantly improve prediction accuracy. Log loss, or cross entropy loss, is used as evaluation metric (**Eq.3**). This function calculates the negative log-likelihood of the true label y, given the predicted probability that a sample equals this true label $\tilde{y}$.

$$-logP_{y|\tilde{y}} = -(y * log(P_{\tilde{y}}) + (1-y) * log(1-P_{\tilde{y}})) \quad (3)$$

5.2.3 Disruption impact prediction

For the prediction of passenger impacts of disruptions, we also apply a supervised learning approach. To quantify passenger delays, we compare the scheduled and realised passenger journey times at the considered PT network. For each journey between a given origin station $i \in S$ and destination station $j \in S$, the realised journey time $jt_{t,ij}$ is obtained from AFC data per time period t. Depending on whether a tap in only or tap in / tap out AFC system is in place, the destination of an AFC transaction is directly available or needs to be inferred using a destination inference algorithm (e.g. Munizaga and Palma, 2012). Besides, transfer inference might be required to connect AFC transactions to journeys, if transfers are made which require an intermediate AFC transaction (e.g. Gordon et al., 2013; Yap et al., 2017). The scheduled journey time for time $t \in T$ is calculated from the timetable. In-vehicle times and station walking times are assumed to be deterministic. The maximum scheduled journey time $\widetilde{jt}_{t,ij}^{max}$ assumes the passenger wait time is equal to the planned headway at that time (i.e. in case a passenger has just missed a train), whilst the minimum scheduled journey time $\widetilde{jt}_{t,ij}^{min}$ assumes a passenger can board the PT vehicle directly (no waiting time). The expected scheduled journey time $E(\widetilde{jt}_{t,ij})$ is then calculated as the average between $\widetilde{jt}_{t,ij}^{min}$ and $\widetilde{jt}_{t,ij}^{max}$. **Eq.5** shows the passenger delay calculation applied in our study, using dummy variable x_1 as defined in **Eq.4**. A journey in time period t is considered delayed if the realised journey time exceeds the maximum scheduled journey time. To prevent underestimating passenger delays in that case, the delay is calculated as the difference between realised and expected scheduled journey time (expressed in minutes) and multiplied by demand $q_{t,ij}$.

$$x_{1,ij} = \begin{cases} x_{1,ij} = 1 & if\ jt_{t,ij} > \tilde{jt}_{t,ij}^{max} \\ x_{1,ij} = 0 & if\ jt_{t,ij} \leq \tilde{jt}_{t,ij}^{max} \end{cases} \quad (4)$$

$$w_t = \sum_{i \in S} \sum_{j \in S} [jt_{t,ij} - E(\tilde{jt}_{t,ij})] \cdot q_{t,ij} \cdot x_{1,ij} \quad (5)$$

The structure of the machine learning model is shown in **Figure 5.1** (upper right). The passenger delay $\tilde{w}$ resulting from disruptions is used as target. It should be noted that this delay cannot be attributed directly to a certain disruption d, as several disruptions can occur spatially and/or temporally close to each other. As disruption end times are not always provided in disruption log data, the disruption duration cannot always be determined. Moreover, even if the end time of a disruption would be known, knock-on effects on passenger delays can persist for up to six times longer than the duration of the initial cause (Malandri et al., 2018). Once a disruption is resolved, there is typically recovery time required to reschedule PT trips and personnel before the origin timetable is restored. Hence, in any case the disruption log data does not provide information when the passenger delay impact for passengers ended. To mitigate this problem, our model is being trained using a rolling horizon where the total passenger delay w from time hour t up to two hours later $[t, t_{+2}]$ is used as target, as function of the considered disruption which started during t together with all other disruptions which started during this time window $[t, t_{+2}]$. This approach implies we consider disruption impacts up to three hours after the moment the disruption occurred. Although this can theoretically underestimate the impact of large disruptions somewhat, the majority of the disruptions on a PT network are typically relatively minor. Therefore, this time horizon is deemed reasonable to capture the complete disruption impacts for the vast majority of all disruptions. If the impact of disruptions which started at t would vanish before t_{+2}, the calculated passenger delay during t_{+2} is expected to be (close to) zero as well, meaning there is no penalty for adopting a relatively long time horizon for smaller disruptions. If a disruption end time would be available from the log data, a more accurate time period could be determined in the rolling horizon. For example, the end of the horizon could be set equal to the logged disruption end time, plus a time period reflecting recovery time as function of the logged disruption duration. We apply the final trained model to a new test dataset where only one disruption per t and s occurs, to predict the pure impact of each disruption separately.

As generic predictors for disruption impact, we use predictors *Time of Day* and *Weekday* (weekday, Saturday or Sunday). For the predictor *Disruption type*, we apply one-hot encoding for all disruption types $d \in D$, where a disruption type is coded as 1 in case this disruption has occurred within the time window $[t, t_{+2}]$. The PT line is also used as predictor. Each line on which a disruption occurred in this same time window $[t, t_{+2}]$ is coded as 1; other lines are coded zero. As it is expected that total passenger delay depends on the total passenger volume using the network at the considered time period, we also use the total demand q_t starting a journey in this time interval as predictor. As disruption impacts can propagate over the total PT network, the total demand summed over all origin-destination stations is used here. We use the *Percentage affected demand for which no (simple) path remains available* in case a disruption blocks services to/from station s as a station-specific predictor. In case of a higher percentage passengers for which no alternative routes remain available in case of a disruption, one might expect a larger passenger delay for the affected passengers. To quantify this predictor g_t, we first calculate the affected demand q_t^a. We represent the total PT network as directional graph $G(V, E)$ with each vertex $v \in V$ representing a stop and each edge $e \in E$ representing a direct PT connection between stops. For each OD pair we calculate the length of the shortest path (expressed in minutes) l_{ij} and the number of simple paths (without cycles) n_{ij} for the

undisrupted scenario. We define the affected demand as the demand between OD pairs for which the shortest path length increases or for which no simple paths remain available if all disrupted stations s^d where a disruption occurred in the time window $[t, t_{+2}]$ are removed from $G(V, E)$ (**Eq.8**). Based on this, g_t can be calculated as value ranging between 0 and 1 (**Eq.9**). For all affected demand for which at least one simple path remains available, the *Expected detour time* on the network (expressed in minutes) can be used as an additional station-specific predictor for the full passenger delay impact. To this end, the increased length of the shortest path can be computed, so that the passenger-weighted average travel time extension $\overline{\Delta h}$ can be quantified as an additional predictor for passenger delays (**Eq.10**). **Eq.6** and **Eq.7** introduce the required dummy variables $x_{2,ij}$ and $x_{3,ij}$.

$$x_{2,ij} = \begin{cases} x_{2,ij} = 1 & if\ \ l_{ij}^d > l_{ij} \\ x_{2,ij} = 0 & if\ \ l_{ij}^d \leq l_{ij} \end{cases} \tag{6}$$

$$x_{3,ij} = \begin{cases} x_{3,ij} = 1 & if\ n_{ij}^d = 0 \\ x_{3,ij} = 0 & if\ n_{ij}^d > 0 \end{cases} \tag{7}$$

$$q_t^a = \sum_{i \in S} \sum_{j \in S} [q_{t,ij} \cdot \max(x_{2,ij}, x_{3,ij})] \tag{8}$$

$$g_t = \frac{\sum_{i \in S} \sum_{j \in S} q_{t,ij} \cdot x_{3,ij}}{q_t^a} \tag{9}$$

$$\overline{\Delta h_t} = \frac{\sum_{i \in S} \sum_{j \in S} [q_{t,ij} \cdot \left(l_{ij}^d - l_{ij}\right) \cdot x_{2,ij}]}{q_t^a \cdot (1 - g_t)} \tag{10}$$

We test different supervised learning regression models to predict passenger delays. We apply a simple linear regression model as baseline, and compare these results with a K-Nearest Neighbours (KNN), Random Forest and Multilayer Perceptron (MLP) regressor. For all these regression models, we apply a randomised 5-fold cross validation and use the *RMSE* (root-mean-squared error) as performance metric. The total number of samples of our models equals the number of disruptions in our database.

5.2.4 Clustering station criticality

The output of the final models which predict disruption exposure and impact is used as input to cluster PT stations based on their expected criticality. The final disruption exposure model is applied to predict disruption probabilities for each disruption type d for one complete year, whilst the final disruption impact model is used to predict the impact for each disruption type at each station s separately for each time period. Multiplication using **Eq.1** results in the expected yearly criticality per disruption type, station and time period, expressed in yearly passenger delay hours. We apply an unsupervised learning method to cluster stations based on this expected criticality (**Figure 5.1**, lower part). This provides insight in differences in susceptibility for different disruption types between stations, and shows clusters of stations with similar disruption exposure and impact patterns.

As our aim is to cluster all stations $s \in S$ without outliers, and no number of clusters k is known *a priori*, we apply hierarchical agglomerative clustering. Input for the clustering is a

matrix consisting of values $\check{c}_{s,d,t}$ with dimensions $(|S|, |D| * |T|)$, which results from our supervised learning models. The distance matrix is determined by calculating the $|D| * |T|$-dimensional Euclidean distance between all points. Ward is used as linkage criterion during the clustering, thereby minimising the within-cluster variance. We use the cophenetic correlation coefficient to assess the degree to which the clustering reflects the input data. The optimal number of clusters k is determined based on visual inspection of the dendrogram and maximising the average silhouette coefficient. The silhouette coefficient for each sample is calculated by taking the difference between the Euclidean distance to the nearest cluster this sample is not part of, and the intra-cluster distance. This difference is then divided by the maximum value of these two. The average silhouette coefficient is obtained by calculating this for all $|S|$ stations.

5.3 Case Study

5.3.1 Case study network

We apply our proposed methodology to the Washington D.C. metro network as case study. The Washington Metro, administered by WMATA, consists of six lines indicated by different colours: the Red line (R), Green line (G), Yellow line (Y), Blue line (B), Orange line (O) and Silver line (S) (**Figure 5.2**). The total length of the metro network is about 190 km. During AM and PM peak hours, the Red line runs 15 trains per hour (tph), of which every other train is a short-turning service to Silver Spring. The other lines run 7.5tph during peak hours. During daytime off-peak periods, all lines run 5tph. The Blue, Orange and Silver line share a substantial part of their routes between Rosslyn and Stadium-Armory. The joint frequency on this trunk section equals 22.5tph during peak hours. At the time of consideration, 95 different metro stations are operational, thus $|S|$=95. We predict disruption probabilities for all distinguished time periods (peak (only for weekdays), daytime off-peak and evening) for a full year. Every week thus consists of 19 time periods (3 time periods for weekdays and 2 time periods for weekend days). Hence, for a complete year $|T|$ equals 991. For our case study network, the total number of samples for the disruption prediction model equals $|S| * |T|$ = 94,145.

Figure 5.2. WMATA metro network

(Map obtained from WMATA: https://www.wmata.com/schedules/maps/upload/2019-System-Map.pdf)

5.3.2 Input data sources

One of the important data sources used as input for our method is incident log data. A 13-month incident database for the Washington metro network is provided by WMATA, which initially consisted of 21,868 records covering all reported incidents from August 1st 2017 to August 31st 2018. The attributes of each record as shown in **Table 5.1**. This shows that each record consist of a start time, incident location, line, train number, disruption category and description. Besides, the minutes of initial train delay (delays for an individual train) and line delay (delay for the entire line) as a result of an incident are indicated. This reflects only the initial delay a certain incident had on the train or line involved, and does not contain any information about the possibly (wider) passenger delay impact following this initial delay due to spill-over effects. The column *Initial incident* indicates if an incident is the result of another incident. The same number in the *Initial incident* column indicates that two logged incidents are related to each other. As the end time of disruptions is not logged for our case study, we use the time window $[t, t_{+2}]$ as rolling horizon in the disruption impact prediction, with t being the hour when the disruption starts (see **Section 5.2.3**). We use all disruptions in the 12-month period from September 1st 2017 to August 31st 2018 as input for our disruption exposure prediction model. Disruption data for August 2017 is used to quantify values for the auto regressor predictor *Disruption frequency previous month* (see **Section 5.2.2**) for disruption probabilities in September 2017.

Table 5.1. Example incident log data

ID	Start Time	Line	Train	Stop	Type	Description	Train delay	Line delay	Initial incident
11	16-08-17 8:30	Blue	419	C07	AIRL	Air leak	5	5	11
12	23-08-17 9:13	Red	231	A11	PUBL	Sick customer	3	0	12

Incident log data of PT systems is generally not primarily intended for vulnerability analysis purposes or to draw policy recommendations from. Instead, this is usually filled out during the real-time control process in the control room when recovering train services. This also entails there might be only a limited degree of consistency in the description and classification of incident notifications, as it strongly depends on manual actions from controllers whose main priority is solving the incident. This was also the case for the Washington dataset provided to us. As a result, it is important to reassure the incident database is fit for our study purpose, for which we perform two data processing steps: a) deriving disruptions from incidents, and b) classifying disruptions.

First, we derive disruptions from incidents in the log file, as this database also contains incidents which did not result in a disruption. For example, a driver not able to perform its duty due to sickness is reported in the incident database, even if a stand-by driver took over the shift without any delays. For our case study, we define a disruption as any incident where either the train delay or line delay is 2 minutes or more. Incidents with both the train and line delay being smaller than 2 minutes are regarded as regular service variability. Additionally, when multiple incidents in the database are related to the same incident, only the initial incident is kept. Other delays can be considered a consequence of this initial incident, rather than separate incidents. When applying our disruption definition, 4,263 distinguishable disruptions remain in the 12-month period from September 2017 to August 2018.

Second, disruptions are classified into a select number of distinctive disruption types. In the provided database, 114 different disruption types are logged. When considering the distribution over different stations $s \in S$ and time periods $t \in T$, there would be an insufficient number of observations per station and time period in the database to develop a prediction model for. Besides, in some cases different definitions were used for the same or very similar disruption types, due to differences in classification by different controllers. For example, in the used

database a train car motor overload is indicated by both disruption type *MOLD* ('motor overload') and *MOLF* ('flashing motor overload'). In these cases, one consistent disruption type is attributed to both disruptions. In some cases, the disruption types in the database did not reflect the root disruption cause. As an illustration, one can find an incident registered as *ONEC* ('operational necessity') with the description 'late dispatch due to door not closing'. In this case, the root cause is a door malfunctioning, resulting in an operational action from the control room. In a manual exercise, all disruptions in the database are classified based on their root cause following their description. Consequently, all disruptions are classified into 15 distinctive types $d \in D$, which occur frequently enough to be able to develop a prediction model for. The distinguished disruption types are visualised by the dark-blue rectangles in **Figure 5.3**. As can be seen in the light-blue rectangles, these disruption types are classified into five main categories *railcar related*, *operations related*, *public related*, *infrastructure related* and *other* disruptions. The category with railcar related disruptions for example consists of *door malfunctioning*, *brake malfunctioning*, *ATC malfunctioning*, *propulsion malfunctioning* and *other* disruption types. Similarly, public related disruptions are categorised as *left unattended item*, *trespassing incident* and *injured/sick/aggressive passenger*.

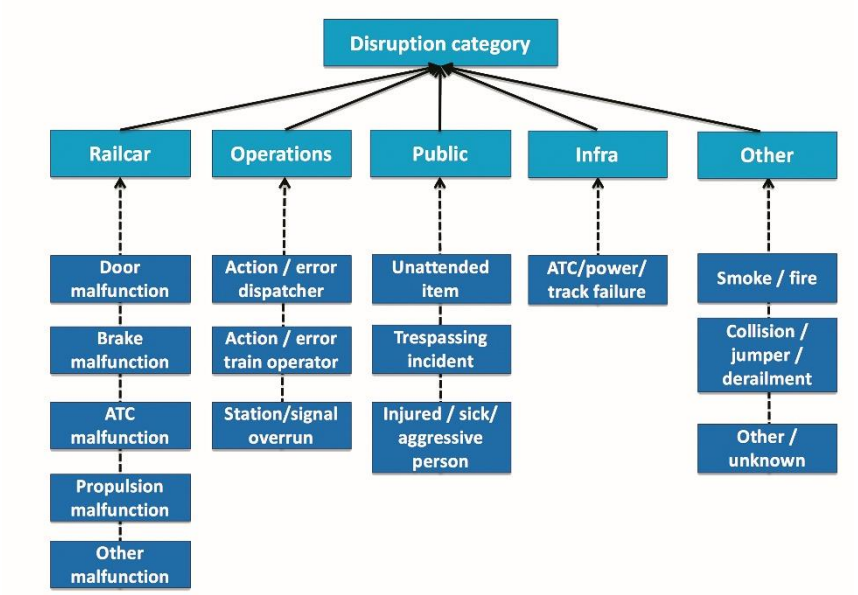


Figure 5.3. Disruption classification

Additional to disruption log data, timetable data about train frequencies and scheduled passenger journey times is provided by WMATA. Besides, individual AFC transactions of each journey made on the metro network in September, November and December 2017, as well as January, February and March 2018 were also available for this study. The Washington metro network is a closed system, where passengers are required to tap in and tap out at gate lines at the stations. For metro-to-metro transfers typically no intermediate tap out and subsequent tap in is required. This means that our case study data directly consists of the journey start time and end time for the metro network, so that no destination or transfer inference was required. Given the availability of 6 months AFC data, our disruption impact prediction model - for which AFC data is required as input for the predictors - is trained based on this 6-month dataset. In this 6-month period, 2,179 disruptions can be distinguished from the dataset after applying the abovementioned data processing steps. This is in contrast with the disruption exposure model, which is trained based on a 12-month disruption log dataset. As we apply a 5-fold cross validation, in each of the five folds 80% of this data is used for model training, whilst the remaining 20% is used for model testing purposes.

5.3.3 Empirical disruption characteristics

Figure 5.4 shows the spatial distribution of disruptions over the Washington metro network. The empirical values show that the weakest stations, being most susceptible to disruptions, can generally be found in the central area of the network where train frequencies and passenger volumes are highest, and at start/terminal stations. The least weak stations are typically intermediate stations (non-terminal and non-transfer stations) at the line branches, often served by one line only. Largo Town Center (red circle in **Figure 5.4**) suffered from most disruptions in the observed 12-month period (160 disruptions). **Figure 5.5** presents the relative frequency of the 15 different disruption types, categorised into the five main categories as set out in **Figure 5.3**. It can be seen that vehicle related disruptions contribute most to the total number of disruptions (45%). In total, vehicle related and passenger related disruptions are responsibe for more than 70% of all disruptions. Infrastructure related disruptions only have a relatively small share in the total number of disruptions. From the individual disruption types d, the most frequently occurring types are injured/sick/aggressive passengers (23%) and vehicle door malfunctioning (15%).

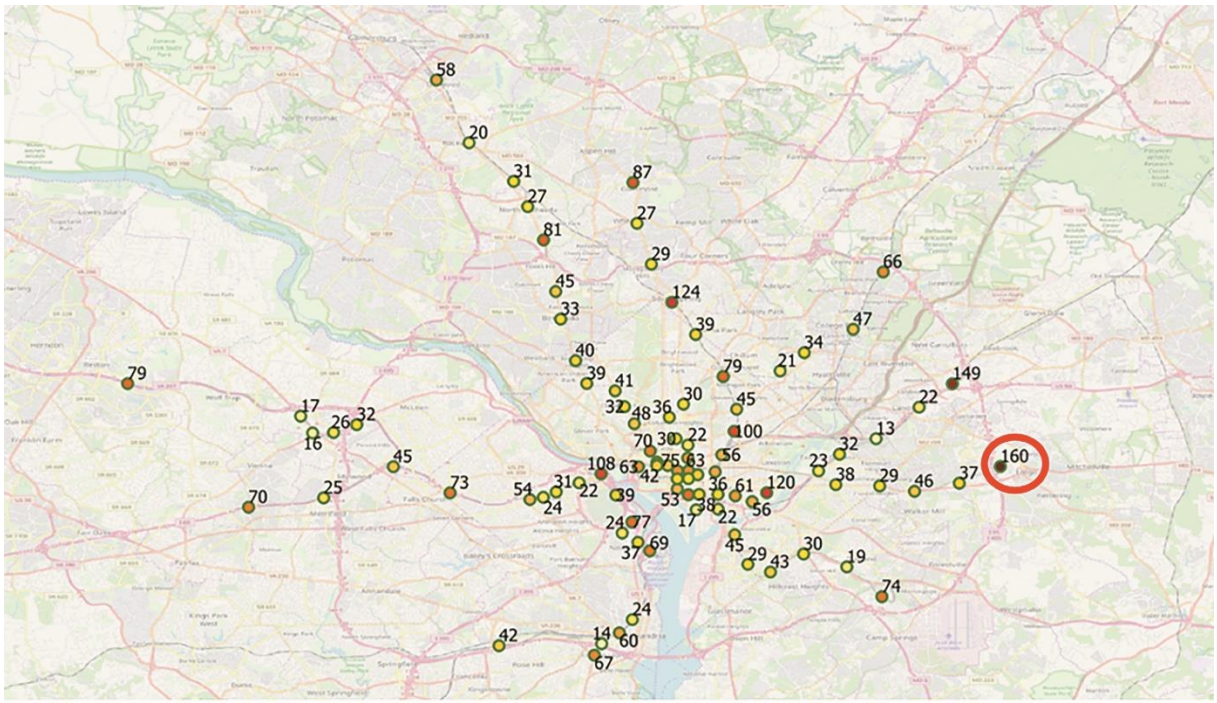


Figure 5.4. Spatial distribution of yearly number of disruptions

5.3.4 Model specification

We use our developed model to predict the probability a certain disruption type occurs at each station during each time period (peak, daytime off-peak and evening) for one full year, applied to the Washington case study network. Based on the number of predictors and one-hot encoding, the final feature matrix for our exposure prediction model consists of (991 distinguished time periods per year * 95 stations) 94,145 samples and 34 columns. The dimension of the target vector is (94,145; 1), respectively (94,145; 16) when dummy coded into the 16 disruption classes (15 disruption types plus no disruption). For the logistic regression model we perform a multiclass regression with a maximum of 200 iterations. *Sag* is used as solver method, as this is fast for relatively large datasets. For the MLP classifier one hidden layer is used. Furthermore, *adam* is used as solver method being fast for large datasets, with the number of iterations being

capped at 200. A logistic sigmoid function is used as activation function for the hidden layer. The number of neurons of the hidden layer is determined by hyperparameter tuning: for all number of neurons between the number of neurons of the input layer and output layer the log loss score (**Eq.3**) is calculated, thereby selecting the number of neurons for the hidden layer which minimises this value. The optimal number of neurons of the hidden layer for the MLP classifier is therefore sought between 16 and 34 neurons. The computation time to predict disruption probabilities for one year for the medium-sized Washington metro network (95 stations) is for both models less than 1 minute on a regular PC.

The final feature matrix of the passenger delay prediction model consists of 2,179 samples (disruptions) and 30 columns (7 one-hot encoded predictors). In the KNN algorithm, we test K-values ranging between 1 and 30 during hyperparameter tuning. For the Random Forest model, we test the number of estimators between 100 and 1,000 with step size 100 for a model which uses the total number of features as maximum feature number, and for a model using the square-root of the number of features as maximum feature number. For the MLP model, we specify a maximum of 10,000 iterations, use *lbfgs* as solver and apply a logistic activation function. During the hyperparameter tuning, we test the number of neurons of the hidden layer between the number of neurons of the output layer (1) and input layer (30). Python is used to execute the machine learning models (Pedregosa et al., 2011). Computation times for these four models to predict disruption impacts range between 1 and 10 minutes on a regular PC.

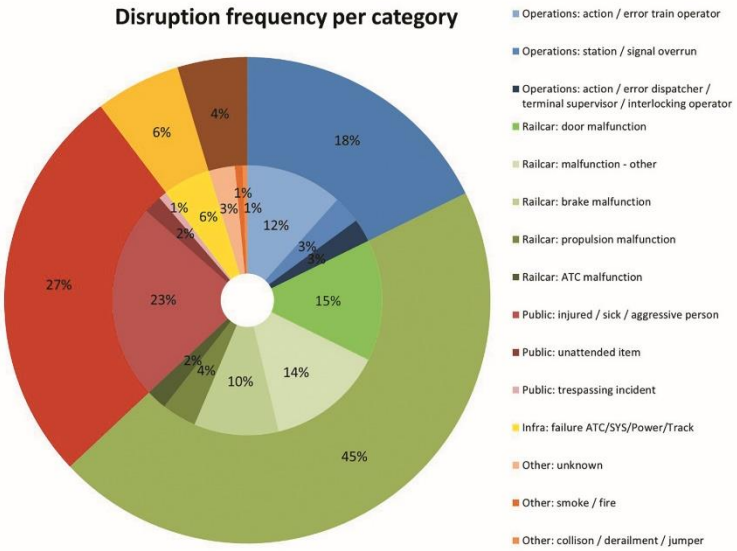


Figure 5.5. Relative frequency of each disruption type

5.4. Results and Discussion

In this section, we first discuss the model estimation and validation results (**Section 5.4.1**). Then, we discuss the prediction results in **Section 5.4.2**, followed by clustering results in **Section 5.4.3**.

5.4.1 Model estimation and validation

Table 5.2 provides an overview of the model specification and performance results for the developed disruption exposure and disruption impact prediction models. Regarding the two tested disrupted exposure prediction models, it can be seen that the log loss score is similar for

the logistic regression and MLP classifier. For this case study we decide to proceed with the results from the MLP classifier, as this model can potentially capture more complex relations between predictors and target. For passenger delay predictions, different machine learning models are compared to a simple linear regression model as baseline. One can conclude that all machine learning models outperform the linear regression model substantially, reducing the RMSE by 95-97%. This indicates the linear regression model is not suitable to capture the complex relation between the predictors and target. When comparing the different machine learning models, especially the Random Forest and KNN regression models result in lower RMSE scores and reasonably high R^2 scores. The Random Forest model, with the total number of features as maximum number of features and using 200 estimators, results is the lowest RMSE score and highest R^2 score of 0.74. We therefore use this model for our final passenger delay predictions.

Table 5.2. Model estimation results

Disruption exposure model	**Model specifications**	**Log loss**	**Improvement**
Logistic regression classifier *(Random 5-fold cross validation)*	Max iterations = 200 Solver = sag	0.268	
Multilayer Perceptron classifier *(Random 5-fold cross validation)*	Max iterations = 200 Solver = adam Activation function = logistic Neurons hidden layer = 30	0.268	0%
Disruption impact model	**Model specifications**	**RMSE (R^2)**	**Improvement**
Simple linear regressor *(Random 5-fold cross validation)*	With intercept	2,536,946 (-551)	
Simple linear regressor *(Random 5-fold cross validation)*	Without intercept	1,575,072 (-292)	37.9%
K-Nearest Neighbours regressor *(Random 5-fold cross validation)*	Number of neighbours = 26	81,950 (0.57)	96.8%
Random Forest regressor *(Random 5-fold cross validation)*	Number of estimators = 200 Max number of features = features	64,722 (0.74)	97.4%
Random Forest regressor *(Random 5-fold cross validation)*	Number of estimators = 900 Max number of features = sqrt(features)	65,533 (0.73)	97.4%
Multilayer Perceptron regressor *(Random 5-fold cross validation)*	Max iterations = 10,000 Solver = lbfgs Activation function = logistic Neurons hidden layer = 25	115,971 (0.18)	95.4%

For model validation purposes of the disruption frequency prediction model, we compare the observed number of disruptions (based on the empirical dataset) with predicted numbers based on the MLP classifier. Predicted values are obtained from the 20% testing sample for each of the five folds in the 5-fold cross validation applied to a 12-month dataset, hence together providing predictions for one complete year. In **Figure 5.6,** a comparison is shown between the predicted and observed disruption frequency for each disruption category, aggregated over stations and time periods for a complete year. There is a high correlation ($>$0.99) between our predicted numbers and observed values. Especially predictions of exposure to door malfunctioning, brake malfunctioning, station overruns and infrastructure related disruptions are highly accurate. Notwithstanding, it can be noted that our prediction model tends to underestimate disruption exposure somewhat. On average the expected number of disruptions is underestimated by 5% using our model, indicating there is still potential for further model improvement.

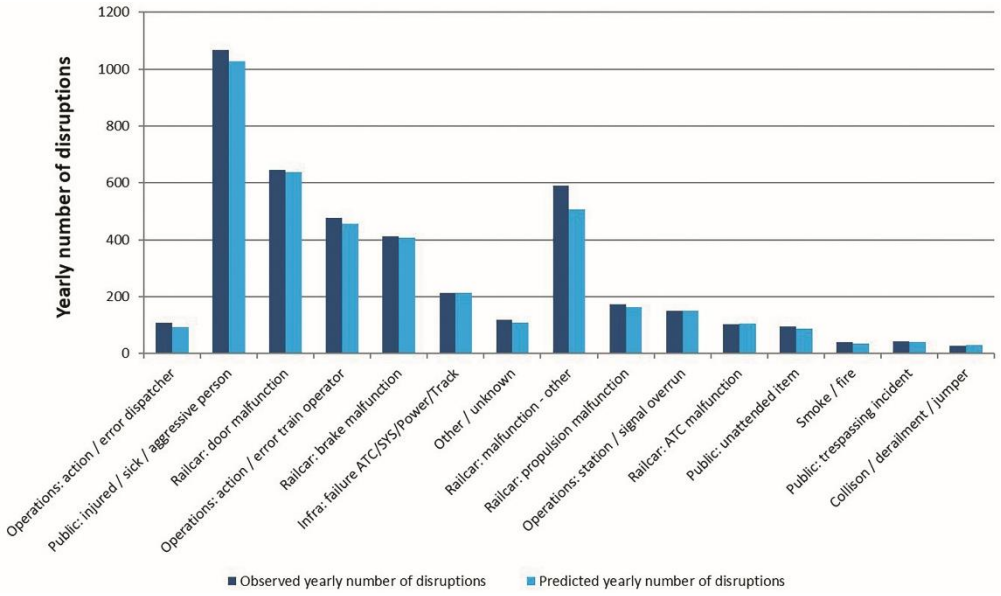


Figure 5.6. Validation MLP classifier

To validate our disruption impact prediction model, we compare the empirical passenger delay hours with predicted passenger delay hours using our Random Forest regression model. Predicted values are based on the 20% testing sample from each of the five folds used in the 5-fold cross validation. This comparison is shown in **Figure 5.7**, where the observed and empirical disruption impacts are aggregated over all stations and all disruptions per month. We can conclude there is a high correlation (>0.99) between predicted and empirical delay hours. Per month, the predicted passenger delay deviates on average 0.6% from the empirical delay hours, with the maximum deviation per month being equal to 5.7% (March). The lowest deviation is observed for November, where the total predicted passenger delay deviates 0.5% from the total observed passenger delay. These results give confidence that our proposed model is able to predict disruptions and passenger delay impacts. Hence, we can apply our models to predict the passenger delay impacts for each station, disruption type and time period. As empirical data about disruption frequency and impact is typically not available for all possible combinations of disruption type, station and time period, our models provide predictions for instances for which no or insufficient historical empirical data is available.

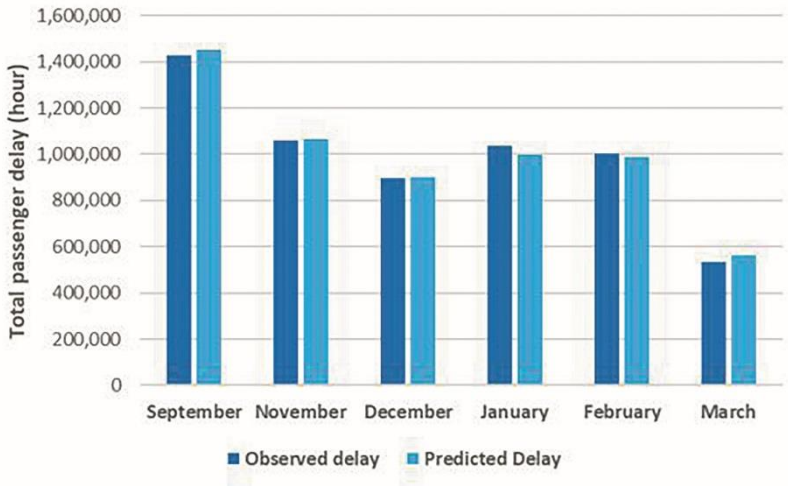


Figure 5.7. Validation Random Forest regressor

5.4.2 Prediction results

Train frequency, day of the week and time of the day are shown to be the three most important predictors for disruption exposure. For the disruption impact prediction model, we find that day of the week (weekday / weekend day) is the most important feature in the Random Forest model. Additionally, the percentage of the affected demand for which no path in the considered PT network remains available in case of a disruption is also an important feature. As station-specific predictors are among the most important predictors in both models, these results indicate the importance of predicting disruptions and their passenger delay impacts for individual stations. When combining the disruption exposure and impact prediction models, these models predict a yearly passenger delay of 5.9 million hours for the total metro network. This value is the sum of the expected station criticality of all metro stations in the network. **Table 5.3** provides an overview of the 10 most and least critical stations with their expected contribution to the yearly passenger delay hours. For the most critical station *Gallery Place* the criticality equals almost 77,000 delay hours, whilst for the least critical station *Stadium-Armory* this value equals almost 43,000 delay hours per year. The 10 most critical stations are all located in the centre of the PT network, where train frequencies and passenger demand are highest. Five of all eight transfer stations in the network are positioned in this top-10 as well. The 10 least critical stations are all located on the eastern branch of the Orange or the Blue/Silver line, and on the western branch of the Silver line (see **Figure 5.2**). None of these stops are start/terminal or transfer locations. Despite the limited number of route alternatives available to passengers when a disruption would occur at a station on one of these branches, the criticality of these stations is relatively low. Stations in the centre of the network are more often exposed to disruptions, and more passengers are affected once a disruption occurs. For stations in the centre section of our case study network, this suggests that the benefit of the availability of multiple route alternatives does not outweigh the costs, namely the more frequent disruption exposure and the larger passenger demand affected by these disruptions.

Table 5.3. Station criticality ranking

Rank	Station (lines)	Criticality (pass-hours per year)	Rank	Station (lines)	Criticality (pass-hours per year)
1	Gallery Place (R)	76,594	86	Landover (O)	52,188
2	Metro Center (R)	74,384	87	Deanwood (O)	49,569
3	Gallery Place (YG)	72,653	88	Benning Road (SB)	49,438
4	Union Station (R)	72,439	89	Greensboro (S)	48,952
5	Metro Center (SOB)	71,926	90	Potomac Ave (SOB)	48,300
6	L'Enfant Plaza (YG)	70,314	91	Spring Hill (S)	48,095
7	Judiciary Square (R)	70,284	92	Cheverly (O)	45,696
8	Farragut West (SOB)	69,750	93	Capitol Heights (SB)	45,552
9	Columbia Heights (YG)	69,683	94	Morgan Boulevard (SB)	45,224
10	NoMa-Gallaudet U (R)	69,426	95	Stadium-Armory (SOB)	42,651

5.4.3 Clustering results

The station clustering results based on predicted criticality is shown in the dendrogram in **Figure 5.8**. The cophenetic correlation coefficient equals 0.70, which can be considered reasonable. From the dendrogram can be seen that the 95 metro stations of the Washington metro network are grouped into five different clusters. **Figure 5.9** shows for each of these clusters the expected yearly number of disruptions per station (left), the average (unweighted) disruption impact (centre), and expected yearly passenger delay per station (right). For stations in cluster 2 both the disruption exposure and impact are highest, resulting in the highest criticality. For this cluster, particular the average disruption impact is high compared to stations

from other clusters. Stations from cluster 3 also have a relatively high disruption exposure, but the impact is smaller than for cluster 2. As a result, expected criticality for stations of cluster 3 is also lower than for cluster 2. The expected disruption impacts for stations from clusters 1, 4 and 5 are similar: these clusters differ primarily in their disruption exposure. Due to the relatively high disruption exposure in cluster 1 compared to clusters 4 and 5, the criticality of cluster 1 is also highest from these three clusters. Stations from cluster 4 are characterised by the lowest weakness, therefore resulting in lowest station criticality from all clusters.

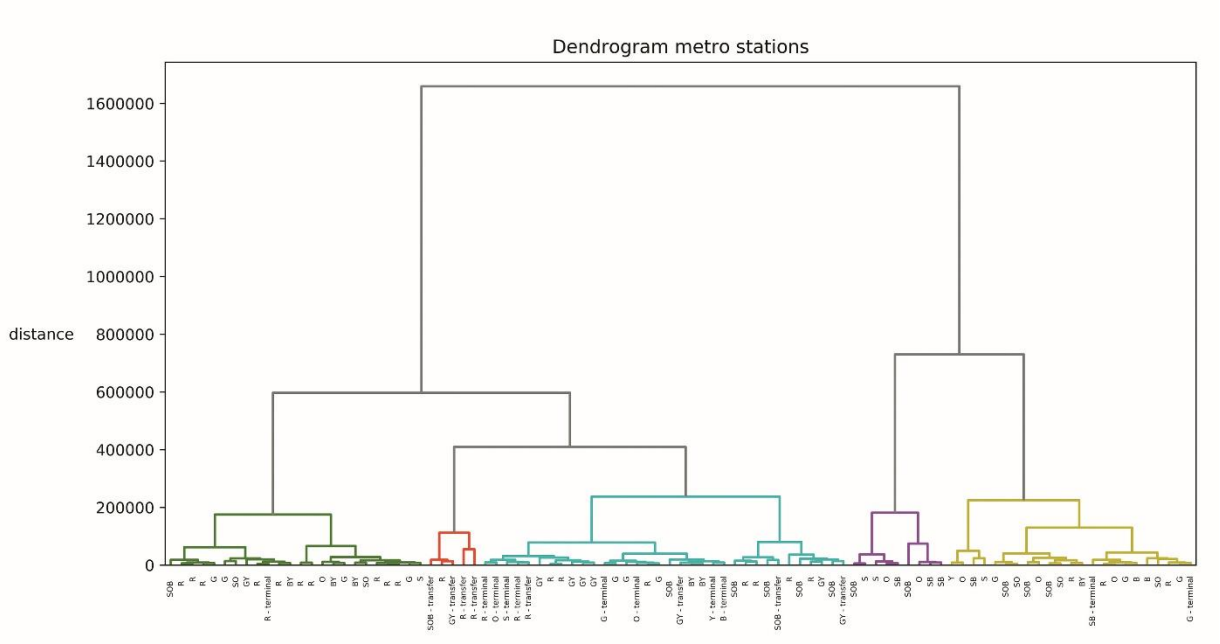


Figure 5.8. Dendrogram with resulting clustering of metro stations

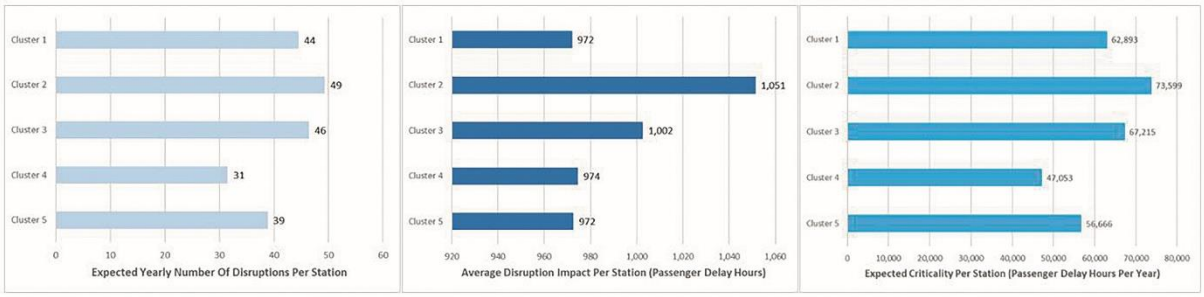


Figure 5.9. Predicted average exposure (left), impact (centre) and criticality (right) per station in cluster

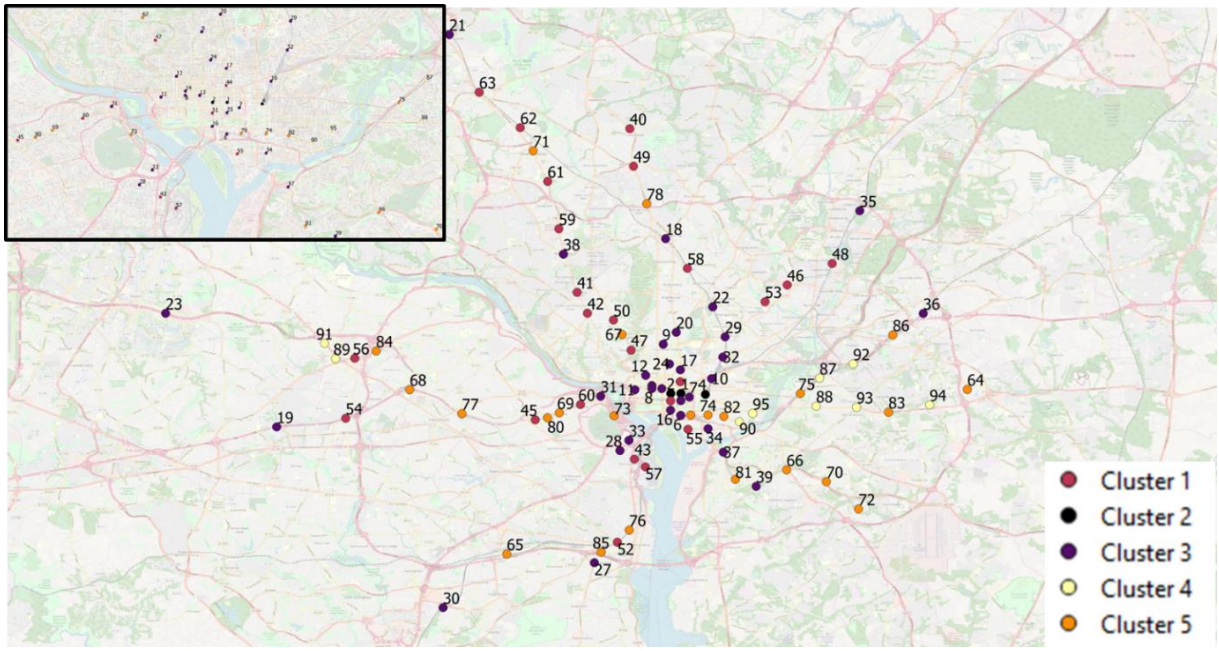


Figure 5.10. Station criticality ranking and clustering
(Numbers refer to station criticality ranking 1-95; inset upper-left zooms into the centre part of the network)

In **Figure 5.10,** we visualise the station ranking and clustering spatially. Each number in the figure shows the ranking of each of the 95 stations in terms of station criticality (see also **Table 5.3**), whereas the colour indicates the cluster to which each station belongs. Cluster 2 (**Figure 5.8**, red) consists of five stations: four transfer stations and the main train station Union Station. These stations are most critical, as exposure and impact are highest. The high passenger delays characterising this cluster can be explained by the relatively high number of passenger related and railcar related disruptions, due to the high train frequency and passenger volumes in this central part of the metro network. These high passenger volumes also result in the highest disruption impacts. Cluster 3 (**Figure 5.8**, blue) is the largest cluster, containing 34 stations. The criticality of these stations is second-highest, after the stations from cluster 2. This cluster consists of all remaining transfer stations and the majority of the start/terminal stations. Besides, most other stations located in the centre area of the PT network are part of this cluster. All these stations are relatively heavily exposed to disruptions. For stations in the centre part of the network, this is mainly caused by the high train frequencies and passenger volumes. For the start/terminal stations, an explanation is that several disruptions often arise at the first station of the line: a railcar malfunctioning when testing the train, a late or sick driver not arriving on time, or a late movement of the train from the yard to the first station. As confirmed from the empirical analysis (**Figure 5.4**), start/terminal stations are more frequently exposed to disruptions than their surrounding stations. As passenger demand at the stations of cluster 3 is lower than for the five busy stations of cluster 2, the expected disruption impact is lower as well. Cluster 4 (**Figure 5.8**, purple) contains the 9 least critical stations. As shown in **Table 5.3**, these stations are located at the end of the western branch of the Silver line, and at the end of the eastern branches of the Orange and Blue/Silver lines. These stations have the lowest disruption exposure from all stations for all disruption types, particularly in relation to passenger related disruptions. One can argue that the relatively low headways combined with relatively low passenger volumes at the end of these lines result in lower passenger impacts, despite the number of available route alternatives in the network also being smaller here. Location-specific characteristics might also play a role here. The stations of cluster 1 (**Figure 5.8**, green) and cluster 5 (**Figure 5.8**, gold) are mainly located between the busiest centre section

of the network and the outer branches of the lines. The 24 stations of cluster 1 are primarily stations on the northern part of the network (Red line, trunk section of the Yellow/Green lines, northern branch of the Green line). The 23 stations of cluster 5 are mostly located on the southern part of the network (trunk section of the Silver/Orange/Blue lines, trunk section of Blue/Yellow lines, southern branch of the Green line). In terms of exposure, these clusters can be positioned between clusters 2/3 on the one hand, and cluster 4 on the other hand. Especially stations from cluster 1 are relatively often exposed to disruptions: this might be caused by the relatively high frequency on the Red line, or by differences in operating stock types. Disruptions at stations of clusters 1 and 5 affect more passengers compared to cluster 4, but some stations offer more route alternatives to these affected passengers. These positive and negative effects seem to cancel out each other on average, as the average disruption impact is similar to stations from cluster 4.

