

Quantifying the Effectiveness of Temporary Railway Timetables under Planned Track Closures

A Data-Driven Multi-Criteria Framework for Temporary Timetable Evaluation

Femke Tamsma

Delft University of Technology

Quantifying the Effectiveness of Temporary Railway Timetables under Planned Track Closures

A Data-Driven Multi-Criteria Framework for Temporary Timetable Evaluation

by

Femke Tamsma

to obtain the degree of Master of Science in Transport, Infrastructure, and Logistics
at the Delft University of Technology,
to be defended publicly on Thursday March 5, 2026 at 11:00 AM.

Student Number: 5122422
Project Duration: September, 2025 – March, 2026
Thesis Committee: Dr. ir. E. Quaglietta, TU Delft, chair
Dr. J. A. Annema, TU Delft, supervisor
Ir. Z. Wang, TU Delft, additional examiner
Dr. D. Middelkoop, ProRail, daily supervisor

Source of Cover: (ProRail, [2017](#))

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

Preface

Dear reader,

My academic journey at Delft University of Technology did not follow a straight line. When I first started studying Civil Engineering at the age of 18 years, I was strongly drawn to mathematical problems, transport networks, and specifically railway systems. However, during the bachelor program I found that large parts of the curriculum focuses on civil engineering parts that did not match my interests. Therefore, I made the difficult decision to stop this study and pursue a bachelor's degree in Earth Sciences instead.

This turned out to be a very good choice. I greatly enjoyed my bachelor's degree and learned a lot about landscapes and the system earth. However, during these years, my interest in transport and logistics never truly disappeared. With a mixture of excitement and hesitation, I visited the master open day of "Transport, Infrastructure, and Logistics". Despite my earlier experience in Delft and Civil Engineering, this immediately felt like the right choice. My fascination with railways only grew stronger when studying this topic in depth. About a year before starting my graduation project, I was introduced to timetable simulation through Egidio Quaglietta and Dick Middelkoop. At that moment, it became clear to me that this was the direction in which I wanted to graduate and I hoped to be able to combine it with an internship at ProRail. Therefore, when I entered the iconic Inktpot building for the first time, this was a very special and existing moment, especially after cycling past it almost daily for years.

Now that my graduation has come to an end, I look back with a lot of excitement and joy. Being part of ProRail taught me a lot about how large organizations work and gave me a glimpse into working life. I am very grateful to Dick Middelkoop for offering me the opportunity to conduct my graduation project in such a professional environment and to supervise me throughout the project. I would also like to thank the I&TV department for their warm welcome, practical tips, words of encouragement when I was feeling down, and all the pleasant company. Special thanks also to my supervisors Egidio Quaglietta and Jan Anne Annema for their guidance, feedback, and support throughout this research. Your insights really challenged me to continuously improve my work. Lastly, I also would like to thank Johannes Lok and Michiel Stabèl, whose expertise on temporary timetable, planned track closures, simulation, and data analysis was invaluable to my thesis. Our discussions greatly improved my understanding of both the technical and operational aspects of railway timetables.

This thesis completes my master's program in Transport, Infrastructure, and Logistics. I am looking forward to step away from timetables and simulations for a while to explore the landscapes of Peru. After that, I hope to continue my career within the railway sector, building further on the knowledge and experience I gained during this internship.

I hope you enjoy reading this thesis!

*Femke Tamsma (5122422)
Delft, February 2026*

Contents

Preface	i
Summary	viii
Samenvatting	ix
Abbreviations	xi
1 Introduction	1
1.1 Background and Context	1
1.2 Problem Definition, Research Gap, and Objectives	2
1.3 Report Outline	3
2 Theoretical Background and Literature Review	4
2.1 Railway Timetable Fundamentals	4
2.1.1 Nominal and Temporary Timetables	4
2.1.2 Stochastic and Deterministic Timetabling	5
2.2 Timetable Definitions and Characteristics	5
2.2.1 Relation Between Timetable Attributes	6
2.3 Timetable Performance Indicators	7
2.3.1 Ex-ante vs. Ex-post	7
2.3.2 Methodological Evaluation Approaches	8
2.3.3 Scientific and Practical Timetable KPIs	8
2.3.4 Temporary Timetable KPIs	10
2.4 Timetable Design Methods	10
2.4.1 Current Timetabling Practice	10
2.4.2 Approaches for Optimized and Automatic Timetabling	12
2.5 Conclusion and Research Scope	13
2.5.1 Conclusion	13
2.5.2 Research Scope and Contributions	13
3 Methodology	15
3.1 Methodological Approach	15
3.1.1 Identification of Relevant Characteristics	15
3.1.2 Defining Effectiveness of Temporary Timetables	15
3.1.3 Translating Characteristics into Indices	15
3.2 Developing a Temporary Timetable Effectiveness Indicator	17
3.2.1 Relevant Characteristics	17
3.2.2 Design Requirements and Definition	17
3.2.3 Defining Indices	18
3.2.4 Towards an Integrated Effectiveness Indicator	24
3.2.5 Stress Testing	24
4 Timetable Design Analysis	25
4.1 Timetable Performance Analysis Method	25
4.1.1 Data Sources	26
4.1.2 Data Processing	26
4.1.3 Data Verification	27
4.2 Qualitative Factor Analysis	27
4.2.1 Factors Based on Planning Norms	27
4.2.2 Factors Specific to Temporary Timetables	28
4.2.3 Inter-dependencies and Combined Effects	30

4.2.4	Implications for Timetable Effectiveness	30
4.3	Quantitative Data-Analysis	30
4.3.1	Data Description	30
4.3.2	PAB Data-Analysis	31
4.3.3	Fitted Disturbance Distributions	33
5	Case Study	35
5.1	Case Definition	35
5.1.1	Study Area	35
5.1.2	Corridor Characteristics	36
5.1.3	Description of PTC	38
5.1.4	Description of Temporary Timetables	38
5.2	Case Study Setup	38
5.2.1	Defining Scenarios and Experiments	39
5.2.2	Driving Behavior and Operational Assumptions	40
5.2.3	Time Horizon and Warm-up Phase	40
5.3	Model Verification and Validation	41
5.4	Results	43
5.4.1	Timetable Assessment per Index	43
5.4.2	Effectiveness Indicator Results	49
5.4.3	Sensitivity Analysis	49
5.4.4	Stress Testing	52
6	Discussion	53
6.1	Interpretation of Results	53
6.2	Reflection on Contributions	53
6.3	Limitations	54
7	Conclusion and Future Research	55
7.1	Practical Recommendations and Guidelines	55
7.2	Conclusion and Future Research	55
7.2.1	Conclusion	55
7.2.2	Future Research	56
	References	57
A	Fitted Distributions	60

List of Figures

2.1	Factors influencing railway timetable robustness (reproduced from (Salido et al., 2008)).	7
3.1	Overview of the methodological process, including methods (blue), and inputs and outputs (yellow).	16
4.1	Overview of data sources and their outputs.	25
4.2	Overview of data processing steps.	26
4.3	Schematic overview of different types of Planned Track Closures (PTCs).	29
4.4	Overview of PAB data set, indicating all variables (columns).	30
4.5	Arrival delay percentiles of PAB data.	31
4.6	Drive through delay percentiles of PAB data.	31
4.7	Arrival delay distribution of PAB data.	32
4.8	Drive through delay distribution of PAB data.	32
4.9	Moving average delays over time of PAB data.	32
4.10	Delays per timetable point and service.	32
5.1	Geographical representation of study area in the southeast Netherlands.	35
5.2	Schematic representation of passenger and freight service plan for the study area.	36
5.3	Schematic representation of infrastructure layout for the corridor Btl-Ehv (Sporenplan.nl, 2025).	37
5.4	Time-Distance diagram of the Boxtel-Eindhoven Centraal corridor.	38
5.5	Schematic representation of parts of the infrastructure between Btl-Ehv during the PTC.	39
5.6	Arrival delay distribution of 'Undisturbed', 'Disturbed', and 'PAB'.	41
5.7	Arrival delays percentiles.	42
5.8	Departure delays percentiles.	42
5.9	Moving average delays over times 'Undisturbed'.	42
5.10	Moving average delays over time 'Disturbed'.	42
5.11	Output delay dynamics under deterministic disturbance scenarios for timetables nominal, AUP-A, AUP-B, and AUP-C.	44
5.12	Box plot of robustness scores under stochastic disturbances for timetables nominal, AUP-A, AUP-B, and AUP-C.	45
5.13	ECDF of robustness scores under stochastic disturbances for timetables nominal, AUP-A, AUP-B, and AUP-C.	45
5.14	Time-distance diagram of the corridor Boxtel (Btl) - Eindhoven Centraal (Ehv) for temporary timetable AUP-A.	46
5.15	Time-distance diagram of the corridor Eindhoven Centraal (Ehv) - Boxtel (Btl) for temporary timetable AUP-A.	46
5.16	Time-distance diagram of the corridor Boxtel (Btl) - Eindhoven Centraal (Ehv) for temporary timetable AUP-B.	46
5.17	Time-distance diagram of the corridor Eindhoven Centraal (Ehv) - Boxtel (Btl) for temporary timetable AUP-B.	47
5.18	Time-distance diagram of the corridor Boxtel (Btl) - Eindhoven Centraal (Ehv) for temporary timetable AUP-C.	47
5.19	Time-distance diagram of the corridor Eindhoven Centraal (Ehv) - Boxtel (Btl) for temporary timetable AUP-C.	47
5.20	Decomposition of operability into infeasible cycle time, total time shift required to solve all conflicts, and remaining slack for timetables nominal, AUP-A, AUP-B, and AUP-C per corridor direction.	48
5.21	Sensitivity of robustness scores to λ for timetables nominal, AUP-A, AUP-B, and AUP-C.	50