5.5 Conclusions

In this study we propose a generic approach to predict how often different disruption types occur at different stations of a PT network, and to predict the impact related to these disruptions as measured in terms of passenger delays. The contribution of our research lies in the development of supervised learning models to predict disruptions and their passenger impacts for each individual station, disruption type and time period, as sufficient empirical disruption observations will not always be available for each location and time period. Besides, our models can predict disruption impacts for all stations and time periods for a medium-sized PT network (consisting of 95 metro stations) within 10 minutes. Hence, our method provides an alternative for existing, computationally more expensive methods to predict passenger delays for a complete PT network. Applied to the Washington metro network, our models predict a yearly passenger delay of 5.9 million hours for the total metro network. Five different types of station are distinguished by clustering stations according to their expected criticality. Stations with high train frequencies and high passenger volumes located at central trunk sections of the network show to be most critical, together with start / terminal and transfer stations. Intermediate stations located on branches of a line are least critical. The lower train frequencies and passenger volumes result in lower disruption exposure and impact, despite less route alternatives typically being available for these passengers when a disruption occurs.

Our study results provide PT authorities and service providers insights into the frequency, location and passenger impact of different disruptions. It provides an overview of the stations which contribute most to the vulnerability of the total PT network. Categorising stations based on their disruption characteristics shows the different station types which can be distinguished based on their contribution to network vulnerability. This supports PT agencies in prioritising what type of disruptions at which location to focus on, to potentially achieve the largest improvements in network robustness. Ranking all stations according to their criticality directly supports decision-makers to target robustness measures at these stations which need it most. The explicit distinction between disruption exposure and impact helps determining what type of measure would be most suitable for each (type of) station. Our method can also be used to quantify the robustness benefits of new infrastructure, such as a new rail link. The model trained for the current PT network can be used to predict the new station criticality in the event of a network adjustment, by updating network-related predictors. This results in a fast and complete quantification of robustness benefits, which can be incorporated in appraisal studies.

We formulate four recommendations for future research. First, we recommend further testing of our model sensitivity in relation to missing disruption duration information. We recommend to apply this method to other case study networks, where the disruption duration - possibly

including recovery time - is provided in the disruption log data, so that the sensitivity of the model performance can be investigated. This might enable a further improvement of the R^2 value of the disruption impact model, which can currently be considered reasonably high. Second, application of our method to a link based or joint node and link based vulnerability analysis is recommended. As disruptions in the dataset provided to us were allocated to stations, we applied a node based vulnerability analysis. Our methodology is however directly applicable for link based analyses using the same predictors. Testing the sensitivity of our model outcomes to this is therefore recommended. Third, whilst our model results in a reasonably accurate prediction of disruption impacts, our model slightly underestimates disruption frequency predictions by 5% on average. Future research is therefore recommended to further improve the accuracy of this prediction model. Fourth, we also recommend incorporating the availability of other modes in the assessment of the number of paths remaining available for passengers, as well as for the indication of passengers' travel time extension, as used as predictors in our disruption impact prediction model. As our model only considers metro lines, robustness resulting from the availability of alternative modes of transport in the network is potentially somewhat underestimated.

6. Identification and Quantification of Link Vulnerability in Multi-level Public Transport Networks: A Passenger Perspective

In the previous **Chapter 5**, we developed a methodology to predict disruption frequencies and impacts for different stops of an urban public transport network. In this chapter, we expand our focus from *urban public transport networks* towards *multi-level public transport networks*, thereby considering the (inter)regional train network, light rail / metro and urban tram and bus network. The scientific contribution of this chapter is the development of a methodology to identify links of a multi-level public transport network, which contribute most to public transport network vulnerability. In this research, we propose an improved method to pre-select the links which are expected to contribute most to vulnerability, based on both expected disruption frequencies and expected disruption impacts for passengers. For these selected links, we predict disruption impacts given the integrated multi-level network available for passengers. Hence, we incorporate how different network levels can potentially mitigate impacts of disruptions occurring on the urban public transport network. Incident log data from different public transport modes, together with outputs from a public transport assignment model, are used as input for this methodology. **Chapter 5** and **Chapter 6** both result in improved methods for performing a public transport vulnerability analysis by incorporating both disruption exposure and impacts for urban networks in a single-level and multi-level perspective, respectively, and thereby provide an answer to Research Question 2 (see **Section 1.3**).

This chapter is based on an edited version of the following article:

Yap, M.D., Van Oort, N., Van Nes, R., Van Arem, B. (2018). Identification and quantification of link vulnerability in multi-level public transport networks: a passenger perspective. *Transportation, 45*, 1161-1180.
© The Author(s) 2018

6.1 Introduction

The operation of public transport services without disturbances is considered a key quality aspect of public transport (PT) by passengers (Golob et al., 1972; Van Oort, 2016). Disturbances can result in longer travel times, more transfers and more crowded vehicles, and thus increase both the nominal and perceived passenger travel costs. This means that during PT disturbances fewer locations and activities can be reached by public transport within a certain travel time, resulting in a decreased accessibility. This relation between network vulnerability and accessibility is among others addressed by Chen et al. (2007), Liao and Van Wee (2017) and Taylor (2017). Therefore, reducing the passenger impact of disturbances is important in order to limit the negative accessibility effects. To gain more insight in these negative accessibility effects of disturbances, it is important to get insight in the frequency with which disturbances occur in public transport, and the impact these disturbances have on passengers. This topic is generally addressed using the concepts of reliability and vulnerability. In scientific literature, different definitions are used to distinguish between these concepts (for example Ziha, 2000; Holmgren, 2007; Van Nes et al., 2007; Tahmasseby, 2009; Korteweg and Rienstra, 2010; Savelberg and Bakker, 2010; Snelder, 2010; Immers et al., 2011; Parbo et al., 2013; Dewilde et al., 2014). An extensive review of definitions and indicators for reliability and vulnerability can be found in Nicholson et al. (2003) and more recently in Oliveira et al. (2016). In this chapter we apply the distinction between reliability and vulnerability as used by Oliveira et al. (2016). Reliability is hereby related to the network performance in relation to recurrent, daily, stochastic fluctuations in supply and demand. Vulnerability, on the other hand, focuses on the network performance related to non-recurrent, infrequent, large events, leading to a partial or full unavailability of one or multiple links of the network. Robustness is inversely related to vulnerability: a network with 0% vulnerability yields 100% robustness, and the other way around (Tahmasseby, 2009; Snelder, 2010).

Despite the importance attributed by passengers to robust public transport, the full passenger impact of public transport network vulnerability is not considered in science and practice yet. Reliability is extensively considered for single-level PT networks: networks on one functional level (e.g. the regional, agglomeration or urban level) usually operated by a single PT operator. Research on improvements of reliability of single-level PT networks is among others conducted by Hollander (2006) and Van Oort and Van Nes (2009) (**Figure 6.1**, upper left quadrant). Examples of measures to improve reliability of single-level PT networks on a strategic, tactical and operational level can be found in Vromans et al. (2006), Delgado et al. (2009), Furth and Muller (2009), Corman et al. (2010), Van Oort et al. (2010), Van Oort and Van Nes (2010) and Xuan et al. (2011). Besides considering reliability of single-level PT networks, research is also conducted to reliability of multi-level PT networks where interactions between different network levels are considered (for example Rietveld et al., 2001; Lee et al., 2014) (**Figure 6.1**, upper right quadrant). For example, such multi-level approach of reliability allows for the incorporation of the consequences of a delay on the train network for transfers to a lower-level tram connection. In addition, studies can be found related to vulnerability and robustness of road networks (e.g. Jenelius et al., 2006) and public transport networks (e.g. Derrible and Kennedy, 2010; Cats and Jenelius, 2014; Cats and Jenelius, 2015). There are several examples of studies to robustness of single-level PT networks, for example Goverde (2005), Kroon et al. (2008), Tahmasseby et al. (2008), Cicerone et al. (2009), Fischetti et al. (2009), Schöbel and Kratz (2009) and Corman et al. (2014) (**Figure 6.1**, lower left quadrant). These studies analyse robustness separately for each PT network on a certain functional network level (single-level perspective), or for a PT network operated by a specific operator (single-operator perspective). However, the interaction between different PT network levels in case of non-recurrent

disturbances is not explicitly considered in these vulnerability studies.

This entails that it is not considered how a certain network level, or PT network operated by operator X_1, can function as backup in case a disturbance occurs on another network level operated by operator X_2. However, when aiming to quantify the full passenger impacts of non-recurrent disturbances, it is important to consider the integrated multi-level PT network, with PT services on other network levels which remain available for passengers. This means that the PT networks on all functional network levels, operated by different operators, should be considered in an integrated approach. In their total door-to-door trip, passengers often use PT services on different network levels, often operated by different PT operators as well. For example, in the period 2006-2009 on average 89.8% of the trips having train as main mode in the Netherlands can be considered multimodal (Van Nes et al., 2014). In case each PT operator only optimises the part of the network it operates, it is likely that different optimised subnetworks lead to a suboptimal total network from a passenger perspective, since interactions between network levels are ignored. In case of large disturbances this leads to suboptimal rescheduling for passenger because network levels of other operators, which might offer potential to function as backup, are not considered. Possible powerful measures on network level X_1, which can reduce the passenger impact of a large disturbance occurring on network level X_2, might not be considered either. This means there is no full and no realistic quantification of passenger impacts of non-recurrent disturbances, since passengers are able to consider the total available multi-level PT network in case of disturbances. We therefore conclude that currently not the full passenger impacts are incorporated in the analysis and quantification of PT network vulnerability.

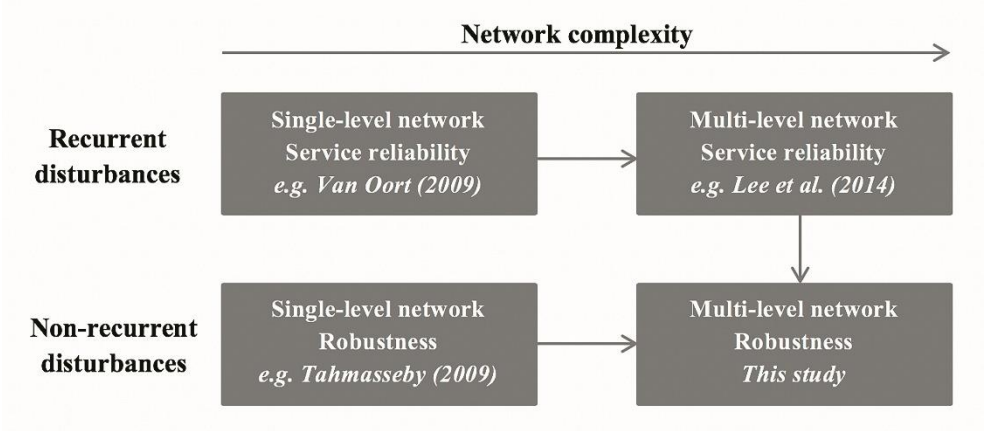


Figure 6.1. Relevance of study focusing on robustness of multi-level public transport networks, including examples of references in other quadrants

Our study explicitly considers public transport network vulnerability from a multi-level perspective (**Figure 6.1**, lower right quadrant). We define a multi-level PT network as an integrated PT network where different network levels – the (inter)regional (train) level, agglomeration (metro / light rail) level and urban (tram) level - are considered simultaneously. Contrary to studies with a multi-modal perspective, by adopting a multi-level public transport perspective we solely consider the public transport network and no other networks such as car and bicycle. We develop a methodology to identify the most vulnerable links in such multi-level public transport network. Based on this method we are able to quantify link vulnerability for these identified links, given the total multi-level PT network available. This allows for the quantification of societal costs of link vulnerability for passengers in a more realistic way, since passengers also consider the multi-level network when looking for route alternatives in case of a disturbance. We adopt a passenger perspective, by aiming to incorporate the full and realistic

passenger impact of disturbances when identifying and quantifying link vulnerability. This methodology is applied to a case study in the Randstad Zuidvleugel area in the Netherlands. In this case study, we compare link vulnerability between different network levels. We use the integrated multi-level PT network to quantify link vulnerability and to quantify the robustness benefits of proposed measures.

The added value of this study is the development of a methodology to identify the most vulnerable links in multi-level public transport networks, thereby incorporating both exposure to disturbances and the impact of disturbances, and to quantify link vulnerability for these identified links. The structure of this chapter is as follows. **Section 6.2** discusses the methodology developed to identify and quantify link vulnerability in multi-level PT networks. In **Section 6.3**, we show the results after applying this methodology to a case study. We finish in **Section 6.4** by formulating conclusions and recommendations for further research.

6.2 Methodology

6.2.1 Link vulnerability

We assume a multi-level PT network represented by a digraph $G(V_n, E_n)$ with nodes $v_n \in V_n$ and directed links $e_n \in E_n$ on PT network level n. Let S represent the total set of public transport stops $s \in S \subseteq V$. The number of stops and number of links are denoted by $|S|$ and $|E|$, respectively. The set of public transport lines is denoted by L. Each public transport line $l \in L$ is defined as ordered sequence of stops $S_l = (s_{l,1}, s_{l,2}..s_{l,|l|})$. The total node set V consists of all stops S and all junctions (intersections, switches etc.) in the network.

Traditionally, link vulnerability c_{e_n} for public transport networks is only assessed based on the impact a non-recurrent disturbance δ has on the network performance. This impact is often expressed as the difference in societal welfare Δw_δ between the undisturbed scenario w_{δ_0} and the scenario with non-recurrent disturbance w_δ occurring on link $e_n \in E_n$. This means that exposure to disturbances is not considered explicitly when determining link vulnerability. This can be explained by the limited historic data available regarding the frequency with which different disturbance types δ_n occur on each PT network level n and their related duration τ_{δ_n}. Instead, a conditional vulnerability is typically applied which calculates the impact of a disturbance given the fact that a certain disturbance has occurred, as expressed by **Eq.1**.

$$c_{e_n} = \Delta w_{e,\delta_n,\tau} | \delta_n \quad (1)$$

When applying **Eq.1**, links where disturbances have the most negative impact Δw_δ on passengers' travel time, costs and comfort are listed as most vulnerable, even if the frequency with which these disturbances occur would be very low. However, from a passenger perspective link vulnerability depends on both the extent to which link e_n is exposed to non-recurrent disturbances δ, and the impact of these disturbances on passengers given the total PT network N available. This is especially relevant when considering multi-level PT networks, where different modes and vehicle types are operating on the different network levels. The frequency and duration of disturbances can differ intrinsically on links of different network levels. These differences can be attributed to different vehicle types (e.g. train versus tram), different infrastructure (e.g. different types of switches, use of signalling systems), different interactions with other traffic (dedicated infrastructure vs. mixed traffic) and different exposure to external events (e.g. operation in tunnels or at grade level). This means that neglecting differences in exposure to disturbances in a link vulnerability analysis can bias the identification of the most vulnerable links in the multi-level PT network based on impact only, when exposure differs

between network levels. Therefore, we propose the incorporation of exposure to disturbances explicitly in link vulnerability identification and quantification: links where the product of exposure to disturbances and the impact of these disturbances is highest are identified as most vulnerable, as expressed by **Eq.2**.

$$c_{e_n} = \sum_{\delta_n} E\left(f_{e,\delta_n}\right) * E\left(\tau_{e,\delta_n}\right) * \Delta w_{e,\delta_n,\tau} \quad (2)$$

The exposure of a link e_n to large disturbances is the product of the frequency f_{e,δ_n} with which different disturbance types δ occur on that link and the duration τ_{e,δ_n} of each disturbance. Both the frequency with which disturbance types occur and the duration of each disturbance are probabilistic variables, which are independent from each other for each disturbance type δ. This results in the multiplication of the expected number of disturbances $E\left(f_{e,\delta_n}\right)$ occurring within a certain time window and the expected duration of each disturbance $E\left(\tau_{e,\delta_n}\right)$, with both f_{e,δ_n} and τ_{e,δ_n} being random variables. $\Delta w_{e,\delta_n,\tau}$ represents the difference in total monetised societal costs for all passengers travelling over all OD pairs affected by that specific disturbance δ_n between the specific disruption scenario and the undisturbed situation.

To be able to incorporate exposure to disturbances explicitly, we use a unique dataset in this study which contains realisation data about the frequency and duration of different types of disturbances δ_n on different PT network levels n (national / interregional / regional / agglomeration / urban level) and for different PT modes (train / metro / light rail / tram), operated by different PT operators in the Netherlands. Historic log-data for train network disturbances is gathered from the train operator Dutch Railways (NS) for the full period of 2.5 years between January 2011 and August 2013. For the urban and agglomeration network level (metro, light rail and tram), realisation data is used from a period of 18 weeks between June and October 2013 from different PT operators. This enables the incorporation of exposure to disturbances on each PT network level explicitly when considering vulnerability. For both the frequency and duration of each disturbance type δ_n, it is statistically tested whether the empirical data fits a theoretical probability density function. By distribution fitting, parameter values f_{δ_n} and τ_{δ_n} are estimated for the probability density functions for each disturbance δ_n. Since particularly the frequency f_{δ_n} with which some disturbances occur can be influenced by the weather, we test whether significant seasonal differences exist in average frequency of each disturbance δ_n. In that case separate parameters are estimated for different seasons. In Cats et al. (2016b) an extensive description of this data analysis can be found, including the distribution of different disturbance types on different network levels.

All disturbances δ_n are categorised in two classes based on the impact a disturbance has on infrastructure availability. Some disturbances usually lead to a partial unavailability of a link (e.g. a train breakdown leading to link blockage in one direction), whereas other disturbances lead to a link being completely unavailable (such as a train-car collision on a level crossing). In the Netherlands, PT operators apply different rescheduling measures depending whether there is a partial or full link blockage. This also means that the passenger impacts $\Delta w_{e,\delta_n,\tau}$ are different in these two scenarios, depending whether PT services on a certain link are partially or completely cancelled. Therefore, it is necessary to distinguish different disruption scenarios S for disturbances leading to partial versus full link unavailability, for which the link vulnerability $c_{e_n}^S$ can be calculated.

6.2.2 Identification of link vulnerability

When aiming to improve PT network robustness, it is important to identify which links are most vulnerable in the multi-level PT network. In scientific literature, two different approaches are applied to identify the most vulnerable network links (Knoop et al., 2012). The first approach uses full computation methods. In these methods, disturbances are simulated on each link $e \in E$ of the network separately to evaluate its vulnerability relative to other links. These methods have a clear advantage in terms of their completeness, since vulnerability of the complete link set E can be assessed and compared. The largest disadvantage is that these approaches can be very time consuming. In the second approach, criteria are specified to pre-select a smaller number of vulnerable links in a network. Disturbances are only simulated on these selected links in a second step. This approach overcomes the disadvantage of very long computation times of full computation methods. However, since pre-selection criteria are used to identify a shortlist of vulnerable links, there is no guarantee that the most vulnerable links are remaining after the pre-selection phase.

Since real-world, complex multi-level PT networks as we consider in this study are usually represented by a large number of links, computation times become unacceptably long when all relevant disruption scenarios would be simulated on each link separately. Therefore, it is necessary to apply a method to pre-select most vulnerable links. In scientific literature various criteria can be found to pre-select the most vulnerable links of road networks. However, there is limited literature where pre-selection criteria are specified for public transport networks. Only some examples can be found, e.g. Cats and Jenelius (2014) using a dynamic vulnerability analysis, and Bell (2003) and Zhang et al. (2010) using game theory. All these methods do not address exposure to disturbances explicitly, which makes them not suitable to apply for multi-level PT networks. Hence, we develop a new methodology to identify the most vulnerable links in multi-level PT networks, which explicitly incorporates exposure to disturbances as well. Input for our new developed methodology is derived from existing methodologies developed to identify vulnerable links for road networks (Jenelius et al., 2006; Li, 2008; Tampère et al., 2007; Immers et al., 2011; Knoop et al., 2012) and is adjusted based on multi-level PT network characteristics.

When analysing the suitability of road network pre-selection criteria for identification of vulnerable links in multi-level PT networks, we can identify four intrinsic differences. First, criteria which consider the probability of disturbances on road networks calculate this probability for each link $e \in E$. For PT networks, we propose to calculate exposure to disturbances per link *segment* y. PT operators usually apply standard rescheduling procedures in case of disturbances: for each location in the network a disruption scenario specifies how PT services are adjusted in case of partial or complete track unavailability. Because rescheduling possibilities for PT services depend on the availability of switches, turning loops, station capacity etc., these procedures are exactly equal for adjacent links with no switches or other rescheduling possibilities in between them. Such procedures are therefore designed per link segment - a set of adjacent links $y = \{e_1, .., e_m\}$ taken together by the PT operator for which one standard disruption scenario applies - instead of per link.

Second, pre-selection criteria for road networks are usually a proxy for disturbance impact, whereas the probability of a disturbance is not or only implicitly or roughly considered. Criteria which only consider the incident impact, implicitly assume an equal probability of disturbances on each link. When incident probabilities are considered for road networks, often one generic predictor (e.g. link length) is used to distinguish incident probabilities for different links. However, Yap (2014) and Yap et al. (2015) show that for identified disturbances δ_n on a multi-level PT network, different predictors (such as link segment length, vehicle-kilometres

per link segment) should be used to distinguish between incident probabilities for different link segments. In addition, it is clear that probabilities of a certain disturbance type δ are different for different PT network levels, given the different characteristics of these network levels. This shows it is not sufficient to assume an equal probability of disturbances for all links in a multi-level PT network, or to use one generic predictor. Pre-selection criteria for multi-level PT networks should therefore be a proxy for both the probability of disturbances (using different predictors for different disturbance types δ and different parameter values f_{δ_n} for different network levels) and the passenger impact of a disturbance.

Third, in road networks some ratio between traffic volume and capacity (such as the Incident Impact Factor or V/C ratio) is often used as proxy for the impact of a disturbance. Since the real incident impact on travel time, costs and comfort $\Delta w_{e,\delta_n,\tau}$ can usually only be quantified after simulation of disturbances, a proxy for this impact has to be used in the identification phase. In PT networks the relation between volume and capacity is less relevant when approximating the impact of a disturbance, as limited congestion occurs between PT vehicles on PT networks - even in case of disturbances - compared to congestion between vehicles on road networks. The impact of a disturbance in PT networks is mainly related to the absolute number of passengers affected, instead of the V/C ratio of PT vehicles on a certain link. For example, a single-track local train line can have a very high V/C ratio if there are limited possibilities for trains to pass each other, whereas a very busy four-track train line might have a lower V/C ratio. In such case, the passenger flow on affected links is a better proxy to represent the impact of an incident than the V/C ratio. This in fact equals the passenger betweenness centrality measure as proposed by Cats and Jenelius (2014).

Fourth, some pre-selection criteria for road networks only focus on the impact of a disturbance on the considered link e_i itself, whereas other criteria also consider spillback effects to adjacent links $e_{j\neq i}$. For road networks, it is clear that disturbances can have spillback effects to other links. However, in PT networks spillback effects of disturbances also occur, though differently compared to road networks. Given the limited congestion between PT vehicles, there are no or only limited direct spillback effects to PT vehicles on adjacent link segments in case of disturbances. However, PT services on other link segments $y_{j\neq i}$ in the network can certainly be affected by a disturbance on link segment y_i. As explained, PT operators apply standard disruption procedures. In these procedures PT lines can be divided into two parts, shortened, rerouted via an alternative track or cancelled. For example, assume a PT line $l\{s_{l1},..,s_{l5}\}$ which is cancelled between s_{l3} and s_{l5} after the occurrence of a disturbance on link segment $y_{s_3-s_4}$, because there is insufficient capacity for short-turning near s_{l4}. This means that not only passengers on link segment $y_{s_3-s_4}$ are affected, but also passengers only travelling over link segment $y_{s_4-s_5}$. This illustrates that PT services on a certain link segment y_i can be affected because of a first-order effect - a disturbance occurring on that link segment y_i itself - and because of a second-order effect. This second-order effect is relevant in case a disturbance occurs on another link segment $y_{j\neq i}$, leading to disruption measures taken by the PT operator or infrastructure manager which also affect PT services on the considered link segment y_i. Except during the transition phase between regular PT operations and the disruption scenario, this spillback effect for PT networks can be considered more static compared to the dynamic spillback effects occurring on road networks.

Based on the differences between PT and road network characteristics, we develop an adjusted methodology to identify the most vulnerable links in multi-level PT networks (**Figure 6.2**).

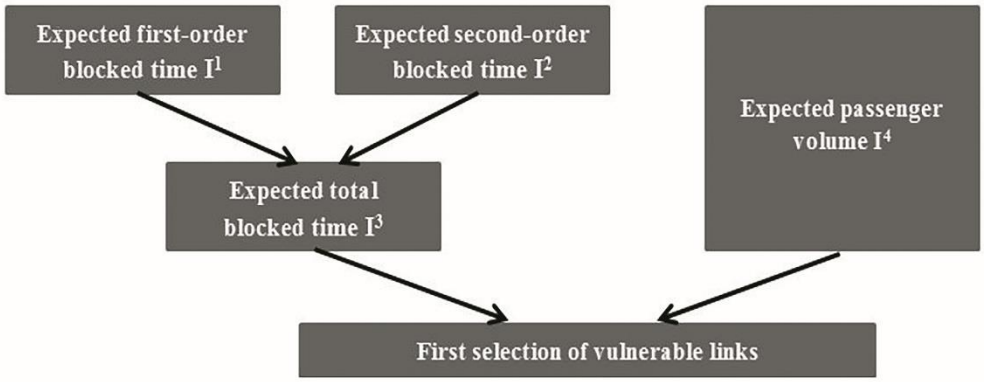


Figure 6.2. Methodology to identify vulnerable links in multi-level PT networks

In this methodology, pre-selection criteria I^1 to I^4 are specified. $I^1_{y_{i_n}}$ (**Eq.3**) reflects the first-order exposure: the expected time that a certain link segment y_{i_n} is exposed to reduced/no PT services because of non-recurrent disturbances occurring on that link segment y_{i_n} itself. This equals the product of the average frequency ${f^*}_{\delta_n,pr,w}$ with which disturbance type δ_n occurs per time period on network level n in season w and the average duration ${\tau^*}_{\delta_n,pr,w}$ of each disturbance δ_n in season $w \in W$. For each δ_n a predictor $pr \in PR$ is determined which enables the transformation of the average frequency with which δ_n occurs per time period on the whole considered network level n (which is known from the database with disturbances we used) to the average frequency per link segment y_{i_n}. This transformation is based on the ratio between the value of this predictor $x_{pr,y_{i_n}}$ on link segment y_i and the value x_{pr,Y_n} summed over all link segments of the total network level. For this criterion only the average frequency f^* and average duration τ^* are used. This prevents the need for Monte Carlo simulation to draw values from the identified distribution functions, resulting in reduced computation times and reduced complexity in this link vulnerability identification phase.

$I^2_{y_{i_n}}$ (**Eq.4**) reflects the second-order exposure effect: the expected time that a certain link segment y_{i_n} is exposed to reduced/no PT services because of non-recurrent disturbances occurring on any other link segment $y_{j_n \neq i_n}$, resulting in measures taken by PT operators which also affect PT operations on the considered link segment y_{i_n}. In this study we used the rescheduling procedures as taken by PT operators in the Netherlands in reality in case of disturbances and assumed these procedures as a given, in order to determine which other link segments $\breve{y}_{j_n \neq i_n}$ affect PT services on link segment y_{i_n} in case of disturbances. $I^3_{y_{i_n}}$ sums the first-order and second-order effects, thus expressing the total expected time a link segment is exposed to non-recurrent disturbances (**Eq.5**).

$$I^1_{y_{i_n}} = \sum_{\delta_n} \sum_{W} {f^*}_{\delta_n,pr,w} * \frac{x_{pr,y_{i_n}}}{x_{pr,Y_n}} * {\tau^*}_{\delta_n,pr,w} \qquad \forall \; y \in Y \qquad (3)$$

$$I^2_{y_{i_n}} = \sum_{\breve{Y}} \sum_{\delta_n} \sum_{W} {f^*}_{\delta_n,pr,w} * \frac{x_{pr,\breve{y}_{j_n \neq i_n}}}{x_{pr,Y_n}} * {\tau^*}_{\delta_n,pr,w} \qquad \forall \; y \in Y \qquad (4)$$

$$I^3_{y_{i_n}} = I^1_{y_{i_n}} + I^2_{y_{i_n}} \qquad \forall \; y \in Y \qquad (5)$$

Where $I^3_{y_{i_n}}$ considers the exposure of each link segment to non-recurrent disturbances explicitly, $I^4_{e_n}$ uses the number of passengers q travelling over the considered link $e \in E$ as

proxy for the impact of a disturbance (**Eq.6**). This value can be determined based on direct passenger counts (e.g. using data from Automated Passenger Count (APC) or Automated Fare Collection (AFC) systems) or after performing an undisturbed passenger assignment using a PT model. This indicator shows the passenger volume travelling over a certain link e in case no disturbances would occur. Because passenger volume can differ for different links $e_i \in y\{e_1, .., e_m\}$, this value is expressed for each link e separately. For all considered links of the multi-level PT network the values of $I^3_{y_{i_n}}$ and $I^4_{e_n}$ can be plotted against each other. Links with the highest value for $I^3|I^4$, or the other way around, appear on the Pareto frontier in this plot. By selecting all links which are plotted on or nearby the Pareto frontier, the most vulnerable links can be identified based on these pre-selection criteria. Adjacent links in the network which all appear on the Pareto frontier can be taken together as one link segment. If one would prefer to further prioritise these identified most vulnerable links, an assessment of the number of available alternative routes in the multi-level PT network could be performed for each of these links based on an expert judgment. Links for which few or no route alternatives are available can then be prioritised.

$$I^4_{e_{i_n}} = q_{e_{i_n}} \qquad \forall \ \ e \in E \tag{6}$$

6.2.3 Quantification of link vulnerability

When the most vulnerable links of the multi-level PT network are identified, vulnerability of these links can be quantified. Quantification of link vulnerability from a passenger perspective can be done using **Eq.2**, which explicitly considers both exposure to disturbances and the impact of these disturbances given the total multi-level PT network N available. As explained in **Section 6.2.1**, different disruption scenarios S can be distinguished for each link based on the impact of a disturbance δ_n on partial or full infrastructure unavailability. Given a chosen time horizon for which link vulnerability is quantified, Monte Carlo simulation is used to generate disturbances δ_n with a certain duration τ_{δ_n} for each distinguished disruption scenario. Based on the estimated parameters for frequency and duration of $\delta_{n,w}$, values are drawn from the identified theoretical distribution functions. In a public transport assignment model the PT network and PT services are adjusted according to each disruption scenario S, based on which a new passenger assignment can be performed. The total monetised societal costs $\Delta w_{e,\delta_n,\tau}$ for all passengers travelling over all OD pairs can then be compared between the undisturbed situation and the specific disruption scenario S.

$$c_t = \sum_{o=1}^{n}\sum_{d=1}^{n} \ (\alpha_a t_a + \alpha_w \sum_{x=1}^{n_t+1} t_{w,x} + \ \alpha_{in} \sum_{y=1}^{n_t+1} t_{in,y} + \ \alpha_{n_t} n_t + \ \alpha_t \sum_{z=1}^{n_t} t_{t,z} + \ \alpha_e t_e) * VoT \tag{7}$$

Eq.7 expresses the calculation of the perceived, monetised travel time effects C_t given a network modelled with n origins O and n destinations D. The different travel time components - access time from origin to a PT stop t_a, waiting time t_w before boarding each PT service (the number of waiting moments x equals the number of transfers $n_t + 1$), in-vehicle time t_{in} for each PT service (the number of used PT services y equals the number of transfers $n_t + 1$), the number of transfers n_t, transfer walking time t_t for each transfer walk z, and the egress time from the PT stop to final destination t_e - are all multiplied by their corresponding weight α as perceived by passengers and monetised using the Value of Time (VoT). The parameter values we used for the different travel time weights and VoT are derived from Bovy and Hoogendoorn-Lanser (2005) and Warffemius (2013). In these studies, the parameter values are derived by discrete choice model estimation based on individual passenger preferences, resulting in

aggregated, average parameter values suitable for the Dutch situation. Here we used a fixed VoT, independent from the amount of delay on a certain OD pair. In addition to travel time effects, the effects of disturbances on travel costs c_c are also evaluated. Incorporating c_c is especially important when considering disturbances in the context of multi-level PT networks. In case of disturbances, passengers sometimes have to use longer routes of another PT operator, potentially increasing travel costs. Another component when quantifying link vulnerability is the societal costs due to reduced travel comfort for seated passengers $c_{comf,seat}$ and standing passengers $c_{comf,stand}$, respectively. This is of relevance, since particularly during disturbances the crowding level on remaining alternative routes in the PT network can increase substantially, thereby reducing the comfort level and increasing passengers' perceived in-vehicle time. These additional societal costs are quantified based on the relation between crowding and perceived in-vehicle time, expressed as in-vehicle time multipliers with the parameter value of this crowding multiplier being in line with values found by Wardman and Whelan (2011). The societal costs of non-facilitated demand c_{non-f} are also quantified, given the possibility that during a disturbance passenger volume on a link of a certain alternative route can exceed the total supplied link capacity (seated plus standing capacity), and passengers have to wait an additional headway due to denied boarding. At last, by applying the rule of half to the generalised travel costs for each affected OD pair, cancellation costs c_{cancel} are quantified for the part of the travellers that have to cancel their PT trip following a disturbance (reflecting a change in either mode choice or trip frequency choice). For a more detailed explanation of the quantification of c_c, $c_{comf,seat}$, $c_{comf,stand}$, c_{non-f} and c_{cancel} we refer to Yap (2014).

$$\Delta w_{y_n,\delta_n,\tau} = \Delta c_t + \Delta c_c + \Delta c_{comf,seat} + \Delta c_{comf,stand} + \Delta c_{non-f} + \Delta c_{cancel} \quad (8)$$

Eq.8 shows all the components based on which the total monetised societal costs $\Delta w_{y,\delta_n,\tau}$ of a disturbance are calculated for link segment y. To incorporate the distinguished disruption scenarios S specified for each link segment y, we adjust **Eq.2** in our proposed methodology to calculate the societal costs of link segment vulnerability within a specified time horizon as shown by **Eq.9**, thereby explicitly incorporating both exposure to disturbances and impact of disturbances.

$$c_{y_n} = \sum_S \sum_{\delta_n} E\left(f^S_{y_n,\delta_n}\right) * E\left(\tau^S_{y_n,\delta_n}\right) * \Delta w^S_{y_n,\delta_n,\tau} \quad (9)$$

6.3 Results

6.3.1 Case study network

The developed methodology for identification and quantification of link vulnerability is applied in a case study to the Randstad Zuidvleugel, the southern part of the most important economic area of the Netherlands (≈2.2 million inhabitants). This area is composed of two main cities The Hague and Rotterdam, and several smaller cities and villages located between and around these cities. This area is selected because of its relatively high PT network density with PT services on different network levels. The interactions between these network levels are especially interesting when considering multi-level PT networks. The PT network is modelled as super-network in a high level of detail with the transport planning software OmniTRANS with 5,791 zones. By selecting the Randstad Zuidvleugel as case study area, we allow for a detailed modelling of PT demand and supply within acceptable computation time by the public transport assignment model. For PT lines $l \in L$ the seat capacity and crush capacity are

specified. A frequency based network representation is applied, meaning that waiting time for each PT line $l \in L$ is assumed to be half of the inter-arrival time between two PT vehicles of that line l, with a user-specified maximum waiting time. Although a schedule based representation results in a higher accuracy, this requires more detailed model input and increases computation times for passenger assignment substantially. Besides, because of the relatively high frequency of PT lines in the Randstad, differences in modelled waiting time between a frequency based and schedule based network representation remain limited.

In our model, four different time periods are distinguished: morning peak 7-8am, morning peak 8-9am, evening peak 4-6pm and the remaining hours of the work day. Especially during the morning peak, PT demand is not uniformly distributed over the two hours of the morning peak in the Netherlands (CBS, 2013). Because we consider societal costs of crowding and non-facilitated demand explicitly, assuming a uniformly distributed PT demand would lead to a biased quantification of these costs. Therefore, the morning peak is split in two separate periods 7–8am and 8–9am with separate OD matrices. The public transport OD matrices are resulting from the regional demand model of this Zuidvleugel area, based on land use input, trip frequencies, trip distribution and modal split. The regional Zuidvleugel demand and PT assignment model we used have been calibrated and validated for 2011 as base year. The Zenith algorithm is applied for performing the passenger assignment in the undisturbed situation and for distinguished disruption scenarios S (Brands et al., 2013). Despite the mentioned importance of comfort and crowding effects, these aspects are not incorporated in the generalised cost function used in the assignment. This is because especially during unplanned disturbances passengers are not expected to know the crowding level of PT services on alternative routes on beforehand, and are therefore not expected to adjust their route choice based on this *a priori*. Therefore, we do not expect that crowding level is dominant as component of the generalised cost function used to predict passenger route choice during disturbances. Besides, incorporating the capacity of PT lines in the assignment would lead to an iterative, capacity-constrained assignment, which increases computation times substantially. The perceived additional disutility because of discomfort on the chosen route is however incorporated afterwards in the evaluation of link vulnerability by $c_{comf,seat}$ and $c_{comf,stand}$, where the difference in perceived in-vehicle time due to crowding between the undisrupted and disrupted scenario is quantified.

6.3.2 Identification of link vulnerability in the Randstad Zuidvleugel network

Vulnerable links are identified for the case study network by using the historic dataset of realised disturbances on the network levels of different PT operators in the Netherlands as input for calculating the first-order, second-order and total link segment exposure. **Figure 6.3** shows the expected first-order, second-order and total exposure to non-recurrent disturbances per year for link segments on the agglomeration (metro / light rail) network level. **Figure 6.4** shows the expected total exposure per year for link segments on the regional (train), agglomeration (metro / light rail) and urban (tram) network level. The metro and light rail network level are grouped together, since these modes operate on the same functional network level. Several insights can be gained from **Figure 6.3** and **Figure 6.4**.

First, **Figure 6.3** shows the importance to incorporate second-order spillback effects when calculating the expected total time a link segment is exposed to large disturbances. **Figure 6.3** clearly shows that the expected total time several link segment y_i are blocked is heavily influenced by disturbances occurring on other link segments $y_{j \neq i}$. Not considering these second-order effects would lead to substantial underestimation of link segment vulnerability.

Second, it becomes clear that the expected total link segment exposure of light rail link segments near The Hague (triangular dots in **Figure 6.3** up to no. 283) is substantially larger

compared to link segments of the Rotterdam metro network (triangular dots in **Figure 6.3** from no. 283). This can be explained by the considerably higher switch density on the Rotterdam metro network compared to the light rail network near The Hague. This results in disturbances remaining more local in Rotterdam, reducing their second-order propagation effect to other link segments. A network with a relatively low switch density means that PT services on more links adjacent to the link where the disturbance actually occurs need to be adjusted, thereby affecting a larger group of passengers. This also means that links will suffer more often from disturbances occurring on other adjacent links, thus increasing the second-order exposure time to disturbances. In the Rotterdam metro network, there are switches available near almost every metro station. This means that when a disturbance occurs on a certain metro link segment, thereby blocking the link in either one or both directions, the operator splits metro services in two separate parts up to both sides of the link segment. The second-order effect then equals the first-order effect, since disturbances occurring on link segment y_{i1} in direction 1 will only affect services on the exact same link segment in the other direction 2 y_{i2} as second-order effect.

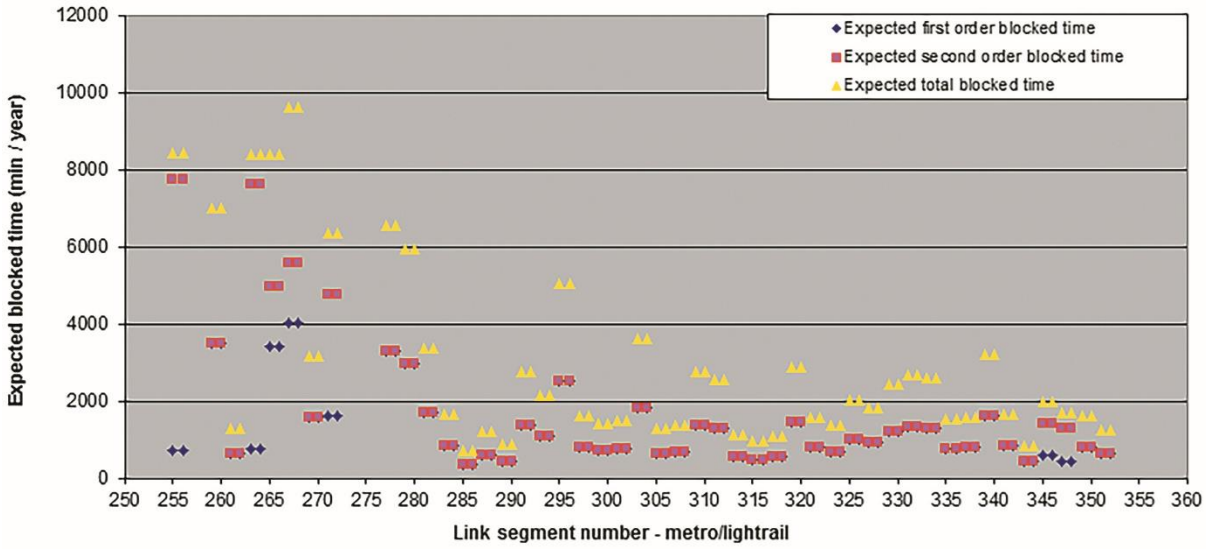


Figure 6.3. Expected first-order, second-order and total exposure per link segment of the light rail / metro case study network

(link segments up to no. 283: The Hague; link segments from no. 283: Rotterdam) (blue dots can be exactly equal to the pink dots in some cases)

Third, **Figure 6.4** shows that train link segments are relatively robust against exposure to disturbances compared to metro / light rail and tram link segments. Possible explanations for this are the own right of way for trains, the availability of a signalling system to prevent train-train collisions and the relatively low train intensity on train links compared to metro, light rail or tram links.

Fourth, in general the link segments of the tram network of The Hague (triangular dots left in **Figure 6.4**) are more vulnerable to exposure to disturbances compared to link segments of the Rotterdam tram network (triangular dots right in **Figure 6.4**). This can partly be explained because in general more parallel (sometimes unused) tram tracks are available in Rotterdam, which can function as backup in case of disturbances and reduce second-order exposure effects. The triangular outlier in the middle of **Figure 6.4** shows the specific link segment Ternoot - Laan van NOI of the tram network of The Hague. This link is located directly before/after the light rail route Laan van NOI – Zoetermeer / Rotterdam, without intermediate rescheduling possibilities. A disturbance on that part of the light rail network therefore also affects PT

services on this tram link segment. Therefore, second-order effects are relatively large on this particular link segment.

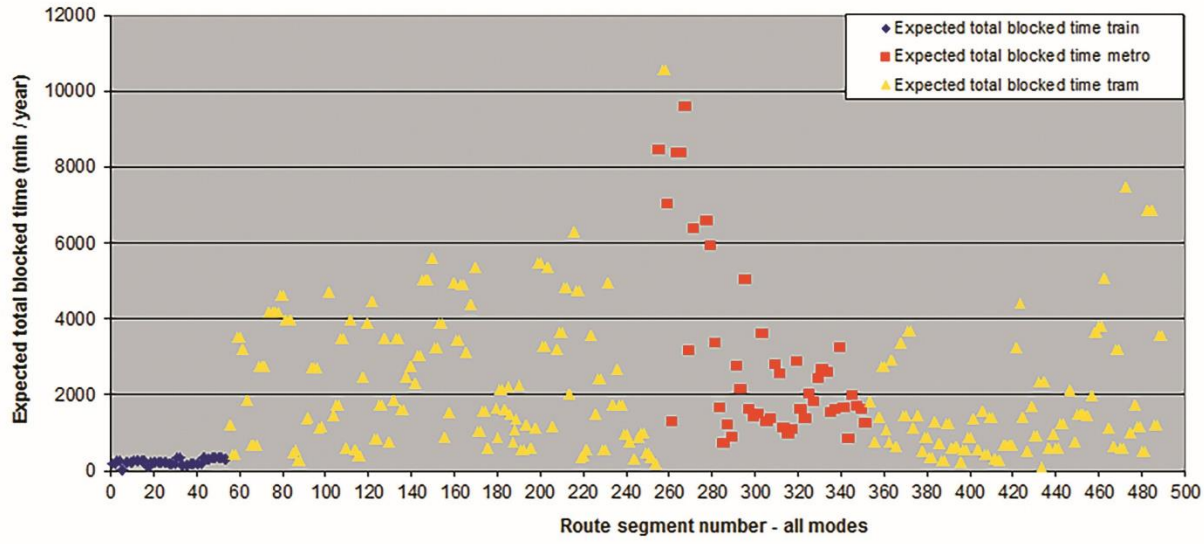


Figure 6.4. Expected total exposure to disturbances per link segment of the multi-level case study network
(yellow triangles left: tram network The Hague; yellow triangles right: tram network Rotterdam)

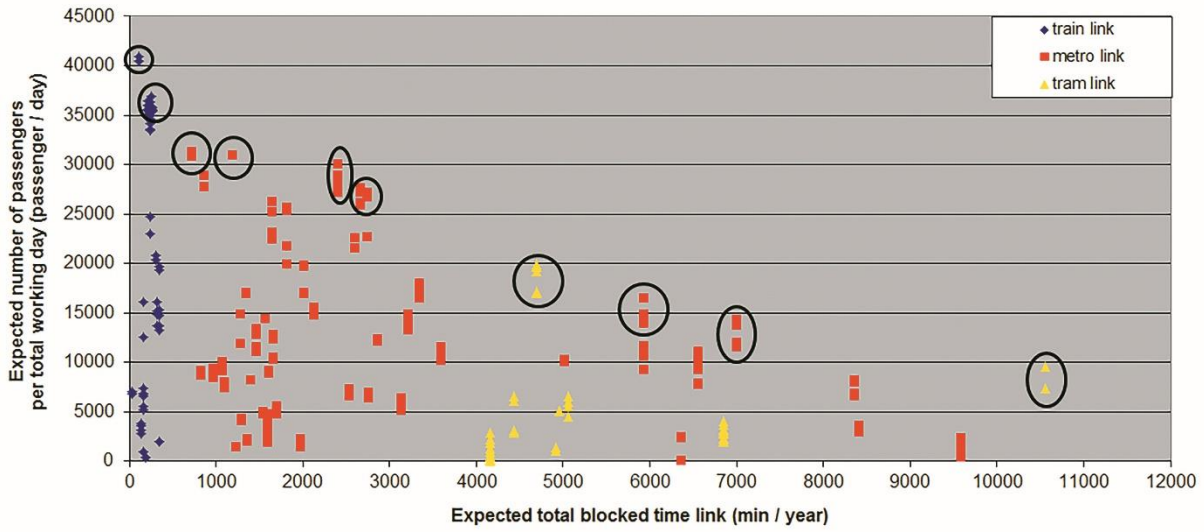


Figure 6.5. Identification of vulnerable links of the multi-level case study network by plotting I^3 against I^4

In **Figure 6.5,** the results for pre-selection criteria I^3 (expected total exposure to disturbances per year) and I^4 (expected passenger volume) are plotted against each other for each link. Based on this figure we stress the importance of using pre-selection criteria in a methodology to identify vulnerable links in a multi-level PT network which capture *both* exposure to disturbances and the impact of each disturbance explicitly. **Figure 6.5** clearly illustrates there are link segments on the Pareto frontier of which the impact of a disturbance is expected to be relatively low, but which are very vulnerable because of relatively heavy exposure to disturbances (see for example the most right triangular dots in **Figure 6.5**). If only the impact of a disturbance would be considered, only the busiest links of the train network would be

identified as most vulnerable. However, given the Pareto frontier where incident probability and impact are both considered, we conclude that there is no network level or mode which clearly contains most vulnerable links. Links on or nearby the Pareto frontier are from the train, metro / light rail and tram network. Train links are especially vulnerable because of the expected large passenger impact of disturbances, whereas metro and tram links are mainly vulnerable when a combination of relatively heavy exposure to disturbances and a relatively large number of affected passengers is expected.

6.3.3 Quantification of link vulnerability of link segment Laan van NOI - Forepark

The developed methodology to quantify link vulnerability is applied to the light rail segment Laan van NOI - Forepark. This link segment is selected from the Pareto frontier as shown in **Figure 6.5** as example. To illustrate our methodology, link vulnerability for this segment is quantified in the current situation without additional measures. This is contrasted with the quantification of link vulnerability when a measure would be applied which potentially reduces the vulnerability of this segment. The results of this case study application are shown in **Figure 6.6**, where monetised link vulnerability is expressed for the current situation without measures (left), and after testing the impact of a robustness measure (right). Public transport services on this link segment are operated by two different operators together: HTM and RET. Disturbances are generated using Monte Carlo simulation for a time horizon of 10 years. Based on this simulation we can conclude that during 10 years the light rail segment Laan van NOI - Forepark is expected to be exposed to non-recurrent disturbances for 964 hours. Assuming on average 18 hours PT operation per day, this means that in 1.5% of the time PT services on that link segment are blocked due to disturbances. In case this link segment is blocked, the applied passenger assignment model shows which alternative routes in the multi-level PT network are used. Link segment vulnerability c_{y_n} is calculated using **Eq.9** in the **Section 6.2**, by multiplying expected exposure to disturbances and expected disturbance impact $\Delta w_{y_n,\delta_n,\tau}$. $\Delta w_{y_n,\delta_n,\tau}$ is calculated using **Eq.8** by summation of the monetised impact of a disturbance on passenger travel time, costs, crowding, non-facilitated demand and trip cancellation costs. The total monetised passenger travel time effects, in turn, are calculated using **Eq.7**, in which all travel time components - access time, in-vehicle time, waiting time, walking time, transfer time and number of transfers - are calculated and summed. In the current situation, the expected total societal costs of disturbances on this link segment in 10 years equal €4.3 million. This value expresses the societal costs of vulnerability of the analysed link segment. **Figure 6.6** (left) shows that additional transfers and their related waiting time and transfer time are the most important contributors to these societal costs. During these 10 years, 787 disturbances are expected on this link segment according to our simulation results. This means that expected average societal costs per disturbance equal €5.4 thousand.

To illustrate our method, we also developed a measure aiming to improve robustness of this link segment. We evaluated this measure using a societal cost-benefit analysis. Since many affected passengers use the local train connection on the more or less parallel train track between Zoetermeer, Ypenburg and The Hague as backup during disturbances, we propose a temporary increase in frequency on this connection. We investigated adding two temporary stops for intercity train services operating on this parallel train track at two already existing local train stops, Zoetermeer and The Hague Ypenburg, *only* in case PT operations on the light rail segment Laan van NOI - Forepark are disrupted. In that case, the frequency of stopping train services between Zoetermeer, Ypenburg and The Hague is effectively doubled during disturbances. This improves transfer possibilities between network levels and improves the backup function of the train network for the disrupted light rail network. A disadvantage

however is that the travel time for through travellers in the intercity service increases with ≈5 minutes caused by the additional stops, which also results in additional operational costs.

After generating disturbances using a pseudo-random generator, we can quantify the robustness effects of this measure (see **Figure 6.6** right). Total societal costs of link segment vulnerability after 10 years now equal €3.9 million. This means that this measure reduces the costs of vulnerability of this link segment by 8%, therefore having a positive Net Present Value. The expected average societal costs per disturbance now equal €5.0 thousand. This measure especially reduces waiting time substantially, at cost of an increase in total in-vehicle time. However, monetised benefits from waiting time reduction outweigh the monetised costs of additional in-vehicle time.

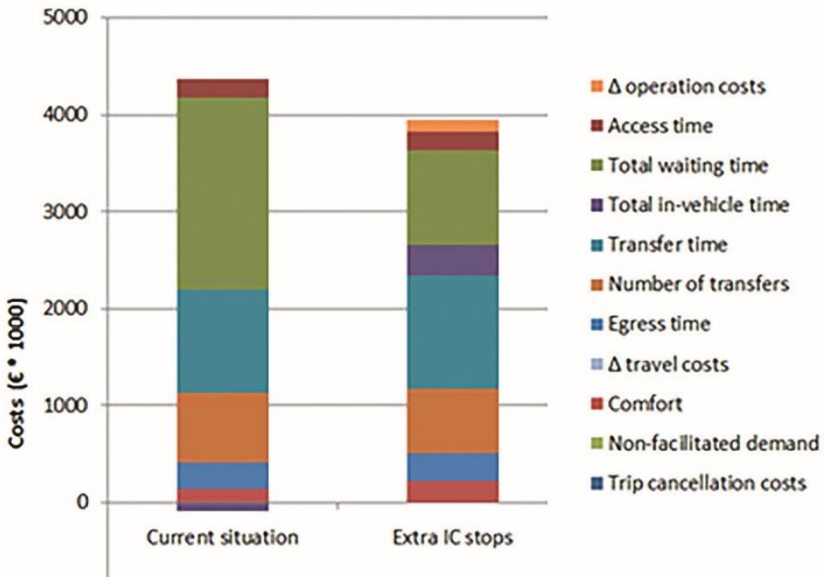


Figure 6.6. Societal costs of link segment vulnerability Laan van NOI - Forepark for the current situation (left) and for the proposed measure (right)

We can formulate some points for discussion, regarding the quantification of the robustness benefits of this specific case study measure. First, for a successful implementation of measures using the multi-level PT network it is important to consider the distribution of financial and societal costs and benefits between stakeholders involved. Most costs of the proposed measure are for the Dutch train operator NS, because of additional timetable hours their trains have to run and additional travel time for train passengers, despite the disturbance occurring on the network operated by the HTM and RET. To implement this measure successfully, it seems likely that (financial) incentives have to be provided to the Dutch Railways by PT authorities or the PT operators HTM and RET. Second, we did not quantify the network wide effects for train passengers caused by the increased travel time in intercity services, for example when connections later on the route are missed. We however did check that sufficient buffer time between conflicting trains was available in the timetable on the specific train track in case running time of intercity trains would be extended by 5 minutes in our measure. Third, for a successful implementation of the proposed measure it is required that the RET and HTM provide passengers information about the temporary doubled frequency of train services stopping at the stations Zoetermeer and The Hague Ypenburg, so that passengers can incorporate this in their route choice accordingly.

6.4 Conclusions and Further Research

Despite the importance of robust public transport networks, the full impact of public transport network vulnerability on passengers has not been considered in scientific literature and practice yet. The added value of this study is the development of a methodology to identify the most vulnerable links in multi-level public transport networks, and to quantify link vulnerability for these identified links. To our best knowledge, this study is the first which addresses full and realistic passenger impacts of disturbances in both the identification and quantification of link vulnerability. Based on our study, we formulate methodological, practical and policy-related conclusions.

In our study we propose a methodology to systematically identify the most vulnerable links in a multi-level public transport network. Based on the identified vulnerable links, our methodology allows for quantification of the full passenger impacts of disturbances on these vulnerable links. Contrary to single-level network perspectives, we consider the integrated, total multi-level PT network, including PT services on other network levels, which remains available after the occurrence of a certain disturbance. This results in a more realistic impact quantification compared to single-level approaches. In our approach, both exposure to large, non-recurrent disturbances and the impact of these disturbances are analysed in a systematic manner in link vulnerability identification and quantification. Our case study results show the importance of incorporating exposure to disturbances explicitly in the identification of vulnerable links, since only considering the impact of disturbances would result in a very different list of most vulnerable links. Based on our results, we also stress the relevance of taking into account second-order spillback effects in this methodology when calculating total link segment exposure to disturbances. Not considering second-order exposure to disturbances can lead to substantial underestimations of link vulnerability.

Our practical case study application shows that train network links are less frequently exposed to disturbances, compared to links of the urban tram network or light rail / metro links on the agglomeration network level. The passenger impact of disturbances on the train network is however relatively large, due to the large passenger flows affected. Our study shows that therefore particularly busy links of the light rail / metro network are vulnerable, given the relatively high disruption exposure and relatively high number of passengers affected. From our case study we estimate that the monetised passenger impact of disturbances on one single, relatively vulnerable light rail / metro link segment equals €4.3 million over 10 years.

Currently, only the costs of measures aiming to improve robustness are known to policy-makers. Based on our methodology we are able to monetise the societal costs of disturbances as well. Besides, applying our methodology enables the quantification of the part of these societal costs which can be reduced by a certain robustness measure. This allows us to express robustness benefits of a certain measure in monetary terms and to compare these with the required costs of that measure, thereby quantifying the value of robustness. This enables policy-makers to make a trade-off between costs of robustness measures and monetised robustness benefits of these measures. From a policy perspective, it is important to realise that the topic of robustness should always be considered as trade-off with other aspects. Some robustness measures (such as the construction of additional switches) can on the one hand reduce the societal costs of a disturbance, once a disturbance occurs, but on the other hand increase link exposure to disturbances. Other measures can reduce the impact of a disturbance for affected passengers, while increasing travel time for other groups of passengers. Some measures might be able to improve robustness substantially on the one hand, but require large investments on the other hand. The result of these trade-offs will be different for different locations in the network and depends on the frequency with which disturbances occur, the impact of disturbances, the number of passengers affected by the disturbance and the extent to which

alternative routes are available in the multi-level PT network. Applying our methodology allows decision-makers to get insight in these trade-offs for each specific location, where all aspects relevant for such trade-off are expressed in the same, monetary units. Therefore, our methodology helps to support and rationalise the decision-making process regarding the implementation of different robustness measures. It provides insights how the additional travel time during disturbances can be reduced by certain robustness measures. This output can be used to quantify the accessibility benefits of different robustness measures, for example by using the number of locations or activities that can be reached by public transport within a certain time during disturbances. This helps prioritising measures based on their contribution to accessibility.

We formulate five recommendations for further research. First, costs for rescheduling and recovery of the planned timetable, vehicle and personnel circulation are not considered in our study. Incorporating this would further increase the financial and societal costs of disturbances. Therefore, the societal costs of disturbances as calculated in this research can be considered a lower bound. Second, we recommend a further extension of our proposed method to identify vulnerable links. In our study, the most vulnerable links from the Pareto frontier can be selected based on a qualitative assessment of the number of available alternative routes for each link. By quantifying this last step, our methodology can be further improved. For example, for each OD pair affected by a disturbance on a certain link, the number of feasible route alternatives and their remaining capacity could be calculated by applying route choice set criteria (see for example Fiorenzo-Catalano, 2007). Third, we recommend to incorporate dynamic en-route passenger route choice in the disrupted passenger assignment based on travel information available to passengers (see for example Van der Hurk et al., 2012; Cats and Jenelius, 2014). For all our assignments we only considered pre-trip route choice, assuming full information about a disturbance during the whole trip. This shows the potential of the multi-level PT network to function as backup in case of a disturbance. However, in reality disturbances are dynamic and there is not always full information available about the disturbance. It is therefore interesting to incorporate the dynamics of disturbances and the role of information provided to passengers, combined with en-route passenger route choice, in the assignment. Fourth, we recommend a further node-based vulnerability analysis next to our performed link-based vulnerability analysis. By explicitly studying which public transport stops are most vulnerable by executing a similar node-based vulnerability analysis, insights can be gained in different levels of vulnerability for different types of public transport stops. In a Dutch context, the Dutch Railways classify all their stations into six categories based on their function and number of passengers, as for example applied by Geurs et al. (2016) and La Paix Puello and Geurs (2016). A node-based vulnerability analysis can provide insights to policy-makers which station type is particularly vulnerable, thereby prioritising the type of stations where robustness measures have most societal value. Fifth, it should be mentioned that a multi-level approach regarding PT robustness is rather complex in terms of collecting revealed data about disturbances occurring on the networks of multiple PT operators, assignment calculation times, and implementation of measures which exceed the borders of the network of a certain PT operator. PT authorities could play a role here by developing passenger-oriented incentives for operators in case of disturbances, which take the total multi-level PT network into consideration.

Part III

Towards Controlling Disruption Impacts

7. Where Shall We Sync? Clustering Passenger Flows to Identify Urban Public Transport Hubs and their Key Synchronisation Priorities

After *measuring disruptions impacts* (Part I of this research) and *predicting the impacts of future disruptions* (Part II of this research), we move in the last stage of this research (Part III) towards developing methods to *mitigate disruption impacts*. This part answers the third research question (as defined in **Section 1.3**): how can we predict and control the direct and propagated impacts of disruptions on the urban public transport network in a multi-level network environment? To this end mitigation measures can be applied to the urban network itself, when controlling impacts of disruptions which occurred on the urban network or on another level of the multi-level public transport network. Besides, mitigation measures can also be applied on a different network level, such as the regional train network level, if this affects the disruption propagation to the urban network. In this chapter, we focus on measures applied to the urban network, whilst **Chapter 8** considers measures applied to the train network level. In this chapter, we focus specifically on public transport synchronisation as mitigation measure applied to the urban network. The contribution of this work is the development of a data-driven clustering method to identify the most important locations and routes to prioritise for optimal synchronisation, when network-wide optimisation becomes computationally expensive or infeasible for large, real-world urban public transport networks. Once these locations and routes are determined, optimal synchronisation can be applied to this selection of locations and routes in a next step, such that disruption and delay impacts can be reduced by holding the appropriate public transport trips.

This chapter is based on an edited version of the following article:

Yap, M.D., Luo, D., Cats, O., Van Oort, N., Hoogendoorn, S.P. (2019). Where shall we sync? Clustering passenger flows to identify urban public transport hubs and their key synchronization priorities. *Transportation Research Part C, 98,* 433-448.
© 2018 Elsevier Ltd. All rights reserved.

7.1 Introduction

Transfers are an inevitable part of public transport (PT) journeys, since it is not economically viable to directly connect all origin-destination pairs in a network. Transfer locations are however potential weak parts of the total passenger journey experience. Empirical studies show that transfers are perceived as one of the most negative components in the public transport journey (e.g. Schakenbos et al., 2016; Van Oort et al., 2016). Improving the transfer experience thus offers potential to improve the attractiveness of public transport and to increase public transport ridership. The passenger transfer experience can be improved by objective and subjective means. Subjective measures focus on improving the waiting time experience at stops during a transfer (e.g. Van Hagen, 2011). On a strategic level, objective measures aim to reduce passenger waiting times by optimising PT lines (e.g. Gkiotsalitis et al., 2019) and optimising headways between services (e.g. Gkiotsalitis and Cats, 2018; Varga et al., 2018). On a tactical and operational level, objective measures relate to transfer synchronisation, thereby aiming to reduce transfer waiting (and possibly walking) time.

Although minimising passengers' transfer waiting time by PT synchronisation is important during tactical planning and real-time operations, there are limits for optimisation in terms of scalability and complexity. The Timetable Synchronisation Problem[1] (TSP), aimed to optimise transfer synchronisation in PT networks, has been addressed earlier in many studies in the context of either tactical planning (timetable design) or real-time control. The latter results from stochasticity or disruptions affecting actual vehicle arrival and departure times. In these studies, solving the TSP is usually applied to a relatively small case study network. For example, Lee et al. (2014) consider the impact of synchronising two lines during tactical planning on service reliability, whereas Gavriilidou and Cats (2019) study real-time synchronisation of two tram lines based on passenger data, both for one specific transfer location. Nesheli and Ceder (2015) compare the effects of different real-time control tactics when solving the TSP, applied to a case study network of three bus lines with two transfer locations. Hadas and Ceder (2010) optimise real-time transfer synchronisation by simulation of different control strategies, applied to a case study network consisting of one train line, three bus lines and five transfer locations. The abovementioned studies are examples of mathematical programming or control theory approaches applied to the TSP. These studies apply their approach either to a selection of PT lines and transfer locations from the total PT network, or to PT networks constituting small- to medium-sized graphs (e.g. metro networks consisting of a few lines and transfer locations). In contrast, to the best of our knowledge, no studies have been successful in solving the TSP optimisation process for large, real-world urban PT networks, often consisting of tens to hundreds of PT lines and transfer locations. Hence, finding or even approximating an optimal solution for the TSP becomes mathematically expensive, if not infeasible.

Solving the TSP is considered NP-hard due to the combinatorial nature of the problem (Desaulniers and Hickman, 2007). For practical problems in larger real-world PT networks, computation time for solving this problem can therefore rise substantially, making it infeasible to solve. Enumeration of all transfer possibilities for a large real-world network would result in a very large number of transfer possibilities, since each transfer location served by $|l|$ unidirectional lines provides $|l| * (|l| - 2)$ transfer possibilities, excluding transfers to the same

[1] The Timetable Synchronisation Problem (TSP) aims to determine the departure times at transfer locations for all trips of a PT network during tactical planning (timetable design) or during real-time operations (i.e. holding), constrained by the headways of each line. The objective is typically to minimise total passenger transfer waiting time, or maximise the number of direct transfers. Decision variables are the departure time from the transfer location, possibly together with the dispatching time from the first stop of the line or the departure time from the previous holding point.

line. For example, enumerating all transfer possibilities for the urban PT network of The Hague, the Netherlands, provides 1,720 transfer possibilities in total between all lines. Solving the TSP for all transfer possibilities of this whole network would become infeasible within reasonable computation times.

To address the computational challenges of solving the TSP for larger, real-world urban PT networks, we propose a methodology for systematically determining where in the network, and for which lines transfer synchronisation should be prioritised. Our study thus introduces two steps preceding solving the TSP: identify key priorities (a) where to synchronise, and (b) which lines to synchronise. These two steps are aimed at reducing the combinatorial complexity of the subsequent TSP, and result in a subset of transfer locations and lines where synchronisation should be prioritised based on passenger transfer flows. Subsequently, the optimisation process to solve the TSP can be applied to this subset, resulting in (an approximation of) optimal synchronisation at the most important locations and amongst the most important lines. The novelty of this study lies in problem definition and in using a combination of approaches which have not been used previously for addressing this key planning and operation challenge.

The first step of our methodology is necessary to find a subset of most important transfer locations from the large number of transfer locations a real-world urban PT network consists of, and to determine the spatial demarcation of these urban transfer locations. Given the large number of transfer locations within a real-world urban PT network, the most important transfer locations - defined as hubs - are prioritised during transfer synchronisation. Determining the geographical boundaries of transfer locations in high-density urban PT networks is however far from a trivial task. In airline networks for example, the spatial demarcation of a hub is usually unambiguous, given the well-defined physical boundaries of airports and the large distance between airports. The same reasoning applies to train stations being part of an (inter-)regional train network as well. Conversely, the spatial demarcation of transfer locations is less clear for urban PT networks, since many stops are located within walking distance from each other. Locations in the urban network where there is a high flow of transferring passengers (e.g. a bus terminal, train station or shopping area) usually have a large number of tram and bus stops in their surroundings. It is however unclear which of these stops can be considered as one coherent transfer location, and which stops do not make up a part of this transfer location. On the one hand, PT stops which share the same public name but are located a bit further away from the other PT stops with this name, could constitute part of one large transfer location, not belong to this transfer location, or possibly form a separate, second transfer location, depending on the passenger transfer flows between all different PT stops. On the other hand, given the relatively small distance between different stops in an urban PT network, there can be substantial transfer flows (involving walking) between stops with different public names, which could mean these stops form one transfer location based on passenger flows. It is therefore necessary to develop a methodology to systematically identify the spatial demarcation of urban PT hubs, being the most important transfer locations, without relying only on local knowledge, geographical information or public stop names.

The second step of our methodology uses the identified hubs with their spatial demarcation from the first step, to determine synchronisation priorities within each of these hubs. Based on the spatial demarcation, it can be determined which lines within a hub should be considered in the prioritising process. When the most important transfer locations for transfer synchronisation are determined, solving the TSP within a specific hub can still be problematic due to the number of transfer possibilities. For example, a hub with 15 bidirectional PT lines already results in 840 transfer combinations, which makes optimising coordination between all lines simultaneously computationally challenging. In order to make solving the TSP feasible, there is a need to identify which PT lines in which direction should be considered as one group

- in the remainder of this chapter coined as one *line bundle* - to be subject to synchronisation efforts simultaneously.

We apply cluster analysis as the unsupervised learning technique to identify the hubs and line bundles to synchronise. Machine learning approaches have been applied in a variety of studies related to understanding travel patterns and predicting passenger flows. Examples of studies applying unsupervised learning techniques to better understand travel patterns are Agard et al. (2007), Ma et al. (2013), Cats et al. (2015b), El Mahrsi et al. (2017) and Luo et al. (2017). Studies performed by Wei and Chen (2012), Ding et al. (2016) and Li et al. (2017b) are exemplary for supervised machine learning applications to better predict passenger flows. The variety of applications of machine learning techniques in public transport demonstrates the potential of using machine learning. Notwithstanding, to the best of our knowledge, machine learning has not yet been applied to identify the spatial boundaries of hubs and to identify line bundles within hubs. This emphasises the need to develop a new machine learning application to be able to address our research statement to prioritise locations and line bundles for synchronisation. In this study, we develop a generic methodology which is data-driven and independent from local network knowledge or a specific network topology by applying unsupervised learning methods. We adopt a passenger perspective, thereby performing our clustering fully based on passenger transfer flow data, rather than using geographical information or incorporating operator constraints in the prioritisation. We apply our proposed methodology to the urban PT network of The Hague, the Netherlands as a case study.

The main contributions of this study are therefore (a) the development of a data-driven methodology to identify hubs with their spatial demarcation in urban PT networks; (b) the determination of line bundles within these hubs that need to be prioritised in transfer synchronisation, based on a graph representation of transfer patterns and a community detection technique adopted from the field of complex network science. **Section 7.2.1** addresses our methodology for identifying the spatial boundaries of transfer locations and selecting hubs from these transfer locations. **Section 7.2.2** describes the approach to identify line bundles to prioritise in transfer synchronisation. In **Section 7.3**, the case study is introduced. **Section 7.4** discusses the results of the successive steps of our methodology, followed by conclusions and further research recommendations in **Section 7.5**.

7.2 Methodology

For the remainder of the chapter, we introduce indices, sets and variables as presented in **Table 7.1**. Furthermore, we introduce the following definitions. A *transfer location* t consists of one or more stops $s \in S$ which are considered one coherent cluster based on passenger transfer flows. Since not all stops $s \in S$ are part of a transfer location, $T \subseteq S$ applies. *Hubs* H are the most important transfer locations amongst T based on the passenger transfer flows between stops, so that $H \subseteq T$ applies. Each public transport line $l \in L$ represents a route with a unique public line number as communicated to passengers in a certain direction and is therefore unidirectional. Each cluster of unidirectional public transport lines which should be considered as one group simultaneously during synchronisation at a hub $h \in H$ is defined as a *line bundle* $c \in C$, with $c = \{l_1^c, l_2^c, \dots, l_n^c\}$.

Figure 7.1 presents a flowchart with all steps of the proposed methodology. It shows the methodology used to identify hubs and their spatial demarcation to prioritise for synchronisation (**Section 7.2.1**), and the methodology used to identify the line bundles to synchronise simultaneously within each identified hub (**Section 7.2.2**).

Table 7.1. Indices and sets, variables and parameters

Indices and sets	
v, V	index for each node of graph G, set of nodes
e, E	index for each edge of graph G, set of links
s, S	index for each public transport stop, set of stops
l, L	index for each unidirectional public transport line, set of lines
i, I	index for matrix rows representing origin nodes, set of origin nodes
j, J	index for matrix columns representing destination nodes, set of destination nodes
t, T	index for selection of stops identified as transfer location, set of transfer locations
h, H	index for selection of transfer locations identified as hub, set of hubs
c, C	line bundle, set of line bundles
Variables	
a_{ij}	element of (weighted) adjacency matrix (Section 7.2.2)
d	scheduled headway
f_{ij}	passenger transfer flow between i and j
k_t	passenger transfer flow between all stops that belong to transfer location t
p_{ij}	element of adjacency matrix of null model (Section 7.2.2)
q	modularity
w_i	the sum of the weights of the edges attached to node i in the definition of modularity
Parameters	
γ_{max}	maximum transfer walking distance
ε	maximum distance function value to form a cluster in DBSCAN (Section 7.2.1)
θ	minimum number of stops required to form a cluster in DBSCAN (Section 7.2.1)
ϑ	walking speed

7.2.1 Identification of transfer location priorities for synchronisation

Infer stop-to-stop transfer flow matrix

The first step of our methodology is to infer the stop-to-stop transfer flow matrix, as visualised by the first row of phase 1 in **Figure 7.1**. As input for our study we use passenger data from Automated Fare Collection (AFC) systems and PT vehicle position data obtained from Automated Vehicle Location (AVL) systems for urban tram and bus services. Each transaction from AFC systems contains at least the passenger smart card id, tap in time, tap in stop, and the vehicle line and trip id of the run a passenger boarded for each journey leg separately. In some cases - such as Seoul, Queensland and the Netherlands - the passenger tap out time and tap out stop are also registered. In case of entry-only AFC systems, the alighting location can be inferred using different destination inference algorithms, of which a trip chaining algorithm is most commonly applied (e.g. Trépanier et al., 2007; Zhao et al., 2007; Nunes et al., 2016; Munizaga and Palma, 2012). This results in AFC transactions with information about the passenger tap in and tap out time and location used as input for our study (see **Table 7.2** as illustrative data format). Each row of the AVL data contains information about the scheduled and realised arrival time and departure time of each PT run with corresponding trip id for each stop (see **Table 7.3** for illustration purposes). Besides, the coordinates of each PT stop of the considered urban PT network are used as input. Each separate platform has a unique stop id and unique coordinates.

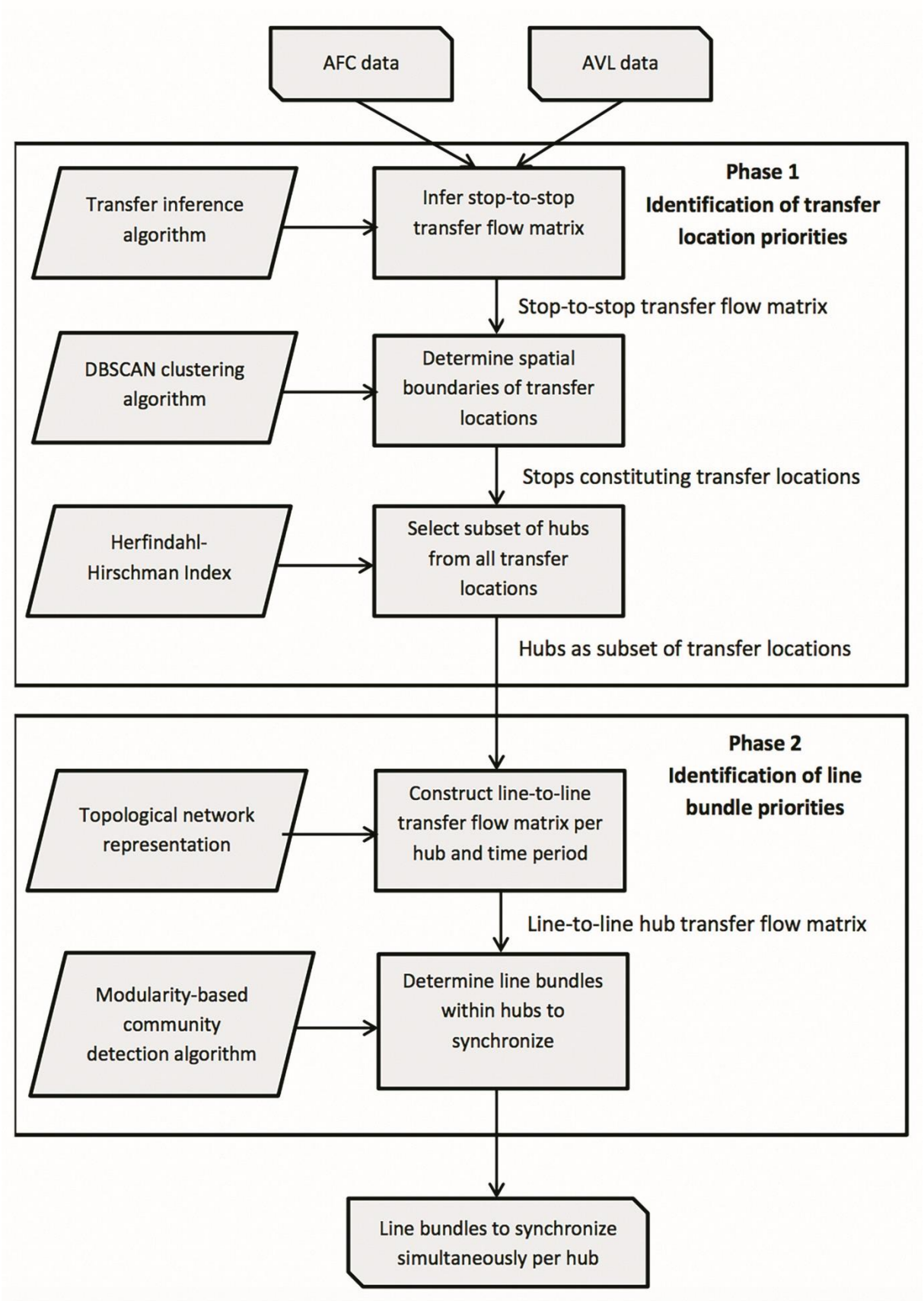


Figure 7.1. Flow chart of the proposed methodology

Table 7.2. Illustration of format AFC data

Tap in date and time	Tap in stop id	Tap in line	Tap out date and time	Tap out stop id	Trip id	Vehicle id	Smart card id
4-3-2018 11:42:37	35309	6	4-3-2018 12:03:19	34997	3423	3050	81675688
4-3-2018 12:15:57	30091	18	4-3-2018 12:23:04	32857	6545	187	81675688

Table 7.3. Illustration of format AVL data

Stop id	Trip id	Scheduled arrival time	Realised arrival time	Scheduled departure time	Realised departure time
1119	4464	2017-01-06 19:22:35	2017-01-06 19:23:25	2017-01-06 19:22:35	2017-01-06 19:23:49
1119	4465	2017-01-06 18:23:48	2017-01-06 18:26:26	2017-01-06 18:23:48	2017-01-06 18:26:44

Since AFC data contains passenger transactions per journey leg for urban tram and bus services, we apply a transfer inference algorithm to determine which transactions constitute one passenger journey. In this study, we apply the transfer inference algorithm proposed by Yap et al. (2017), which elaborates on previous work by Gordon et al. (2013). This algorithm distinguishes transfers from final destinations using assumptions on passenger behaviour during both undisrupted and disrupted scenarios. Using this transfer inference algorithm has therefore the advantage that no *a priori* data cleaning or data classification as disrupted or undisrupted is needed, making the need to demarcate the time the passenger impact of a disruption starts and ends obsolete.