5.22	Sum of differences of R_r between nominal and temporary timetables AUP-A, AUP-B, and AUP-C to λ .	50
5.23	Output delay dynamics under deterministic disturbance scenarios for timetables nominal, AUP-A, AUP-B, and AUP-C with varying driving behavior MOP and IOV.	51
5.24	Distribution of robustness scores under stochastic disturbances for timetables nominal, AUP-A, AUP-B, and AUP-C with varying driving behavior MOP and IOV	51
5.25	Timetable effectiveness dynamics under decreasing reference periods for timetables nominal, AUP-A, AUP-B and AUP-C.	52
A.1	Histogram with theoretical PDF, CDF and Q-Q plot of A6400 train at Best (Bet).	60
A.2	Histogram with theoretical PDF, CDF and Q-Q plot of A6400 train at Boxtel (Btl).	60
A.3	Histogram with theoretical PDF, CDF and Q-Q plot of A6400 train at Eindhoven Strijp-S (Ehs).	61
A.4	Histogram with theoretical PDF, CDF and Q-Q plot of A6400 train at Eindhoven Centraal (Ehv).	61
A.5	Histogram with theoretical PDF, CDF and Q-Q plot of B6400 train at Best (Bet).	61
A.6	Histogram with theoretical PDF, CDF and Q-Q plot of B6400 train at Boxtel (Btl).	62
A.7	Histogram with theoretical PDF, CDF and Q-Q plot of B6400 train at Eindhoven Strijp-S (Ehs).	62
A.8	Histogram with theoretical PDF, CDF and Q-Q plot of B6400 train at Eindhoven Centraal (Ehv).	62
A.9	Histogram with theoretical PDF, CDF and Q-Q plot of E1100 train at Eindhoven Centraal (Ehv).	63
A.10	Histogram with theoretical PDF, CDF and Q-Q plot of E3500 train at Eindhoven Centraal (Ehv).	63
A.11	Histogram with theoretical PDF, CDF and Q-Q plot of F3500 train at Eindhoven Centraal (Ehv).	63
A.12	Histogram with theoretical PDF, CDF and Q-Q plot of G1100 train at Eindhoven Centraal (Ehv).	64
A.13	Histogram with theoretical PDF, CDF and Q-Q plot of H3500 train at Eindhoven Centraal (Ehv).	64
A.14	Histogram with theoretical PDF, CDF and Q-Q plot of H6400 train at Best (Bet).	64
A.15	Histogram with theoretical PDF, CDF and Q-Q plot of H6400 train at Boxtel (Btl).	65
A.16	Histogram with theoretical PDF, CDF and Q-Q plot of H6400 train at Eindhoven Strijp-S (Ehs).	65
A.17	Histogram with theoretical PDF, CDF and Q-Q plot of H6400 train at Eindhoven Centraal (Ehv).	65
A.18	Histogram with theoretical PDF, CDF and Q-Q plot of I2700 train at Eindhoven Centraal (Ehv).	66
A.19	Histogram with theoretical PDF, CDF and Q-Q plot of I6500 train at Best (Bet).	66
A.20	Histogram with theoretical PDF, CDF and Q-Q plot of I6500 train at Boxtel (Btl).	66
A.21	Histogram with theoretical PDF, CDF and Q-Q plot of I6500 train at Eindhoven Strijp-S (Ehs).	67
A.22	Histogram with theoretical PDF, CDF and Q-Q plot of I6500 train at Eindhoven Centraal (Ehv).	67
A.23	Histogram with theoretical PDF, CDF and Q-Q plot of J2700 train at Eindhoven Centraal (Ehv).	67
A.24	Histogram with theoretical PDF, CDF and Q-Q plot of J6500 train at Best (Bet).	68
A.25	Histogram with theoretical PDF, CDF and Q-Q plot of J6500 train at Boxtel (Btl).	68
A.26	Histogram with theoretical PDF, CDF and Q-Q plot of J6500 train at Eindhoven Strijp-S (Ehs).	68
A.27	Histogram with theoretical PDF, CDF and Q-Q plot of J6500 train at Eindhoven Centraal (Ehv).	69
A.28	Histogram with theoretical PDF, CDF and Q-Q plot of K2700 train at Eindhoven Centraal (Ehv).	69
A.29	Histogram with theoretical PDF, CDF and Q-Q plot of K6500 train at Best (Bet).	69

A.30 Histogram with theoretical PDF, CDF and Q-Q plot of K6500 train at Boxtel (Btl).	70
A.31 Histogram with theoretical PDF, CDF and Q-Q plot of K6500 train at Eindhoven Strijp-S (Ehs).	70
A.32 Histogram with theoretical PDF, CDF and Q-Q plot of K6500 train at Eindhoven Centraal (Ehv).	70
A.33 Histogram with theoretical PDF, CDF and Q-Q plot of L2700 train at Eindhoven Centraal (Ehv).	71
A.34 Histogram with theoretical PDF, CDF and Q-Q plot of L6500 train at Best (Bet).	71
A.35 Histogram with theoretical PDF, CDF and Q-Q plot of L6500 train at Boxtel (Btl).	71
A.36 Histogram with theoretical PDF, CDF and Q-Q plot of L6500 train at Eindhoven Strijp-S (Ehs).	72
A.37 Histogram with theoretical PDF, CDF and Q-Q plot of L6500 train at Eindhoven Centraal (Ehv).	72

List of Tables

2.1	Overview of key literature on timetable design and evaluation methods.	14
3.1	Mathematical notation for the TTE framework	18
4.1	Comparison of key factors influencing timetable effectiveness under nominal and PTC conditions.	29
4.2	Overview of fitted disturbance distributions, derived from PAB analysis.	34
5.1	Capacity-related metrics for nominal and alternative temporary timetables.	43
5.2	Operability-related metrics for nominal and alternative temporary timetables.	48
5.3	Integrated timetable effectiveness results under different weight compositions.	49

Summary

Railway operations increasingly rely on temporary timetables to accommodate planned track closures (PTCs) due to growing maintenance needs, infrastructure renewal programs, capacity pressure, and stricter European regulations regarding early timetable publication. During such closures, infrastructure capacity is temporarily reduced, and service patterns are adapted through the implementation of temporary timetables. These temporary schedules operated under more constrained conditions than nominal timetables, making them more vulnerable to infeasibility, delay propagation, and performance degradation. In practice, multiple temporary timetable designs are possible for a single PTC, each reflecting different operational trade-offs. The selection between these alternatives is currently based on expert judgment and experience, without a transparent quantitative framework that accounts for stochastic operational variability.

The majority of the literature provides a wide range of indices and evaluation methods to assess the performance of nominal railway schedules. These approaches are developed for stable operating conditions and deterministic planning assumptions. Therefore, their direct application to temporary timetables is limited. Temporary schedules are designed according to planning norms and design principles to ensure acceptable performance under constrained conditions. However, there are no specified indices or evaluation methods to assess the performance level of temporary timetables. This reveals a clear scientific and practical knowledge gap: the lack of a systematic framework to quantify the effectiveness of temporary timetables operating under PTC conditions.

To address this gap, this research develops a methodological framework to quantify the effectiveness of temporary railway timetables. The approach follows a three-step methodology. First, effectiveness is formally defined as the ability of a temporary timetable to maintain operable, stable, and robust train operations under deterministic and stochastic operational variability. Second, existing timetable attributes capacity, stability and robustness are reformulated to reflect the operational reality of temporary timetables, and a new attribute 'operability' is introduced. Third, these attributes are translated into quantitative indices. The capacity index captures preserved throughput through a corridor while accounting for traffic heterogeneity. The stability index measures the ability to absorb deterministic input delays and prevent delay propagation within limited timetable cycles. The robustness index evaluates the ability to withstand stochastic operational variability, using replicated simulation and activity deviation times. The operability index quantifies the degree to which potentially infeasible timetables remain executable within a fixed reference period by measuring the time required to resolve blocking-time conflicts. The indices are aggregated into a data-driven multi-criteria performance quality indicator, Temporary Timetable Effectiveness (*TTE*). Stochastic variability is embedded in the framework by fitting theoretical distributions to empirical delay data. These delay distributions represent realistic day-to-day operational variability and are applied to the microsimulation. This enables the evaluation of timetable performance under both deterministic and continuous stochastic disturbances.

The proposed methodological framework is applied to a case study on the Boxtel-Eindhoven corridor, located in the southern part of the Dutch railway network. For a real-world planned track closure, three alternative temporary timetables are evaluated. The index results show clear trade-offs between performance dimensions and the aggregation into the *TTE* indicator reveals how these trade-offs interact and how different design choices influence timetable effectiveness. Sensitivity analyses and stress testing further demonstrate that the framework gives consistent, interpretable, and comparable results under varying modeling assumptions and increasing operational pressure.