To execute this algorithm, we fuse the AFC and AVL datasets based on the trip id variable both data systems have in common. For urban tram and bus networks with 100% smart card penetration rate, in-vehicle devices and registered or inferred tap out location, the stop-to-stop vehicle occupancies for each trip id can be directly obtained from fusion of AFC and AVL data. In case the smart card penetration rate is not 100%, the obtained occupancies should be increased by the percentage non-card users obtained from alternative data sources, such as passenger counts. Below we shortly describe the applied transfer inference algorithm. For a more extensive explanation we refer the reader to Yap et al. (2017). An alighting is considered a transfer if it satisfies the following three conditions:

- Temporal condition: an alighting passenger boards the first vehicle of the subsequent line after the first reasonable passenger arrival time at the next boarding stop - based on the transfer walking distance and assumed walking speed ϑ - where the vehicle occupancy does not exceed the norm capacity.
- Spatial condition: the next boarding location does not exceed a maximum transfer walking distance threshold γ_{max} from the previous alighting stop.
- Line based condition: the next boarding line is not the same line as previously alighted from, or the next boarding line is the first run of this same line after the alighted run, in order to incorporate the impact of possible rescheduling measures during disruptions, such as short-turning or deadheading.

After applying this transfer inference, a stop-to-stop transfer flow matrix can be constructed from the alighting stop and next boarding stop for each alighting which is considered a transfer.

Determine the spatial boundaries of transfer locations

In the next step (visualised by the second row of phase 1 in **Figure 7.1**), we determine the spatial demarcation of transfer locations by applying a clustering algorithm. This entails determining which urban PT stop id's form a coherent cluster of stops between which passenger transfer flows occur. To this end, we apply a density based clustering technique. This data-

driven approach for determining the spatial boundaries of transfer locations implies that the number of stops each cluster is composed of is not known *a priori*. The total number of clusters is thus also not pre-determined, making k-means / k-medoid clustering not suitable for this purpose. Moreover, our clustering should also not be collectively exhaustive: the considered PT network includes transfer locations consisting of one stop id only based on passenger transfer flows, such as one platform being served my multiple PT lines with transferring passengers between these lines. Consequently, clustering only needs to be applied for locations where transfers occur between multiple PT stops (multiple platforms). This requirement makes hierarchical agglomerative clustering, a collectively exhaustive clustering technique, not suitable for purpose. A density based clustering technique, which allows for partial clustering without a pre-defined number of clusters (e.g. DBSCAN, OPTICS) fulfils all of the aforementioned requirements (Tan et al., 2004). We apply DBSCAN, the most commonly applied density based clustering technique. For an in-depth explanation of the algorithm of DBSCAN we refer to Ester et al. (1996).

In traditional geographical applications of DBSCAN, a geographical measure - such as the Euclidean or geodesic distance between nodes - is applied as distance measure for clustering. This means that the closer two nodes are geographically positioned to each other, the more likely that these nodes are being grouped into the same cluster. A larger distance value thus indicates a lower clustering likelihood. However, to identify synchronisation priorities based on passenger demand, in our study we cluster purely based on passenger flows rather than geographical distances. We define $f_{s_i s_j}$ as the obtained transfer flow between origin PT stop i and destination PT stop j, where each urban PT stop $s \in S$ indicates a unique stop code in the considered urban PT network and $|S|$ indicates the number of stops (see **Table 7.1**). To create a symmetrical distance matrix, we first sum transfer flows $f_{s_i s_j}$ and $f_{s_j s_i}$ to $g_{s_i s_j}$ (**Eq.1**). In this context, contrary to traditional geographical distance measures in DBSCAN, a higher transfer flow between stops thus *increases* the likelihood of these stops being clustered together. This means that in our case the regular distance measure cannot be applied for DBSCAN. Values $g_{s_i s_j}$ are therefore transformed into $g'_{s_i s_j}$, such that a higher value decreases the clustering likelihood. By subtracting $g_{s_i s_j}$ from the maximum value $\max_{s_i s_j \in S} g_{s_i s_j}$, all values remain non-negative (**Eq.2**). Clustering based on distance measure $g'_{s_i s_j}$ entails stops being clustered if there are strong transfer flows between them. This implies that stops which are geographically close to each other do not necessarily have to be clustered together, if there is no substantial transfer flow between these stops.

$$g_{s_i s_j} = f_{s_i s_j} + f_{s_j s_i} \qquad \forall\, s_i, s_j \in S \tag{1}$$

$$g'_{s_i s_j} = \max_{s_i s_j \in S} g_{s_i s_j} - g_{s_i s_j} \qquad \forall\, s_i, s_j \in S \tag{2}$$

Two parameters θ and ε need to be specified in the DBSCAN algorithm. θ indicates the minimum number of stops required to form a cluster. We set this value to one, meaning that the inclusion of at least one other stop (two stops in total) is required for the minimum cluster size. Since we perform a partial clustering where not all stops need to be clustered, each cluster is composed of at least two stops. ε indicates the maximum distance function value required when forming a cluster. For our formulated distance measure $g'_{s_i s_j}$, this means setting a minimum requirement for the transfer flow between stops that form a cluster. In line with the recommendation by Ester et al. (1996), we plot the θ-distance graph for different values of ε to determine the knee in the plot to identify a suitable parameter value. This total distance measure

is calculated by averaging the minimum distance value of all identified clusters for a given value of ε (**Eq.3**), where $t \in T$ is a transfer location resulting from DBSCAN and T is the set of all identified transfer locations (see **Table 7.1**). After applying DBSCAN it can be determined which PT stops constitute a certain transfer location, which shows the spatial boundaries of each transfer location.

$$dist = \frac{\sum_{t \in T} \min_{s_i s_j \in t} g' s_i s_j}{|T|} \tag{3}$$

Select subset of hubs from all transfer locations

Not all identified transfer locations are considered a hub, given substantial differences in magnitude of transfer flows within each transfer location. The aim of this step is to identify the set of hubs H, which is a subset of all transfer locations $H \subseteq T$. The transfer locations with the largest transfer flows between stops are considered a hub (as visualised in the third row of phase 1 in **Figure 7.1**). We apply a method used to determine hubs in airline networks based on Costa et al. (2010), since no comparable method has yet been applied to identify hubs in urban PT networks. First, the total intra-transfer location transfer flow k_t between all stops is calculated for each identified transfer location t (**Eq.4**). Based on the work of Costa et al. (2010), we apply the Herfindahl- Hirschman Index (HHI) to calculate the market concentration based on the market share of each transfer location $t \in T$ in terms of passenger transfer flow as a share of the total transfer flow in the considered PT network. The integer number of hubs $|H|$ can then be determined based on n, which equals the inverse of the HHI (**Eq.5**). In **Eq.5**, n can be an integer or a decimal value. When all transfer locations are ranked in decreasing order based on k_t, only the top $|T| < n$ clusters are considered hubs, based on which the set of hubs H is determined. The steps of phase 1 of our proposed methodology for identifying hubs are illustrated in **Figure 7.2**.

$$k_t = \sum_{s_i \in S_t} \sum_{s_j \in S_t} f_{s_i s_j} \qquad \forall\, t \in T \tag{4}$$

$$n = \frac{1}{\sum_{t \in T} \left[\left(\frac{k_t}{\sum_{t \in T} k_t} \right)^2 \right]} \tag{5}$$

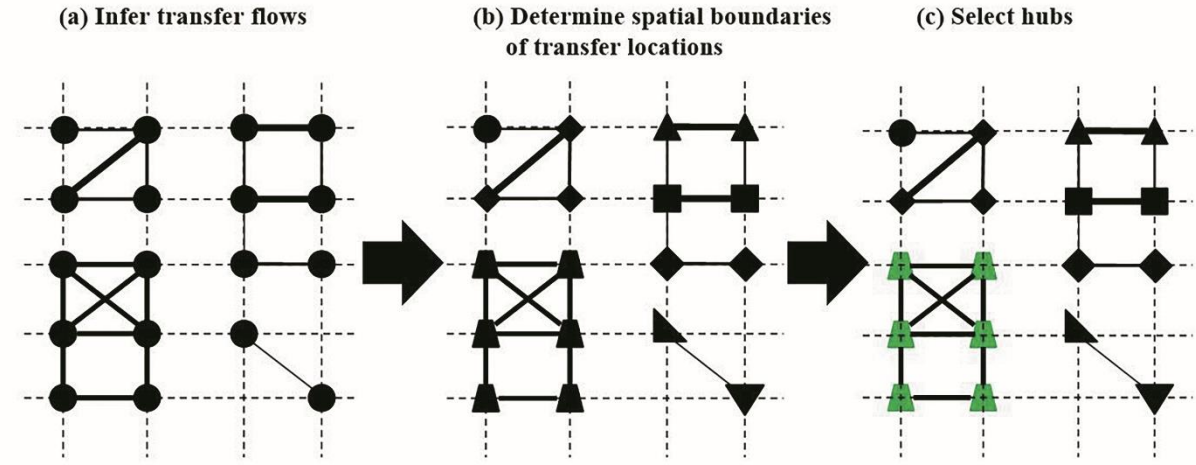


Figure 7.2. Illustration of phase 1 of our methodology for a single hypothetical PT network

Nodes reflect stops; the width of the edge represents the strength of passenger transfer flow. Based on the transfer flows between stops (a), DBSCAN identifies which stops form a cluster of transfer stops (indicated in (b) by the same shape). Applying the HHI identifies hubs as most important transfer locations in (c) (marked green).

7.2.2 Identification of line bundle priorities for synchronisation

A complex-network theoretic approach is developed to identify line bundles - in literature sometimes referred to as cliques or communities - within a hub to prioritise simultaneously for synchronisation. The proposed approach consists of two steps: (1) establishing the topological representation of the transfer pattern among unidirectional lines within the identified hubs from phase 1; (2) identifying the line bundles within each hub using a community detection technique. This approach provides a data-driven solution that is automated, intuitive and scalable. The details of these two components are presented in the following sections.

Topological representation of the transfer pattern within hubs

In this step of our methodology (as visualised in the first row of phase 2 in **Figure 7.1**) the transfer topology within an identified hub is represented as a directed graph $G = (V, E)$. This representation is inspired by the C-space topological representation of PT networks where individual lines are represented as nodes and are connected via an edge only if they share common transfer stops (see von Ferber et al. (2009) for a detailed description of the C-space representation of PT networks). In our case, each node $v \in V$ corresponds to a PT line in a certain direction $l \in L$, whereas each edge $e \in E$ represents the observed transfer activity between two lines in certain directions within an identified hub. An illustration of such topological representation is sketched in **Figure 7.3**. Furthermore, the graph G is represented by a weighted adjacency matrix A where a_{ij} denotes the weight of the edge between i and j. We consider two different types of weights in this study, namely the passenger transfer flow (**Eq.6**) and the passenger transfer waiting time (**Eq.7**). The objective of applying two different link weights is to compare clustering results between the case where only passenger transfer flows are considered, and the case when the expected transfer time is incorporated as well.

The first type of link weight corresponds to the number of passengers transferring between two lines in a certain direction at hub $h \in H$, which is defined using **Eq.6**. In this equation $f^h_{l_i l_j}$ denotes the observed transfer flow from unidirectional line l_i to line l_j within a hub. Values for $f^h_{l_i l_j}$ are derived from the stop-to-stop transfer flow matrix obtained by fusion of AFC and AVL data (as explained in **Section 7.2.1**), for which transfer flows between stops which are part of the same hub between lines l_i and l_j are summed. The second type of link weight relates to the total expected transfer waiting time between two unidirectional lines, which is calculated using **Eq.7**. Variables $f^h_{l_i l_j}$ and $d^h_{l_j}$, respectively, denote the observed transfer flow between line i and j as calculated using **Eq.6**, and the planned headway of line l_j at hub h. Given our focus on high-frequent urban PT networks, the assumption of random passenger arrivals at the stop can be justified. In future work a more advanced passenger arrival and waiting time distribution, as proposed by Ingvardson et al. (2018), can potentially be applied to our method.

$$a^h_{ij} = f^h_{l_i l_j} \tag{6}$$

$$a^h_{ij} = \frac{d^h_{l_j} * f^h_{l_i l_j}}{2} \tag{7}$$

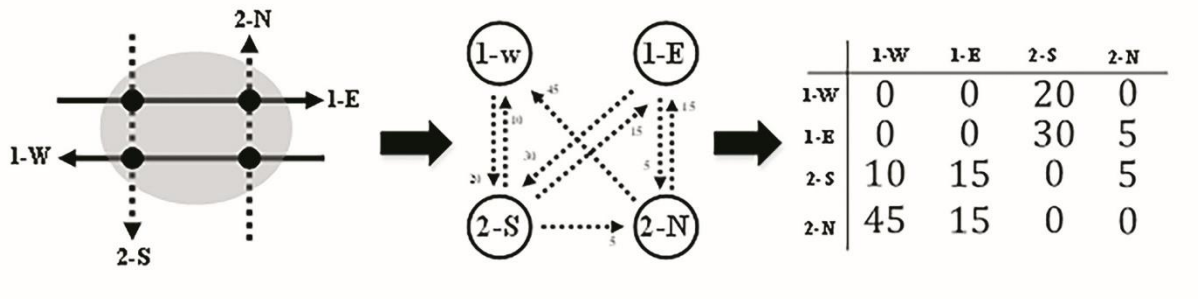


Figure 7.3. Illustration of the topological representation for the hub transfer pattern
The original layout of the identified hub (shaded area) is presented on the left with four directed lines marked, i.e. 1-E, 1-W, 2-S and 2-N. The transfer pattern is then represented as a graph (middle). The weighted adjacency matrix is displayed on the right.

Determine hub line bundles for synchronisation

Based on the graph representation with the link weighted by transfer flow or expected transfer waiting time, a community detection technique from the field of complex network science is applied to identify line bundles within a hub (see the second row of phase 2 in **Figure 7.1**). In essence, the problem that community detection intends to address is to partition a network into communities of densely connected nodes, with the nodes belonging to different communities being only sparsely connected. Such technique is for example applied by Yildirimoglu and Kim (2018) to identify a community structure in urban mobility networks. In our application, line bundles will thus become the partitioning result based on passenger transfer flows or transfer waiting time used as the link weight, in which intra-community transfer connections are maximised while inter-community values are minimised

Given our aforementioned objective, an optimisation-based method called the Louvain method is adopted to identify hub line bundles. Proposed by Blondel et al. (2008), the Louvain method is a heuristic method based on modularity optimisation. As a class of community detection methods that has received the greatest attention from researchers, the optimisation technique aims at finding an extremum - usually the maximum - of a function indicating the quality of a clustering, over the space of all clustering possibilities (Fortunato and Hric, 2016). The most popular quality function is the *modularity* proposed by Newman and Girvan (2004), which estimates the quality of a partition of the network in communities. The essential idea of this measure is to reveal how non-random the network structure is by comparing the actual structure and its randomisation where network communities are destroyed. The value of modularity q varies between −1 and 1, which measures the density of links inside communities as opposed to links between communities. Its general expression is formulated by **Eq.8**.

$$q = \frac{1}{2|E|}\sum_{ij}(a_{ij} - p_{ij})\delta_{c_i,c_j} \tag{8}$$

In this equation, $|E|$ represents the number of edges of the graph. The summation runs over all pairs of nodes i and j, in which a_{ij} and p_{ij} denote the element of the adjacency matrix and the randomised null model term, respectively. Derived by randomising the original graph, the term p_{ij} indicates the average adjacency matrix of an ensemble of networks to preserve some of its features. c_i indicates the community to which node i is assigned. The Kronecker delta function δ_{c_i,c_j} is defined using **Eq.9**.

$$\delta_{c_i,c_j} = \begin{cases} 1, & if\ c_i = c_j \\ 0, & otherwise \end{cases} \tag{9}$$

The regular modularity function of the Louvain method only uses the adjacency matrix to perform the community detection. However, for our study purpose we consider not only the question whether two nodes are connected in the graph or not, but also how many passengers are transferring between these two nodes. To incorporate passenger transfer flows or transfer waiting time in the community detection, a_{ij} reflects the weighted element of the adjacency matrix as computed by **Eq.6** and **Eq.7**. The modularity function is adjusted using these weighted links as shown by **Eq.10**, where $w_i = \sum_j a_{ij}$ denotes the sum of the weights of the edges attached to node i (Newman, 2004).

$$q = \frac{1}{2|E|}\sum_{ij}(a_{ij} - \frac{w_i w_j}{2|E|})\delta_{c_i,c_j} \tag{10}$$

The modularity measures essentially how different the original graph is from a randomised graph. The Louvain method is adopted because it has been recognised as one of the best-performing clustering algorithms after a comparative evaluation (Lancichinetti and Fortunato, 2009). The Louvain method has several advantages. First, the algorithm is intuitive and easy to implement. Second, the outcome is unsupervised and computationally light, which requires the link label matrix as the only input. The essence of this method is a greedy optimisation of q in a hierarchical manner. It assigns each node to the community of their neighbours that can yield the largest q, and thus creates a smaller weighted super-network whose nodes are the clusters already found. Therefore, partitions found on this super-network consist of clusters that contain previous ones as well, resulting in a higher hierarchical level of clustering. This procedure is not stopped until the largest possible modularity value is reached. A visualisation of the steps of this algorithm is presented in **Figure 7.4**.

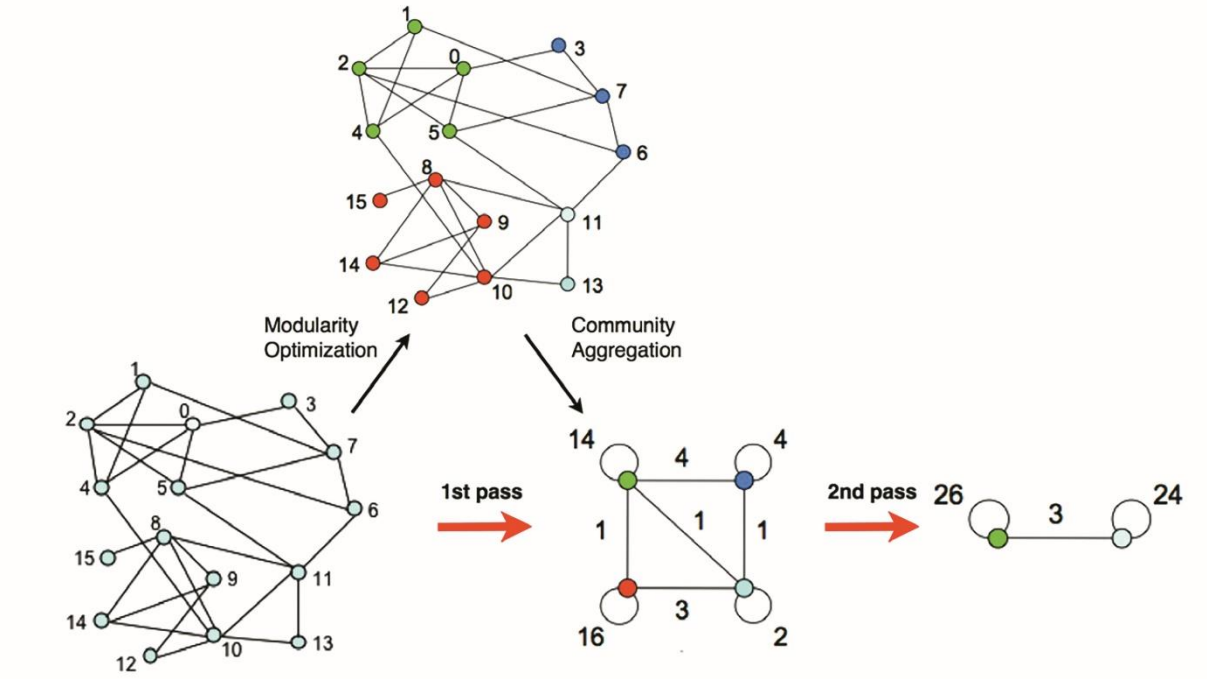


Figure 7.4. Visualisation of the steps of the Louvain method

Each pass consists of two phases. In the first phase, the modularity is optimised by allowing only local changes of communities; in the second one, the communities found are aggregated in order to build a new network of communities. The passes are repeated iteratively until no increase in modularity is possible (Blondel et al., 2008).

7.3 Case Study

We apply our methodology to the urban PT network of the city of The Hague, the Netherlands, operated by HTM (**Figure 7.5**). The Hague has more than 500,000 inhabitants and is one of the main cities of the so-called Randstad, the most important economic area in the western part of the Netherlands. The urban PT network of The Hague consists of 12 tram lines and 10 urban bus lines at the time of consideration (November 2015). We only consider the urban PT network, meaning that services on the (inter)regional train network level and regional bus services are not incorporated. On an average working day, there are more than 300,000 AFC transactions within the case study network.

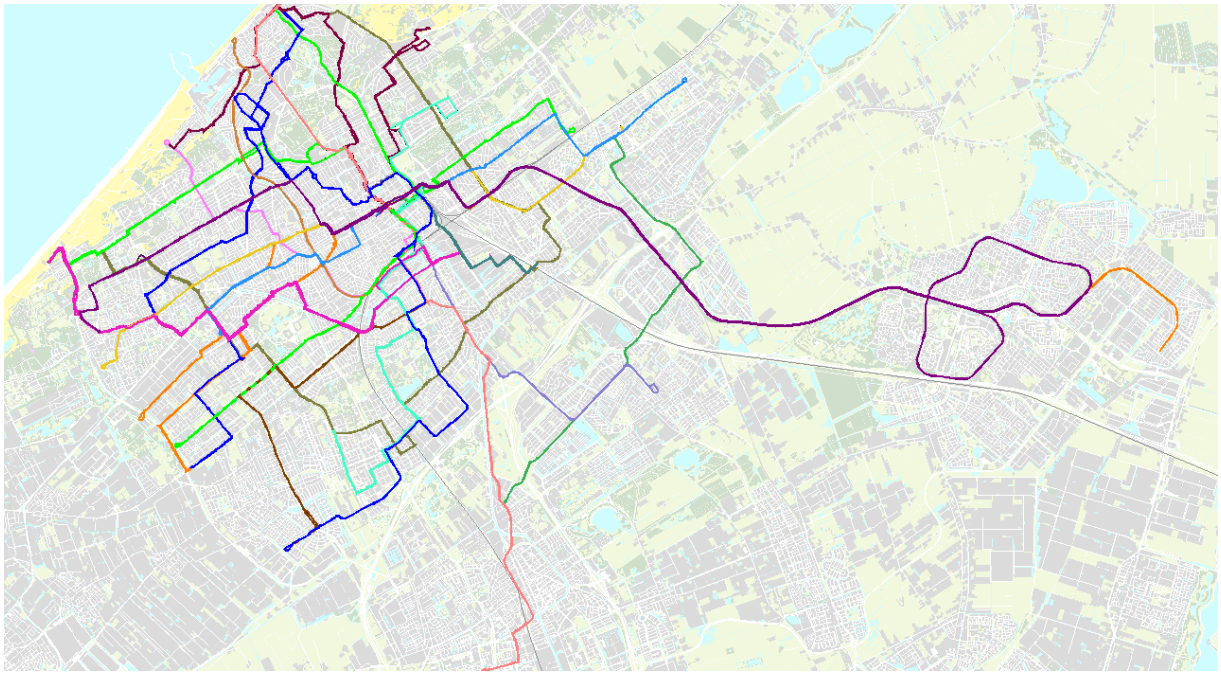


Figure 7.5. Overview of urban tram and bus services of the case study network in The Hague

As initial input, we use all AFC transactions on the case study network for all 20 working days between November 2nd and November 29th, 2015, together with all AVL transactions for this period. For identification of both hubs and line bundles within hubs, we only incorporated AFC data for the AM peak and PM peak of working days in which no large disruptions occurred during these time periods. Since disruptions can result in adjusted passenger route choice including transfer line choice, this can introduce bias when determining the key hubs and line bundles. Based on disruption log-data provided by the operator of this network, we removed data from 10 working days. Our resulting dataset thus contains AFC data from 10 working days of the abovementioned period (2 Mondays, 1 Tuesday, 3 Wednesdays, 2 Thursdays and 2 Fridays). No destination inference is needed for our case study, given that passengers are required to tap in and tap out using devices located within urban trams and buses in The Hague. Incomplete AFC records (1.3%) and AFC records where system errors occurred (<0.4%) have been removed. The resulting dataset contains a total of 3.04 million AFC transactions to which we applied our transfer inference algorithm. Since we only focus on AM and PM peak journeys, transactions part of journeys with a starting time outside the interval 07:00-09:00 or 16:00-18:00 have been removed after this step from our dataset. This resulted in 1.1 million AFC transactions remaining for our analysis. The steps we applied in the data cleaning process are summarised in **Table 7.4**.

Table 7.4. Data cleaning process

Data cleaning process	transactions	% transactions
Initial AFC transactions for 10 working days	3,086,453	100%
Incomplete AFC transactions (missing tap out)	-40,195	-1.30%
System error transactions	-11,162	-0.36%
AFC transactions part of journey started outside AM or PM	-1,930,711	-62.6%
Complete AFC transactions part of journey started in AM or PM	1,104,385	35.8%

In phase 1 of our proposed methodology, we identify the set of hubs H and their spatial demarcation for our case study network. For this phase, we use the combined AM and PM transfer flows over the 10 working days, since the definition of a hub is considered independent from the period of the day. In phase 2 of our methodology, we determine line bundles to prioritise in synchronisation for each identified hub. Since passenger (transfer) flows can differ substantially over the day, synchronisation priorities might differ as well. Therefore, we apply the second phase of our methodology for each time period separately. We compare the clustering results in phase 2 when using transfer flow or transfer waiting time as link weights, which results in four cases per hub. **Table 7.5** shows an overview of all cases considered in phase 2 of our study.

Table 7.5. Overview of cases in study phase 2 for line bundle identification

	Link weight: passenger transfer flow		**Link weight: passenger transfer waiting time**	
Hub	*AM*	*PM*	*AM*	*PM*
h_1	Case 1A	Case 1B	Case 1C	Case 1D
h_2	Case 2A	Case 2B	Case 2C	Case 2D
$h_{\|n\|}$	Case nA	Case nB	Case nC	Case nD

7.4 Results and Discussion

This section discusses the results and implications of the hub identification phase (**Section 7.4.1**) and line bundle identification phase (**Section 7.4.2**) of our proposed methodology.

7.4.1 Hub identification

We applied the proposed transfer inference algorithm to all 1,104,385 AFC transactions using the Euclidean distance between stops to compute transfer walking distances, the 2.5th percentile walking speed ϑ from a normal distributed walking speed function $N(1.34,0.34)$ based on Hänseler et al. (2016), and γ_{max} of 400 Euclidean metres. The value for the maximum transfer walking distance threshold γ_{max} is obtained from Yap et al. (2017), where this showed to be the optimal value when validating the destination inference algorithm applied in that study. This results in 150,792 alighting transactions which are considered a transfer. The sum of the obtained stop-to-stop transfer flow matrix for the AM and PM period together for the considered 10 working days thus equals 150,792. We performed a sensitivity analysis with respect to γ_{max}. Increasing this value by 50% (600 Euclidean metres) and 100% (800 Euclidean metres) reduces the number of identified separate journeys by merely 1% and 2%, respectively. We thus conclude that our results are robust to different values of γ_{max}. While using on-street distances rather than Euclidean distances can further improve the accuracy of the transfer inference algorithm, it might require use of data sources and software which are not always easily accessible. Using on-street distances might be particularly relevant if the methodology would be applied to a network with large variations in street lay-out, such as an application which considers both urban and inter-urban PT networks rather than urban PT networks only.

All 150,792 classified transfers occur between 754 different stop id's, resulting in a 754×754 transfer flow matrix. We performed DBSCAN to identify the geographical boundaries of clusters of transfer locations and set θ equal to 1 (see **Section 7.2.1**). The θ-distance plot is shown as function of average distance $dist$ (**Figure 7.6**) and as function of the number of identified clusters (**Figure 7.7**). These functions are used to determine the optimal parameter value of ε, which reflects a minimum requirement for the transfer flow between stops that form a cluster. Since data from 10 working days are included, we use a step size of 50 for values of ε in the 1-distance plot to maintain a meaningful interpretation of this value (using daily integer transfer flows with step size 5, starting with $\varepsilon = \max_{s_i s_j \in S} g_{s_i s_j} = 5516$). As can be observed from **Figure 7.6**, the shape of the graphs is opposite of the regular shape of the θ-distance plot when applying DBSCAN. This is because in our application a higher transfer flow between stops increases the probability of being clustered together, in contrast to traditional distance measures where a larger value entails a smaller clustering probability. In both **Figure 7.6** and **Figure 7.7,** the knee in the graph can be observed for $\varepsilon = 5166$ and 23 identified clusters of transfer locations. From all 754 transfer stops in the case study network, 11% (81 stops) is part of a cluster of at least two stops. The average cluster size equals 3.54; the median cluster size is 3. The largest cluster consists of 11 stops. The other 673 transfer stops are not clustered by DBSCAN and form a transfer location consisting of one stop only. This means that a total of 696 transfer locations with their geographical boundaries are identified for the case study network. **Figure 7.8** presents all identified transfer locations, highlighting the 23 clusters.

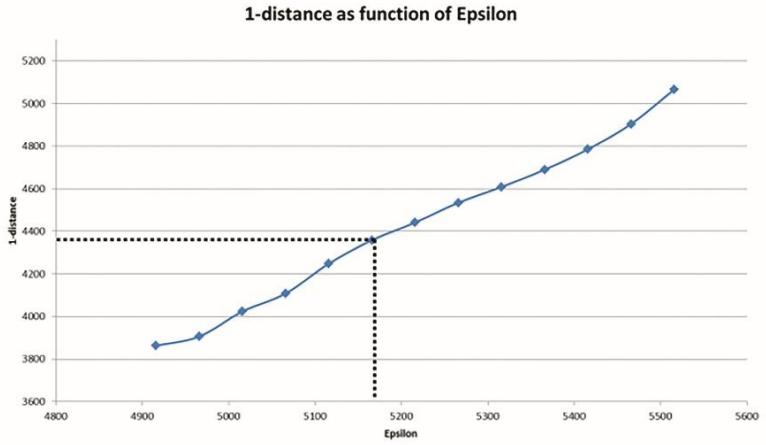


Figure 7.6. 1-distance plot as function of ε

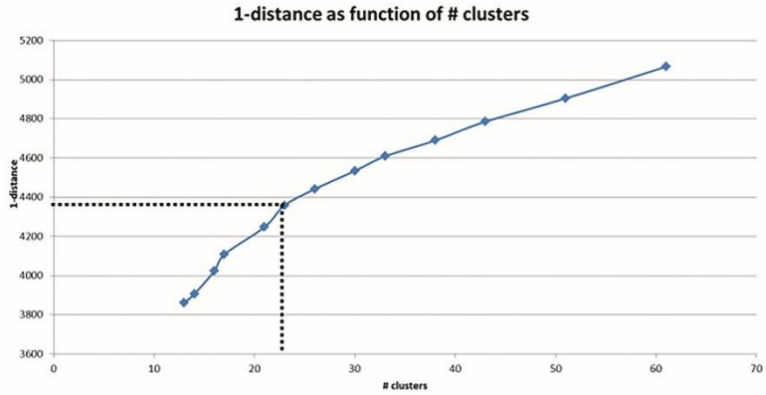


Figure 7.7. 1-distance plot as function of the number of identified clusters

After applying the HHI, six transfer locations with the highest number of intra-cluster transfer flows are identified as hub: Central Station (hub 1), Centre (hub 2), Station Hollands Spoor (hub 3), Leidschenveen (hub 4), Leyenburg hospital (hub 5) and Brouwersgracht (hub 6). The identified hubs 1, 2, 3 and 5 are locations where a large number of tram and bus lines intersect near a train station (1 and 3), the city centre (2) and a hospital (5). Hubs 4 and 6 are served by fewer lines: these hubs are mainly characterised by large transfer flows between a corridor of high-frequency tram lines and one intersecting tram line (4) or bus line (6). The stops constituting these hubs are shown in **Figure 7.8**. In the box of **Figure 7.8**, hub 3 (Station Hollands Spoor) is shown in greater detail. This hub illustrates a key difference between the stops which would be considered as one hub purely based on the geographical location or public name of the stop, and the stops found to constitute a hub based on passenger flows. At the north-side of this station, there are several tram stops located relatively close to each other (star-shaped blue nodes indicated by 'A'), whereas a few bus stops of this station are located at the south-side of the station, about 3 minutes walking from each other (star-shaped blue nodes indicated by 'B'). Besides, there are stops belonging to another public stop name 'Rijswijkseplein', about 5 minutes walking from the tram stops of the station itself. Our clustering results show that - from a passenger perspective - some of the stops of Rijswijkseplein are part of one large hub (star-shaped blue nodes indicated by 'C'), whereas they would be considered a separate transfer location if clustered based on geographical location or public stop name. Some other stops of Rijswijkseplein form a separate transfer location but are not part of hub 3 (circular red nodes indicated by 'D').

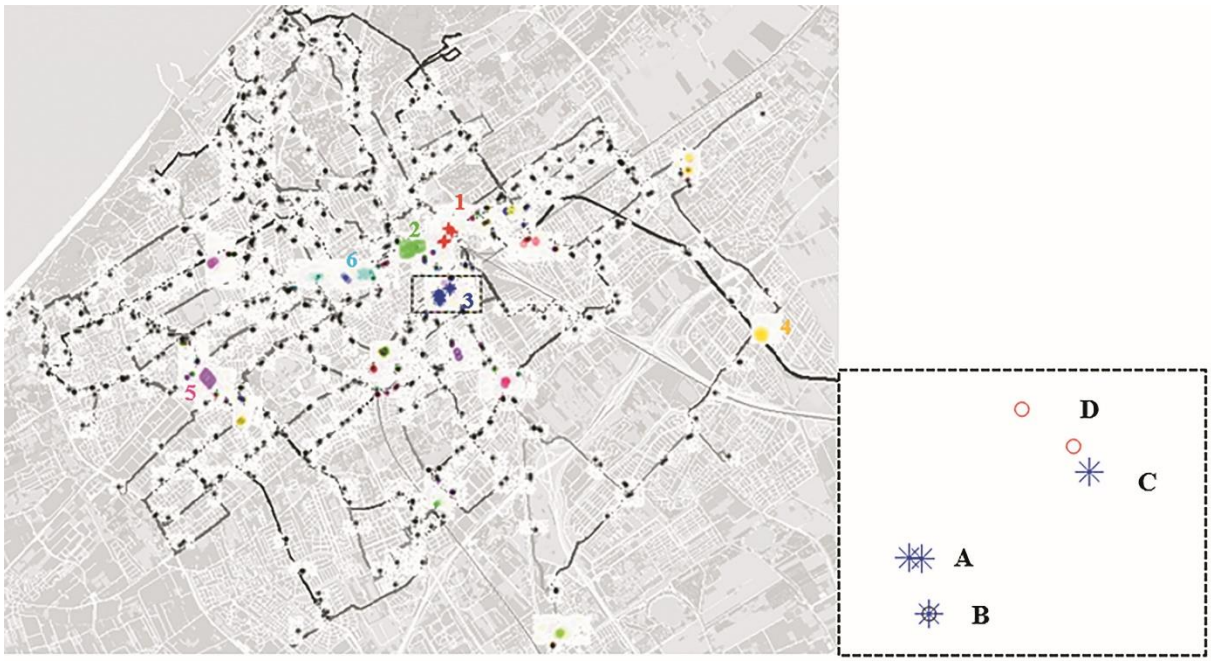


Figure 7.8. Identified transfer locations with 23 transfer clusters (all groups of coloured stops) and the six identified hubs (coloured stops indicated by numbers 1-6)
The box on the right-hand side of the figure zooms into the stops around hub 3 (station Hollands Spoor): blue stops form one hub; red stops (with public name 'Rijswijkseplein') form another transfer location and are not part of the hub.

Figure 7.9 (left) shows the cumulative distribution function (CDF) of all transferring passengers k_t for the 696 transfer locations of the case study network. In **Figure 7.9** (right) the Lorenz curve is plotted, where the realised transfer flow distribution over the transfer locations is contrasted with the hypothetical scenario if transfers would be equally distributed over all

transfer locations. It can clearly be observed that transfer patterns are not uniformly distributed over all locations. In contrast, a sharp concentration of transfers at a few locations can be observed. The calculated Gini-coefficient of 95.7% from **Figure 7.9** (right) confirms the unequal spatial distribution of passenger transfers over the urban PT network. The total intra-hub transfer flows, summed over the six identified hubs, represent 70.1% of the transfers k_t within all transfer locations, whereas 78.4% of all network transfers are within one of the 696 transfer locations. Our results thus show that more than 70% of the transfers within all transfer locations can be captured in the optimisation, while only 6 out of 696 (0.9%) identified transfer locations need to be incorporated. Given our aim not to be exhaustive, these results show that the complexity of solving the TSP can be reduced substantially, against relatively limited costs. As can be seen from **Figure 7.9** (left), considering the top-6 transfer locations is an effective means to capture a large part of the total network transfer flow. Given the unequal spatial transfer distribution, the efficiency of each additional transfer location added to the optimisation problem will decrease: the complexity of the TSP increases, whereas the number of additional captured transfers only decreases compared to the previous ranked transfer location.

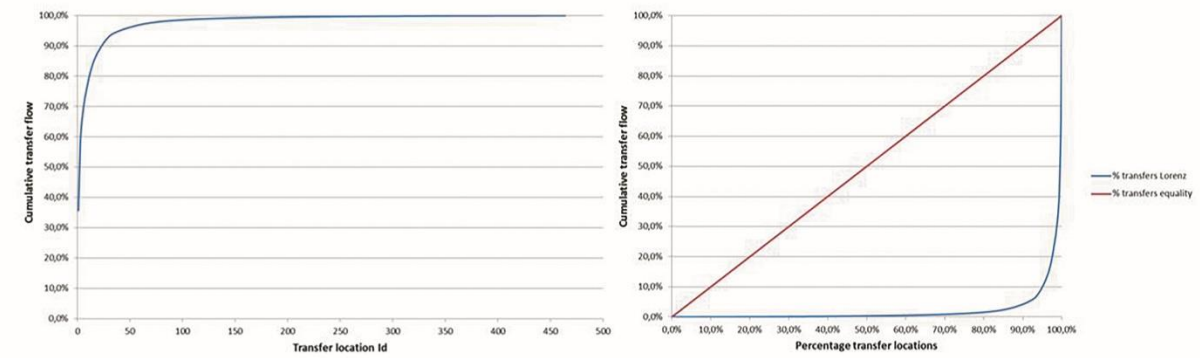


Figure 7.9. CDF (left) and Lorenz curve (right) for transfer flow distribution over the identified transfer locations

100% equals the total within-cluster transfer flow; the figure does not consider between-cluster transfer flows.

7.4.2 Line bundle identification

Table 7.6 summarises the statistics for the line bundle identification phase of our methodology. Given that six hubs were identified, we performed the line bundle identification for 6*4 cases in total: for the AM and PM peak, using transfer flow and transfer waiting as link weight (see **Table 7.5**). **Table 7.6** shows the number of unidirectional lines, the modularity value q resulting from this community detection technique (**Eq.10**), the number of line bundles (communities) identified, and for each hub the within-community transfer flow as share of the total hub transfer flow. The latter is used as additional indicator for the clustering performance.

The modularity value remains at a similar level for the hubs Central Station (1), Centre (2) and Leidschenveen (4), regardless the time period or link weight used. For the hubs Hollands Spoor (3), Leyenburg (5) and Brouwersgracht (6), the modularity is relatively insensitive to the link weight used (transfer flow or transfer waiting time). This can be explained by the relatively high and similar frequency of most urban PT services for the considered hubs. This makes the use of transfer waiting time less distinctive from the use of transfer flow as link weight. However, for these hubs 3, 5 and 6, substantial differences in modularity can be observed between AM and PM, indicating different transfer patterns between these time periods. For the majority of the scenarios two different line bundles are identified. Only for the hub Centre (AM, based on transfer waiting time link weight) and Leyenburg (PM), three line bundles are identified. The percentage within-community transfer flow ranges between 54% and 95% over

all scenarios. For most hubs, this percentage is rather stable when a different time period or link weight is used. Similar to the modularity value, this percentage shows to be more sensitive to the time period than the assigned link weight, particularly for hub 5 (Leyenburg).