Based on these findings, this research concludes that the effectiveness of temporary railway timetables can be systematically quantified using the proposed framework. By conceptualizing temporary timetables as a distinct research domain, reformulating existing performance attributes, introducing operability as a new attribute, and integrating stochastic simulation-based evaluation, this study provides a methodology able to assess and compare temporary timetables under PTC conditions.

Samenvatting

Spoorwegoperaties maken in toenemende mate gebruik van tijdelijke dienstregelingen om geplande buitendienststellingen (Planned Track Closures, PTC's) te faciliteren. Deze ontwikkeling wordt gedreven door groeiende onderhouds- en vernieuwingsprojecten, toenemende capaciteitsdruk en strengere Europese regelgeving met betrekking tot het vroegtijdig publiceren van dienstregelingen. Tijdens deze buitendienststellingen is de beschikbare infrastructuurcapaciteit tijdelijk beperkt en worden dienstpatronen aangepast via tijdelijke dienstregelingen. Deze dienstregelingen opereren onder meer beperkte omstandigheden dan nominale dienstregelingen, waardoor ze gevoeliger zijn voor onuitvoerbaarheid, vertragingen, vertragingsoverdracht en prestatiedegradatie. In de praktijk bestaan er voor één buitendienststelling vaak meerdere mogelijke tijdelijke dienstregelingen. Deze verschillen in operationele afwegingen. De keuze tussen deze alternatieven is momenteel grotendeel gebaseerd op expert judgment en ervaring, zonder een transparant kwantitatief kader dat rekening houdt met stochastische operationele variabiliteit.

De bestaande wetenschappelijke literatuur biedt een breed scala aan indicatoren en evaluatiemethoden om de prestaties van nominale dienstregelingen te beoordelen. Deze methodes zijn echter ontwikkeld voor stabiele operationele omstandigheden en deterministische planningsaanname. Hierdoor is hun directe toepasbaarheid op tijdelijke dienstregelingen beperkt. In de praktijk worden tijdelijke dienstregelingen ontworpen aan de hand van planningsnormen en ontwerpprincipes om een acceptabel prestatieniveau onder beperkende omstandigheden te waarborgen. Desondanks ontbreken specifieke prestatie-indicatoren en evaluatiemethoden om de kwaliteit van tijdelijke dienstregelingen te beoordelen. Dit leidt tot een duidelijke wetenschappelijke en praktische kennis kloof: het ontbreken van een gestructureerd kader om de effectiviteit van tijdelijke dienstregelingen tijdens buitendienststellingen systematisch te kwantificeren.

Om deze kloof te adressen, ontwikkelt dit onderzoek een methodologisch raamwerk om de effectiviteit van tijdelijke dienstregelingen te kwantificeren. De voorgestelde aanpak volgt een driestapsmethodologie. Ten eerste wordt effectiviteit formeel gedefinieerd als het vermogen van een tijdelijke dienstregeling om onder deterministische en stochastische operationele variabiliteit uitvoerbare, stabiele en robuuste treinoperaties te behouden. Ten tweede worden bestaande dienstregeling attributen capaciteit, stabiliteit en robuustheid herformuleerd om beter aan te sluiten bij de operationele condities van tijdelijke dienstregelingen. Ook wordt een nieuw attribuut geïntroduceerd, uitvoerbaarheid. Ten derde worden deze attributen vertaald naar kwantitatieve indices. De capaciteitsindex beschrijft de behouden doorstroming over corridors, rekening houdend met verkeersheterogeniteit. De stabiliteitsindex meet het vermogen om deterministische verstoringen te absorberen en vertragingsoverdracht binnen beperkte dienstregelingscycli te voorkomen. De robuustheidsindex evalueert de weerstand tegen stochastische variabiliteit op basis van herhaalde simulaties en afwijkingen in activiteit tijden. De uitvoerbaarheidsindex kwantificeert in welke mate potentieel onuitvoerbare dienstregelingen alsnog uitvoerbaar zijn binnen een vaste referentie periode, door de benodigde tijd voor het oplossen van sectiebezettingsconflicten te meten. Deze indices worden samengevoegd tot een data-gedreven multi-criteria prestatie-indicator, Temporary Timetable Effectiveness (*TTE*). Stochastische variabiliteit wordt expliciet in het raamwerk geïntegreerd door theoretische verdelingen te fitten op empirische vertragingen data, die de dagelijkse operationele variatie realistisch representeren. Deze vertragingen verdelingen worden toegepast in microsимуlaties om de prestatie van dienstregelingen te evalueren onder zowel deterministische als continue stochastische verstoringen.

Het voorgestelde raamwerk is toegepast op een casestudy op de corridor Boxtel-Eindhoven, gelegen in het zuiden van het Nederlandse spoorwegennet. Voor een reële geplande buitendienststelling zijn drie alternatieve tijdelijke dienstregelingen geanalyseerd. De resultaten op indexniveau laten duidelijke trade-offs zien tussen verschillende prestatie dimensies. De aggregatie tot de *TTE*-indicator geeft inzicht in hoe deze trade-offs samenkomen en hoe ontwerpkeuzes de algehele effectiviteit van de dienstregeling beïnvloeden. Aanvullende gevoeligheidsanalyses en stresstests tonen aan dat het raamwerk consistente,

interpreteerbare en vergelijkbare resultaten oplevert onder uiteenlopende modelleringsaannames en toenemende operationele druk.

Op basis van deze bevindingen concludeert dit onderzoek dat de effectiviteit van tijdelijke dienstregelingen systematisch kan worden gekwantificeerd met behulp van het voorgestelde raamwerk. Door tijdelijke dienstregelingen te conceptualiseren als een afzonderlijk onderzoek domein, bestaande prestatie-attributen te herformuleren, uitvoerbaarheid als nieuw attribuut te introduceren en stochastische simulatie te integreren, biedt deze studie een methodologie om tijdelijke dienstregelingen tijdens geplande buitendienststellingen te evalueren en onderling te vergelijken.

Abbreviations

The following table provides an overview of the abbreviations used throughout this report for ease of reference.

Abbreviation	Definition
At	Acht
AUEC	Area Under the Effectiveness Curve
AUP	Alternatief Uur Patroon (Alternative Hour Pattern)
Bd	Breda
BD	Basis Dag (Basic Day)
Bet	Best
Btl	Boxtel
BUP	Basis Uur Patroon (Basic Hour Pattern)
BW	Basis Week (Basic Week)
CDF	Cumulative Density Function
DONS	Designer Of Network Schedules
DONNA	Designer Of Network Schedules for NATIONAL use
Ehs	Eindhoven Strijp-S
Ehv	Eindhoven Centraal
Ehst	Eindhoven Stadion
ERTMS	European Rail Traffic Management System
FRISO	Flexibele Rail Infra Simulatie Omgeving
GDR	Freight service/train
HSL	High Speed Line
IC	Intercity service/train
IOV	Inspelen Op Vertragingen (Make up for delay)
KPI	Key Performance Indicator
Lpe	Liempde
MOP	Mik Op Pad (Stick to the plan)
Nm	Nijmegen
NS	Nederlandse Spoorwegen
PAB	Prestatie Analyse Bureau
PESP	Periodic Event Scheduling Problem
PTC	Planned Track Closure
Q-Q	Quantile-Quantile
Rm	Roermond
RMS	Root Mean Square
SAHR	Sum of Arrival Headway Reciprocals
SD	Specifieke Dag (Specific Day)
SPR	Sprinter service/train
SSHR	Sum of Shortest Headway Reciprocals
TDL	Timetable Design Levels
TSB	Tijdelijke Snelheids Beperking (Temporary Speed Restriction)
TTE	Temporary Timetable Effectiveness
VI	Venlo
Zbm	Zaltbommel

Introduction

This chapter introduces the context, research questions and objectives of this research. The operational background of railway timetabling is outlined in [Section 1.1](#), with a specific focus on the growing importance of temporary timetables used during planned track closures. [Section 1.2](#) defines the problem addressed in this research from both a scientific and a practical perspective, leading to the identification of the research gap and questions. Finally, the structure of the report and the reading guide are discussed in [Section 1.3](#).

1.1. Background and Context

Rail transportation plays an important role in modern transportation systems. Trains offer a mode of transport that has high capacity, energy efficiency, and reliability. In the Netherlands, railways are the backbone of public transport. The largest passenger operator, NS (Nederlandse Spoorwegen), carries over one million passengers daily on a national network of approximately 2,100 km of mainline track (NS, [n.d.](#)). This national network is in the hands of ProRail, which is responsible for maintaining and managing the railway infrastructure and allocating the track capacity (Rijksoverheid, [n.d.](#)). In total, ProRail manages over 7,000 km of track and nearly 400 stations (ProRail, [2024b](#)).

In Europe, rail networks are expanding due to the growing demand for capacity and the increasing maintenance needs, although the current demand remains below pre-COVID levels (ProRail, [2024a](#)). For a compact and densely populated country like the Netherlands, having a high-functional railway system is crucial to meet environmental, economic, and social goals. At the same time, the Dutch rail network is under continuous pressure. Population growth, increasing mobility demand, and freight traffic expansion drive the need for more train services, while the infrastructure requires regular maintenance and inspections and the unavailability of specialized technical staff is increasing. Furthermore, major infrastructure replacements are needed in the coming years due to end-of-life assets and renew technology (ProRail, [2024a](#)). These activities often require temporary track closures and restrictions on parts of the network. For example, ProRail announced that in 2025 more than 400 construction projects were planned, many of which have already started, inevitably causing changes in the schedule (ProRail, [2025](#)).