Table 7.6. Summary statistics community detection technique

Hub	*Central Station (1)*	*Centre (2)*	*Station Hollands Spoor (3)*	*Leidschenveen (4)*	*Leyenburg (5)*	*Brouwersgracht (6)*
# lines	28	21	16	6	10	8
Modularity						
Flow AM	0.28	0.27	0.13	0.04	0.17	0.12
Flow PM	0.29	0.22	0.28	0.04	0.25	0.40
Waiting AM	0.29	0.25	0.13	0.06	0.13	0.16
Waiting PM	0.28	0.21	0.30	0.06	0.26	0.38
# line bundles						
Flow AM	2	2	2	2	2	2
Flow PM	2	2	2	2	3	2
Waiting AM	2	3	2	2	2	2
Waiting PM	2	2	2	2	3	2
share within-community transfer flow / total hub transfer flow						
Flow AM	0.85	0.77	0.83	0.59	0.75	0.95
Flow PM	0.82	0.77	0.79	0.54	0.59	0.93
Waiting AM	0.86	0.61	0.82	0.58	0.75	0.95
Waiting PM	0.82	0.75	0.80	0.57	0.60	0.94

In **Figure 7.10,** the results of the community detection algorithm are visualised for all 24 cases. Each plot shows the lines with their corresponding direction (north-, south-, east- or westbound) which are grouped together as one line bundle (indicated by the same colour). The link width represents the magnitude in terms of transfer flow or transfer waiting time. These case study results display which bundles of lines should be prioritised simultaneously when devising tactical and real-time synchronisation measures. Generally, the results show to be intuitive with lines heading in the same direction(s) being grouped together, while producing clusters that could not be formed merely based on grouping each direction. Also in line with expected travel patterns, lines heading in opposite directions (e.g. west- and eastbound lines, or north- and southbound lines) are generally not grouped together. For Central Station (hub 1), one line bundle clearly reflects south-/west-bound passenger journeys, whereas the other bundle reflects north/east-bound journeys. For hubs Centre and Station Hollands Spoor (2 and 3), there is a dominance of northbound and westbound lines being grouped together, and southbound and eastbound lines grouped together. For Leyenburg (hub 5) in the AM, two separate line bundles with eastbound and westbound lines can be detected. During the PM, it can be seen that the westbound lines are grouped into two separate clusters. Probably due to a different mixture of passengers and their corresponding trip purpose and travel patterns, a separate line bundle can be detected between bus lines 23 and 26 in the westbound direction during the PM. At Leidschenveen (hub 4), there is a dominant transfer flow between intersecting tram line 19 southbound and particularly tram line 4 westbound in the AM. However, during the PM a large transfer flow can also be observed between line 19 southbound and line 4 eastbound. At Brouwersgracht (hub 6) is a clear transfer flow between eastbound bus line 25 - from a residential area headed for the city centre - and the tram corridor served by lines 2, 3 and 4 towards the main train station. During the PM an opposite pattern can be observed, with a clear line bundle consisting of tram lines 2, 3 and 4 and westbound bus line 25 bound for a residential area of the city.

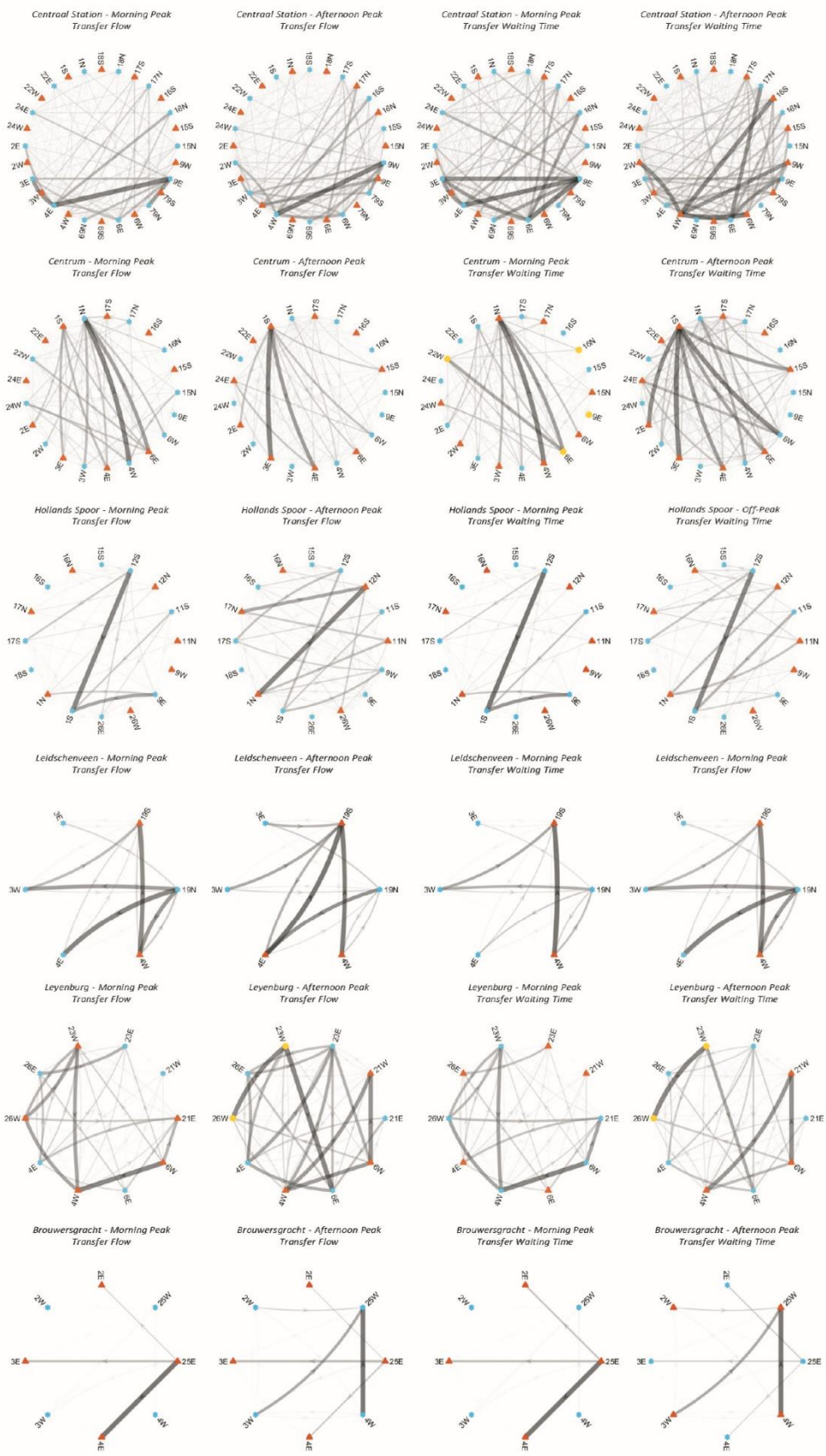


Figure 7.10. Results line bundle identification for six identified hubs

7.5 Conclusions

In this research, we developed a data-driven, generic and passenger-oriented methodology for systematically determining where in the network, and for which lines transfer synchronisation should be prioritised in the TSP, so that the TSP becomes solvable for larger, real-world urban PT networks. Our study thus introduces two steps preceding solving the TSP: identify key priorities (a) where to synchronise, and (b) which lines to synchronise. To this end, our method identifies hubs and their spatial boundaries in urban public transport networks, and determines line bundles within these hubs to be prioritised simultaneously when devising transfer synchronisation measures at either tactical or operational planning phases. The proposed non-supervised learning techniques enable the identification of hubs and line bundles based on passenger transfer flows, independent from local knowledge or the geographic location of the urban public transport stops. Our results show that hubs can be composed of a different set of stops when applying DBSCAN clustering, compared to the set which would result when clustered purely based on geographical information or public stop name. Our clustering results shape the spatial boundaries of public transport transfer locations as used and experienced by passengers. The application of a modularity based community detection technique shows intuitive lines being grouped together to prioritise during transfer synchronisation. Our results illustrate the necessity of synchronising different line bundles during different periods of the day, depending on the travel patterns prevailing during the relevant time period. The clustering results for our case study show to be relatively insensitive to the use of passenger transfer flows or transfer waiting times as link weight, which can be explained by the relatively similar headways associated with all urban PT lines serving a certain hub. If lines with more varying headways would be clustered, a higher sensitivity of clustering results to the used link weight can be expected. Our methodology and study results support public transport operators in timetable design and real-time control, such as holding, by determining where and which lines to synchronise. Moreover, public transport agencies can use these study results to determine where to invest in measures for improving the design of a seamless transfer experience (e.g. amenities, physical environment, island vs. side platforms). Contrary to simply prioritising pairs of lines with the largest transfer flow between them in the synchronisation process, our partitioning approach yields different bundles of lines which should be synchronised simultaneously.

Our approach is able to capture more than 70% of all transfers within identified transfer locations, while only requiring 0.9% of these transfer locations, thus reducing the complexity of solving the TSP substantially at a relatively low cost. In a next step, the optimisation process to solve the TSP can be applied to this subset of transfer locations and lines from the total considered real-world PT network. As input for the TSP at the tactical planning phase, our method ranks the transfer locations and the lines at these locations to be prioritised based on passenger flow data. Depending on the network characteristics, the number of lines serving the different transfer locations, the used optimisation method and accepted computation time, a PT operator can select the $|T|$ most important transfer locations with corresponding line bundles to incorporate in the TSP, so that the TSP becomes solvable within acceptable computation time. The number of transfer locations which can be considered simultaneously may have to be constrained in the TSP in order to make the TSP solvable depending on the approach used. When adopting an approach which relaxes the synchronisation to allow for a pre-defined time window (Ibarra-Rojas and Rios-Solis, 2012) or minimises the total passenger transfer time without pre-setting synchronisation requirements (Knoppers and Muller, 1995), the number of transfer locations of interest does not have to be constrained for mid-size urban networks. In case of real-time transfer synchronisation decisions in response to an early or late arrival of a PT service of line l_i at a certain transfer location t_i, our method proposes for which lines -

clustered within the same line bundle - transfer synchronisation should be considered. In a next stage, the optimal holding time decision for different lines can be taken by minimising the predicted additional travel time for all affected passenger segments, such as transferring passenger, downstream waiting passengers, and downstream transferring passengers (see Gavriilidou and Cats, 2019). In this control framework, the predicted impact of synchronisation at considered transfer location t_i on potentially missed transfers at downstream transfer locations $t_{j\neq i}$ can be incorporated.

We formulate four recommendations for further research. First, we recommend coupling the optimisation process to an assignment model or variable demand model, particularly for networks with relatively many low-frequent services. Since passenger demand depends on the quality of the public transport supply, the results from the transfer synchronisation following the identified synchronisation priorities may influence passenger route choice and, possibly, mode choice. This can result in changes in passenger transfer flows, which in turn can re-set the synchronisation priorities. Especially if PT frequencies are relatively low, substantial changes in transfer flows may result from the synchronisation process. In particular for PT networks with lower frequencies, it is therefore recommended to couple an assignment model or variable demand model to the optimisation process into an iterative supply setting - demand forecasting approach. Second, we recommend experimenting with different clustering techniques in the hub identification phase of our proposed methodology. We used DBSCAN as a density based, partial clustering technique without a pre-defined number of clusters, which requires two different input parameters, namely ε and θ. Contrary to θ, which can be obtained from the context of the application, ε needs to be determined from the θ-distance plot by applying DBSCAN for a large number of instances. Therefore, we recommend testing and comparing the use of other techniques such as OPTICS, in which no distance parameter ε needs to be specified explicitly, thus potentially attaining computational gains (Tan et al., 2004). Third, we recommend extending the line bundle identification phase in our study by applying a link based clustering technique rather than node based clustering. In our modularity based community detection technique, the nodes - i.e. lines in a certain direction - are clustered. However, when the transfer links between nodes would be clustered, one would be able to distinguish between transfer flows from line l_i in direction a to l_j in direction b, and flows from l_j in direction b to l_i in direction a. Incorporating the transfer direction between two lines, next to the lines itself, enables deriving further recommendations for timetable planning and real-time coordination by specifying the desired sequence of arrivals. Fourth, further developments may examine how properties of the optimisation process, such as choices related to the type of optimisation method and type of graph representation, can assist the settings of our methodology.

8. Quantification and Control of Disruption Propagation in Multi-level Public Transport Networks

This chapter forms the second component of Part III of this research, aiming to reduce disruption impacts for passengers on the urban public transport network level. In the previous **Chapter 7**, our objective was to enable synchronisation being applied to urban public transport services, to mitigate urban network disruption impacts. In this chapter**,** we move from an urban network level perspective towards a multi-level network perspective. This chapter puts the emphasis on the urban network impacts of disruptions originating at another level of the multi-level public transport network. We develop a methodology to predict how the impact of a train network disruption propagates to the urban network, using a combination of a railway optimisation model and a simulation-based public transport assignment model. We study how we can develop and evaluate different real-time control strategies applied at the *train network level*, where the disruption originates, which reduce disruption impact propagation to the *urban network level*. This contributes to answering Research Question 3 as defined in **Section 1.3**: how can we predict and control the direct and propagated impacts of disruptions on the urban public transport network in a multi-level network environment?

This chapter is based on an edited version of the following articles:

Yap, M.D., Cats, O., Törnquist Krasemann, J., Van Oort, N., Hoogendoorn, S.P. (under review). Quantification and control of disruption propagation in multi-level public transport networks.

Yap, M.D., Cats, O., Törnquist Krasemann, J., Van Oort, N., Hoogendoorn, S.P. (2020). Quantification and control of disruption propagation in multi-level public transport networks. Presented at the 99th Annual Meeting of the Transportation Research Board (TRB), Washington, DC.

8.1 Introduction

8.1.1 Study relevance

Quantifying and minimising the impacts of public transport (PT) disruptions is important from the perspectives of both service users and service providers. PT disruptions can negatively affect passengers' nominal and perceived journey time as a result of longer in-vehicle times, additional transfers, and longer waiting times in case of missed connections. More severe crowding levels on remaining services also increase perceived in-vehicle times and can potentially result in an increase in the number of passengers being denied boarding (see for example Hörcher et al., 2017, Tirachini et al., 2017 and Yap et al., 2018a). Over a longer time horizon PT disruptions can influence the mode choice of travellers, reducing the PT share in the modal split and reducing revenues for the PT service provider (Yap et al., 2018b). For a PT operator to provide an attractive and a competitive public transport service to passengers, it is thus of utmost importance to understand and limit the impacts of PT disruptions.

The integrated PT network consists of different functional network levels - such as the (inter)national train network level, the regional train network level, and the urban tram and bus network level - which are hierarchically connected to each other. In this study, we use the term *multi-level network* to refer to the entire PT network consisting of these different network levels. As disruption impacts can spill-over from one network level to another network level, it is important to quantify and mitigate the disruption impacts for the total multi-level PT network, instead of limiting the considerations to the disruption impact for the PT network level where this particular disruption occurs. The impact of a PT disruption on a certain network level can propagate to another network level in two different ways: via primary and secondary effects. First, a primary effect relates to the direct impact of a disruption on journeys of passengers who travel over the different network levels during one journey. For example, a passenger travelling on the regional train network level might miss the scheduled connection to the urban tram network level, due to a delayed train arrival at the transfer stop following a disruption on this train network level. Second, a secondary effect is experienced by passengers travelling on a lower PT network level who are affected indirectly by a disruption on a higher network level. For example, a disruption on the regional train network level might result in several delayed trains arriving almost simultaneously at the transfer stop. Consequently, this results in a sudden increase in transfer volume from the regional train network towards the urban network level, thus increasing crowding levels in the first urban trips serving this transfer location. Passengers making a journey only on the urban network level using one of these trips will experience higher levels of discomfort due to crowding, caused by a disruption on another network level. Such secondary effects have been found by Malandri et al. (2018), where impacts of simulated disruptions were observed at segments more than 10-15 km away from the disruption location.

The abovementioned examples illustrate that disruption impacts do not stop at the border of the network level on which the disruption occurs, as often assumed, but can propagate to another network level as well. For a full understanding of the impact of a disruption and how to potentially mitigate its impact, one should therefore consider the impact a disruption has on the multi-level PT network as a whole, including its propagation. Meanwhile, it is also important to consider the role other PT network levels can play in mitigating disruption impacts by increasing network robustness (as studied by Jenelius and Cats, 2015; Yap et al., 2018c).

8.1.2 State-of-the-art and problem definition

Studies focusing on predicting and mitigating PT disruption impacts can broadly be classified as optimisation-based or simulation-based approaches. Several optimisation-based approaches

propose a mathematical programming framework to determine the optimal vehicle holding time to regulate PT services or to synchronise services for transferring passengers. For example, Delgado et al. (2009) and Delgado et al. (2012) test vehicle holding, potentially combined with setting boarding limits, to regulate bus services on a PT corridor with a deterministic mathematical programming model. Sanchez-Martinez et al. (2016) formulate a deterministic holding control model which incorporates dynamic running times and demand. Hadas and Ceder (2010) develop a dynamic programming model which minimises passengers' total travel time to synchronise PT services in an optimal way. Optimisation-based approaches are also commonly used to solve the railway traffic rescheduling problem, in case disruptions occur on the railway network. Selected examples of the extensive research performed in this area are Törnquist Krasemann (2012) proposing a greedy algorithm for train-rescheduling, D'Ariano et al. (2007) using a branch-and-bound algorithm, Corman et al. (2010) testing a tabu search algorithm and Dollevoet et al. (2014) proposing an iterative optimisation framework for delay management and train rescheduling. Binder et al. (2017) propose an integer linear program to solve the multi-objective railway rescheduling problem. For a comprehensive literature overview of algorithms proposed for real-time railway rescheduling, we refer the reader to Cacchiani et al. (2014).

Simulation-based approaches on the other hand are used for disruption impact prediction and for testing rule-based strategies for disruption management. For example, Cats and Jenelius use the dynamic agent-based PT assignment model BusMezzo to quantify the robustness value of spare capacity (Cats and Jenelius, 2015), the value of real-time information provision (Cats and Jenelius, 2014), and the impact of partial link closures (Cats and Jenelius, 2018) for high frequent urban PT networks. Leng et al. (2018) and Paulsen et al. (2018) use MATsim as agent-based simulation software to predict passenger delay impacts from rail disruptions in the metropolitan areas of Zürich and Copenhagen, respectively. Younan and Wilson (2010) develop a rule-based controller to support a real-time holding decision between two connecting bus routes based on expected impact on passengers' net travel time. Daganzo and Anderson (2016) use simulation to test a rule-based holding control strategy for transfer synchronisation between metro and bus, whilst Laskaris et al. (2018) use a simulation-based dynamic PT assignment model to test a multiline holding control strategy for transit corridors, applied to a selection of bus lines in Stockholm, Sweden. Gavriilidou and Cats (2019) propose a rule-based holding controller for urban PT services which considers capacity constraints and on-board crowding levels using a dynamic PT assignment model. For an extensive literature review of holding control strategies we refer to Gavriilidou and Cats (2019).

Few studies have combined the aforementioned approaches using a simulation-based optimisation approach. For example, Shakibayifar et al. (2017) use a simulation-based optimisation model with the objective of minimising total train delay times during train disruptions. Schmaranzer et al. (2019) combine a discrete event simulation model and metaheuristic optimisation model to optimise headways for urban PT systems.

The abovementioned literature review illustrates that optimisation-based approaches are typically applied for disruption management on the train network level; whereas simulation-based approaches are primarily used for predicting disruption impacts and testing disruption management strategies for the urban PT network level or for metropolitan PT networks where it is important to account for passenger flow re-distribution. Optimisation-based approaches, often using microscopic or mesoscopic models, result in (an approximation of) optimal rescheduling, retiming and rerouting of train services in response to a disruption. As the PT rescheduling problem for larger, real-world PT networks is considered NP-hard (Desaulniers and Hickman, 2007), these optimisation-based approaches typically account only for limited stochasticity in PT demand and supply. These studies predominantly employ deterministic

mathematical programming models and generally do not consider stochastic passenger route choice over the PT network, stochastic demand patterns, or stochasticity related to vehicle running times or dwell times. Dynamic interactions between demand and supply, such as bunching, are typically not considered. Simulation-based methods, often using agent-based mesoscopic PT models, are able to capture dynamics in PT demand and supply and their interactions. For example, these models can consider stochastic running times, flow-dependent dwell times and stochastic and dynamic passenger route choice when being confronted with a disruption. These methods allow for testing rule-based control strategies or for testing the impact of several disruption scenarios, albeit without resulting in optimal disruption control strategies.

Railway networks do experience less stochasticity than urban PT networks on average, as train running times are not influenced by mixed traffic operation. Furthermore, the lower network density of train networks reduces the route choice alternatives passengers realistically have. This results in deterministic route choice assumptions being less problematic for train networks, compared to relatively high-density urban PT networks which offer more route redundancy. Moreover, the typically lower train frequencies combined with the prevention of early departures from most train stations do reduce the dynamic interaction between demand and supply which can result in bunching, as often observed for urban PT services. Due to the more complex interaction between PT demand and supply on the urban PT level, simulation-based dynamic assignment models are often necessary for sufficiently realistic predictions of the impact of disruptions and disruption management strategies when considering larger, real-world urban PT networks.

Quantifying and controlling the effects of a train network disruption beyond merely the train network poses two methodological challenges. First, a disruption on the train network level is typically solved by an optimisation-based train rescheduling model resulting in an updated train timetable. The extent to which a train disruption propagates to the urban network is thus a function of the train rescheduling optimisation model. This entails that the train optimisation model needs to be considered, when quantifying propagated disruption impacts to the urban network - typically using a simulation model - adequately. Second, this train rescheduling is based on the characteristics of the train network level only, and does not consider the impact of this rescheduling strategy on disruption propagation to the urban PT network. Passenger trips on the urban level can be subject to control strategies in response to this updated train timetable afterwards. This however implies that urban network control strategies in response to a train network disruption are performed in a sequential way, where first services on the train network level are optimised for this network level only, after which services on the urban network level can only be controlled taken the train network rescheduling as a given. This sequential approach may yield sub-optimal rescheduling solutions, as the disruption impacts are not considered for the integrated multi-level PT network simultaneously. Incorporating the impact of train network disruption management on the urban PT network level however requires considering the stochasticity and dynamics of the urban PT network. These dynamics are difficult to incorporate in an optimisation-based rescheduling model while still maintaining acceptable computation times.

8.1.3 Research contribution

In this study, we develop an integrated methodology to predict the impact of a disruption occurring on the train network for the PT network as a whole by accounting for delay propagation, i.e. cross-network spill-over effects. Our objective is to assess the delay impact alternative train rescheduling strategies have for train passengers on the disrupted network level, along with the disruption propagation impacts to the urban network level (see **Figure**

8.1). Given our research objective, it is necessary to test optimal train rescheduling strategies obtained from an optimisation-based method, and to assess the impact of each strategy on the integrated PT network including the urban level, calling for a simulation-based evaluation approach. We therefore propose a simulation-based optimisation framework to predict disruption impact propagation from the train network to the urban network level. Using our proposed methodology, we test how different train rescheduling strategies can be used to mitigate disruption propagation to the urban network level. In our study, we only control train trips to mitigate disruption propagation: controlling urban PT trips subsequently (e.g. by applying holding control strategies tailored for the disruption conditions) falls outside our research scope. Our study contribution is threefold:

- Development of a methodology to predict disruption impact propagation from train network to urban PT network level.
- Integration of a simulation-based and an optimisation-based approach into one modelling framework to predict disruption propagation impacts.
- Test the impact of different train rescheduling strategies on controlling disruption impacts for the total multi-level PT network.

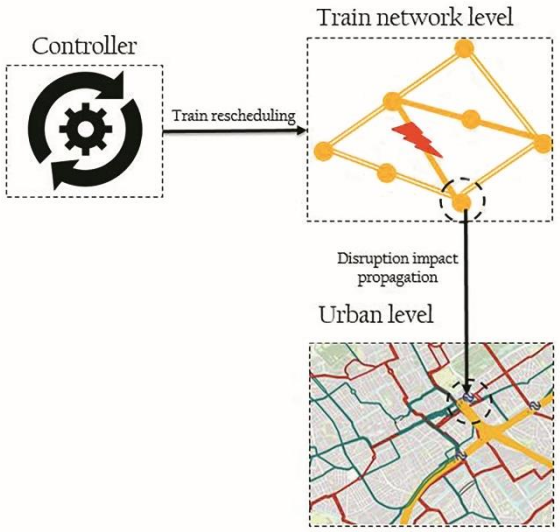


Figure 8.1. Illustration of propagation of train network disruption to urban network

The remainder of this paper is structured as follows. **Section 8.2** discusses our proposed modelling methodology to quantify and control disruption propagation. We apply this methodology to a real-world case study in The Hague, the Netherlands, which is introduced in **Section 8.3**. Results of this case study application are discussed in **Section 8.4**. **Section 8.5** provides conclusions and recommendations for future research directions.

8.2 Methodology

This section discusses our proposed methodology to quantify and control disruption impact propagation over the multi-level PT network. **Section 8.2.1** introduces our proposed modelling framework. **Section 8.2.2** and **Section 8.2.3** provide more details of the dynamic PT assignment model and the train rescheduling model, respectively, that we employ in this modelling framework. **Table 8.1** introduces the notations used in the dynamic PT assignment model, whilst **Table 8.2** lists the notations used in the train rescheduling model.

Table 8.1. List with sets, indices, variables and parameters for the dynamic PT assignment model

Sets and indices	
s, S	public transport stop as node of graph G, set of stops
a, A	edge of graph G, set of links
l, L	unidirectional public transport line, set of lines
k, K	public transport trip, set of trips
o, O	public transport stop representing origin node of G, set of origin nodes
d, D	public transport stop representing destination node of G, set of destination nodes
g, G	passenger route choice action, set of actions
dw	index for dwell time
r	index for running time
s	index for scenario
t	index for regional train network level
u	index for urban public transport network level
ivt	index for in-vehicle time
wkt	index for walking time
wtt	index for waiting time
$wtt - d$	index for waiting time due to denied boarding
$on - board$	index for passengers on-board a public transport trip
$alight$	index for alighting passengers
$board$	index for boarding passengers
tf	index for transferring passengers
arr	index for trip arrival
dep	index for trip departure
Variables	
$\tilde{h}$	scheduled headway of a public transport line
n	number of passengers
t	time
v	generalised passenger journey cost
Parameters	
δ	dwell time coefficient
ε	weights for passenger perception coefficients of travel time components
ζ	threshold convergence criterion 1
η	threshold convergence criterion 2

8.2.1 Modelling framework

The main contribution of this study is the development of a simulation-based optimisation modelling framework as a methodology for predicting the propagation of a train network disruption to the urban PT network. This modelling framework combines two different models: a simulation-based dynamic PT assignment model and an optimisation-based train rescheduling model. Using a dynamic PT assignment model which represents the entire multi-level PT network enables the quantification of the impact of a train network disruption for the PT network as a whole, thereby incorporating the dynamic and stochastic demand and supply characteristics particularly relevant for the urban network level (as mentioned in **Section 8.1.2**). The train rescheduling model is required, as the total disruption impact depends on the train rescheduling strategy applied to train services in response to a disruption, typically derived from an optimisation-based rescheduling model.

We represent the PT network using a directed graph $G(S, A)$ with S being the set of all stops and train stations and A the set of links. The train network level and urban network level are represented by subgraphs $G^t(S^t, A^t)$ and $G^u(S^u, A^u)$ respectively, with $G^t \in G$ and $G^u \in G$. Passenger demand n^{od} is defined from each origin stop $o \in S$ to each destination stop $d \in S$. The integrated modelling framework we propose in this study is shown in **Figure 8.2**. This framework consists of three modelling steps, which needs to be performed to adequately quantify the disruption propagation impact subject to different train rescheduling strategies.

Table 8.2. List with sets and indices, variables and parameters for the train rescheduling model

Sets and indices	
i, T	index for train trip, set of all train trips
j, B	index for rail infrastructure segment of train network, set of segments
k, E	index for time slot request event by train for a rail infrastructure segment, set of events
p, P	index for track for each train infrastructure segment, set
$alight$	index for alighting passengers
tf	index for transferring passengers
$begin$	arrival at segment
end	departure from segment
$initial$	scheduled event time
$static$	event time during disruption
Variables	
b	start time of event
d	minimum segment running time
e	end time of event
$\tilde{h}$	scheduled headway of a public transport line
n	number of passengers
o	point of origin of event
q	binary variable indicating if an event uses a certain track
r	binary variable indicating if an event occurs before another event
s	binary variable indicating if an event is rescheduled to occur after another event
w	train arrival time deviation
x	time at segment
z	train delay
Parameters	
α	weight for penalising track changes
β	weight for alighting passengers in transfer-alighting based delay minimisation strategy
γ	weight for transferring passengers in transfer-alighting based delay minimisation strategy
δ^m	minimum time between two trains driving in opposite direction
δ^f	minimum time between two trains driving in the same direction
μ	track or platform initially intended to be used by an event

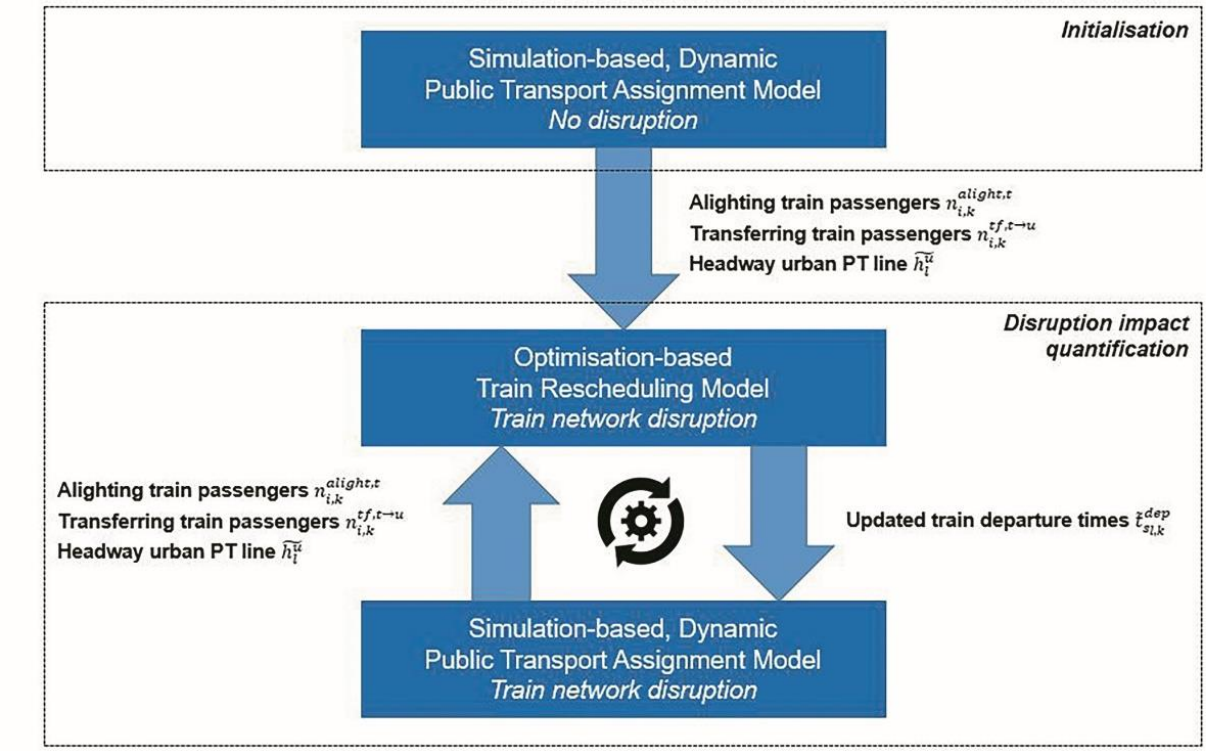


Figure 8.2. Integrated modelling framework

The first step is the model initialisation, where the dynamic PT assignment model (more details are provided in **Section 8.2.2**) is used to assign the total PT passenger demand n (for train and urban network level) over the multi-level PT network for a scenario without disruption s_0. This results in passenger route choice over the total multi-level PT network in case there would be no disruption, yet subject to recurrent service variations. Based on this, passengers' generalised travel costs v_{s_0} in the steady-state condition can be computed by summation of the different travel time components (walking time t^{wkt}, waiting time t^{wtt}, in-vehicle time t^{ivt}, waiting time due to denied boarding t^{wtt-d}, number of transfers n^{tf}) multiplied by their corresponding weights ε and value of time (VoT) (**Eq.1**).

$$v_s^{od} = \left(\varepsilon_{ivt} \cdot t^{ivt,od} + \varepsilon_{wkt} \cdot t^{wkt,od} + \varepsilon_{wtt} \cdot t^{wtt,od} + \varepsilon_{wtt-d} \cdot t^{wtt-d,od} + \varepsilon_{tf} \cdot n^{tf,od}\right) * VoT \quad (1)$$

Second, the train rescheduling model (more details are provided in **Section 8.2.3**) is applied to perform an optimal train rescheduling for a given train network disruption s_i. The train rescheduling model requires the number of alighting passengers from each train trip at each train station $n_{i,k}^{alight,t}$ as input, as well as the number of transferring passengers $n_{i,k}^{tf,t\rightarrow u}$ from the train network level to the urban network level. Depending on the rescheduling strategy applied in this model, the scheduled headway of each urban PT line $\widetilde{h_l^u}$ where passengers transfer to is also required as input. These three variables are outputs from the assignment process of the dynamic PT assignment model and are fed into the train rescheduling model as input for the objective function of a certain train rescheduling strategy. A train network disruption is coded as input for the train rescheduling model, after which the train rescheduling problem is solved for a certain rescheduling strategy. As output, the train rescheduling model provides an updated train timetable with rescheduled train departure times $\tilde{t}_{s_{l,k}}^{dep}$. The scheduled train departure times for each station in the undisrupted case are equal for the train trips modelled in the train rescheduling model and in the dynamic PT assignment model. Due to the difference in granularity between the two models, only updated departure times from commercial train stations are fed back into the dynamic PT assignment model, i.e. updated departure times from other timetable time points such as movable bridges are not fed back from the train rescheduling model to the PT assignment model.

Third, the dynamic PT assignment model is applied again for each rescheduling strategy applied in the train rescheduling model. The updated train departure times from the train rescheduling model in response to the modelled disruption on the train network are used as input. Using this updated train timetable, the total PT demand is re-assigned over the multi-level PT network, based upon which the generalised travel costs v_{s_i} can be computed (**Eq.1**). Due to the stochastic nature of the dynamic PT assignment model, multiple replications of this model are required in both step 1 and step 3 of the model sequence. The generalised travel costs are then averaged over the number of replications.

Next, step 2 and step 3 are repeated iteratively. This is of relevance as the reassigned train passenger flows in the dynamic assignment model can update the number of alighting and transferring train passengers, which can affect the rescheduling results in the optimisation model as a consequence. This iterative process between step 2 and step 3 terminates when convergence is reached. We define two different convergence criteria in this study, of which at least one needs to be satisfied to consider results as converged. The first criterion compares the generalised journey costs v_{s_i} for the total multi-level network between two iterations, as can be computed from the outputs of the PT assignment model (**Eq.2**). When the difference of v_{s_i} - as average over the multiple replications within each iteration - between two subsequent iterations j and $j-1$ is smaller than a predefined threshold ζ, convergence is reached after j iterations.

The second criterion compares the passenger volume assigned for each train trip $i \in T$ on each track segment between two iterations. If at least 95% of the train segment passenger volumes in the model differ by less than a predefined threshold η from the volumes in the previous iteration, convergence is reached (**Eq.3**). We use two convergence criteria in this study, as this relates to the two models used. If the total assignment results between two iterations do not differ more than ζ, the PT assignment model results can be considered stable (first convergence criterion). If the train passenger volume used as input for the train rescheduling model does not differ more than η from the previous iteration, it indicates that the updated train departure times resulting from the train rescheduling model will be stable. As these are used to update the PT assignment model, consequently the results of the assignment model will be stable as well (second convergence criterion). Hence, satisfying one of these criteria is sufficient to consider the model results as converged.

$$\Delta v_{s_i} = \sum_{o \in S} \sum_{d \in S} (v_{s_i}^{od,j} - v_{s_i}^{od,j-1}) \,/\, \sum_{o \in S} \sum_{d \in S} (v_{s_i}^{od,j-1}) \tag{2}$$

$$\Delta n_{i,k}^{on-board,t} = (n_{i,k}^{on-board,t,j} - n_{i,k}^{on-board,t,j-1}) \,/\, n_{i,k}^{on-board,t,j-1} \quad \forall\, i \in T, k \in E \tag{3}$$

Eq.4 quantifies the total passenger disruption impact Δv, expressed as generalised passenger delay costs. The generalised journey costs resulting from disruption s_i after convergence v_{s_i} are compared with these costs when there is no disruption v_{s_0}. We distinguish between journeys with their origin and destination at the train network level or urban network level, which results in four different passenger segments. This enables the quantification of the impact of a train network disruption on this disrupted network level, as well as the spill-over impacts due to propagation to the urban PT network level. The disruption impact of a train network disruption on the disrupted train network level Δv^t relates to the increase in generalised travel costs for passengers starting and terminating their journey at the train network level G^t. The disruption propagation to the urban network level Δv^u relates to the additional generalised journey costs for passengers with their journey starting and/or terminating at the urban network level G^u.

$$\Delta v = \sum_{o \in t,u} \sum_{d \in t,u} (v_{s_i}^{od} - v_{s_0}^{od}) \tag{4}$$

8.2.2 Dynamic PT assignment model

This section discusses the properties of the dynamic PT assignment model employed in this study in more detail. We use a mesoscopic, simulation-based dynamic PT assignment model to represent the multi-level PT network. The train network level, as well as the urban tram and bus network level are represented in this model. In terms of granularity each node corresponds to a PT stop, and each link is the direct connection between two stops $s \in S$. These links typically represent a PT connection between two adjacent stops, whereas they represent a walk connection between stops located close to each other, for example within a single PT hub. We use an agent-based simulation model to mimic the emerging order from interactions among numerous vehicles and passengers. To be able to reflect stochastic demand patterns due to day-to-day variation, the arrival rate of passengers at the origin stop for each OD pair is assumed to follow a Poisson distribution. The arrival rate parameter of the Poisson distribution can typically be estimated from Automated Fare Collection (AFC) data.

PT supply dynamics

The set of PT lines is denoted by L, with $|L|$ representing the total number of lines. Each line $l \in L$ is defined by a sequence of stops $l = \{s_{l,1}, s_{l,2}..s_{l,j}\}$ with $K_l = \{k_{l,1}, k_{l,2}..k_{l,j}\}$ denoting the set of scheduled trips on line $l \in L$. The scheduled headway of a line is denoted by $\tilde{h}_l$, which can be time-dependent. The total time $t_{l,k}$ it takes a vehicle to complete trip k of line l equals the summation of all running times $t^r_{s_l}$ from stop s_l to stop s_{l+1} and dwell times $t^{dw}_{s_l}$ at each stop s_l, as expressed by **Eq.5**.

$$t_{l,k} = \sum_{s=1}^{s-1} t^r_{s_{l,k}} + \sum_{s=1}^{s-1} t^{dw}_{s_{l,k}} \qquad \forall\, k_l \in K_l, l \in L \tag{5}$$

Running times $t^r_{s_{l,k}}$ can be assumed deterministic, using the scheduled times obtained from the timetable, or can be stochastic. In the latter case, typically a lognormal or log-logistic distribution function is best fitted to characterise the empirical running times obtained from Automated Vehicle Location (AVL) data. In our study, we use deterministic running times for the train network, as these running times are relatively stable given the limited interactions with other traffic. For urban tram and bus lines, we fit a lognormal or log-logistic distribution to the empirical AVL data, to capture the predominantly stochastic running times within an urban environment.