The railway timetable is a fundamental component of rail transportation, ensuring efficient operations and optimizing the occupation of the infrastructure. Traditionally, railway schedules are designed using nominal timetables, which operate under stable and predefined conditions. However, real-world railway operations frequently require temporary timetables, implemented during planned maintenance or unforeseen disruptions. These adjustments can be predefined and prepared well in advance or made as real-time modifications in response to unforeseen disruptions in daily operations. Here, the available preparation time, and thus the resulting quality of the temporary timetable, differs significantly. These schedules have adjusted arrival and departure times, routes, track allocations, or train frequencies to accommodate various additional constraints while minimizing impacts. A particular and recurring form of such adjustments is a Planned Track Closure (PTC), the temporary unavailability of one or more railway tracks, during which train traffic is not allowed (SpoorData ProRail, [n.d.](#)). Prior to these PTCs, timetables must be modified to avoid conflicts between trains while maintaining system performance.

Ensuring system performance can be done by designing effective timetables. Effectiveness refers to a timetable that is operationally executable, stable, and robust. Operationally executable means that, despite conflicts from simultaneous infrastructure use, the timetable is executable. Stability indicates

that the timetable can absorb delays and avoid delay propagation. Robustness relates to the timetable's performance under small stochastic variations in process time that do not require service cancellations. Timetable effectiveness is critical during PTCs, as constrained conditions due to limited track availability and compressed schedules can increase the likelihood of infeasibility and delay propagation.

European regulations require all infrastructure managers to publish temporary timetables further in advance. For ProRail, this particularly affects the established planning process, which now needs to accommodate these earlier publication deadlines, making it more difficult to anticipate and incorporate operational uncertainties. Moreover, in recent years, Dutch railway operations have become increasingly complex due to shifting passenger travel patterns, increasing infrastructure maintenance needs, and increasing service disruptions (NS, 2025). Both of these trends highlight the importance of understanding how temporary timetables perform under constrained conditions. Ensuring effective performance under these PTC conditions is therefore essential for maintaining service performance and operational reliability.

In summary, the Dutch railway sector faces a triple challenge: maintaining and renewing an increasingly saturated infrastructure and ensuring reliable services during planned maintenance, while publishing timetables earlier due to stricter regulations. This combination makes it important to understand and assess how temporary timetables perform under constrained and uncertain conditions. A focus on timetable effectiveness during PTCs is therefore both a theoretical and a practical necessity.

1.2. Problem Definition, Research Gap, and Objectives

Although nominal timetables are carefully designed and constructed under stable operating assumptions, temporary schedules implemented during PTCs often lack the same level of systematic evaluation. Under nominal conditions, feasibility is ensured beforehand, and stability and robustness are managed by applying deterministic rules, such as fixed headway buffers and running time supplements. Although these rules ensure acceptable performance levels under regular conditions, they do not fully reflect the operational reality of PTCs. During such constrained conditions, parts of the infrastructure are temporarily unavailable and service patterns might change. As a result, available margins are often reduced and the risk of delay propagation increases. Furthermore, the European requirement to publish temporary timetables further in advance, limits the flexibility to adjust schedules closer to the actual track closure. This makes temporary schedules more sensitive to stochastic operational variations such as small delays or knock-on effects from other disturbances.

From a scientific perspective, the literature on railway timetables has developed a range of methods to define and measure feasibility, stability, and robustness, as well as stochastic approaches for timetable design. However, most of this research focuses on nominal timetables that operate under stable conditions. In contrast, temporary timetables have been studied substantially less. The concept of timetable effectiveness and the individual indicators such as feasibility, stability, or robustness have received limited attention in timetable research. Moreover, the extent to which stochastic simulation-based methods can support the design and evaluation of temporary timetables remains largely unexplored. From a practical perspective, ProRail faces the challenge of accommodating an increasing number of infrastructure works in the coming years, each requiring adjustments to the nominal schedule. In such situations, multiple alternative temporary timetables are developed to reflect different operational considerations. The selection between these alternatives is currently based on expert judgment and experience and may vary between perspectives. There is a lack of transparent and reproducible decision support to assess how different timetables perform across multiple performance dimensions under stochastic operational conditions. Therefore, developing a systematic, data-driven framework that quantifies and compares the effectiveness of temporary timetables is highly relevant.

This dual perspective highlights a clear research gap: while the performance of nominal timetables has been extensively studied and supported by various standardized metrics, the effectiveness of temporary timetables under PTC conditions remains underexplored and lacks systematic quantification. To the best of our knowledge, there is currently no integrated framework that enables the structured evaluation and comparison of temporary timetables while explicitly accounting for stochastic operational behavior. Together, this indicates the need for a clear definition, metric, and method to incorporate effectiveness and stochasticity into temporary timetable assessment. Therefore, this research aims to define the

effectiveness of the schedule in the context of temporary railway operations, to develop a quantitative framework to assess the performance of temporary timetables, and to explore the role of stochastic methods in the evaluation of timetables. The main research question is as follows:

How can the effectiveness of temporary railway timetables be systematically quantified, in the context of planned track closures (PTCs)?

In order to answer this main question, five sub-questions have been determined.

1. *What are the main characteristics of temporary railway timetables?*
Through a literature review and expert consultations, the key properties of temporary timetables are identified. Extensive research exists in the field of nominal timetables, while temporary timetables differ fundamentally.
2. *How can the effectiveness of temporary railway timetables be quantified?*
This sub-question establishes what timetable effectiveness means in the context of PTCs. Building on existing theoretical concepts, a framework and multi-criteria performance indicator that evaluates temporary timetable effectiveness is developed.
3. *What are factors and their relationship, that affect the effectiveness of temporary railway timetables during real-life operations?*
This question identifies factors that influence the performance of temporary timetables during PTCs. A combined qualitative and quantitative analysis reveals how various factors influence timetable effectiveness.
4. *How can the effectiveness of real-world temporary railway timetables of PTCs be quantified?*
To assess the effectiveness of temporary schedules in practice, the framework and indicator are applied in a case study. Using realized operational data, multiple temporary timetables are evaluated and compared under stochastic variability using micro-simulation.
5. *What practical recommendations and guidelines can be provided to support the quantification of effective temporary railway timetables?*
Based on insights from the literature, expert consultations, factor analysis, stochastic simulations, and indicator evaluation, this question synthesizes practical guidelines for the evaluation and comparison of timetable effectiveness.

1.3. Report Outline

The remainder of this thesis is structured as follows. [Chapter 2](#) presents the theoretical background and literature review. Existing concepts related to railway timetabling, including feasibility, stability, and robustness are discussed, and a comparison is made between nominal and temporary timetables. In addition, insights are provided into current timetable design and evaluation practices in the Dutch railway context. The definition of timetable effectiveness and the conceptual framework are introduced in [Chapter 3](#). This chapter defines the performance dimensions relevant for temporary timetables and presents the multi-criteria effectiveness indicator, including the underlying indices and their interpretation. A qualitative factor analysis in [Chapter 4](#) provides empirical insights into the factors that affect the effectiveness of the schedule. Also, to model real world variations, the fitted stochastic disturbance distributions are defined by a quantitative data-analysis. [Chapter 5](#) applies the proposed framework and methodology to a real-world case study. Multiple alternative temporary timetables are evaluated and compared using stochastic micro-simulation. The results of the data-analysis and case study and the implications for the evaluation and comparison of temporary railway timetables are reflected in [Chapter 6](#). Moreover, the strengths and limitations of the proposed research are discussed. Lastly, [Chapter 7](#) concludes this research by answering the main research question and outlining directions for future research.

2

Theoretical Background and Literature Review

This chapter provides relevant background information about railway timetabling. To the best of our knowledge, no research has been done with a specific focus on temporary timetables. Therefore, the theoretical concepts and models discussed are mostly derived from the context of nominal timetables. In [Section 2.1](#) the theoretical basis of timetabling is explained, including differences between nominal and temporary timetables and stochastic and deterministic methods. [Section 2.2](#) gives an overview of important timetable attributes and how they relate to each other. [Section 2.3](#) provides various performance indicators for timetable evaluation. [Section 2.4](#) focuses on timetable design methods and gives insight into the current practice at ProRail. This chapter ends with a conclusion of the most important insights and the research scope for the framework of timetable effectiveness, given in [Section 2.5](#).

2.1. Railway Timetable Fundamentals

Designing a timetable is about finding a conflict-free schedule for arrival, departure, and passing through events of trains at stations and other timetable points (Correia Duarte et al., 2025). The timetable coordinates the use of limited infrastructure capacity, aligning passenger services, freight operations, and maintenance activities and is defined at three decision levels: strategic, tactical, and operational (Correia Duarte et al., 2025). The strategic level focuses on long-term decision-making such as line planning and network design. At the tactical level, the timetable itself is designed, including train paths, headways, dwell times, etc. At the operational level, short-term adjustments to real-world train dynamics are made to mitigate disturbances and delays.

2.1.1. Nominal and Temporary Timetables

It is necessary to make a distinction between nominal and temporary timetables, as both serve different purposes within railway operations. Nominal timetables are timetables designed under ideal circumstances, where infrastructure capacity constraints are satisfied (Cacchiani and Toth, 2012). Such baseline timetables aim at maximizing efficiency (e.g. minimizing passenger travel times, maximizing passenger satisfaction, or minimizing operational costs) where the objective depends on the railway infrastructure manager or operator ((Polinder et al., 2021) & (Cacchiani and Toth, 2012)). These schedules incorporate long-term planning considerations and allow for negotiation of operator wishes to define a carefully balanced timetable.

Temporary timetables are revised timetables implemented during periods of interrupted operations, such as maintenance activities. Such temporary timetables are not used for every-day, as they are constrained by reduced infrastructure availability, partly canceled routes, or different operational constraints. A recurring example of this is a PTC, where specific track(s) (segments) are temporarily taken out of service. Due to these additional constraints, nominal timetables can no longer be applied in their original form, and thus an adjusted timetable must be created. Such modified temporary timetables typically have less slack and represent a trade-off; ensuring sufficient service performance while working within limited infrastructure availability and avoiding excessive inefficiencies.

2.1.2. Stochastic and Deterministic Timetabling

Deterministic and stochastic approaches represent two fundamental ways to model timetabling problems. Deterministic timetable design approaches assume fixed inputs for normal service operations (Salido et al., 2012). Because it is assumed that inputs are known in advance and do not consist of distribution functions, uncertainty and stochastic variation are not explicitly taken into account. However, stability, resilience, robustness and other timetable evaluation parameters can be taken into account by adding deterministic values of time supplements in the schedule. The general aim of deterministic design methods is to find a conflict-free timetable under ideal conditions.

By contrast, stochastic methods incorporate uncertainty directly in the timetable design phase (Salido et al., 2012). Such approaches assume variability in inputs by probability distributions, where variation in past timetables is taken into account in future timetables. The goal of these methods is to design a conflict-free schedule that also performs well in a range of possible disturbances. In real-world operations, train delays arise due to stochastic disturbances such as fluctuating running and dwell times, and incorporate variability into the system (Salido et al., 2012). Stochastic methods embed uncertainty directly into the design process by representing probability distributions for various sources of variability. Including such buffers and distributions allow timetables to be both conflict-free and more reliable across a range of possible disturbance scenarios.

2.2. Timetable Definitions and Characteristics

In railway planning, the timetable specifies when and where trains operate, under which constraints, and with what margins. Schedules translate the long-term demand for train services into an operational plan that assigns to each train a conflict-free path, including track allocations, over the available infrastructure within a given period ((Hansen and Pachl, 2008), (Cacchiani and Toth, 2012), & (Correia Duarte et al., 2025)). In timetabling, various attributes are taken into account. The most important ones are:

- **Feasibility:** Whether the timetable is conflict-free and operationally executable under ideal conditions (Goverde and Hansen, 2013).
- **Stability:** Whether small delays can be absorbed without spreading or requiring rescheduling ((Goverde and Hansen, 2013) & (Goverde, 2010)).
- **Resilience:** Whether the system can recover after disruptions through system adaptation or rescheduling measures (CORDIS: Publications Office of the European Union, n.d.).
- **Efficiency:** Whether the timetable achieves optimal or near-optimal use of resources under nominal conditions (Cacchiani and Toth, 2012).

These attributes are closely related to the concept of rail capacity, as they depend on how the available infrastructure can accommodate train movements. Capacity is generally defined as the number of trains per unit of time on a part of the infrastructure. Different variants of capacity are distinguished. Theoretical capacity defines the maximum number of trains (upper bound) in a given network section, assuming a perfectly homogeneous traffic flow without any buffers and slack ((Besinovic and Goverde, 2018) & (Jensen et al., 2020)). The maximum number of trains that can be scheduled, given the actual traffic heterogeneity, without headway buffers, is the maximum capacity (Jensen et al., 2020). The practical capacity reflects the effective capacity that can be scheduled in a robust and stable manner, including headway buffers ((Besinovic and Goverde, 2018) & (Jensen et al., 2020)). The actual time that infrastructure is occupied by a set of trains is defined as the capacity occupation (Jensen et al., 2020). Lastly, the capacity occupation rate is the share of the available time window in which the infrastructure is occupied. It is the ratio between capacity occupation and a time period (Jensen et al., 2020).

In addition to feasibility, stability, resilience, efficiency, and capacity, a frequently used timetable attribute is robustness. Although the other attributes have relatively well-established and clear definitions in railway timetabling, the definition of robustness remains vague. In the existing literature, no single unified definition of robustness has been defined. As Lusby et al. (2018) propose, a single and clear definition of robustness is lacking in the literature, because the detailed definition depends on the perspective. Solinen (2025) states that existing definitions of robustness remain vague, especially regarding the magnitude of initial delays a schedule should be able to handle.

A broad network perspective indicates that robustness reflects the influenceability of the system. A robust railway system can continue to function under difficult circumstances, whereas a non-robust system will quickly experience delay propagation (Vromans, 2005). From an operational standpoint, a robust timetable is one that prevents the uncontrolled propagation of small delays, ideally without requiring heavy dispatching measures (Jensen et al., 2014). Solinen (2025) proposes an operationally oriented definition of robustness as a timetable where trains can run according to their planned paths and arrive punctual at intermediate stations and junctions and final destinations, regardless of small delays and without causing knock-on delays. From a passenger-oriented perspective, robustness is associated with the degree of passenger dis-utility, based on the experienced discomfort, number of transfers and waiting time (Takeuchi et al., 2007), the maximum possible initial delay before violating passenger transfer possibilities (Schobel and Kratz, 2009) and minimizing the total weighted real travel times of passengers (Dewilde et al., 2014).

The EU-funded ON-TIME project introduced the Timetable Design Levels (TDL) framework (CORDIS: Publications Office of the European Union, n.d.) which distinguishes between various concepts related to robustness, namely:

- **TDL 0 (Low Quality):** No conflict detection or stability analysis.
- **TDL 1 (Stable):** Includes delay absorption analysis.
- **TDL 2 (Feasible):** Conflict detection and stability analysis.
- **TDL 3 (Robust):** Robustness analysis to withstand statistical variations.
- **TDL 4 (Resilient):** Enables efficient real-time rescheduling measures.

From this framework, it becomes clear that robustness is an extension beyond feasibility and stability, including statistical variability (CORDIS: Publications Office of the European Union, n.d.). The most simply stated is robustness as ‘the ability to resist imprecision’ (Salido et al., 2008). In line with this view, Goverde and Hansen (2013) emphasize that robustness is the ability of a timetable to cope with design inaccuracies, fluctuations in parameters, and varying operational conditions. Robust timetabling highlights the objective to minimize delay propagation by including variations in operational conditions in the design phase (Cacchiani and Toth, 2012). This typically involves the addition of larger or better distributed time supplements to absorb potential minor disturbances (Polinder et al., 2019). However, the literature emphasizes an important trade-off where buffers and time supplements increase the ability to recover from small delays, but simultaneously reduce capacity consumption and operational efficiency. Solinen et al. (2017) therefore define robustness as the ability of railway timetables to maintain their scheduled time slots, regardless of small disturbances, without causing unrecoverable delays to other trains, while still allowing recovery within existing time supplements (Solinen et al., 2017).

2.2.1. Relation Between Timetable Attributes

Despite some overlap between timetable attributes, each concept has its own emphasis and relevance. Timetable feasibility is achieved when all train movements are possible without conflicts and trains follow their planned routes (Goverde and Hansen, 2013). According to Goverde and Hansen (2013), a timetable is feasible when operational processes can be completed within the scheduled process times and the scheduled train movements do not interfere, allowing uninterrupted paths. In other words, feasibility is whether a timetable can function as planned under *ideal* conditions, while robustness extends this by addressing how well a timetable performs under *non-ideal* conditions.

When feasibility is achieved, the stability of the timetable can be evaluated (Besinovic et al., 2016). Stability is the capability of a schedule to absorb delays, without the need for traffic management and rescheduling ((Goverde and Hansen, 2013) & (Goverde, 2010)). This concept is closely linked to infrastructure occupation and buffer times, as higher infrastructure occupation leads to less time allowances and thus to lower stability. Where timetable stability concerns the presence of sufficient time margins, robustness focuses on how these margins are distributed across all train paths to preserve performance when deviations from scheduled paths occur (Goverde and Hansen, 2013).

Resilience refers to a system’s ability to withstand initial disturbances and restore functionality quickly afterward (Chung et al., 2025). Resilience ensures proactive and reactive system recovery, while

robustness focuses on preventing performance degradation (Chung et al., 2025). In the context of railway timetables, resilience is the ability of a timetable to cope with stochastic operational variations after implementing real-time traffic management strategies (CORDIS: Publications Office of the European Union, n.d.), such as re-timing, re-ordering, re-routing, re-tracking, and re-scheduling. This differs from robustness, which does not take into account system adaptations. Chung et al. (2025) emphasize that robustness and resilience should complement each other, because if a timetable is optimized for robustness to well-known disruptions, the system could collapse if the operational variation exceeds expectations. In contrast, if a timetable is too resilient, it may risk inefficiency due to extensive preparedness and built-in redundancy.