The dwell times $t^{dw}_{s_{l,k}}$ for each trip $k_l \in K_l$ at each stop $s \in S$ are dependent on the number of boarding and alighting passengers $n^{board}_{s_{l,k}}$ and $n^{alight}_{s_{l,k}}$. The flow-dependent dwell time function used in this study assumes a linear relation between the number of boarding and alighting passengers and the required dwell time, whilst the model also allows for adding a non-linear effect of on-board crowding on dwell times based on Weidman (1994). As crowding levels for our case study network (see **Section 8.3**) are relatively low, even when subject to the disruption types we consider, we deem using a simple linear function sufficient and beneficial in terms of computation times. For PT networks or disruption types where severe crowding occurs, the use of a dwell time function with non-linear crowding effect is however recommended. In case separate doors of a vehicle are used for boarding and alighting, the dwell time depends on the maximum of the number of boarding and alighting passengers, multiplied by the related dwell time coefficient δ which reflects the required boarding or alighting time per passenger (**Eq.6**). When all doors are used for both boarding and alighting, the dwell time is calculated using **Eq.7**. The dwell time function is calibrated for different vehicle types (e.g. high-floor trams, low-floor trams and buses) by executing a regression analysis predicting the realised dwell times based on the boarding and alighting volumes obtained from AFC and AVL data. A separate dwell time function is calibrated for each vehicle type. If different bus vehicle types (e.g. buses with a different number of doors or with different boarding regimes) would be used for different lines, different coefficients need to be calibrated. Similarly, if a tram line would be operated by longer trams (for example, two coupled tram carriages), separate coefficients need to be estimated for this line due to the different number of total doors available for boarding and alighting.

$$t^{dw}_{s_{l,k}} = \delta_0 + \max\left(\delta_1 \cdot n^{board}_{s_{l,k}}, \delta_2 \cdot n^{alight}_{s_{l,k}}\right) \quad \forall\, k_l \in K_l, l \in L, s \in S \tag{6}$$

$$t^{dw}_{s_{l,k}} = \delta_0 + \delta_1 \cdot n^{board}_{s_{l,k}} + \delta_2 \cdot n^{alight}_{s_{l,k}} \qquad \forall\, k_l \in K_l, l \in L, s \in S \tag{7}$$

The departure time of a trip $t_{s_{l,k}}^{dep}$ depends on the arrival time at that stop $t_{s_{l,k}}^{arr}$ and the required dwell time $t_{s_{l,k}}^{dw}$. In case a stop is a holding point and a schedule-based holding control regime is employed, the departure time can never be earlier than the scheduled departure time from that specific stop $\tilde{t}_{s_{l,k}}^{dep}$. For urban PT networks, a select number of stops are usually holding points, whereas all train network stations are holding points as trains are generally not able to depart ahead of schedule from a station. **Eq.8** shows the departure time calculation for a stop which is not a holding point; **Eq.9** shows the calculation for a holding point stop.

$$t_{s_{l,k}}^{dep} = t_{s_{l,k}}^{arr} + t_{s_{l,k}}^{dw} \qquad \forall\, k_l \in K_l, l \in L, s \in S \tag{8}$$

$$t_{s_{l,k}}^{dep} = \max\left(t_{s_{l,k}}^{arr} + t_{s_{l,k}}^{dw}, \tilde{t}_{s_{l,k}}^{dep}\right) \qquad \forall\, k_l \in K_l, l \in L, s \in S \tag{9}$$

PT demand dynamics

The number of boarding and alighting passengers is obtained from a successive number of choices each individual passenger makes during the journey. At each stop a passenger can make a boarding decision to board a certain trip or to wait, or make a connection decision to walk to another PT stop. When boarded a certain trip, a passenger can make an alighting decision at each downstream stop whether to alight from this vehicle or to stay on-board. These decisions can be made en-route and in a dynamic way if the expected utility of a certain choice changes during a journey, for example in response to high crowding levels or to information provided about a downstream disruption. These successive decisions are based on the expected utility of a path v_g corresponding to a certain action g as logsum over the path set $A_g \in A_{od}$ associated with this action (**Eq.10**). The probability of passenger n choosing this action g is calculated using a multinomial logit (MNL) model (**Eq.11**), which results in stochastic route choice over the network. The structural part of the utility function is calculated based on the sum product of the expected values of the different travel time attributes and the weights of the corresponding coefficients. The model considers in-vehicle time (nominal and perceived in-vehicle time caused by crowding), walking time, waiting time (regular waiting time as well as waiting time caused by denied boarding in case of crowding) and the number of transfers. For different travel time components, different coefficients are used reflecting the perceived time by passengers, as well as a fixed transfer penalty for each transfer. To alleviate potential violations of the IIA assumption of the MNL model, common stops and lines are merged into hyper-paths. A single non-equilibrium assignment procedure without day-to-day learning is applied. A more extensive description of the dynamic PT assignment model proposed for this study is provided by Cats et al. (2016a).

$$v_{n,g} = ln \sum_{a \in A_g} e^{v_{n,a}} \qquad \forall\, n \in N, g \in G \tag{10}$$

$$p_{n,g} = \frac{e^{v_{n,g}}}{\sum_{g \in G} e^{v_{n,g}}} \qquad \forall\, n \in N, g \in G \tag{11}$$

8.2.3 Train rescheduling model

Model formulation

We use a mesoscopic optimisation model for train rescheduling in response to a train network disruption. The model is a modified version of the model proposed by Törnquist and Persson

(2007). This model only represents the train network level $G^t \in G$, which is a subset of the total multi-level PT network G considered in this study. In this model, individual train trips are represented. Each node corresponds to a train station or infrastructure junction, such as a movable bridge or track merging; each separate track between two nodes is represented by an individual link. In this train rescheduling model G^t is represented with a higher granularity than in the dynamic PT assignment model, to allow for optimal rescheduling of each individual train trip in case of a disruption.

Let T represent the set of all train trips in the selected train network level and let B denote the set of segments that defines the rail infrastructure for the train network level. E denotes the set of events, where an event can be seen as a time slot request by a train for a specific network segment. The index i is associated with a specific transport service in the set T (i.e. $i \in T$), while the index j is associated with a specific network segment ($j \in B$), and index k is associated with an event ($k \in E$). An event is associated with a combination of a network segment and a transport service. The set $K_i \subseteq E$ is an ordered set of events for each transport service i, while $L_j \subseteq E$ is an ordered set of events for each network segment j. Each segment j in B has a number of parallel tracks, with each track indicated by $p_j \in P_j$. Each track requires a separation in time between subsequent events (i.e. the minimum time required between one train leaving the track and the next train entering the same track). The latter is reflected by δ_j^m for the minimum time between trains driving in the opposite direction, and by δ_j^f for trains following each other in the same direction.

This train rescheduling model focuses primarily on train delay minimisation but allows for weighting the delay of different trains based on the number of passengers on-board the trains, in order to adopt a more passenger-oriented approach in the train delay minimisation. It should however be noted that passengers and their dynamic route choice are not explicitly modelled here, since incorporating demand- and supply-related stochastics and dynamics of both network levels of a real-world PT network is computationally expensive (as addressed in **Section 8.1**). The objective function of this model in its most basic form is therefore the minimisation of the sum of all delays for all train trips (**Eq.12**). The optimisation is formulated as mixed integer linear programming (MILP) problem. The decision variables reflecting the decisions to be made during the train rescheduling are reflected by **Eq.13-15**.

$$minimise \sum_{i \in T} \sum_{k \in K_i} z_{i,k} \tag{12}$$

$$q_{i,k,p} = \begin{cases} 1, \text{if event } k \text{ uses track } p, k \in K_i, k \in L_j, i \in T, p \in P_j, j \in B \\ 0, \text{otherwise} \end{cases} \tag{13}$$

$$r_{k,\hat{k}} = \begin{cases} 1, \text{if event } k \text{ occurs before event } \hat{k}, k \in L_j, j \in B: \mathrm{k} < \hat{k} \\ 0, \text{otherwise} \end{cases} \tag{14}$$

$$s_{k,\hat{k}} = \begin{cases} 1, \text{if event } k \text{ is rescheduled to occur after event } \hat{k}, k \in L_j, j \in B: \mathrm{k} < \hat{k} \\ 0, \text{otherwise} \end{cases} \tag{15}$$

Constraints

The optimisation is subject to several constraints related to the timing and sequence of events and the capacity and safety limitations of the infrastructure. We introduce the following notations specifically related to the model constraints. The variables x_k^{begin} and x_k^{end} reflect the arrival time at a segment for a specific train, and the departure time from this segment,

respectively. The scheduled start and end time of each event are reflected by $b_k^{initial}$ and $e_k^{initial}$, whilst a disruption is modelled by changing the start and end time of selected events to b_k^{static} and e_k^{static}. The minimum running time of each trip for each segment $d_{i,k}$ is provided as model input. The constraints reflected by **Eq.16-21** are related to train restrictions. Each event of a specific train trip needs to be followed directly by the next event of this trip (**Eq.16**). Events which started before the disruption starts, but are not finished yet when the disruption start, should start as planned (**Eq.17-18**). The duration of each event for a certain segment should at least be equal to the minimum running time required for this segment (**Eq.19**), whilst events are not allowed to start before their original scheduled departure time (**Eq.20**). In **Eq.21**, the train delay exceeding threshold τ minutes is calculated for each event.

$$x_{i,k}^{end} = x_{i,k+1}^{begin}, k \in K_i, i \in T: k \neq |K_i| \quad (16)$$
$$x_{i,k}^{begin} = b_{i,k}^{static}, k \in K_i, i \in T: b_k^{static} > 0 \quad (17)$$
$$x_{i,k}^{end} = e_{i,k}^{static}, k \in K_i, i \in T: e_k^{static} > 0 \quad (18)$$
$$x_{i,k}^{end} \geq x_{i,k}^{begin} + d_{i,k}, k \in K_i, i \in T \quad (19)$$
$$x_{i,k}^{begin} \geq b_{i,k}^{initial}, k \in K_i, i \in T \quad (20)$$
$$x_{i,k}^{end} - e_{i,k}^{initial} - \tau \leq z_{i,k}^{+\tau}, k \in K_i, i \in T \quad (21)$$

The constraints as formulated in **Eq.22-28** concern the permitted interactions between trains, given the capacity limitations of the infrastructure (including safety restrictions). First, each event must use exactly one track per segment (**Eq.22**). **Eq.23-27** make sure that if two events using the same track within a segment, this can only occur if the first event has finished and the minimum required time δ_j^m or δ_j^f has passed (depending whether these subsequent trains are running in the same or opposite direction). o_k refers here to the point of origin of event k, which enables determining whether two subsequent events are using a segment in the same or in opposite direction. M is a large positive constant. **Eq.28** guarantees that an event k cannot be scheduled both before and after event $\hat{k}$. As a last set of constraints, the train rescheduling model allows for the incorporation of preferences of the PT service provider, such as guaranteed connections between specific trips or giving different weights to connections with different levels of importance.

$$\sum_{p \in P_j} q_{i,k,p} = 1, k \in K_i, k \in L_j, i \in T, p \in P_j, j \in B \quad (22)$$
$$q_{i,k,p} + q_{i,\hat{k},p} - 1 \leq r_{k,\hat{k}} + s_{k,\hat{k}},$$
$$k, \hat{k} \in L_j, k \in K_{i,}, \hat{k} \in K_{\hat{\imath},} p \in P_j, j \in B, i, \hat{\imath} \in T: k < \hat{k} \quad (23)$$
$$x_{\widehat{i,k}}^{begin} - x_{i,k}^{end} \geq \delta_j^m r_{k,\hat{k}} - M(1 - r_{k,\hat{k}}),$$
$$k, \hat{k} \in L_j, k \in K_{i,}, \hat{k} \in K_{\hat{\imath},} p \in P_j, j \in B, i, \hat{\imath} \in T: k < \hat{k}, o_{\hat{k}} \neq o_k \quad (24)$$
$$x_{\widehat{i,k}}^{begin} - x_{i,k}^{end} \geq \delta_j^f r_{k,\hat{k}} - M(1 - r_{k,\hat{k}}),$$
$$k, \hat{k} \in L_j, k \in K_{i,}, \hat{k} \in K_{\hat{\imath},} p \in P_j, j \in B, i, \hat{\imath} \in T: k < \hat{k}, o_{\hat{k}} = o_k \quad (25)$$
$$x_{i,k}^{begin} - x_{\widehat{i,k}}^{end} \geq \delta_j^m s_{k,\hat{k}} - M(1 - s_{k,\hat{k}}),$$
$$k, \hat{k} \in L_j, k \in K_{i,}, \hat{k} \in K_{\hat{\imath},} p \in P_j, j \in B, i, \hat{\imath} \in T: k < \hat{k}, o_{\hat{k}} \neq o_k \quad (26)$$
$$x_{i,k}^{begin} - x_{\widehat{i,k}}^{end} \geq \delta_j^f s_{k,\hat{k}} - M(1 - s_{k,\hat{k}}),$$
$$k, \hat{k} \in L_j, k \in K_{i,}, \hat{k} \in K_{\hat{\imath},} p \in P_j, j \in B, i, \hat{\imath} \in T: k < \hat{k}, o_{\hat{k}} = o_k \quad (27)$$
$$r_{k,\hat{k}} + s_{k,\hat{k}} \leq 1, k, \hat{k} \in L_j, j \in B: k < \hat{k} \quad (28)$$

Rescheduling strategies

Although passengers are not explicitly modelled within the train rescheduling model, it is possible to add different weights to the delays of different trains in the objective function based on the number of passengers in each train. This gives, for example, a heavier penalty to a delay of a busy train compared to a delayed train which is less busy. In our study, we test four different train weightings, which are aimed to control the propagation of train disruption impacts to the urban PT network. These four different weights result in four different objective functions applied to the train rescheduling model, taking the objective function as introduced by **Eq.12** as base. We tested the following train rescheduling strategies:

- Passenger based delay minimisation (**Eq.29**) (default): minimise train delays larger than two minutes $z_{i,k}^{+2}$, where each train is weighted by the expected number of passengers leaving each train $n_{i,k}^{alight,t} + n_{i,k}^{tf,t\to u}$ (both alighting and transferring passengers).
- Transfer based delay minimisation (**Eq.30**): minimise train delays larger than two minutes $z_{i,k}^{+2}$, where each train is weighted by the expected number of transferring passengers from the train network level to the urban PT network level $n_{i,k}^{tf,t\to u}$. This implies that trains are only weighted according to the number of transferring passengers to the urban level.
- Transfer-time based delay minimisation (**Eq.31**): minimise train delays larger than two minutes $z_{i,k}^{+2}$, where each train is weighted by the number of transferring passengers to the urban PT network level multiplied with the headway of the urban PT service where is transferred to $n_{i,k}^{tf,t\to u} \cdot \widetilde{h_l^u}$. This reflects the expected passenger waiting time for transferring passengers in case a scheduled transfer from train to urban network level would be missed due to a delayed train arrival at the transfer stop.
- Weighted transfer-alighting based delay minimisation (**Eq.32**): minimise train delays larger than two minutes $z_{i,k}^{+2}$, where each train is weighted based on the number of alighting passengers $n_{i,k}^{alight,t}$ and the number of transferring passengers from train to urban PT network level $n_{i,k}^{tf,t\to u}$, with different weights β and γ respectively being applied to the two passenger segments. This reflects that the impact of a delayed train arrival can potentially be more severe for transferring passengers, when a connection to the urban network level would be missed, than for alighting passengers reaching their final destination. This typically results in a higher weight $\gamma > \beta$ being applied to the number of transferring passengers.

The objective functions corresponding to the different train rescheduling strategies are shown in **Eq. 29-32**. $q_{i,k,p}$ is a binary variable related to the use of rail infrastructure by an event k and is defined in **Eq.13**.

$$minimise \sum_{i\in T}\sum_{k\in K_i}[(n_{i,k}^{alight,t} + n_{i,k}^{tf,t\to u})\cdot z_{i,k}^{+2} + w_{i,k}] \;+\; \sum_{i\in T}\sum_{k\in K_i}\sum\nolimits_{p\in P_j: p\neq \mu_{i,k}^{train}} \alpha\cdot q_{i,k,p} \tag{29}$$

$$minimise \sum_{i\in T}\sum_{k\in K_i}[n_{i,k}^{tf,t\to u}\cdot z_{i,k}^{+2} + w_{i,k}] \;+\; \sum_{i\in T}\sum_{k\in K_i}\sum\nolimits_{p\in P_j: p\neq \mu_{i,k}^{train}} \alpha\cdot q_{i,k,p} \tag{30}$$

$$minimise \sum_{i\in T}\sum_{k\in K_i}[(n_{i,k}^{tf,t\to u}\cdot \widetilde{h_l^u})\cdot z_{i,k}^{+2} + w_{i,k}] \;+\; \sum_{i\in T}\sum_{k\in K_i}\sum\nolimits_{p\in P_j: p\neq \mu_{i,k}^{train}} \alpha\cdot q_{i,k,p} \tag{31}$$

$$minimise \sum_{i \in T} \sum_{k \in K_i} [(\beta \cdot n_{i,k}^{alight,t} + \gamma \cdot n_{i,k}^{tf,t \to u}) \cdot z_{i,k}^{+2} + w_{i,k}] + \sum_{i \in T} \sum_{k \in K_i} \sum_{p \in P_j : p \neq \mu_{i,k}^{train}} \alpha \cdot q_{i,k,p} \quad (32)$$

The following train rescheduling actions are permitted in this model:

- Retiming: changing the departure and arrival times, while respecting the initial earliest departure time and minimum dwell times at commercial stops and running times of the trains.
- Reordering: permitting shift of train order and overtaking to neighbouring stations, while respecting the safety constraints in the network.
- Local rerouting: allowing change of track and platform assignment at a train station.

The model permits for example trains to run faster than scheduled and to run ahead of schedule at certain stretches (i.e. arriving before the scheduled arrival time) in order to enable trains to catch-up from delays and make way for other trains quicker. The model also permits trains to change tracks and platforms at stations as well as to overtake and meet at other locations than initially planned, if that leads to a reduction of knock-on delays. Hence, in order to ensure that only such beneficial 'delay-reducing' re-scheduling decisions are adopted, those need to be associated with a smaller penalty corresponding to e.g. one minute delay. Therefore, in addition to delay minimisation, the objective function also minimises other arrival and departure time deviations and track changes. The objective is to minimise train delays larger than two minutes $z_{i,k}^{+2}$, weighted by a passenger component depending on the rescheduling strategy, arrival time deviations $w_{i,k}$ and track changes $q_{i,k,p}$. The parameter $\mu_{i,k}^{train}$ specifies the track or platform that was initially intended to be used by event k belonging to train i. The parameter α specifies the weight used for penalising track changes. If the time-related variable values are given in e.g. seconds, the value of α needs to be set quite high in order to balance the trade-off between reducing train delays and keeping the timetable intact as much as possible with respect to the planned routes of the trains through/within the stations.

As our model does not incorporate (full or partial) train cancellations, global rerouting (i.e. rerouting trains via a completely different route) or the supply of rail-replacement bus services in the optimisation, this has implications for the type and magnitude of disruptions this model can be applied for. This specific study focuses on disruptions which do not result in the blockage of certain rail infrastructure. For example, one can think of vehicle or infrastructure related disruptions (e.g. a signal failure, a faulty train) which result in delays, but which do not result in the complete unavailability of a certain infrastructure segment. When infrastructure becomes unavailable, train cancellations, short-turning or rerouting trains are measures commonly applied. Additionally, our method focuses primarily on unplanned disruptions with a short to medium-long duration (up to a couple of hours). For planned disruptions and for long-lasting unplanned disruptions - for example a disruption which lasts for multiple days - supply of rail-replacement buses can be expected. In these cases, a wider demand response than only rerouting can be expected as well, as passengers might also change their mode choice, destination choice or trip frequency choice.

8.3 Case Study

This section discusses the case study for which our methodology is applied. **Section 8.3.**1 introduces the case study network of The Hague, the Netherlands. Subsequently, **Section 8.3.2** describes the tested disruption scenario.

8.3.1 Case study network

We apply our methodology to the multi-level public transport network of The Hague, the Netherlands. The Hague is the third largest city in the Netherlands, located in the main economic area of the Netherlands called the Randstad in the western part of the country. The population size of the city is over 500,000 inhabitants. The urban agglomeration of The Hague including its surrounding cities covers an area of 405 sq.km with more than 1 million inhabitants.

The case study multi-level PT network encompasses the complete urban PT network of The Hague consisting of 12 tram lines and 8 bus lines, and all train services calling at The Hague as depicted in **Figure 8.3**. The tram and bus lines are operated by HTM, the urban public transport operator of The Hague. Two tram lines are light rail lines connecting the main city of The Hague with the satellite city of Zoetermeer. The other 10 tram lines function on the urban network level providing connections between different areas within The Hague and neighbouring municipalities. The eight considered bus lines all belong to the urban concession area of HTM in The Hague. The case study network consists of 498 bus, tram and light rail stops. All train services from/to the directions Leiden, Gouda and Delft starting at, terminating at, or serving one of the train stations of The Hague are incorporated in our case study. Both intercity train services, serving only larger cities, and local train services stopping at all stations are simulated. The train network is cordoned at the stations Leiden, Gouda and Delft Zuid, meaning that these stations are modelled as gate nodes for the parts of train services extending beyond the boundaries of the case study network. The cordoned train network consists of 16 stations, of which 10 stations allow passengers to transfer between the (inter)regional train network level and the urban tram and bus network of The Hague.

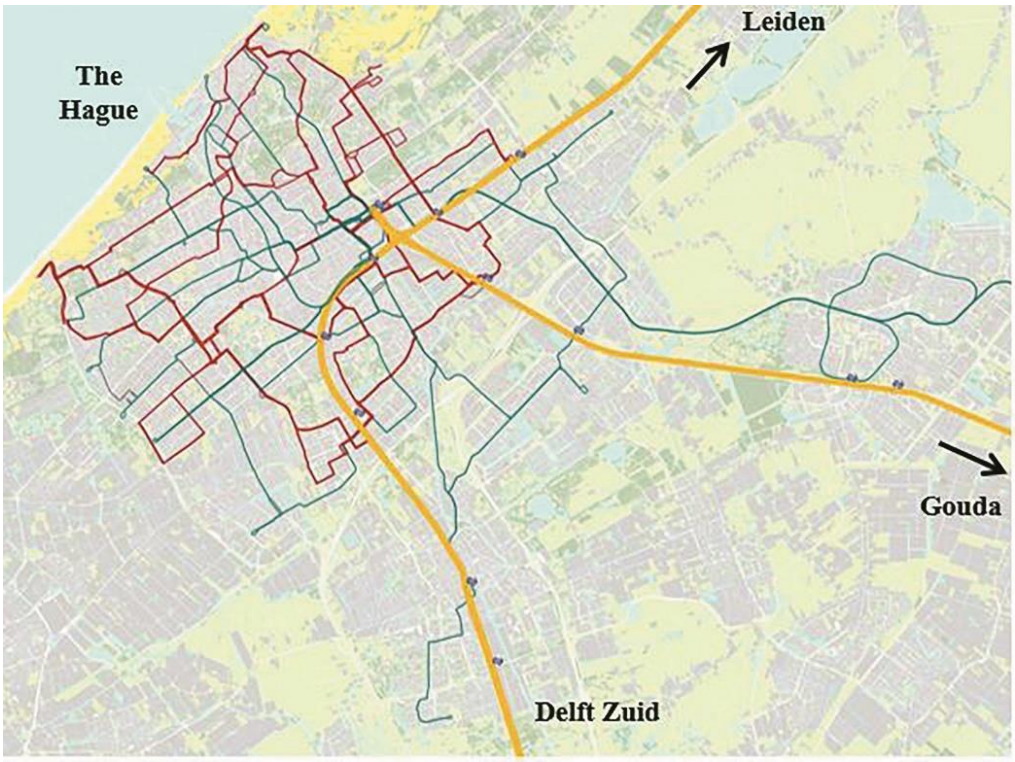


Figure 8.3. Case study public transport network
(yellow: train services / green: tram and light rail services / red: bus services)

Passenger demand is obtained from Automated Fare Collection (AFC) data from 20 working days between 5 March and 30 March 2018. For the urban tram and bus network in The Hague,

a distance based fare system applies where passengers are required to tap in and tap out at in-vehicle devices for each journey leg. This means that each complete AFC transaction consists of a tap in time, stop, line and vehicle ID, as well as a tap out time and stop (see also Van Oort et al., 2017). The dataset consists of 6.48 million AFC transactions solely for the urban tram and bus network, equating ≈325,000 AFC transactions per average working day made on the urban PT network. 29,271 AFC transactions (0.5%) were incomplete due to an error in the AFC system and removed from the dataset. Due to the on-board tap in and tap out devices, destination inference is not required for complete AFC transactions. In case of an incomplete AFC transaction where a passenger (un)deliberately did not tap out, a trip chaining algorithm is applied to infer the most plausible tap out stop (Munizaga and Palma, 2012). If there is only one AFC transaction made by a certain card ID on the day of the incomplete transaction, or if no candidate alighting stop is found within a plausible walking distance of 400 Euclidean meter from the next registered boarding stop, no destination inference is performed. Consequently, another 43,427 (0.7%) AFC transactions were removed from the dataset. For all remaining 6.39 million AFC transactions on the urban PT network, a transfer inference algorithm is applied to construct stop-to-stop journeys based on Gordon et al. (2013) and Yap et al. (2017), thereby using both the AFC and AVL (open) data corresponding to this 20 working days period.

To construct a multi-level stop-to-stop OD matrix, the OD matrix generated solely for the urban PT network is amended based on information about transfers between the train and urban PT network. As the train and urban PT network are operated by different PT operators, AFC systems of these network levels are generally not linked together. Therefore, no direct multi-level OD matrix is available. However, the relative distribution of transferring passengers between the three case study train corridors (directions Leiden, Gouda or Delft) and between intercity and local train services, and the urban PT network was provided to us for each multi-level transfer location in The Hague. These transfer flows are distributed proportionally over the different urban PT stops as origins and destinations, thereby replacing the multi-level transfer location as origin/destination for the original urban PT journey. This results in an OD matrix for the total multi-level PT network. It should be noted that this complete OD matrix could alternatively be obtained from a strategic transport model rather than using direct empirical data, depending on data availability for the considered case study area.

In our case study, we focus on the disruption impacts for AM peak journeys with starting time between 7-9AM. Alongside simulating PT demand and supply between 7-9 AM, demand and supply are also simulated between 6-7AM and 9-10AM as warm-up and cooling-down period. This is necessary to make sure all passengers starting their journey between 7-9AM have PT supply available at all locations to start and finish their journey. It is also necessary to reflect crowding levels in PT services adequately by incorporating passengers starting their journey outside the AM peak, who affect crowding levels of passengers who started their journey within the AM peak. After applying the abovementioned transfer inference algorithm, there are about 104,000 journeys simulated for the multi-level PT network starting between 7-9AM. About 49,000 journeys start and/or end at the urban PT network level and can potentially benefit from our integrated approach, whilst approximately 55,000 journeys are only using the train network level.

8.3.2 Disruption scenario

We illustrate our proposed modelling framework by applying it to a disruption scenario. For this scenario we quantify how the impact of a train network disruption propagates to the urban PT network level, after applying optimised rescheduling and control strategies to train services on the disrupted (inter)regional train network. We simulate an infrastructure failure - such as a signal failure or switch failure - at a certain (fixed) location, resulting in lower speeds and thus

delays for all passing trains. The disruption is simulated just before Leiden for all inbound trains towards The Hague coming from Schiphol Airport (see **Figure 8.4**). In this figure, the four most important transfer stations between the train network and the urban PT network are indicated (The Hague Central, The Hague HS, The Hague Laan van NOI and Delft). Potential disruption propagation from train to urban network occurs mainly via these stations. The disruption is simulated to last from 6AM to 9AM during the simulation period. The simulation hour from 9AM to 10AM is used for service recovery. It is assumed that all trains passing this disruption location between 6-9AM obtain a random delay drawn from a distribution function with an average of 15 minutes.

In our experiments we use BusMezzo as dynamic PT assignment model (Cats et al., 2010). The number of replications required given the stochastic nature of this model is calculated based on Burghout (2004), such that the allowable percentage error does not exceed 5%. The optimisation model for train rescheduling is built in Java and solved using Gurobi. The exchange of inputs and outputs between the two models is performed automatically using a model integration tool built in Java (Obrenovic, 2019). We test the four different train rescheduling strategies as outlined in **Section 8.2.3**. **Table 8.3** provides an overview of the parameter values used for our case study. The coefficients of the travel time components of the generalised cost function (**Eq.1**) are obtained from a Revealed Preference study performed by Yap et al. (2018a) utilising smart card data records. The Value of Time is in line with values typically applied in the Netherlands. We use ζ=|0.005| and η=|0.10| as thresholds for our convergence criteria. This entails that convergence is reached if the passenger journey costs for the total network do not change more than 0.5% between two iterations, or if for at least 95% of all train segments the passenger load does not change more than 10% between two iterations.

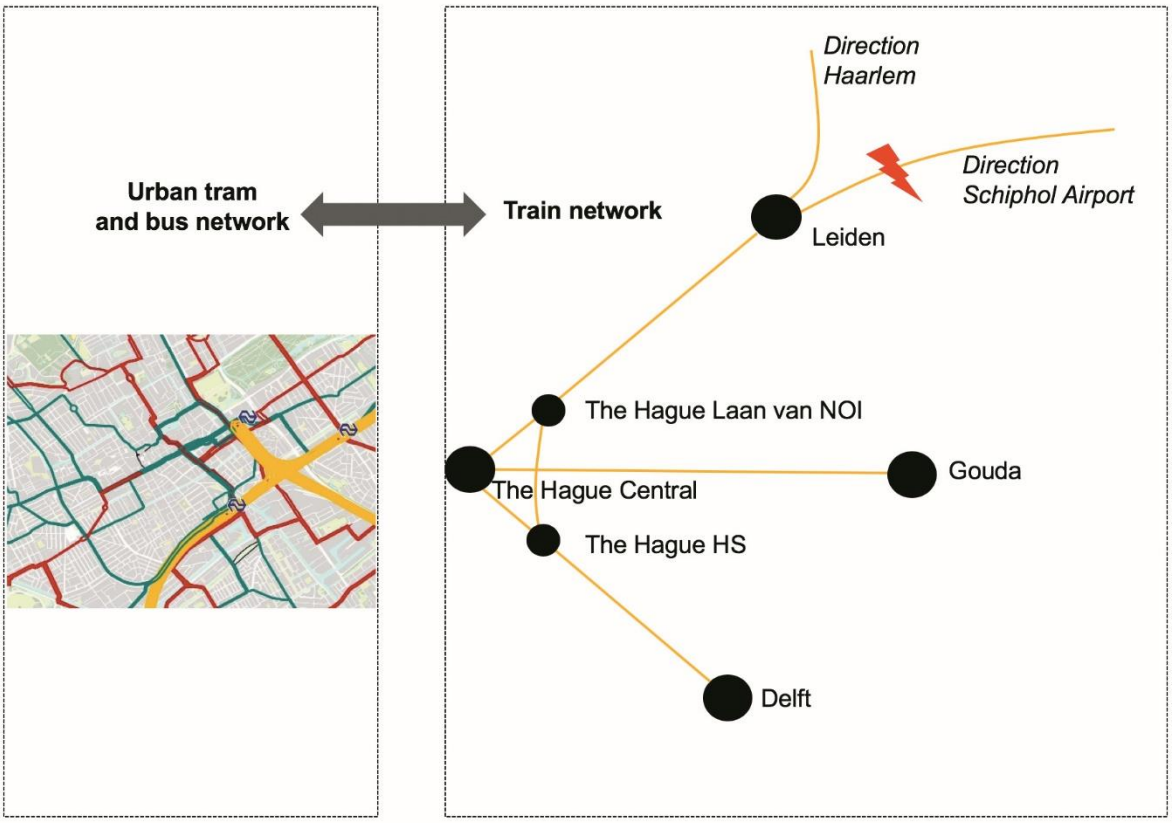


Figure 8.4. PT network with train disruption before Leiden from direction Schiphol Airport
The yellow lines in the figure left correspond to the train network as shown in the figure right.

Table 8.3. Parameter values for case study

Parameter	***Parameter function***
α=60	Weight for penalising track changes
β=1	Weight for alighting passengers
γ=3	Weight for transferring passengers
δ_0=20.4 / 16.2 / 19.8	Dwell time constant high floor tram / low floor tram / bus
δ_1=0.188 / 0.178 / 0.313	Boarding coefficient high floor tram / low floor tram / bus
δ_2=0.218 / 0.119 / 0.177	Alighting coefficient high floor tram / low floor tram / bus
ε_{ivt}=1 / ε_{wkt}=1.58 / ε_{wtt}=1.58 / ε_{wtt-d}=3.5 / ε_{tf}=4.8	Coefficients in generalised travel cost function
ζ=|0.005|	Threshold for first convergence criterion
η=|0.10|	Threshold for second convergence criterion
VoT=€9 / hour	Value of Time

8.4 Results and Discussion

This section discusses the train rescheduling results (**Section 8.4.1**) and the disruption impact results (**Section 8.4.2**). For discussion of our case study results, we refer to the four different train rescheduling strategies as follows: S1 refers to total passenger based train delay minimisation (the default strategy), S2 to transferring passenger based train delay minimisation, S3 to transfer-time based train delay minimisation, and S4 to weighted alighting-transferring passenger based train delay minimisation.

8.4.1 Train rescheduling results

Based on the convergence criteria we adopted in this study, scenarios S1, S2, S3 and S4 require 7, 3, 5 and 5 iterations, respectively, to reach convergence. For each iteration of the dynamic PT assignment model, 15 replications were required to capture the stochasticity in PT demand and supply for this case study. One replication of the dynamic PT assignment model takes about 3 minutes on a regular PC, whilst solving the train rescheduling problem requires roughly 10 minutes. Therefore, one complete iteration of both the train rescheduling model and PT assignment model requires ≈55 minutes.

The results of the train rescheduling model show different updated timetables in response to the disruption, when different control strategies are applied. **Figure 8.5** provides the arrival delay of the affected train trips which arrive at The Hague Central, which is a terminal station for all train services and the most important transfer location between train and urban network level for our case study. It can be seen that the most severe delayed trains suffer from ≈22 minutes delay when arriving at the destination. For a couple of train trips the different rescheduling strategies result in the same arrival delay at The Hague Central. Overall, we see that control strategies which incorporate the number of alighting (non-transferring) passengers result in a more similar train rescheduling: the results of strategy S1 (passenger based control) and strategy S4 (weighted alighting-transfer based) are similar for most trains. On the other hand, control strategies which are only based on the number of transferring passengers (strategies S2 and S3) result in comparable rescheduling decisions as well. In case relatively large passenger volumes transfer from a certain train to an urban tram or bus with a relatively low frequency, this train gets prioritised in strategy S3 compared to strategy S2, as can be seen for trains *2246001* and *2118001*. Compared to strategy S2, strategy S3 results in higher train arrival delays for trains with fewer passengers interchanging to urban lines with relatively low frequencies (e.g. trains *2446001* and *2263001*). When comparing strategies S1 and S4 on the one hand, and strategies S2 and S3 on the other hand, we conclude that the transfer(-time) based

strategies S2 and S3 result in fewer trains arriving late at The Hague Central, with the average arrival delay being slightly smaller than for strategies S1 and S4. The (weighted) passenger based strategies S1 and S4 tend to distribute the delays over more trains, resulting in delays for a larger number of trains. This confirms that strategies which only incorporate transferring passenger volumes in the weighted train delay minimisation, tend to result in fewer trains arriving delayed at the important transfer stations.

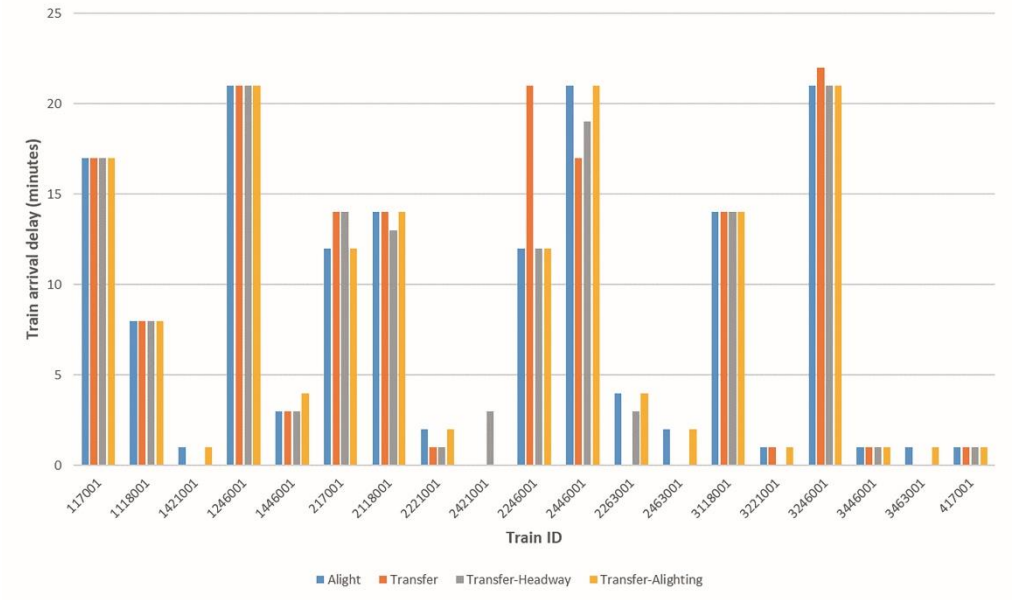


Figure 8.5. Train arrival delay at The Hague Central for different train rescheduling strategies

8.4.2 Disruption impact results

The results from the dynamic PT assignment model allow for quantifying the disruption impact on the disrupted train network level Δv^t and the spilled-over disruption propagation to the urban PT network level Δv^u. **Table 8.4** provides the monetised disruption impact in Euros (left), and the relative impact compared to the default passenger based delay minimisation strategy S1 (right). Depending on the rescheduling strategy applied, the propagated disruption costs make up 6-8% of the total passenger disruption costs. For this case study, our results thus show that neglecting disruption propagation to the urban network results in an underestimation of 6-8% of the total disruption costs for passengers. The delayed train arrivals caused by this disruption influence journeys starting at the train network and terminating at the urban network (and in the opposite direction) due to potential missed connections or prolonged waiting times. In addition, the shifted train arrival trains can result in less uniform transfer volumes to the urban PT trips, thereby resulting in higher average crowding levels for urban PT trips. This can have a negative impact on journeys entirely made on the urban network as well.

When comparing the different control strategies, one can see that disruption propagation costs differ substantially between the strategies. When the default rescheduling strategy S1 is applied, the forecast propagated disruption costs are €6,900. However, rescheduling strategies S2-S4 which explicitly account for transferring passengers to the urban network in the prioritisation of trains during the rescheduling, are all able to reduce the spilled-over disruption impact to the urban network. Strategies S3 and S4 result in spilled-over disruption impacts of €5,800-€5,900, whilst a further reduction in disruption propagation is predicted for strategy S2 (€5,000). This means that strategies S2-S4 are able to reduce disruption propagation to the urban network by up to 27%. The total passenger disruption

impact from strategies S2-S4 is reduced by up to 3.3%. The direct passenger disruption costs for journeys entirely made on the train network do however remain stable for all strategies S1-S4 (~€79,000-€80,000). This suggests that it is possible to reduce disruption impact propagation to the urban network, without increasing disruption costs for the train network, indicating that strategy S1 may yield suboptimal rescheduling results from a total network perspective.

Table 8.4. Disruption impact for different train rescheduling strategies

	Disruption impact (Euro)				*Relative disruption impact*		
	Disrupted level (train)	*Spilled-over level (urban)*	*Total*		*Disrupted level (train)*	*Spilled-over level (urban)*	*Total*
Strategy S1 – passenger	80,353	6,857	*87,210*		100%	100%	*100%*
Strategy S2 - transfer	79,353	5,001	*84,354*		-1.2%	-27.1%	*-3.3%*
Strategy S3 - transfer time	80,435	5,755	*86,190*		0.1%	-16.1%	*-1.2%*
Strategy S4 - alighting-transfer	79,941	5,884	*85,825*		-0.5%	-14.2%	*-1.6%*

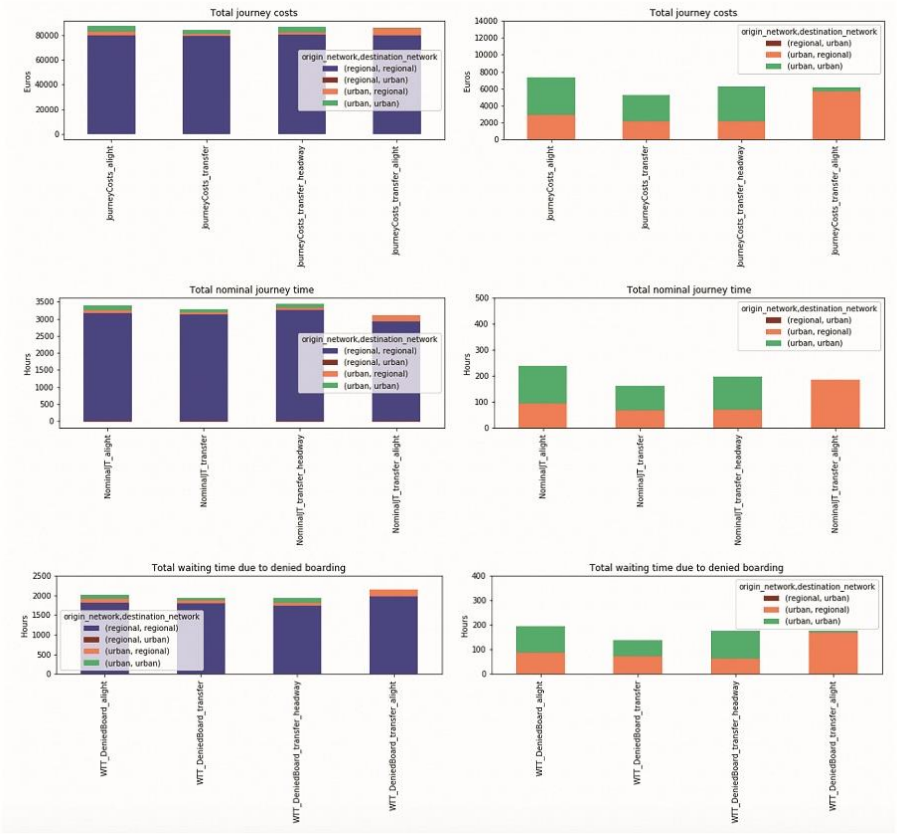


Figure 8.6. Disruption impact propagation to urban PT network expressed in nominal and perceived passenger delays

Figure 8.6 separates the total (left) and propagated (right) disruption impacts for the four passenger segments we distinguish in this research, depending on whether the passenger journey starts and/or ends at the train network or urban network level. From the left side of this figure, we can confirm that the majority of the disruption impact applies to journeys made exclusively using the train network (dark blue). On the right side of the figure, we zoom in to journeys which start and/or end at the urban network. The propagated disruption impact

primarily affects journeys entirely made on the urban network, and journeys from urban to train network. Strategy S2 - which weights train delays based on the number of transferring passengers to the urban network only - results in the largest decrease in additional nominal journey time, waiting time and waiting time due to denied boarding, and therefore results in the largest total journey cost reduction compared to default strategy S1. Whilst the total spilled-over disruption costs of S3 and S4 are comparable, it can be seen that strategy S3 influences both urban-urban and urban-train journeys. Strategy S4 on the other hand, which attaches 3 times more weight to transferring passengers than to alighting passengers in the rescheduling, almost only affects urban-train journeys. This implies that strategy S4 distributes the disruption impacts over a smaller number of passengers, for which the average disruption impact is larger than for affected passengers by strategies S2 or S3.