Lastly, efficiency concerns the optimal performance under nominal conditions (Cacchiani and Toth, 2012). Robustness, stability, and resilience conflict with efficiency (Polinder et al., 2019) as a very robust, stable, and resilient timetable is too conservative to protect against delays, using increased time supplements which increase total travel times. To address this trade-off, the price of robustness is measured, which is the ratio between the cost (e.g. financial or capacity) of a robust timetable and an optimal (i.e. most efficient) timetable (Andersson et al., 2013b).

Although these timetable attributes can be conceptually distinguished, they are strongly interrelated. Improving performance with respect to one attribute often comes at the expense of another, resulting in inherent trade-offs. As can be seen in Figure 2.1, the higher the capacity, efficiency, or heterogeneity, the higher the risks of delays and delay propagation, thus the robustness decreases. A higher capacity utilization improves service frequency, but reduces buffers, which increases sensitively to small disturbances. Similarly, optimizing the timetable to minimal travel times, or other efficiency objectives, will push the timetable towards a tight planning with minimal supplements. Therefore, robustness is sacrificed for higher efficiency. Lastly, a schedule with mixed train types with different speeds and stopping patterns will require more buffer time to avoid conflicts and overtakes. Higher heterogeneity thus decreases timetable robustness.

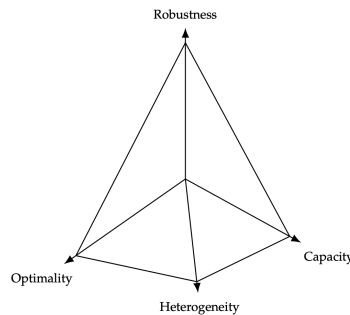


Figure 2.1: Factors influencing railway timetable robustness (reproduced from (Salido et al., 2008)).

2.3. Timetable Performance Indicators

The performance of a railway schedule can be assessed through a range of Key Performance Indicators (KPIs). These indicators provide insights into timetable attributes, where each attribute highlights a different dimension of timetable quality. In practice, these timetable characteristics are closely related and depend on each other.

2.3.1. Ex-ante vs. Ex-post

In the literature, a distinction is made between ex-ante metrics, properties embedded in the timetable itself, and ex-post indicators, derived from traffic performances. Ex-ante measures are related to timetable characteristics and can be determined without knowledge of traffic performances or disturbances (Andersson et al., 2013b). Examples are planned running time supplements, headway buffer times, or the distribution of slack along a train path. These ex-ante metrics provide quick insights at early planning stages, yet often at the cost of (over-)simplifying dynamic interactions.

Ex-post indicators evaluate timetables by simulation or observing their execution in real-world conditions.

Commonly used ex-post indicators include punctuality, total delay, secondary delay, or cumulative delay propagation patterns. Jensen et al. (2014) distinguish between metrics and indicators by showing that indicators integrate multiple characteristics into higher-level constructs. For these approaches, traffic performance data are necessary, which require larger amounts of input data, computational resources, and model calibration and validation. However, these methods can capture disturbances and interactions in more detail, providing a more realistic evaluation of schedule performance and improved calibration of future timetables.

2.3.2. Methodological Evaluation Approaches

Analytical evaluation approaches use mathematical formulations without requiring simulation. Classical methods include timetable compression and blocking time theory, which quantify infrastructure capacity consumption based on deterministic microscopic calculations (Hansen and Pachl, 2008). Goverde et al. (2013) define a new method for dynamic infrastructure occupation, under disturbed (stochastic) conditions, to evaluate capacity utilization of timetables. The infrastructure occupation is again an algebraic-based method, yet the introduction of dynamic infrastructure occupation is done by running multiple simulations and determining the average infrastructure occupation of all replications. Furthermore, max-plus algebra is a mathematical way of modeling discrete-event systems. In such cases, the timing of one event depends on the timing of other events, and the state changes only when events occur (Goverde, 2007). The focus is on critical paths that have the least accumulated slack times and to identify whether delays are likely to propagate throughout the network (Goverde, 2007).

Secondly, empirical evaluation techniques make use of simulation-based approaches. Microsimulation tools provide high precision by capturing interactions at the level of individual train movements and detailed modeling of routes and traffic management concepts. However, validity depends on whether the tools provide an accurate representation of reality for the purpose of the study (Tielman, 2017). Lindfeldt (2008) shows an example of a microscopic simulation in RAILSYS to study the effects of delays on a congested mixed traffic corridor. The focus is on model validation for the isolation of primary delays. When data or computational effort are limited, macroscopic simulation offers a more aggregated but less detailed alternative. These simplified simulations are particularly suitable for evaluating timetables in earlier planning stages, where reduced precision is accepted (Jensen et al., 2014). Empirical methods allow for what-if analyses to understand how the system reacts to unplanned events. Simulation techniques are useful in analyzing the average delays caused by disturbances and the recovery time needed to return to the planned schedules ((Salido et al., 2008) & (Wolniewicz, 2024)).

2.3.3. Scientific and Practical Timetable KPIs

Feasibility

Feasibility depends on the absence of conflicts between train paths. According to Goverde and Hansen (2013), an ex-ante indicator for feasibility is the scheduled process time. If the scheduled time of individual train processes exceeds the minimum process time, the timetable is considered infeasible. An ex-post KPI is the number of scheduled train path conflicts (Goverde and Hansen, 2013). A conflict is detected when at least one train is forced to brake, stop, or deviate from its scheduled movements to avoid unsafe interactions with other trains. The number of scheduled train conflicts should ideally be zero. This can be achieved by ensuring sufficient separation. The minimum spacing required must be equal to the minimum headway time between successive trains (Hansen and Pachl, 2008). Timetable feasibility can also refer to the ability of a schedule to ensure that all scheduled events comply with logical and technical sequencing constraints (Polinder et al., 2021). This means that departures can only occur after the corresponding arrival and minimum dwell times, constrained by feasible running times between stations. From an event-sequencing perspective, a KPI for timetable feasibility is the absence of logical inconsistencies in event sequencing. This ensures that trains cannot depart or arrive earlier than physically possible and that the dependency between trains are satisfied (Polinder et al., 2021).

Stability

The stability of the schedule refers to the ability to absorb initial and primary delays such that trains can return to their scheduled train paths without requiring active dispatch or rescheduling interventions ((Sparing and Goverde, 2017) & (Goverde and Hansen, 2013)). More stable timetables will return to their scheduled event times more quickly than less stable schedules. The absorptive ability of a timetable is dependent on the time supplements. Commonly used ex-ante indicators for stability are therefore

the amount and distribution of running time supplements and buffer time. Various ex-post KPIs for stability exist as well, such as the magnitude of disturbances (Goverde and Hansen, 2013). From a periodic-timetable perspective, stability is linked to the concept of cycle times (Sparing and Goverde, 2017). A timetable is considered stable if the minimum cycle time (i.e., the closed sequence of events with the least mean time allowance ((Goverde, 2007) & (Goverde, 2010))) is shorter than the nominal timetable period (e.g., one hour for cyclic timetables). In other words, the minimum cycle time of a schedule is an ex-ante indicator of the stability of railway schedules. The main idea is that a stable timetable absorbs all delays within a predefined number of timetable cycles. Other ex-post KPIs are focused on weighted travel time extension or heterogeneity indices (e.g., SSHR and SAHR), which quantify how headways and traffic mixes impact stability ((Jensen et al., 2014) & (Dewilde et al., 2014)).

Resilience

Resilience is the ability of a timetable to prevent secondary delays through real-time traffic management interventions (e.g., re-timing, re-ordering, re-routing) (Goverde and Hansen, 2013). These interventions are influenced by the availability of operational resources such as rolling stock and crew. Resilience focuses on the adaptive capacity of a timetable to cope with larger disturbances. KPIs for resilience focus on the satisfied passenger demand, the speed of service recovery, or number of passengers at the disturbed station (Janic, 2018). Goverde and Hansen (2013) evaluate the resilience of the timetable using ex-post indicators including mean delay, punctuality, and maximum secondary delay. In addition, resilience can be assessed through more operationally oriented indicators that capture how well dispatching measures mitigate disruption impacts. Examples are flexibility of train orders and routes, reduction of secondary delays, recovery time, and service continuity under disruption ((Janic, 2018), (Goverde and Hansen, 2013) & (CORDIS: Publications Office of the European Union, n.d.)).

Efficiency

In general, efficiency refers to the ratio of input and output. Specifically for railway operations, efficient schedules focus on optimal performance under nominal conditions (Cacchiani and Toth, 2012). An often used ex-ante KPI is the total time supplement in the schedule (Lee et al., 2017). The total slack is the sum of running time supplements and buffer times (Scheepmaker and Goverde, 2015). The supplements are expressed as percentages of minimum running times or headways. Other indicators use capacity utilization (ex-ante), measured as the infrastructure occupation rate (Goverde and Hansen, 2013) and the average travel time of passengers (ex-post) (Lee et al., 2017). The infrastructure occupation rate captures how much of the available track capacity for a chosen infrastructure scope (input), is consumed by a given timetable (output), and thus indicates the utilization (Goverde and Hansen, 2013).