8.5 Conclusions

In this research we propose a methodology to predict the propagation of the impact of a train network disruption to the urban PT network level, subject to different train rescheduling control strategies. We propose an integrated modelling framework where we combine a dynamic PT assignment model and an optimisation-based train rescheduling model in an iterative process. We incorporate the number of transferring passengers to the urban network level in the optimisation process by weighting train delays accordingly. This allows the train rescheduling model to incorporate potential disruption propagation to the urban PT network level in the decision which trains to prioritise for retiming, reordering or rerouting.

We applied our developed modelling framework to a case study in the Netherlands. The case study application illustrates the relevance of quantifying the propagation of disruption impacts to other network levels: the modelling results show that neglecting disruption propagation to the urban network results in 6-8% underestimation of the total disruption costs for passengers. This can potentially influence the outcomes of appraisal studies when evaluating measures to improve PT robustness. For this specific case study disruption, our findings illustrate that incorporating the number of transferring passengers to the urban network level in the objective function of the train rescheduling model can reduce propagation of passenger delays to the urban network level by 14-27%, without increasing passenger delays on the train network level. This demonstrates that train rescheduling strategies which do not consider disruption propagation impacts can result in suboptimal rescheduling from a total passenger perspective.

Based on our findings we recommend train network managers to consider how control decisions can result in train disruption impacts propagating to the urban PT network with which they interface. Whilst operators in practice often only consider the trips and passengers on the part of the network they are assigned to monitor and control, our research offers evidence that it can be made beneficial for all passenger groups to consider the wider PT network in control decisions. This can potentially reduce the disruption impact for passengers on the network level where the disruption occurs, as well as for passengers travelling on another PT network level.

The main purpose of our case study application is illustrating that our modelling framework can be applied to large, real-world public transport networks and that it provides plausible results within reasonable computation times. For future research we recommend testing the disruption (propagation) impacts for more disruptions, locations and time periods, and exploring the use of different weights for control strategy S4 using our proposed modelling framework. This can provide a more systematic insight into the relation between different train control strategies and their impact on controlling disruption propagation, and hence provide more generalizable conclusions based on the case study outcomes. In this research only the passenger costs that are associated with a disruption are quantified. Other disruption costs for

the service provider, such as crew-related costs, rescheduling costs or reduced revenues, are not calculated in this study as our focus is on the passenger disruption impacts. Notwithstanding, calculating the disruption costs for the PT service provider is a relevant topic we recommend to incorporate in future research. At last, in our study we only consider train trips to be subject to control interventions to mitigate disruption propagation. It should however be mentioned that additional control can be applied to urban PT trips to further alleviate disruption propagation impacts. Therefore, for future research we recommend exploring how control to train trips and urban PT trips can be applied simultaneously to further control disruption propagation.

9. Conclusions

9.1 Main Findings

The main objective of this research is to improve methods to measure, predict and control disruption impacts for urban public transport (PT) networks. Based on the previous chapters, we provide answers to the three research questions as formulated in **Section 1.3** and summarise the scientific contribution of our work.

1. How can we measure and characterise the behavioural and demand response of passengers during planned and unplanned urban public transport disruptions?
As PT disruptions can have severe impacts for passengers, PT service providers and authorities, it is important to be able to measure the impact of PT disruptions. Empirical data from passive data systems is an important source to measure these disruption impacts. For a passenger-oriented impact measurement, passengers' generalised journey costs need to be inferred and compared between a disrupted and an undisrupted journey. As a first step for this, we develop a robust transfer inference algorithm with the ability to infer passenger journeys from individual smart card transactions during disrupted and undisrupted circumstances (**Chapter 2**). This algorithm uses data from Automated Vehicle Location (AVL) and Automated Fare Collection (AFC) systems (such as smart card data) of the urban network as input. It relaxes existing state-of-the-art transfer inference algorithms to incorporate the atypical passenger behaviour that can be observed during a PT disruption. An alighting is considered a transfer if it satisfies the following temporal, spatial and binary criteria. The temporal criterion states that an alighting is a transfer if a passenger boards the first reasonable vehicle arriving at a transfer location, thereby incorporating required transfer walking time, crowding levels and potential denied boarding for the consecutive vehicle. The spatial criterion indicates that the maximum transfer walking distance should not exceed a threshold of 400 Euclidean metres, which was calibrated for our case study network of The Hague, the Netherlands, unless a passenger uses PT services of another network level (where no AFC data is available for) as intermediate journey stage during a disruption. The binary criterion states that a transfer to the same line is only possible when made to the next vehicle of this same line, to incorporate the effect of operational

measures as short-turning or deadheading possibly being applied. The relaxation increases the average number of trips per journey from 1.18 to 1.21, when applied to our case study public transport network. A partial validation shows that our algorithm improves inference during disruptions, without compromising inference results during undisrupted circumstances. Our results confirm that the atypical passenger behaviour during disruptions can be incorporated in the transfer inference algorithm, whilst there are no journeys resulting for which it is obvious that the inference would have been incorrect (based on the ratio between travelled and Euclidean distance between origin and destination).

A second step when measuring disruption impacts is to infer how passengers perceive the different journey components, especially in relation to crowding (**Chapter 3**). Compared to previous studies on crowding valuation in public transport, our study results are entirely based on Revealed Preference (RP) route choice observations. Observed routes are obtained from AFC data, whilst the route attributes are derived from AVL data (in-vehicle time, transfer time), and by fusion of both data sources (crowding). Based on the estimated discrete choice model with panel effects, we find that the average in-vehicle time crowding multiplier for urban trams and buses equals 1.16 when all seats are occupied and no passengers are standing. This implies that passengers perceive one minute travelling as 1.16 minute, when the vehicle occupancy equals the seat capacity. In case occupancies increase to an average standing density of 3 passengers per m^2, this in-vehicle time multiplier equals 1.34. For frequent passengers, these two values equal 1.31 and 1.75, respectively. Hence, on average this passenger segment perceives 12 minutes travelling in a PT vehicle with on average 3 passengers standing per m^2 as 21 minutes. Infrequent passengers do not incorporate crowding in their route choice, due to the lack of prior knowledge about crowding levels. Our estimated crowding multipliers are lower than values found in previous Stated Preference (SP) experiments. For example, previous SP studies found in-vehicle time multipliers ranging between 1.6-1.9 when the occupancy equals the seat capacity (e.g. MVA Consultancy, 2008), whereas our study suggests values up to 1.31 under these circumstances. This gives evidence for the tendency of SP experiments to overestimate values of coefficients, compared to RP based studies.

A third step is to infer passengers' demand response in the event of planned disruptions such as maintenance work, as demand suppression can be expected for communicated service disruptions (**Chapter 4**). This demand suppression indicates that passengers temporarily accept an alternative for their affected PT journey, for example by changing their mode choice, destination choice or trip frequency choice. For this study purpose, we calibrate route and mode choice parameters of a PT ridership prediction model based on empirical data from two planned disruptions. A three-step rule-based procedure is developed for this. The calibrated parameter set is validated using two different planned disruptions. In our approach, we compare the predicted and empirical relative impacts of a planned disruption on the number of passengers and passenger-kilometres for affected PT lines and time periods, thereby correcting for seasonality effects. Our results suggest that passengers perceive in-vehicle time in replacement buses about 11% more negatively compared to the tram line being replaced. Waiting time perception for rail-replacement buses is ≈30% higher than for regular trams or buses, potentially caused by limited facilities at temporal bus stops and by uncertainty about service headways and reliability. Besides, when modelling replacement bus services, it is recommended to use the frequency of the initial tram line if the frequency of rail-replacement services is higher and crowding is not incorporated in the assignment process, as passengers do not seem to perceive this theoretical benefit of higher frequencies. The new parameter set improves prediction accuracy up to 13% compared to the default parameter set, and shows to be a robust and valuable tool for public transport operators.

2. How can we incorporate disruption frequency and impact predictions in a public transport vulnerability analysis for urban and multi-level public transport networks?
To identify the components of a PT network which contribute most to its vulnerability, it is necessary to consider both the expected exposure to disruptions of each component, as well as the disruption impact once a disruption occurs. Focusing solely on vulnerability in relation to disruption impacts can incorrectly put the emphasis on locations where very severe yet very rare disruptions occur, and therefore result in an incorrect identification of the most vulnerable locations in a PT network. We develop a full scan method (**Chapter 5**) and a pre-selection method (**Chapter 6**) for PT vulnerability analyses, which both incorporate disruption frequency and impact.

For the pre-selection method (**Chapter 6**), we develop four criteria as proxy for the contribution of each link to PT network vulnerability: the expected direct disruption exposure (I), the expected indirect disruption exposure in case disruptions elsewhere on the network result in service adjustments for the considered link segment (II), the total expected disruption exposure (III: I+II) and the expected number of affected passengers as proxy for disruption impact (IV). We use disruption log data as input to fit statistical functions to predict the frequency with which different disruption types occur on the considered network. We translate network-wide disruption expose to individual links via simple predictors, such as the number of train-kilometres or the number of switches per link. The multi-level PT network is explicitly considered, by comparing how links of different network levels contribute to total network vulnerability. We apply this method to the multi-level PT network of the southern part of the Randstad in the Netherlands as case study. The results indicate that links of the metro / light rail and tram network are relatively often exposed to disruptions, possibly caused by the influence of mixed traffic. On the other hand, the impact of train network disruptions is expected to be relatively high due to the larger passenger volumes affected, once a disruption occurs. Therefore, busy links of the metro / light rail network generally have the largest contribution to vulnerability of the multi-level PT network, as both exposure and the number of affected passengers are relatively high. Our study results show the relevance of incorporating disruption frequencies in vulnerability analyses, as the list of most critical links differs substantially from a list based only on expected disruption impacts.

We also develop a data-driven full scan methodology to identify the most critical stations of a PT network within reasonable computation times (**Chapter 5**). A supervised learning approach is developed to predict the probability of each disruption type for each individual station. We use spatial variables (lines serving each station), temporal variables (day of the week, time of the day, season), network characteristics (whether a station is a terminal or transfer location), demand (passenger volume obtained from AFC data) and an auto regressor (disruption exposure in the previous month obtained from disruption log data) as predictors. A supervised learning model is also developed to predict passenger delay impact of each disruption type for each station, based on demand predictors, temporal predictors and network topology predictors. In a last step, stations are clustered using unsupervised learning based on their expected contribution to network vulnerability. This improves the transferability of our case study results, as this indicates which types of stations contribute most to PT vulnerability. We apply our methodology to the Washington, D.C. metro network. Case study results show that stations with high train frequencies and high passenger volumes on central trunk sections are most critical, together with transfer stations and terminals (where disruptions often arise). Intermediate stations located on branches of a line are least critical. The lower train frequencies and passenger volumes result in lower disruption exposure and impact, despite less route alternatives typically being available for these passengers once a disruption occurs.

Our proposed pre-selection method and full scan method can be used separately or simultaneously during PT vulnerability analyses. If the study purpose is to predict passenger

delay impacts for a large number of disruptions, our data-driven full scan methodology is appropriate as this enables a fast calculation of delay impacts for many disruption instances. The pre-selection method can provide more detailed insights in the spatial impacts of disruptions. Identifying a select number of critical links for which a PT assignment model is used to predict disruption impacts allows for a detailed assessment of the spatial distribution of disruption impacts over the network. It provides more insights in passenger rerouting strategies and indicates routes and stations where more crowding might be expected. Besides, using pre-selection criteria combined with a PT assignment model allows using more complex metrics for disruption impacts than only the nominal passenger delay impact, for example by incorporating impacts from increased crowding and denied boarding. Both methods could also be used simultaneously. A PT assignment model could be used to provide a more detailed insight in the spatial impacts of disruptions and the impacts for individual PT routes, by simulating disruptions for the stations or links which show to be most critical in our data-driven full scan methodology.

3. How can we predict and control the direct and propagated impact of disruptions on the urban public transport network in a multi-level network environment?

To control disruption impacts for the urban PT network, one can apply control to urban PT services (**Chapter 7**), or apply control to train services to mitigate disruption propagation to the urban network level (**Chapter 8**) or to mitigate the impact of urban network disruptions (**Chapter 6**).

In relation to applying control to urban PT services, our research focuses on enabling synchronisation for large PT networks. Optimal PT synchronisation is currently applied to relatively small case study networks, as the problem becomes difficult to solve within reasonable times for larger networks. Our contribution is therefore not the development of an improved synchronisation strategy as such, but the development of a generic, preparatory method to reduce dimensionality by identifying key locations and routes to prioritise for synchronisation (**Chapter 7**). First, using the transfer flow matrix derived from AFC data as input, we apply a density based clustering technique to determine the subset of PT hubs where synchronisation needs to be prioritised. Second, we represent the transfer patterns within each hub using a C-space inspired topological network representation. By using a community detection algorithm, groups of lines are identified for which it is recommended to synchronise them simultaneously. When applied to the urban PT network of The Hague as case study, results show that 70% of all transfers occurring within identified transfer locations would be captured by considering less than 1% of all transfer locations for synchronisation, thus substantially reducing the complexity of solving the optimal synchronisation problem.

When considering the multi-level network environment of the urban PT network, we can conclude there can be a substantial propagation of disruption impacts from a train network disruption to the urban PT network level. Testing a disruption scenario for the multi-level PT network of The Hague as case study illustrates that disruption propagation costs are responsible for up to 8% of the total passenger disruption costs. We develop a simulation-based optimisation framework to control this disruption propagation from the train to the urban network level (**Chapter 8**). In this approach, we combine a train rescheduling optimisation model - in which only the train network is represented - and a dynamic PT assignment model where the total multi-level network is represented. We incorporate the number of transferring passengers from the train to the urban network level in the objective function of the train rescheduling model, and test the impact of a control strategy using the dynamic assignment model based on updated train timetables from the optimisation process. The train rescheduling model is then iteratively updated based on train passenger volumes resulting from the PT assignment model. This allows the train rescheduling model to incorporate potential disruption

propagation to the urban PT network level in the decisions which trains to prioritise for retiming, reordering or rerouting. In our case study, the propagation of passenger delays to the urban network could be reduced by up to 14-27% by incorporating transferring passengers to the urban network in the train rescheduling optimisation process, without increasing delays for passengers on the train network.

We also illustrate how the train network level can be used as means to reduce the impact of disruptions occurring on the urban level, using the multi-level PT network of the southern part of the Randstad (the Netherlands) as case study (**Chapter 6**). We test the impact of temporarily stopping intercity services at two local train stations in the event of a disruption on the parallel light rail route, to provide a better alternative for affected passengers. A societal cost-benefit analysis framework is established to predict the impact for the different passenger segments. This case study confirms that this measure can reduce the total disruption impacts by 8%, the longer train running times included, thus illustrating the potential of the multi-level network to mitigate disruption impacts. As we used a static assignment model for the latter case study, these case study results mainly apply for planned or longer-lasting disruptions for which passenger awareness can be assumed.

9.2 Implications for Practice

Based on the results of this research, we formulate several implications for the public transport industry.

Improved passenger predictions during disruptions
This research allows public transport authorities and service providers to improve the accuracy of their passenger predictions during planned and unplanned disruptions. For many metropolitan PT systems, each year there are many planned and unplanned disruptions. When predicting passengers' route and mode choice response can be improved during these disruptions, this can result in a better product for the passenger and lead to higher customer satisfaction. Our study results in an updated crowding in-vehicle time multiplier and calibrated coefficients reflecting the passenger perception of rail-replacement buses and the demand suppression during planned disruptions. These values, for example if applied within transport planning software, support PT authorities and service providers to predict more accurately how passengers will react during disruptions, and what the implications on service quality might be. It can provide better insights whether additional capacity needs to be provided on alternative routes, or if excessive station or vehicle crowding can be expected. These values also help to improve ridership predictions for rail-replacement buses, so that PT operators can better align their bus supply with the expected demand. Additionally, this improves the revenue predictions for agency or operator during a disruption and thus supports their business plan development.

Improved quantification of disruption impacts
In practice, many PT authorities and service providers still apply vehicle-oriented or simplified passenger-oriented metrics as measure to quantify disruption impacts on a PT network. Methods developed in this research can result in an improved quantification of disruption impacts based on empirical data. Our research improves and eases the calculation of a comprehensive passenger-oriented reliability metric, which compares nominal and perceived journey costs for passengers between scheduled and realised circumstances using empirical data. The improved transfer inference algorithm, which can infer journeys for both disrupted and undisrupted scenarios, enables journey inference without the need to explicitly demarcate the disrupted and undisrupted periods. This implies that journey inference can be automated and applied to large empirical AFC and AVL datasets. The updated perception coefficients for

the different journey time components (e.g. (crowded) in-vehicle time, waiting time and transfer time for trams, buses and rail-placement services), derived from realised route and mode choices, enable a more accurate quantification of such performance metric.

Prioritisation of robustness measures

Our study supports transport policy makers in prioritising the locations and disruption types for which to develop and implement robustness measures. Our developed methods identify the nodes or links contributing most to PT network vulnerability. This provides decision-makers the locations where measures aimed to reduce disruption frequencies and/or disruption impacts should be prioritised. Our research also helps policy makers to devise appropriate measures. Depending on the type of station (for example: a transfer or terminal station), the location on the network (e.g. a station located on a trunk section or branch), the network level (regional train or urban tram and bus level) and the predicted disruption exposure and impact, one can decide whether measures should be focused on reducing the number of disruptions (of certain disruption types), or on reducing disruption impacts and improving PT network resilience. Our cost-benefit analysis framework helps quantifying the robustness benefits of potential measures and compare these with expected costs. This enables decision-makers to evaluate different robustness measures and prioritise measures with the highest expected (societal) benefit-cost ratio for implementation, thereby rationalising the decision-making process.

Improved control to mitigate disruption impacts

Our research results can improve the real-time control decisions taken by controllers to mitigate disruption impacts, hence reducing the passenger impact of disruptions. An important insight from our study is the value a multi-level PT network can provide when devising real-time control measures to mitigate disruption impacts for the urban network. Our research shows that incorporating the number of transferring passengers from train to urban network in the train rescheduling strategies can substantially reduce disruption propagation, without increasing the impact for train passengers. Other results show how control applied to the train network can mitigate the impact of an urban disruption, where the net impact for passengers shows to be positive. Whilst PT controllers traditionally only consider the part of the PT network they are responsible for, our study shows the value of developing overarching, integrated control strategies which consider the total multi-level PT network which is available and used by passengers. Our methods can help predicting the impact of different integrated control strategies, and therefore support controllers in their decision-making during disruptions. Additionally, our research supports controllers to identify the locations and routes for which transfer synchronisation and holding control can be prioritised. For larger networks with many possible locations for synchronisation, our research gives controllers an overview where their actions are expected to benefit most transferring passengers, thereby rationalising the decision-making process under difficult circumstances. The latter could also be applied in the tactical planning phase when designing timetables, by prioritising synchronisation at the identified locations and between the identified routes.

In summary, the objective of the developed methods in this research is to strengthen the passenger perspective when measuring, predicting or controlling disruption impacts. We develop methods which ease the measurement and prediction of passenger impacts of disruptions rather than vehicle impacts, and methods which incorporate more accurately how passengers perceive the different journey components during disruptions. In addition, we propose methods to prioritise locations where to apply robustness measures, such that the contribution to network vulnerability is largest for passengers. Furthermore, our methods enable the application of control strategies during disruptions which can further mitigate disruption

impacts for affected passengers. The application of our methods to different case study networks confirms our methods can be applied in practice. Although results might differ from case to case, our case study results suggest that passenger benefits can be realised when applying our approaches. Our research provides generic methods and tools for the public transport industry to apply to their specific public transport network. We recommend a close cooperation between science and the public transport industry, to implement methods and results from this research in the daily business of the public transport sector. This has the potential to improve the public transport product delivered to passengers.

9.3 Recommendations for Future Research

In this section, we formulate several recommendations for directions of future research resulting from our work.

A detailed study towards passengers' demand response during planned disruptions
In **Chapter 4**, we developed a rule-based approach to calibrate route and mode choice parameters which reflect passengers' behavioural and demand response during planned disruptions. A limitation of this method is that coefficients are not obtained via an optimal fitting procedure, such as maximum likelihood, but were obtained rule-based instead. In our case, this was caused by the lack of individual AFC transactions available for the different planned disruptions we considered. Although these individual transactions were used as input for our method, these were already processed and aggregated by the PT operator when provided. This meant that a maximum likelihood procedure - comparable to the method we adopted to infer crowding valuation in **Chapter 3** - was not possible. Estimation of a RP based discrete choice model with different observed route choice alternatives, together with a base alternative of not using public transport during a disruption, could improve the estimation of these coefficients. Incorporating segmentation between passengers with different trip purposes or during different time periods (e.g. peak versus off-peak, or weekday versus weekend day) is also recommended. Our study can therefore be considered a first exploratory approach to find better fitting - yet not optimal - coefficients.

Besides, we recommend a detailed study towards the demand response of passengers during planned disruptions. In our study, we calibrated a generalised cost elasticity coefficient to reflect the decrease in PT ridership during planned disruptions. However, it is recommended to investigate what this ridership reduction implies. Based on long-term ticketing data, mobile phone data and data from ride-hailing services or bicycle providers, one could investigate to which extent passengers change their departure time choice, destination choice or mode choice. It provides insights in the willingness of passengers to use alternative transport services, such as ride-hailing services or bicycle-sharing schemes, in different parts of the network, for different passenger segments (e.g. for different trip purposes or different ages), or during different times of the day (e.g. daytime vs. evenings). Availability of data from different transport modes also enables studying how passengers react once the planned disruption has been resolved. For example, one could quantify at what pace public transport demand recovers after the planned disruption is terminated, or which percentage of passengers keeps using alternative modes of transport.

Investigate passenger behaviour during unplanned disruptions
We recommend a more thorough investigation how the modelling of passenger behaviour during unplanned disruptions can be improved. Due to the dynamic nature of disruptions, the use of static PT assignment models is not recommended for unplanned disruptions. Our results from **Chapter 6**, where we test how control to the train network can contribute to reducing

urban disruption impacts, are therefore primarily applicable for planned or longer-lasting disruptions. The simulation-based optimisation approach as proposed in **Chapter 8** uses an agent-based simulation-based dynamic PT assignment model, and is therefore suitable to reflect en-route passenger route choice decisions during unplanned disruptions. The limitation of our study is that dynamic en-route passenger route choice parameters are not calibrated and validated against empirical data. Hence, we recommend such calibration and validation exercise to investigate how to model passengers' dynamic route choice decisions. For example, route choice parameters in this dynamic PT assignment model could be calibrated based on observed behaviour for a specific case study, or segmented for different passenger groups. Another direction for future research is investigating the impact of real-time information and flexible working arrangements on passengers' route and mode choice. The availability of real-time information during unplanned disruptions might result in passenger behaviour which resembles behaviour during planned disruptions. An increased acceptance of flexible working arrangements, such as working from home, can also affect passengers' demand response during unplanned disruptions. The traditional assumption of fixed demand during unplanned disruptions can be questioned in countries where flexible working is commonly accepted, as passengers might decide to work from home or from a different location in the event of a disruption (Shires et al., 2018). More research to this topic - for example using ticketing data, mobile phone data and surveys - is therefore recommended.

Investigate the role of information provision during disruptions

For future research, we recommend explicitly incorporating the role of information provision to passengers before and during disruptions in relation to the prediction and control of disruption impacts. In the PT assignment model used in **Chapter 4**, passenger route choice is based on prior expectations of the disutility of different route alternatives. In **Chapter 8**, the PT assignment model used to predict propagated disruption impacts does allow for dynamic route choice during a journey based on updated expectations of the disutility of different routes, for example as function of real-time information provision or passenger observations (e.g. observing an overcrowded train or platform). In these studies we did however not explicitly account for the extent to which information about a disruption was provided and received by different passenger segments, and how this influenced their mode or route choice. Incorporating information provision in route and mode choice models (as for example proposed by Chorus et al., 2013) can potentially improve the accuracy of predictions of passenger behaviour during disruptions. Information provision can also play a role in disruption mitigation. As passengers attach a certain value to travel information (e.g. Chorus et al., 2006; Lu et al., 2017), information provision can help reducing perceived disruption impacts. Additionally, information provision can result in a better distribution of passengers over the available alternative routes during disruptions, hence reducing disruption impacts as well. Integrating the role of information provision in frameworks to quantify and control disruption impacts is therefore recommended.

Improved method to decouple disruptions and connect disruptions and delays

In **Chapter 5**, we developed a method to predict the passenger delay impact caused by different disruptions. One of the challenges here is to attribute observed passenger delays (from AFC systems) to individual disruptions. For typical metropolitan PT systems, it is common that many (often smaller) disruptions occur simultaneously, such as different train delays or cancellations. Besides, the impact of one disruption can propagate over the PT network, such that it might interfere with another disruption occurring at another network location or at a later moment in time (Malandri et al., 2018). The main limitation in this part of our research is the time horizon we imposed for which we predicted delay impacts. In our research, we only considered the delay impact of a disruption up to two hours after the hour the disruption started, to limit the

potential interaction between disruption impacts of different disruptions. However, a more advanced method to decouple disruptions and to connect delays to individual disruptions is recommended. For example, a machine learning approach as proposed by Marra and Corman (2019) to link delays to disruptions could be used as a starting point.

Compare synchronisation priorities with network-wide synchronisation

In **Chapter 7**, we developed a method to identify key locations and routes in urban PT networks to prioritise for synchronisation. This reduces the dimensionality of the transfer synchronisation problem and therefore enables performing optimal synchronisation for this selection of locations and routes in a next step. The main limitation of this two-step approach is that no network-wide optimisation is applied, as optimal synchronisation is only applied for the selected locations and routes. Recently, Gkiotsalitis et al. (2019) proposed an approach to solve the network-wide bus synchronisation problem by relaxation of the original minimax problem. Although a relaxation of the original problem is applied, this approach performs network-wide bus synchronisation in one step, rather than our proposed two-step approach. It is therefore recommended to compare the performance and required computation time of our two-step approach with results from this recent study. For example, both methods could be applied to one case study public transport network. Once both methods generate the timetables for all PT trips, the generalised journey time for all passengers can be calculated and compared. This enables a comparison between the performance of the different methods, and might highlight how different methods propose different synchronisation priorities for different parts of the public transport network.

References

Abenoza, R.F., Cats, O., Susilo, Y.O. (2017). Travel satisfaction with public transport: Determinants, user classes, regional disparities and their evolution. Transportation Research Part A, 95, 64-84.

Abenoza, R.F., Cats, O., Susilo, Y.O. (2019). Determinants of traveler satisfaction: Evidence for non-linear and asymmetric effects. Transportation Research Part F, 66, 339-356.

Agard, B., Morency, C., Trépanier, M. (2007). Mining public transport user behaviour from smart card data. Montréal, Canada: CIRRELT-2007-42.

Alsger, A., Assemi, B., Mesbah, M., Ferreira, L. (2016). Validating and improving public transport origin-destination estimation algorithm using smart card fare data. Transportation Research Part C, 68, 490-506.

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

Arentze, T.A., Molin, E.J.E. (2013). Travelers' preferences in multimodal networks: Design and results of a comprehensive series of choice experiments. Transportation Research Part A, 58, 15-28.

Balcombe, R., Mackett, R., Paulley, N., Preston, J., Shires, J., Titheridge, H., Wardman, M., White, P. (2004). The demand for public transport: a practical guide. TRL Report TRL 593, Crowthorne, UK.

Batarce, M., Munoz, J.C., Ortuzar, J. (2016). Valuing crowding in public transport: Implications for cost-benefit analysis. Transportation Research Part A, 91, 358-378.

Batarce, M., Munoz, J.C., Ortuzar, J., Raveau, S., Mojica, C., Rios, R.A. (2015). Valuing crowding in public transport systems using mixed stated/revealed preference data: the case of Santiago. Transportation Research Record, 2535, 73-78.

Bates J., Polak J., Jones P., Cook A. (2001). The valuation of reliability for personal travel. Transportation Research Part E, 37, 191–229.

Bell, M.G.H. (2003). The use of game theory to measure the vulnerability of stochastic networks. IEEE Transactions on Reliability, 52, 63-68.

Bierlaire, M. (2003). BIOGEME: A free package for the estimation of discrete choice models. Proceedings of the 3rd Swiss Transportation Research Conference, Ascona, Switzerland.

Binder, S., Maknoon, Y., Bierlaire, M. (2017). The multi-objective railway timetable rescheduling problem. Transportation Research Part C, 78, 78-94.

Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E. (2008). Fast unfolding of communities in large networks. Journal of Statistical Mechanics, 2008, P10008.

Bovy, P.H.L., Hoogendoorn-Lanser, S. (2005). Modelling route choice behaviour in multi-modal transport networks. Transportation, 32, 341-368.

Brands, T., De Romph, E., Veitch, T., Cook, J. (2013). Modelling public transport route choice with multiple access and egress modes. European Transport Conference, Frankfurt, Germany.

Briand, S.A., Come, E., Trépanier, M., Oukhellou, L. (2017). Smart card clustering to extract typical temporal passenger habits in transit network. Two case studies: Rennes in France and Gatineau in Canada. 3rd International Workshop and Symposium: Research and applications on the use of passive data from public transport (TransitData), Santiago, Chile.

Bunschoten, T., Molin, E.J.E., Van Nes, R. (2013). Tram or bus; does the tram bonus exist? European Transport Conference, Frankfurt, Germany.

Burghout, W. (2004). A note on the number of replication runs in stochastic traffic simulation models. Unpublished.

Cacchiani, V., Huisman, D., Kidd, M., Kroon, L., Toth, P., Veelenturf, L., Wagenaar, J. (2014). An overview of recovery models and algorithms for real-time railway rescheduling. Transportation Research Part B, 63, 15-37.

Cats, O., Abenoza, R.F., Liu, C., Susilo, Y.O. (2015a). Identifying priority areas based on a thirteen years evolution of satisfaction with public transport and its determinants. Transportation Research Record, 2323, 99-109.

Cats, O., Burghout, W., Toledo, T., Koutsopoulos, H.N. (2010). Mesoscopic modelling of bus transportation. Transportation Research Record, 2188, 9-18.

Cats, O., Jenelius, E. (2014). Dynamic vulnerability analysis of public transport networks: Mitigation effects of real-time information. Networks and Spatial Economics, 14, 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, 81, 47-61.

Cats, O., Jenelius, E. (2018). Beyond a complete failure: The impact of partial capacity degradation on public transport network vulnerability. Transportmetrica B, 6, 77-96.

Cats, O., Wang, Q., Zhao, Y. (2015b). Identification and classification of public transport activity centers in Stockholm using passenger flow data. Journal of Transport Geography, 48, 10-22.

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

Cats, O., Yap, M.D., Van Oort, N. (2016b). Exposing the role of exposure: Public transport network risk analysis. Transportation Research Part A, 88, 1-14.

CBS (2013, Oct. 23). Mobiliteit in Nederland; mobiliteitskenmerken en motieven, regio's. Verplaatsingen per persoon per dag (in Dutch). Retrieved from http://statline.cbs.nl/StatWeb/publication/?VW=T&DM=SLNL&PA=81123NED&D1=0&D2=0&D3=a&D4=0,25-37&D5=0&D6=l&HD=130830-1202&HDR=T,G4,G1,G5,G2&STB=G3.

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, 241-256.

Chorus, C.G., Arentze, T.A., Molin, E.J.E., Timmermans, H.J.P., Van Wee, B. (2006). The value of travel information: Decision strategy-specific conceptualizations and numerical examples. Transportation Research Part B, 40, 504-519.

Chorus, C.G., Walker, J.L., Ben-Akiva, M. (2013). A joint model of travel information acquisition and response to received messages. Transportation Research Part C, 26, 61-77.

Cicerone, S., D'Angelo, G., Di Stefano, G., Frigioni, D., Navarra, A., Schachtebeck, M., Schöbel, A. (2009). Recoverable robustness in shunting and timetabling. In R.K. Ahuja, R.H. Möhring and C.D. Zaroliagis (eds.), Robust and online large-scale optimization: Models and techniques for transportation systems. Lecture notes in computer science (pp. 28-60). Berlin Heidelberg, Germany: Springer.

Corman, F., D'Ariano, A., Hansen, I.A. (2014). Evaluating disturbance robustness of railway schedules. Journal of Intelligent Transport Systems, 18, 106-120.

Corman, F., D'Ariano, A., Pacciarelli, D., Pranzo, M. (2010). A tabu search algorithm for rerouting trains during rail operations. Transportation Research Part B, 44, 175-192.

D'Ariano, A., Pacciarelli, D., Pranzo, M. (2007). A branch and bound algorithm for scheduling trains in a railway network. European Journal of Operational Research, 183, 643–657.

Daganzo, C. F., Anderson, P. (2016). Coordinating Transit Transfers in Real Time. Institute of Transportation Studies, UC Berkeley. https://escholarship.org/uc/item/25h4r974.

De Jonge, B., Scarf, P.A. (2019). A review on maintenance optimization. European Journal of Operational Research, in press.

De Souza, F., Verbas, O., Auld, J. (2019). Mesocopic traffic flow model for agent-based simulation. Procedia Computer Science, 151, 858-863.

Delgado, F., Munoz, J.C., Giesen, R. (2012). How much can holding and/or limiting boarding improve transit performance? Transportation Research Part B, 46, 1202-1217.

Delgado, F., Munoz, J.C., Giesen, R., Cipriano, A. (2009). Real-time control of buses in a transit corridor based on vehicle holding and boarding limits. Transportation Research Record, 2090, 59-67.

Derrible, S., Kennedy, C. (2010). The complexity and robustness of metro networks. Physica A, 389, 3678-3691.

Desaulniers, G., Hickman, M.D. (2007). Public Transit. In C. Barnhart and G. Laporte (eds.), Handbook in OR & MS (pp. 69-127). Amsterdam, the Netherlands: Elsevier.

Dewilde, T., Sels, P., Cattrysse, D., Vansteenwegen, P. (2009). Improving the robustness in railway station areas. European Journal of Operational Research, 235, 276-286.

Ding, C., Wang, D., Ma, X., Li, H. (2016). Predicting short-term subway ridership and prioritizing its influential factors using gradient boosting decision trees. Sustainability, 8, 1-16.

Dinh, T.N., Thai, M.T. (2014). Network under joint node and link attacks: Vulnerability assessment methods and analysis. IEEE/ACM Transactions on Networking, 23, 1001-1011.

Dollevoet, T., Corman, F., D'Ariano, A., Huisman, D. (2014). An iterative optimization framework for delay management and train rescheduling. Flexible Services and Manufacturing Journal, 26, 490-515.

Douglas, N., Karpouzis, G. (2005). Estimating the cost to passengers of station crowding. Proceedings of the 28th Australasian Transport Research Forum, Sydney, Australia.

El Mahrsi, M.K., Come, E., Oukhellou, L., Verleysen, M. (2017). Clustering smart card data for urban mobility analysis. IEEE Transactions on Intelligent Transportation Systems, 18, 712-728.

Engelson, L., Fosgerau, M. (2011). Additive measures of travel time variability. Transportation Research Part B, 45, 1560–1571.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. University of Munich, Munich, Germany.

Fiorenzo-Catalano, M.S. (2007). Choice set generation in multi-modal transportation networks (Ph.D. Thesis). Delft University of Technology, Delft, the Netherlands.

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

Fortunato, S., Hric, D. (2016). Community detection in networks: A user guide. Physics Reports, 659, 1-44.

Furth, P.G., Muller, T.H.J. (2009). Optimality conditions for public transport schedules with timepoint holding. Public Transport, 1, 87-102.

Gavriilidou, A., Cats, O. (2019). Reconciling transfer synchronization and service regularity: Real-time control strategies using passenger data. Transportmetrica A, 15, 215–243.

Gentile, G., Florian, M., Hamdouch, Y., Cats, O., Nuzzolo, A. (2016). The theory of transit assignment: Basic modelling frameworks. In G. Gentile and K. Noekel (eds.), Modelling public transport passenger flows in the era of intelligent transport systems (pp. 287-386). Cham, Switzerland: Springer International Publishing.

Geurs, K.T., La Paix, L., Van Weperen, S. (2016). A multi-modal network approach to model public transport accessibility impacts of bicycle-train integration policies. European Transport Research Review, 8, 1-15.

Gkiotsalitis, K., Cats, O. (2018). Reliable frequency determination: Incorporating information on service uncertainty when setting dispatching headways. Transportation Research Part C, 88, 187-207.

Gkiotsalitis, K., Eikenbroek, O.A.L., Cats, O. (2019). Robust network-wide bus scheduling with transfer synchronizations. IEEE Transactions on Intelligent Transportation Systems. DOI: 10.1109/TITS.2019.2941847.

Golob, T.F., Canty, E.T., Gustafson, R.L., Vitt, J.E. (1972). An analysis of consumer preferences for a public transportation system. Transportation Research, 6, 81-102.

Gordon, J.B., Koutsopoulos, H.N., Wilson, N.H.M., Attanucci, J.P. (2013). Automated inference of linked transit journeys in London using fare-transaction and vehicle location data.

Transportation Research Record, 2343, 17-24.

Goverde, R.M.P. (2005). Punctuality of railway operations and timetable stability analysis (Ph.D. Thesis). Delft University of Technology, Delft, the Netherlands.

GVB Holding NV (2019, Nov. 8). Jaarverslag 2018 (in Dutch). Retrieved from https://jaarverslag.gvb.nl/.

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

Hamdouch, Y., Ho, H.W., Sumalee, A., Wang, G. (2011). Schedule-based transit assignment model with vehicle capacity and seat availability. Transportation Research Part B, 45, 1805-1830.

Hänseler, F.S., Bierlaire, M., Scarinci, R. (2016). Assessing the usage and level-of-service of pedestrian facilities in train stations: A Swiss case study. Transportation Research Part A, 89, 106-123.

Haywood, L., Koning, M. (2015). The distribution of crowding costs in public transport: new evidence from Paris. Transportation Research Part A, 77, 182–201.

Hendren, P., Antos, J., Carney, Y., Harcum, R. (2015). Transit travel time reliability: Shifting the focus from vehicles to customers. Transportation Research Record, 2535, 35-44.

Hofmann, M., O'Mahony, M. (2005). Transfer journey identification and analyses from electronic fare collection data. Proceedings of the 8th IEEE International Conference on Intelligent Transportation Systems, Vienna, Austria.

Hollander, Y. (2006). Direct versus indirect models for the effects of unreliability. Transportation Research Part A, 40, 699-711.

Holmgren, Å.J. (2007). A framework for vulnerability assessment of electric power systems. In A.T. Murray and T.H. Grubesic (eds.), Critical infrastructure: Reliability and vulnerability (pp. 31-55). Berlin Heidelberg, Germany: Springer-Verlag.

Hörcher, D., Graham, D.J., Anderson, R.J. (2017). Crowding cost estimation with large scale smart card and vehicle location data. Transportation Research Part B, 95, 105-125.

HTM Personenvervoer N.V. (2016, July 21). Jaarverslag 2015 (in Dutch). Retrieved from https://www.htm.nl/media/364324/jaarverslag-2015.pdf.

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

Idris, A., Habib, K., Shalaby, A. (2015). An investigation on the performances of mode shift models in transit ridership forecasting. Transportation Research Part A, 78, 551-565.

Immers, B., Egeter, B., Snelder, M., Tampère, C. (2011). Reliability of travel times and robustness of transport networks. In M. Kutz (ed.), Handbook of transportation engineering (pp. 3.1-3.30). New York City, NY: McGraw-Hill.

Ingvardson, J.B., Nielsen, O.A., Raveau, S., Nielsen, B.F. (2018). Passenger arrival and waiting time distributions dependent on train service frequency and station characteristics: A smart card data analysis. Transportation Research Part C, 90, 292-306.