Robustness

Unlike the timetable attributes mentioned above, robustness lacks an universally accepted definition in railway timetabling. Its meaning depends on the context, scope, and perspective from which timetable performance is evaluated. As a result, a wide range of KPIs have been proposed. Robustness is commonly associated with the ability of a schedule to withstand small operational disturbances without causing uncontrolled delay propagation. A central design principle underlying robustness are timetable margins. Buffer times can be added to ensure that delay propagation and tight dependencies between successive trains are avoided (Andersson et al., 2013a). Running time supplements can be used to adjust the minimum running times to account for variations in actual driving behavior (Andersson et al., 2013a). The allocation of slack along a train path is also an ex-ante metric, as poor allocation can lead to ineffective delay absorption and recovery despite sufficient margins (Andersson et al., 2013b). Dutch practice applies several operational steering indicators such as PINT Compliance which defines the difference between planned running times and the technically calculated PINT-times (i.e., benchmark for minimum running times) and long line norms which require that dwell time buffers are distributed over long train services to prevent delay accumulation of small delays over large distances. Ex-post indicators evaluate realized performance under disturbed conditions and include measures such as punctuality (Goverde and Hansen, 2013), secondary delays (Dewilde et al., 2014), total cumulative delays (Salido et al., 2008), and percentile-based process times (Goverde and Hansen, 2013). Other commonly used KPIs are heterogeneity indices such as SSHR and SAHR (Jensen et al., 2014), infrastructure and timetable complexity (Jensen et al., 2014), and weighted travel time extensions (Dewilde et al., 2014). In practice, the delayed arrivals at key nodes (i.e., major interchange stations such as Utrecht Centraal and Amersfoort Centraal) are used as ex-post robustness indicator.

2.3.4. Temporary Timetable KPIs

Quantification of timetable performance is based on ex-ante characteristics and ex-post performance measures. The choice of quantification methods depends on the context. Nominal timetables are assessed with standardized efficiency and capacity-oriented measures, whereas temporary timetables require situation-specific evaluation under constrained infrastructure and higher operational variability.

The KPIs described previously provide a systematic way to benchmark nominal timetables, where conditions are stable and slack is included to increase stability and robustness. However, quantifying performance for temporary timetables is more challenging. Due to decreased availability of infrastructure, rerouting or cancellation of services, and changes in rolling stock and crew availability, capacity and margins decrease while heterogeneity might increase. Quantification of timetable performance therefore must emphasize delay propagation and service continuity under constrained conditions. Limited studies have quantified performance for temporary schedules. Based on a structured literature review using keywords related to timetable performance, effectiveness, and temporary timetables, to the best of our knowledge no research has provided measures to benchmark temporary timetables. As shown in [Table 2.1](#), existing studies predominantly focus on nominal timetables, while standardized performance evaluation frameworks for temporary schedules under PTC conditions are lacking. In practice, the performance of temporary timetables is often assessed ad hoc and experience-based. This highlights an important research gap: the lack of systematic, data-driven quantification methods for temporary timetables under PTCs. Beyond this scientific gap, there is also a clear practical gap in use of these performance indicators. Numerous academic approaches and KPIs have been proposed that provide valuable theoretical insights, but their practical adoption is limited. Infrastructure managers typically rely on more straightforward rules and norms, which are easier and faster to apply in operational contexts. This dependence becomes even more pronounced for temporary timetables, where the absence of standardized theoretical quantification methods makes performance evaluation largely indirect through experience.

2.4. Timetable Design Methods

In railway operations, there is a distinction between timetable design and evaluation (Solinen et al., 2017). The former focuses on creating schedules for trains, whereas the latter is about assessing how a given timetable performs. These two stages are closely connected as a feasible timetable (design stage) may perform poorly under disturbances (evaluation stage).

2.4.1. Current Timetabling Practice

In Dutch railway planning practice, timetables are structured around recurring patterns and adapted whenever infrastructure availability or operational needs require it. The starting point for the yearly nominal timetable is the Basic Hour Pattern (BUP). Yet, due to European regulatory requirements, this process is shifting towards earlier timetable development, with a Basic Day (BD)-type starting point. This nominal schedule is designed under stable conditions and integrates long-term strategic concepts and operator capacity requests. In this process, new operator requirements and requests (e.g. additional services, running with new rolling stock, or changes in stopping patterns) are considered, provided that they can be accommodated within (potentially changing) infrastructure capacity and other norms.

The current nominal timetable design process is deterministic where the Basic Hour Pattern (BUP) is expanded into a Basic Day (BD) and a Basic Week (BW), based on re-occurring patterns. The new schedule for the coming year is presented in April of the current year. Annual minor changes to the existing BUP are made to optimize the timetable and accommodate long-term construction work that are so severe that they impact the timetable year-long. In contrast, more substantial adjustments occur once every five to seven years. These larger changes are driven by structural developments such as new infrastructure, innovations in rolling stock and train protection systems (e.g. ERTMS), the adoption of new planning philosophies (e.g. a train every 10 minutes), and changing travel demand.

To evaluate whether the modified BUP is conflict-free, various software is used. Solvers and algorithms are extensively studied in the literature and show potential for producing efficient and robust timetables. However, their practical applicability remains limited. A key reason is that these methods are often perceived as a 'black box'. The optimization process is difficult to interpret, making it hard for planners to understand why and how certain solutions are found. In contrast, current conflict detection methods

allow planners to directly visualize and resolve conflicts during the design process. These models strictly satisfy all constraints, while in practice expert judgment and experience are used to relax certain constraints slightly, in order to arrive at more workable solutions. Optimization-based solvers are currently used as a supportive tool to generate suggestions or provide insights, rather than as prescriptive guidelines in the practical timetable design process.

The ability of the timetable to absorb small operational variations is taken into account by deterministic plan norms and design criteria. For example, between two successive trains, a fixed buffer of 30 seconds is placed, a 4-8% driving time allowance is added to all running times, and fixed values for turning times per rolling stock type and length are used. These deterministic plan norms have been calibrated on performance data. Although these calibrated norms and criteria improve overall network punctuality under nominal conditions, these deterministic norms do not account for stochastic variability in specific contexts. In particular, temporary timetables during PTCs may face higher heterogeneity and different disturbance patterns, for which the adequacy of fixed plan norms and design criteria remains uncertain.

Besides nominal timetables, ProRail defines many temporary timetables, which are used during PTCs. In these cases, the nominal BUP is adjusted to an Alternative Hour Pattern (AUP), which defines the repetitive hourly service pattern. For a Specific Day (SD), this AUP serves as a template, where trains can be generated in bulk according to the predefined pattern. Unique in the Netherlands, is the degree of involvement of railway operators on the design of temporary timetables. In most other European countries, the timetables used during PTCs are determined by the infrastructure managers and imposed on the operators. However, ProRail asks operators what their wishes are, including the passenger's and freight's interests, and under their supervision, all operators try to complete the puzzle taking into account as many wishes as possible.

In the annual publication of the BUP, it is presented which PTCs will take place and when. However, in this publication the temporary timetable and AUP for each specific PTC will not yet be known. 23 weeks before the start of the PTC, the design of the temporary timetable starts. The first step is to identify the wishes and demands of all operators and try to combine these into a unified alternative timetable. During these discussions with infrastructure managers and railway operators, experts are also present. These experts assess whether certain wishes are feasible or not, based on expert judgment and experience. In practice, the BUP is uploaded into planning software which detects conflicts of the infrastructure, on a section level. Thereafter, the PTC is plotted in the model, so that it becomes clear which infrastructure is (un)available. Trains running on unavailable tracks and track sections will be manually rerouted onto other tracks, canceled, turned around, etc.

Manual adjustment of the BUP is largely experience-based and typically begins by identifying the bottleneck created by the PTC. Adaptations are made to maintain feasible service patterns around this bottleneck constraint. A key principle is that each passenger train service should run at least twice per hour. As a result, the first priority in designing an AUP is to ensure that these minimum service frequencies are maintained, especially for lines that already operate only twice per hour under nominal conditions. Once these minimum frequencies are secured, planners assess the remaining capacity to determine which additional services can be operated and whether re-routing or other adjustments are required. These decisions are supported by guidelines, where the use of buses as replacement service is considered a last resort. Another important consideration is the preservation of critical transfer connections, particularly outside the PTC area. These critical transfers are not formally defined, but are identified based on historical passenger behavior and operational experience. Planners aim to ensure that at least one reliable transfer per half hour remains available on key interchange connections, even if these means reducing other higher-frequency connections. Furthermore, keeping the effects mostly local is an important guideline. This is done by measures such as splitting train services, introducing short-turn operations, or partially canceling low-demand services.

Norms and principles are used to absorb small operational variations when planning an AUP. Norms are microscopic planning standards, applied at the same modeling level as conflict-detection, that must be met. Principles are applied to make explicit choices to ensure sufficient transport capacity. Extra emphasis is put on checking that there is no (unnoticed) accumulation of multiple agreed 'tightnesses' that increase the risk of disruption. Tightness is defined as the deviation from planning norms. This verification relies on expert judgment within the conflict-detection environment. According to these norms, train paths must be planned according to 'simultaneous infrastructure use', 'planned in

buffer', and 'running time tolerance'. Each train is planned with a headway buffer of 30 seconds. This buffer is based on a worst-case scenario, assuming, for example, that the following train immediately receives a green signal, a bridge opens, or a very long and heavy freight train is involved. In such conditions, this 30 seconds buffer should provide sufficient margin. However, in practice, this buffer is not always strictly applied as it is accepted to plan train paths within this buffer to increase capacity. Another method to increase capacity is by making use of insertion speeds ('inlegsnelheden'), lower than permitted maximum speeds, imposed on some corridors to prevent a train from arriving too early at the next timetable point and thus approaching a red signal. The consequence of such interventions is that (unnoticed) tightness will be introduced in the schedule. An accumulation of such (unnoticed) tightness must be avoided as they can lead to non-operable timetables. As a result, temporary timetables are often not conflict-free. Various internal studies at ProRail have shown that a considerable share of delays are not a consequence of external disturbances but from initial conflicts within the timetable. Such conflicts are tolerated after extensive consultation with various parties and when their occurrence and consequences are limited. A transparent and systematic evaluation method can support this process by making trade-offs explicit. Certain temporary timetables are thus approved despite containing conflicts, making them theoretically infeasible. Ensuring timetable performance therefore becomes more challenging, as performance measures presuppose a feasible baseline schedule.