Jang, W. (2010). Travel time and transfer analysis using transit smart card data. Transportation Research Record, 2144, 142-149.

Jenelius, E. (2007). Incorporating dynamics and information in a consequence model for road network vulnerability analysis. Proceedings of the 3rd International Symposium on Transport Network Reliability (INSTR), The Hague, the Netherlands.

Jenelius, E., Cats, O. (2015). The value of new public transport links for network robustness and redundancy. Transportmetrica A, 11, 819-835.

Jenelius, E., Petersen, T., Mattsson, L.-G. (2006). Importance and exposure in road network vulnerability analysis. Transportation Research Part A, 40, 537-560.

Killeen, P., Ding, B., Kiringa, I., Yeap, T. (2019). IoT-based predictive maintenance for fleet management. Procedia Computer Science, 151, 607-613.

Knoop, V.L., Hoogendoorn, S.P., Van Zuylen, H.J. (2008). The influence of spillback modelling when assessing consequences of blockings in a road network. European Journal of Transportation and Infrastructure Research, 8, 287–300.

Knoop, V.L., Snelder, M., Van Zuylen, H.J., Hoogendoorn, S.P. (2012). Link-level vulnerability indicators for real-world networks. Transportation Research Part A, 46, 843-854.

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

Korteweg, J.A., Rienstra, S. (2010) De betekenis van robuustheid. Robuustheid in kosten-batenanalyses van weginfrastructuur (in Dutch). The Hague, the Netherlands: Kennisinstituut voor Mobiliteitsbeleid.

Kroes, E., Kouwenhoven, M., Debrincat, L., Pauget, N. (2014). On the value of crowding in public transport for Ile-de-France. Transport Research Arena, Paris, France.

Kroon, L., Maroti, G., Retel Helmrich, M., Vromans, M.J.C.M., Dekker, R. (2008). Stochastic improvement of cyclic railway timetables. Transportation Research Part B, 42, 553-570.

La Paix Puello, L., Geurs, K.T. (2016). Integration of unobserved effects in generalised transport access costs of cycling to railway stations. European Journal of Transport Infrastructure Research, 16, 385-405.

Lancichinetti, A., Fortunato, S. (2009). Community detection algorithms: A comparative analysis. Physical Review E - Statistical, Nonlinear, and Soft Matter Physics, 80, 56117.

Laskaris, G., Cats, O., Jenelius, E., Rinaldi, M., Viti, F. (2018). Multiline holding based control for lines merging to a shared corridor. Transportmetrica B, 7, 1062-1095.

Lee, A., Van Oort, N., Van Nes, R. (2014). Service reliability in a network context: impact of synchronizing schedules in long headway services. Transportation Research Record, 2417, 18-26.

Leng, N., De Martinis, V., Corman, F. (2018). Agent-based simulation approach for disruption management in rail schedule. Proceedings of the 14th Conference on Advanced Systems in Public Transport (CASPT), Brisbane, Australia

Li, M. (2008). Combining DTA approaches for studying road network robustness (Ph.D. Thesis). Delft University of Technology, Delft, the Netherlands.

Li, H., Gao, K., Tu, H. (2017a). Variations in mode-specific valuations of travel time reliability and in-vehicle crowding: Implications for demand estimation. Transportation Research Part A, 103, 250-263.

Li, Z., Hensher, D. (2011). Crowding and public transport: A review of willingness to pay evidence and its relevance in project appraisal. Transport Policy, 18, 880-887.

Li, Y., Wang, X., Sun, S., Ma, X. (2017b). Forecasting short-term subway passenger flow under special events scenarios using multiscale radial basis function networks. Transportation Research Part C, 77, 306-328.

Liao, F., Van Wee, B. (2017). Accessibility measures for robustness of the transport system. Transportation, 44, 1213-1233.

Liu, R., Sinha, S. (2007). Modelling urban bus service and passenger reliability. Proceedings of the 3rd International Symposium on Transport Network Reliability (INSTR), The Hague, the Netherlands.

Lord, D., Washington, S.P., Ivan, J.N. (2005). Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: Balancing statistical fit and theory. Accident Analysis & Prevention, 37, 35-46.

Lu, H., Burge, P., Heywood, C., Sheldon, R., Lee, P., Barber, K., Phillips, A. (2017). The impact of real-time information on passengers' value of bus waiting time. Transportation Research Procedia, 31, 18-34.

Lu, H., Fowkes, A.S., Wardman, M. (2008). Amending the incentive for strategic bias in stated preference studies: a case study in users' valuation of rolling stock. Transportation Research Record, 2049, 128-135.

Luo, D., Bonnetain, L., Cats, O., Van Lint, J.W.C. (2018). Constructing spatiotemporal load profiles of transit vehicles with multiple data sources. Transportation Research Record, 2672, 175-186.

Luo, D., Cats, O., Van Lint, J.W.C. (2017). Constructing transit origin-destination matrices with spatial clustering. Transportation Research Record, 2652, 39-49.

Ma, X., Wu, Y., Wang, Y., Chen, F., Liu, J. (2013). Mining smart card data for transit riders' travel patterns. Transportation Research Part C, 36, 1-12.

Malandri, C., Fonzone, A., Cats, O. (2018). Recovery time and propagation effects of passenger transport disruptions. Physica A: Statistical Mechanics and its Applications, 505, 7-17.

Markolf, S.A., Hoehne, C., Fraser, A., Chester, M.V., Shane Underwood, B. (2019). Transportation resilience to climate change and extreme weather events - Beyond risk and robustness. Transport Policy, 74, 174-186.

Marra, A.D., Corman, F. (2019). From delay to disruption: the impact of service degradation on public transport network. Proceedings of the 8th Symposium of the European Association for Research in Transportation (hEART), Budapest, Hungary.

Maslow, A.H. (1948). 'Higher' and 'lower' needs. Journal of Psychology, 25, 433-436.

McDaniels, T., Chang, S., Cole, D., Mikawoz, J., Longstaff, H. (2008). Fostering resilience to extreme events within infrastructure systems: Characterizing decision contexts for mitigation and adaptation. Global Environmental Change, 18, 310-318.

Metropoolregio Rotterdam Den Haag (MRDH) (2016, July 12). Openbaar vervoer (in Dutch). Retrieved from http://mrdh.nl/project/openbaar-vervoer.

Munizaga, M.A., Devillaine, F., Navarrete, C., Silva, D. (2014). Validating travel behaviour estimated from smart card data. Transportation Research Part C, 44, 70-79.

Munizaga, M.A., Palma, C. (2012). Estimation of a disaggregate multimodal public transport origin-destination matrix from passive smartcard data from Santiago, Chile. Transportation Research Part C, 24, 9-18.

Murray, A.T., Matisziw, T.C., Grubesic, T.H. (2008). A methodological overview of network vulnerability analysis. Growth and change: A journal of urban and regional policy, 39, 573-592.

MVA Consultancy (2008). Valuation of overcrowding on rail services. Report prepared for the UK Department of Transport.

Nash, A., Huerlimann, D. (2004). Railroad simulation using Open Track. Advances in Transport, 15, 45-54.

Nazem, M., Trépanier, M., Morency, C. (2011). Demographic analysis of route choice for public transit. Transportation Research Record, 2214, 71-78.

Nederlandse Spoorwegen (2019, Nov. 8). Geld terug bij vertraging (in Dutch). Retrieved from https://www.ns.nl/klantenservice/geld-terug/geld-terug-reguliere-reizen.html.

Nesheli, M.M., Ceder, A. (2015). A robust, tactic-based, real-time framework for public transport transfer synchronization. Transportation Research Part C, 60, 105-123.

Newman, M.E.J. (2004). Analysis of weighted networks. Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics, 70, 9.

Newman, M.E.J., Girvan, M. (2004). Finding and evaluating community structure in networks. Physical Review E - Statistical, Nonlinear, and Soft Matter Physics, 69, 26113.

Nicholson, A., Schmöcker, J.-D., Bell, M.G.H., Lida, Y. (2003). Assessing transport reliability: Malevolence and user knowledge. Proceedings of the 1st International Symposium on Transport Network Reliability (INSTR), Kyoto, Japan.

Nijënstein, S., Bussink, B. (2015). Combining multimodal smart card data: exploring quality improvements between multiple public transport systems. European Transport Conference, Frankfurt, Germany.

Nunes, A.A., Dias, T.G., eCunha, J.F. (2016). Passenger journey destination estimation from automated fare collection system data using spatial validation. IEEE Transactions on Intelligent Transportation Systems, 17, 133-142.

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

Obrenovic, N. (2019). TRANS-FORM D3.3: Tool for evaluating real-time strategies. TRANS-FORM project report. Lausanne: Switzerland.

Oliveira, E.L., Portugal, L.S., Junior, W.P. (2016). Indicators of reliability and vulnerability: Similarities and differences in ranking links of a complex road system. Transportation Research Part A, 88, 195-208.

Olsson, L.E., Friman, M., Pareigis, J., Edvardsson, B. (2012). Measuring service experience: Applying the satisfaction with travel scale in public transport. Journal of Retailing and Consumer Services, 19, 413-418.

Parbo, J., Nielsen, O.A., Landex, A., Prato, C.G. (2013). Measuring robustness, reliability and punctuality within passenger railway transportation - a literature review. KGS Lyngby, Denmark: Technical University of Denmark.

Paulsen, M., Rasmussen, T.K., Anker Nielsen, O. (2018). Modelling railway-induced passenger delays in multi-modal public transport networks. Proceedings of the 14th Conference on Advanced Systems in Public Transport (CASPT), Brisbane, Australia.

Pedregoas, F., Varoquaux, G., Gramfort, A., Michel, V. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825-2830.

Pel, A., Bel, N., Pieters, M. (2014). Including passengers' response to crowding in the Dutch national train passenger assignment model. Transportation Research Part A, 66, 111-126.

Pelletier, M.P., Trépanier, M., Morency, C. (2011). Smart card data use in public transit: A literature review. Transportation Research Part C, 19, 557-568.

Pimm, S.L. (1984). The complexity and stability of ecosystems. Nature, 307, 321-326.

Prud'homme, R., Koning, M., Lenormand, L., Fehr, A. (2012). Public transport congestion costs: the case of the Paris subway. Transport Policy, 21, 101-109.

Rietveld, P., Bruinsma, F.R., Van Vuuren, D.J. (2001). Coping with unreliability in public transport chains: a case study for Netherlands. Transportation Research Part A, 35, 539-559.

Rodriguez-Nunez, E., Garcia-Palomares, J.C. (2014). Measuring the vulnerability of public transport networks. Journal of Transport Geography, 35, 50-63.

Roelofsen, D., Cats, O., Van Oort, N., Hoogendoorn, S.P. (2018). Assessing disruption management strategies in rail-bound urban public transport systems from a passenger perspective. Proceedings of the 14th Conference on Advanced Systems in Public Transport (CASPT), Brisbane, Australia.

Rose, A. (2007). Economic resilience to natural and man-made disasters: Multidisciplinary origins and contextual dimensions. Environmental Hazards, 7, 383-398.

Saberi, M., Ghamami, M., Gu, Y., Shojaei, M.H., Fishman, E. (2018). Understanding the impacts of a public transit disruption on bicycle sharing mobility patterns: A case of Tube strike in London. Journal of Transport Geography, 66, 154-166.

Sánchez-Martinez, G.E. (2017). Inference of public transportation trip destinations by using fare transaction and vehicle location data: Dynamic programming approach. Transportation Research Record, 2652, 1-7.

Sanchez-Martinez, G.E., Koutsopoulos, H.N., Wilson, N.H.M. (2016). Real-time holding control for high-frequency transit with dynamics. Transportation Research Part B, 83, 1-19.

Savelberg, F., Bakker, P. (2010). Betrouwbaarheid en robuustheid op het spoor (in Dutch). The Hague, the Netherlands: Kennisinstituut voor Mobiliteitsbeleid.

Schakenbos, R., La Paix, L., Nijënstein, S., Geurs, K. (2016). Valuation of a transfer in a multimodal public transport trip. Transport Policy, 46, 72-81.

Schmaranzer, D., Braune, R., Doerner, K.F. (2019). Population-based simulation optimization for urban mass rapid transit networks. Flexible Services and Manufacturing Journal. DOI: 10.1007/s10696-019-09352-9.

Schmöcker, J.-D., Bell, M.G.H. (2002). The PFE as a tool for robust multi-modal network planning. Traffic Engineering Control, 44, 10-14.

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

Schöbel, A., Kratz, A. (2009). A bicriteria approach for robust timetabling. In R.K. Ahuja, R.H. Möhring and C.D. Zaroliagis (eds.), Robust and online large-scale optimization: Models and techniques for transportation systems. Lecture notes in computer science (pp. 119-144). Berlin Heidelberg, Germany: Springer.

Scott, D.M., Novak, D.C., Aultman-Hall, L., Guo, F. (2006). Network Robustness Index: A new method for identifying critical links and evaluating the performance of transportation networks. Journal of Transport Geography, 14, 215-227.

Seaborn, C., Attanucci, J., Wilson, N.H.M. (2009). Analyzing multimodal public transport journeys in London with smart card fare payment data. Transportation Research Record, 2121, 55-62.

Shakibayifar, M., Sheikholeslami, A., Corman, F. (2017). A simulation-based optimization approach to reschedule train traffic in uncertain conditions during disruptions. Scientia Iranica: International Journal of Science & Technology, 25, 646-662.

Shires, J.D., Ojeda-Cabral, M., Wardman, M. (2018). The impact of planned disruptions on rail passenger demand. Transportation. DOI: 10.1007/s11116-018-9889-0.

Significance, VU University, John Bates Services, TNO, NEA, TNS NIPO, PanelClix (2013). Values of time and reliability in passenger and freight transport in the Netherlands (Project No. 08064). The Hague, the Netherlands: Significance.

Snelder, M. (2010). Designing robust road networks. A general design method applied to the Netherlands (Ph.D. Thesis). Delft University of Technology, Delft, the Netherlands.

Stone, M., Broughton, J. (2003). Getting off your bike: Cycling accidents in Great Britain in 1990–1999. Accident Analysis Prevention, 35, 549-556.

Straits Times (2017, Aug. 7). Hong Kong's MTR faces $3.5m fine for 10-hour train delay. Retrieved from https://www.straitstimes.com/asia/east-asia/hks-mtr-faces-35m-fine-for-10-hour-train-delay.

Sullivan, J.L., Novak, D.C., Aultman-Hall, L., Scott, D.M. (2010). Identifying critical road segments and measuring system-wide robustness in transportation networks with isolating links: A link-based capacity-reduction approach. Transportation Research Part A, 44, 323-336.

Susilo, Y.O., Cats, O. (2014). Exploring key determinants of travel satisfaction for multi-modal trips by different traveller groups. Transportation Research Part A, 67, 366-380.

Tahmassby, S. (2009). Reliability in urban public transport network assessment and design (Ph.D. Thesis). Delft University of Technology, Delft, the Netherlands.

Tahmasseby, S., Van Oort, N., Van Nes, R. (2008) The role of infrastructures on public transport service reliability. Proceedings of the First International IEEE Conference on Infrastructure Systems and Services: Building Networks for a Brighter Future (INFRA), Rotterdam, the Netherlands.

Tampère, C.M.J., Stada, J., Immers, B., Peetermans, E., Organe, K. (2007). Methodology for identifying vulnerable sections in a national road network. Transportation Research Record, 2012, 1-10.

Tan, P.N., Steinbach, M., Kumar, V. (2004). Cluster analysis: Basic concepts and algorithms. In M. Kantardzic (ed.), Data mining: Concepts, models, methods, and algorithms (pp. 487-568). Hoboken, NJ: Wiley Press.

Taylor, M. (2017). Vulnerability analysis for transportation networks. Amsterdam, the Netherlands: Elsevier.

Tirachini, A., Hurtubia, R., Dekker, T., Daziano, R.A. (2017). Estimation of crowding discomfort in public transport: Results from Santiago de Chile. Transportation Research Part A, 103, 311-326.

Tirachini, A., Sun, L., Erath, A., Chakirov, A. (2016). Valuation of sitting and standing in metro trains using revealed preferences. Transport Policy, 47, 94-104.

Tonnelier, E., Baskiotis, N., Guigue, V., Gallinari, P. (2018). Anomaly detection in smart card logs and distant evaluation with Twitter: a robust framework. Neurocomputing, 298, 109-121.

Törnquist Krasemann, J. (2012). Design of an effective algorithm for fast response to the re-scheduling of railway traffic during disturbances. Transportation Research Part C, 20, 62-78.

Törnquist, J., Persson, J.A. (2007). N-tracked railway traffic re-scheduling during disturbances. Transportation Research Part B, 41, 342-362.

TRANS-FORM (2019, Nov. 15). Smart transfers through unravelling urban form and travel flow dynamics. Retrieved from http://www.trans-form-project.org/.

Transport for London (2019a, Oct. 17). Tube and DLR delays. Retrieved from https://tfl.gov.uk/fares/refunds-and-replacements/tube-and-dlr-delays.

Transport for London (2019b, Oct. 17). Underground service performance. Retrieved from https://tfl.gov.uk/corporate/publications-and-reports/underground-services-performance.

Trépanier, M., Tranchant, N., Chapleau, R. (2007). Individual trip destination estimation in a transit smart card automated fare collection system. Journal of Intelligent Transportation Systems, 11, 1-14.

Trépanier, M., Yamamoto, T. (2015). Workshop synthesis: System based passive data streams systems; Smart cards, phone data, GPS. Transportation Research Procedia, 11, 340-349.

Tu, W., Cao, R., Yue, Y., Zhou, B., Li, Q., Li, Q. (2018). Spatial variations in urban public ridership derived from GPS trajectories and smart card data. Journal of Transport Geography, 69, 45-57.

Turnquist, M.A., Bowman, L.A. (1980). The effects of network structure on reliability of transit service. Transportation Research Part B, 14, 79-86.

Van Exel, N.J.A., Rietveld, P. (2001). Public transport strikes and traveller behaviour. Transport Policy, 8, 237-246.

Van Hagen, M. (2011). Waiting experience at train stations (Ph.D. Thesis). University of Twente, Enschede, the Netherlands.

Van der Hurk, E., Kroon, L.G., Maróti, G. (2018). Passenger advice and rolling stock rescheduling under uncertainty for disruption management. Transportation Science, 52, 1391-1411.

Van der Hurk, E., Kroon, L.G., Maroti, G., Vervest, P.H.M. (2012). Dynamic forecasting model of time dependent passenger flows for disruption management. Proceedings of the 12th Conference on Advanced Systems in Public Transport (CASPT), Santiago, Chile.

Van Nes, R., Hansen, I.A., Winnips, C. (2014). Potentie multimodaal vervoer in stedelijke regio's (in Dutch). Delft, the Netherlands: DBR.

Van Nes, R., Marchau, V., Van Wee, G.P., Hansen, I.A. (2007). Reliability and robustness of multimodal transport network analysis and planning: towards a new research agenda. Proceedings of the 3rd International Symposium on Transport Network Reliability (INSTR), The Hague, the Netherlands.

Van Oort, N. (2011). Service reliability and urban public transport design (Ph.D. Thesis). TRAIL PhD Thesis Series, Delft, the Netherlands.

Van Oort, N. (2016). Incorporating enhanced service reliability of public transport in cost-benefit analyses. Public Transport, 8, 143-160.

Van Oort, N., Boterman, J.W., Van Nes, R. (2012). The impact of scheduling on service reliability: trip-time determination and holding points in long-headway services. Public Transport, 4, 39-56.

Van Oort, N., Brands, T., De Romph, E. (2015a). Short-term prediction of ridership on public transport with smart card data. Transportation Research Record, 2535, 105-111.

Van Oort, N., Brands, T., De Romph, E., Yap, M.D. (2016). Ridership evaluation and prediction in public transport by processing smart card data: A Dutch approach and example. In F. Kurauchi and J.-D. Schmöcker (eds.), Public Transport Planning with Smart Card Data (pp. 197-224). Boca Raton, FL: CRC Press.

Van Oort, N., Drost, M., Brands, T., Yap, M.D. (2015b). Data-driven public transport ridership prediction approach including comfort aspects. Proceedings of the 13th Conference on Advanced Systems in Public Transport (CASPT), Rotterdam, the Netherlands.

Van Oort, N., Van Nes, R. (2009). Regularity analysis for optimizing urban transit network design. Public transport, 1, 155-168.

Van Oort, N., Van Nes. R. (2010). Impact of rail terminal design on transit service reliability. Transportation Research Record, 2146, 109-118.

Van Oort, N., Wilson, N.H.M., Van Nes, R. (2010). Reliability improvement in short headway transit services: schedule-based and headway-based holding strategies. Transportation Research Record, 2143, 67-76.

Varga, B., Tettamanti, T., Kulcsár, B. (2018). Optimally combined headway and timetable reliable public transport system. Transportation Research Part C, 92, 1-26.

Von Ferber, C., Holovatch, T., Holovatch, Y., Palchykov, V. (2009). Public transport networks: Empirical analysis and modeling. The European Physical Journal B, 68, 261-275.

Vromans, M.J.C.M. (2005). Reliability of railway systems (Ph.D. Thesis). Erasmus University, Rotterdam, the Netherlands.

Vromans, M.J.C.M., Dekker, R., Kroon, L.G. (2006). Reliability and heterogeneity of railway services. European Journal of Operational Research, 172, 647-655.

Wang, W., Attanucci, J.P., Wilson, N.H.M. (2011). Bus passenger origin-destination estimation and related analyses using automated data collection systems. Journal of Public Transportation, 14, 131-150.

Wardman, M. (2004). Public transport values of time. Transport Policy, 11, 363-377.

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

Warffemius, P. (2013). De maatschappelijke waarde van kortere en betrouwbare reistijden (in Dutch). The Hague, the Netherlands: Kennisinstituut voor Mobiliteitsbeleid.

Washington Metropolitan Area Transit Authority (WMATA) (2019, Oct. 17). Rush Hour Promise. Retrieved from https://www.wmata.com/fares/smartrip/rush-hour-promise.cfm.

Wei, Y., Chen, M. (2012). Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks. Transportation Research Part C, 21, 148-162.

Weidmann, U. (1994). Der Fahrgastwechsel im öffentlichen Personenverkehr (in German). IVT no. 99. Zürich: Switzerland.

Whelan, G., Crockett, J. (2009). An investigation of the willingness to pay to reduce rail overcrowding. Proceedings of the 1st International Conference on Choice Modelling, Harrogate, England.

Xuan, Y., Argote, J., Daganzo, C.F. (2011). Dynamic bus holding strategies for schedule reliability: Optimal linear control and performance analysis. Transportation Research Part B, 45, 1831-1845.

Yap, M.D. (2014). Robust public transport from a passenger perspective: A study to evaluate and improve the robustness of multi-level public transport networks (M.Sc. Thesis). Delft University of Technology, Delft, the Netherlands.

Yap, M.D., Cats, O. (2019). Analysis and prediction of disruptions in metro networks. Proceedings of the 6th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), Krakow, Poland.

Yap, M.D., Cats, O., Törnquist Krasemann, J., Van Oort, N., Hoogendoorn, S.P. (2020). Quantification and control of disruption propagation in multi-level public transport networks. Presented at the 99th Annual Meeting of the Transportation Research Board (TRB), Washington, DC.

Yap, M.D., Cats, O., Van Arem, B. (2018a). Crowding valuation in urban tram and bus transportation based on smart card data. Transportmetrica A.
DOI: 10.1080/23249935.2018.1537319.

Yap, M.D., Cats, O., Van Oort, N., Hoogendoorn, S.P. (2017). A robust transfer inference algorithm for public transport journeys during disruptions. Transportation Research Procedia, 27, 1042-1049.

Yap, M.D., Luo, D., Cats, O., Van Oort, N., Hoogendoorn, S.P. (2019). Where shall we sync? Clustering passenger flows to identify urban public transport hubs and their key synchronization priorities. Transportation Research Part C, 98, 433-448.

Yap, M.D., Nijënstein, S., Van Oort, N. (2018b). Improving predictions of public transport usage during disturbances based on smart card data. Transport Policy, 61, 84-95.

Yap, M.D., Van Oort, N. (2018). Driver schedule efficiency vs. public transport robustness: A framework to quantify this trade-off based on passive data. Proceedings of the 14th Conference on Advanced Systems in Public Transport (CASPT), Brisbane, Australia.

Yap, M.D., Van Oort, N., Van Nes, R., Van Arem, B. (2015). Robustness of multi-level public transport networks: A methodology to quantify robustness from a passenger perspective. Proceedings of the 6th International Symposium on Transportation Network Reliability (INSTR), Nara, Japan.

Yap, M.D., Van Oort, N., Van Nes, R., Van Arem, B. (2018c). Identification and quantification of link vulnerability in multi-level public transport networks: a passenger perspective. Transportation, 45, 1161-1180.

Yildirimoglu, M., Kim, J. (2018). Identification of communities in urban mobility networks using multi-layer graphs of network traffic. Transportation Research Part C, 89, 254-267.

Younan, B., Wilson, N.H.M. (2010). Improving transit service connectivity: The application of operations planning and control strategies. 12th World Conference on Transport Research Society (WCTRS), Lisbon, Portugal.

Zhang, X., Guo, C., Wang, L. (2010). Using game theory to reveal vulnerability for complex networks. Proceedings of the 10th IEEE International Conference on Computer and Information Technology, Bradford, UK.

Zhang, F., Zhao, J., Tian, C., Xu, C., Liu, X., Rao, L. (2016). Spatiotemporal segmentation of metro trips using smart card data. IEEE Transactions on Vehicular Technology, 65, 1137-1149.

Zhao, J., Frumin, M., Wilson, N.H.M., Zhao, Z. (2013). Unified estimator for excess journey time under heterogeneous passenger incidence behavior using smartcard data. Transportation Research Part C, 34, 70-88.

Zhao, J., Rahbee, A., Wilson, N.H.M. (2007). Estimating a rail passenger trip origin-destination matrix using automatic data collection systems. Computer-Aided Civil and Infrastructure Engineering, 24, 376-387.

Zhu, Y., Koutsopoulos, H.N., Wilson, N.H.M. (2017). A probabilistic Passenger-to-Train Assignment Model based on automated data. Transportation Research Part B, 104, 522-542.

Ziha, K. (2000). Redundancy and robustness of systems of events. Probabilistic Engineering Mechanics, 15, 347-357.

Zou, X., Yue, W.L. (2017). A Bayesian network approach to causation analysis of road accidents using Netica. Journal of Advanced Transportation, 2017.
https://doi.org/10.1155 /2017/2525481.

About the Author

Menno Yap (1989) is born in The Hague, the Netherlands. In 2008, he started his Bachelor study Systems Engineering, Policy Analysis and Management at the Delft University of Technology, faculty of Technology, Policy and Management. He obtained his B.Sc. degree in 2011 *cum laude*. In 2014, he received his M.Sc. degree (*cum laude*) in Transport, Infrastructure & Logistics from the Delft University of Technology, where he specialised in the design of public transport networks. Menno also completed a Minor programme in Social and Organisational Psychology at the Leiden University. Besides, he completed his Propaedeutic exam in Medicine at the Leiden University in 2008. In 2013, he spent two months in Guangzhou, China, working on an interdisciplinary research project towards the sustainability of the Port of Guangzhou.

Menno is passionate about public transport. His aim is to understand the challenges the public transport industry faces, develop scientific methods to contribute solving these, and to make sure solutions can be implemented back in the industry. In his professional life, Menno is active in both academia and the public transport industry. After finishing his studies, he worked as public transport consultant for Goudappel Coffeng B.V. from 2014 to 2017. In this role, he was involved in many projects for different public transport operators in the Netherlands. In April 2016 he started his Ph.D. research at the Delft University of Technology, which he fulfilled in a 0.5 FTE part-time appointment. In 2017 he moved to London, where he worked for Atkins Ltd as public transport modelling consultant. He joined Transport for London (TfL) in 2018 as public transport planner and modeller in the Transport Modelling team, whilst still doing his Ph.D. research in the Netherlands. He completed his Ph.D. research in February 2020.

In his spare time, Menno is active as volunteer at The Hague's public transport museum. In this role, he was responsible for the realisation and implementation of a historic hop-on hop-off tram in The Hague. He is also a qualified tram driver for historical trams in The Hague, thereby maintaining a close connection with the operational side of public transport.

List of Publications

Book Chapters

- Van Oort, N., Brands, T., De Romph, E., Yap, M.D. (2016). Ridership evaluation and prediction in public transport by processing smart card data: A Dutch approach and example. In F. Kurauchi and J.-D. Schmöcker (eds.), *Public Transport Planning with Smart Card Data* (pp. 197-224). Boca Raton, FL: CRC Press.

Journal Articles

- Yap, M.D., Luo, D., Cats, O., Van Oort, N., Hoogendoorn, S.P. (2019). Where shall we sync? Clustering passenger flows to identify urban public transport hubs and their key synchronization priorities. *Transportation Research Part C, 98,* 433-448.
- Yap, M.D., Munizaga, M. (2019). Workshop report 8: Big data in the digital age and how it can benefit public transport users. *Research in Transport Economics, 69,* 615-620.
- Yap, M.D., Cats, O., Van Arem, B. (2018). Crowding valuation in urban tram and bus transportation based on smart card data. *Transportmetrica A.* DOI: 10.1080/23249935.2018.1537319.
- Yap, M.D., Van Oort, N., Van Nes, R., Van Arem, B. (2018). Identification and quantification of link vulnerability in multi-level public transport networks: a passenger perspective. *Transportation, 45,* 1161-1180.
- Yap, M.D., Nijënstein, S., Van Oort, N. (2018). Improving predictions of public transport usage during disturbances based on smart card data. *Transport Policy, 61,* 84-95.
- Scheltes, A.F., Yap, M.D., Van Oort, N. (2017). Reizigerspotentie en -voorkeuren ten aanzien van zelfrijdende voertuigen op de last-mile in een openbaar vervoer reis (in Dutch). *Tijdschrift Vervoerwetenschap, 53,* 3-8.

- Lee, A., Yap, M.D., Van Oort, N. (2017). Overstappen en onbetrouwbaarheid in het OV: Een methode voor kwantificering van reizigerseffecten in een netwerkcontext (in Dutch). *Tijdschrift Vervoerwetenschap, 53*, 36-53.
- Yap, M.D., Correia, G., Van Arem, B. (2016). Preferences of travellers for using automated vehicles as last mile public transport of multimodal train trips. *Transportation Research Part A, 94*, 1-16.
- Cats, O., Yap, M.D., Van Oort, N. (2016). Exposing the role of exposure: Public transport network risk analysis. *Transportation Research Part A, 88*, 1-14.
- Yap, M.D., Van Oort, N., Van Nes, R., Van Arem, B. (2015). Robuustheid van multi-level openbaar vervoer netwerken: Een methodologie om (on)robuustheid te kwantificeren vanuit een reizigersperspectief (in Dutch). *Tijdschrift Vervoerwetenschap, 51*, 82-99.

Under Review

- Yap, M.D., Cats, O. (under review). Predicting disruptions and their passenger delay impacts for public transport stops.
- Yap, M.D., Cats, O., Törnquist Krasemann, J., Van Oort, N., Hoogendoorn, S.P. (under review). Quantification and control of disruption propagation in multi-level public transport networks.

Peer-reviewed Conference Contributions

- Yap, M.D., Cats, O. (2020). Predicting disruption exposure and impact to assess station criticality in a public transport vulnerability analysis. *Presented at the 99th Annual Meeting of the Transportation Research Board (TRB)*, Washington, DC.
- Yap, M.D., Cats, O., Törnquist Krasemann, J., Van Oort, N., Hoogendoorn, S.P. (2020). Quantification and control of disruption propagation in multi-level public transport networks. *Presented at the 99th Annual Meeting of the Transportation Research Board (TRB)*, Washington, DC.
- Yap, M.D., Cats, O. (2019). Analysis and prediction of disruptions in metro networks. *Proceedings of the 6th International Conference on Models and Technologies for Intelligent Transport Systems (MT-ITS)*, Krakow, Poland.
- Yap, M.D., Cats, O. (2019). Predicting and clustering station vulnerability in urban networks. *Presented at the 5th International Workshop and Symposium of TransitData*, Paris, France.
- Yap M.D., Cats, O., Van Oort, N., Hoogendoorn, S.P. (2018). Controlling the propagation of passenger disruption impacts in multi-level public transport networks. *Presented at the International Conference on Operations Research (OR2018)*, Brussels, Belgium.
- Yap M.D., Cats, O., Van Oort, N., Hoogendoorn, S.P. (2018). Controlling the propagation of passenger disruption impacts in multi-level public transport networks. *Presented at the 7th European Symposium on Quantitative Methods in Transportation Systems (hEART)*, Athens, Greece.
- Yap, M.D., Van Oort, N. (2018). Driver schedule efficiency vs. public transport robustness: A framework to quantify this trade-off based on passive data. *Proceedings of the 14th Conference on Advanced Systems in Public Transport (CASPT)*, Brisbane, Australia.

- Yap, M.D., Luo, D., Cats, O. (2018). Using passenger flows to determine key interchange connections for public transport synchronization. *Proceedings of the 14th Conference on Advanced Systems in Public Transport (CASPT),* Brisbane, Australia.
- Brands, T., Van Oort, N., Yap, M.D. (2018). Automatic bottleneck detection using AVL data: A case study in Amsterdam. *Proceedings of the 14th Conference on Advanced Systems in Public Transport (CASPT),* Brisbane, Australia.
- Yap, M.D., Cats, O., Van Oort, N., Hoogendoorn, S.P. (2017). Robust transfer inference: a transfer inference algorithm for public transport journeys during disruptions. *Transportation Research Procedia, 27,* 1042-1049.
- Yap, M.D., Cats, O., Yu, S., Van Arem, B. (2017). Crowding valuation in urban tram and bus transportation based on smart card data. *Presented at the 15th International Conference on Competition and Ownership in Land Passenger Transport (Thredbo 15),* Stockholm, Sweden.
- Yap, M.D., Nijënstein, S., Van Oort, N. (2017). Improving predictions of the impact of disturbances on public transport usage based on smart card data. *Proceedings of the 96th Annual Meeting of the Transportation Research Board (TRB),* Washington, DC.
- Yap, M.D., Correia, G., Van Arem, B. (2015). Valuation of travel attributes for using automated vehicles as egress transport of multimodal train trips. *Transportation Research Procedia, 10,* 462-471.
- Van Oort, N., Drost, M., Brands, T., Yap, M.D. (2015). Data-driven public transport ridership prediction approach including comfort aspects. *Proceedings of the 13th Conference on Advanced Systems in Public Transport (CASPT),* Rotterdam, the Netherlands.
- Yap, M.D., Van Oort, N., Van Nes, R., Van Arem, B. (2015). Robustness of multi-level public transport networks: A methodology to quantify robustness from a passenger perspective. *Proceedings of the 6th International Symposium on Transportation Network Reliability (INSTR),* Nara, Japan.
- Cats, O., Yap, M.D., Van Oort, N. (2015). Exposing the role of exposure: Identifying and evaluating critical links in public transport networks. *Proceedings of the 6th International Symposium on Transportation Network Reliability (INSTR),* Nara, Japan.

Professional Magazines and Newspapers

- Yap, M.D. (2018). OV-reis lijkt fors langer bij drukte (in Dutch). *OVPro.nl*
- Yap, M.D. (2018). Wat kost een OV-storing de reiziger en maatschappij werkelijk? (in Dutch) *OVPro.nl*
- Van Oort, N., Yap, M.D. (2018). AI-technieken winnen terrein (in Dutch). *OV Magazine.*
- Yap, M.D. (2018). Wachten op de cijfers van Menno (in Dutch). *AD Newspaper.*
- Van Oort, N., Cats, O., Yap, M.D. (2017). Vervoer op afroep is niet meer te stuiten (in Dutch). *OV Magazine.*
- Yap, M.D., Van Oort, N. (2014). Robuust OV, uitgedrukt in euro's (in Dutch). *OV Magazine.*

Awards, Nominations and Recognitions

- Honourable Mention Michael Beesley Award: Best workshop paper presented by a young researcher. *Awarded at the 5th International Conference on Competition and Ownership in Land Passenger Transport (Thredbo15, 2017),* Stockholm, Sweden.
  Awarded paper: Yap, M.D., Cats, O., Yu, S., Van Arem, B. (2017). Crowding valuation in urban tram and bus transportation based on smart card data.
- 1st price for best paper CVS 2016. *Awarded at the Colloquium Vervoersplanologisch Speurwerk 2016: Hoe slim is smart nu eigenlijk (in Dutch)? (CVS, 2016),* Zwolle, the Netherlands.
  Awarded paper: Scheltes, A.F., Yap, M.D., Van Oort, N. (2017). Het verbeteren van de last-mile in een OV reis met automatische voertuigen: Een Delftse case studie en stated preference onderzoek gecombineerd (in Dutch).
- Runner-up Cuperusprijs for best Master Thesis in the field of transport planning. *Awarded at the Nationaal Verkeerskunde Congres (NVC, 2016),* Zwolle, the Netherlands.
  Awarded work: Yap, M.D. (2014). Robust public transport from a passenger perspective: A study to evaluate and improve the robustness of multi-level public transport networks.

TRAIL Thesis Series

The following list contains the most recent dissertations in the TRAIL Thesis Series. For a complete overview of more than 250 titles see the TRAIL website: www.rsTRAIL.nl.

The TRAIL Thesis Series is a series of the Netherlands TRAIL Research School on transport, infrastructure and logistics.

Yap, M.D., *Measuring, Predicting and Controlling Disruption Impacts for Urban Public Transport*, T2020/3, February 2020, TRAIL Thesis Series, the Netherlands

Luo, D., *Data-driven Analysis and Modeling of Passenger Flows and Service Networks for Public Transport Systems*, T2020/2, February 2020, TRAIL Thesis Series, the Netherlands

Erp, P.B.C. van, *Relative Flow Data: New opportunities for traffic state estimation*, T2020/1, February 2020, TRAIL Thesis Series, the Netherlands

Zhu, Y., *Passenger-Oriented Timetable Rescheduling in Railway Disruption Management*, T2019/16, December 2019, TRAIL Thesis Series, the Netherlands

Chen, L., *Cooperative Multi-Vessel Systems for Waterborne Transport*, T2019/15, November 2019, TRAIL Thesis Series, the Netherlands

Kerkman, K.E., *Spatial Dependence in Travel Demand Models: Causes, implications, and solutions*, T2019/14, October 2019, TRAIL Thesis Series, the Netherlands

Liang, X., *Planning and Operation of Automated Taxi Systems*, T2019/13, September 2019, TRAIL Thesis Series, the Netherlands

Ton, D., *Unravelling Mode and Route Choice Behaviour of Active Mode Users*, T2019/12, September 2019, TRAIL Thesis Series, the Netherlands

Shu, Y., *Vessel Route Choice Model and Operational Model Based on Optimal Control,* T2019/11, September 2019, TRAIL Thesis Series, the Netherlands

Luan, X., *Traffic Management Optimization of Railway Networks,* T2019/10, July 2019, TRAIL Thesis Series, the Netherlands

Hu, Q., *Container Transport inside the Port Area and to the Hinterland,* T2019/9, July 2019, TRAIL Thesis Series, the Netherlands

Andani, I.G.A., *Toll Roads in Indonesia: transport system, accessibility, spatial and equity impacts,* T2019/8, June 2019, TRAIL Thesis Series, the Netherlands

Ma, W., *Sustainability of Deep Sea Mining Transport Plans,* T2019/7, June 2019, TRAIL Thesis Series, the Netherlands

Alemi, A., *Railway Wheel Defect Identification,* T2019/6, January 2019, TRAIL Thesis Series, the Netherlands

Liao, F., *Consumers, Business Models and Electric Vehicles,* T2019/5, May 2019, TRAIL Thesis Series, the Netherlands

Tamminga, G., *A Novel Design of the Transport Infrastructure for Traffic Simulation Models,* T2019/4, March 2019, TRAIL Thesis Series, the Netherlands

Lin, X., *Controlled Perishable Goods Logistics: Real-time coordination for fresher products*, T2019/3, January 2019, TRAIL Thesis Series, the Netherlands

Dafnomilis, I., *Green Bulk Terminals: A strategic level approach to solid biomass terminal design*, T2019/2, January 2019, TRAIL Thesis Series, the Netherlands

Feng, F., *Information Integration and Intelligent Control of Port Logistics System,* T2019/1, January 2019, TRAIL Thesis Series, the Netherlands

Beinum, A.S. van, *Turbulence in Traffic at Motorway Ramps and its Impact on Traffic Operations and Safety,* T2018/12, December 2018, TRAIL Thesis Series, the Netherlands

Bellsolà Olba, X., *Assessment of Capacity and Risk: A Framework for Vessel Traffic in Ports,* T2018/11, December 2018, TRAIL Thesis Series, the Netherlands

Knapper, A.S., *The Effects of using Mobile Phones and Navigation Systems during Driving,* T2018/10, December 2018, TRAIL Thesis Series, the Netherlands

Varotto, S.F., *Driver Behaviour during Control Transitions between Adaptive Cruise Control and Manual Driving: empirics and models,* T2018/9, December 2018, TRAIL Thesis Series, the Netherlands