Although planners adhere to these norms and principles, the actual performance of an AUP is not formally quantified. Instead, it is largely based on experience and expert judgment. It is therefore important to note that there is no single 'correct' solution for an AUP. Different planners may design different AUPs for the same PTC, reflecting personal trade-offs regarding performance, capacity, efficiency, and passenger needs. Although there is no standard quantification for timetable performance, operational variability during (temporary) timetables is managed through deterministic planning rules and manual adjustments. ProRail relies mostly on fixed norms and experience-based guidelines. These rules and norms implicitly address variability, but do not explicitly quantify timetable performance. The dependency on expert judgment for performance quantification, rather than systematic, consistent, data-driven approaches, identifies the further need to study the role of stochastic methods in temporary timetable design and evaluation. This conclusion is based on discussions with several experts within ProRail, including planning, process managers, advisors in capacity management, and performance analysis specialists.

2.4.2. Approaches for Optimized and Automatic Timetabling

Although current Dutch practice in timetable design is based on deterministic planning norms and constraint satisfaction, the scientific literature offers a broader range of methodological approaches. The main difference between the practice and other methods lies in their orientation. Practical timetabling emphasizes feasibility and operational workability within strict horizons, whereas other approaches pursue theoretical maxima through formal optimization and simulation techniques.

Deterministic scientific approaches use mathematical optimization or constraint satisfaction methods. There are many studies that use optimization models ((Serafini and Ukovich, 1989), (Dewilde et al., 2014), (Kroon et al., 2008), & (Polinder et al., 2019)) with deterministic input variables to identify the optimal value of an objective function (Hansen, 2009). Such optimization models are often based on the Periodic Event Scheduling Problem (PESP). In these formulations, the objective is to find a cyclic, conflict-free timetable that satisfies all constraints, often combined with optimization objectives (Serafini and Ukovich, 1989). For large-scale problems, when exact optimization is too computationally complex, (meta)heuristics are applied. Another method is that of constraint satisfaction, which refers to the process of finding a solution by assigning values to a set of variables while satisfying a set of predefined constraints (Hoos and Stutzle, 2005). Currently, ProRail adopts a constraint satisfaction method to design deterministic timetables, rather than relying on a PESP-based formulation. In this approach, the schedule is guided by a set of deterministic variables and constraints that are formulated at a higher level of abstraction than the real-world operational timetable. The train paths are planned iteratively, ensuring that all constraints are met. There is no explicitly determined objective function that needs to be optimized. Specialized software tools are used to verify the feasibility of planned train movements. The emphasis lies on designing a conflict-free timetable, by satisfying predefined rules and constraints, rather than optimizing an objective function.

In stochastic timetabling, optimization methods are also used. Stochastic programming uses mathematical models to find optimal solutions for input variables with a probability distribution, and thus incorporates randomness into the model (Fischetti and Monaci, 2009). Robust optimization has the objective of finding a solution that performs well under the conditions of the worst-case scenario (Fischetti and Monaci, 2009). Such a solution aims to guarantee feasibility of all outcomes within the set of uncertainties, depending on how this set and worst-case scenario are specified. A more recent approach, light robustness, is a hybrid approach that sets a maximum allowable deterioration of the objective value, together with robustness requirements. This allows for controlled violations of robustness constraints, leading to a balanced outcome of solution quality and robustness of the timetable (Fischetti and Monaci, 2009). Other stochastic approaches focus on increased schedule redundancy, by making use of genetic algorithms that optimize the delay sensitivity of job shop problems (Leon et al., 1994). In addition, redundancy increases by introducing time slacks between events, which improves the timetable's ability to absorb small disturbances and reduces the number of conflicts (Davenport and Beck, 2001). Another technique involves defining a set of complementary solutions, allowing the system to switch to the most appropriate alternative timetable based on the current state (Drummond et al., 1994). Preparing multiple feasible solutions in advance improves the flexibility of scheduling in changing conditions, enhancing timetable performance.

2.5. Conclusion and Research Scope

This chapter has outlined the theoretical and practical foundations necessary to understand the performance of railway timetables, and in particular the challenges and differences associated with temporary timetables in the context of PTCs.

2.5.1. Conclusion

The literature shows that timetable attributes such as capacity, feasibility, efficiency, and robustness have been extensively studied under nominal conditions. However, their interpretation and quantification become more complex under constrained conditions. Across the literature, as shown in [Table 2.1](#), numerous ex-ante and ex-post indicators exist to quantify nominal timetable performance, yet there is no widely accepted theoretical framework for evaluating and quantifying the performance of temporary timetables. Practical approaches rely heavily on tacit knowledge and deterministic norms, but lack systematic and data-driven quantification of schedule performance. As a result, temporary schedules are implicitly evaluated rather than measured explicitly, and no existing definition of timetable performance captures the specific operational challenges of temporary schedules. These insights reveal a clear scientific and practical knowledge gap, namely the lack of an integrated, context-specific framework for evaluating timetable performance under PTC conditions. This highlights the need for a data-driven quality indicator, specifically for temporary timetables, integrating capacity, feasibility, stability, and robustness.

2.5.2. Research Scope and Contributions

Based on insights from the literature and current planning practice, this research positions temporary timetables as a distinct research domain within rail timetabling. Temporary timetables are studied as a fundamentally different planning product, requiring specified indicators and assessment methods. The main objective of this research is to develop a data-driven multi-criteria performance quality indicator for temporary timetable evaluation under constrained conditions. Timetable effectiveness is introduced as an overarching performance concept that integrates the timetable attributes capacity, feasibility, stability, and robustness. A formal and context specific definition of timetable effectiveness is proposed. From this definition, a framework is developed to translate timetable performance into quantitative indices. Existing timetable attributes are reformulated where necessary to ensure applicability under constrained conditions. In addition, a new attribute is introduced to capture the degree to which a timetable remains executable in practice, even when theoretical infeasibilities exist. This framework makes it possible to compare alternative temporary timetables, which reduces the reliance on experience-based decision-making in practice and enables planners to systematically assess trade-offs between different design choices.

Table 2.1: Overview of key literature on timetable design and evaluation methods.

Reference (APA)	Design / Evaluation	Ex-ante / Ex-post	Method	Eval type	Short description
Andersson, Peterson & Krasemann (2013a)	Evaluation	Ex-ante	Deterministic	Algebraic	Robustness in Critical Points (RCP) to identify bottlenecks.
Andersson, Peterson & Krasemann (2013b)	Evaluation	Ex-ante	Deterministic	Algebraic	Quantifies robustness in critical points and shows where and how robustness should be increased.
Bešinović, Goverde, Quaglietta & Roberti (2016)	Des. & Eval.	Ex-ante	Stochastic	Simulation	Integrated micro-macro framework, selecting robust timetables via ILP and delay checks.
Bešinović & Goverde (2018)	Evaluation	Ex-ante	Deterministic	Algebraic	Algebraic capacity assessment at corridor, station and network level to identify bottlenecks.
Cacchiani & Toth (2012)	Design	—	Det. & Stoc.	—	Survey and trade-offs of robust timetabling models.
Cats (2016)	Evaluation	Ex-post	Stochastic	Simulation	Passenger-oriented robustness evaluation of development plans.
Correia Duarte et al. (2025)	Design	—	Det. & Stoc.	—	Review of 50 years of OR in robust timetabling.
Davenport & Beck (2001)	Evaluation	Ex-ante	Deterministic	Simulation	Slack-based techniques to improve schedule robustness.
Dewilde et al. (2014)	Des. & Eval.	Ex-post	Det. & Stoc.	Simulation	Proposes WTTE to balance robustness and travel time.
Fischetti & Monaci (2009)	Design	—	Stochastic	—	LR as midway between SP and RO, uses slack to cap objective loss while maintaining robust.
Goverde (2007)	Evaluation	Ex-ante	Deterministic	Algebraic	Max-plus spectral analysis to compute stability margins and recovery times.
Goverde (2010)	Evaluation	Ex-ante	Deterministic	Algebraic	Delay propagation algorithm assessing network-wide robustness of periodic timetables.
Goverde, Corman & D'Ariano (2013)	Evaluation	Ex-ante	Stochastic	Alg. & Sim.	Extends UIC406 with dynamic infrastructure occupation under disturbed conditions.
Goverde & Hansen (2013)	Evaluation	Ex-ante	Det. & Stoc.	Indicators	Defines a comprehensive timetable performance indicator set.
Hansen & Pacht (2008)	Evaluation	Ex-ante	Deterministic	Algebraic	Blocking-time theory: headways, margins, compression for capacity consumption.
Jensen, Landex & Nielsen (2014)	Evaluation	Ex-ante	Deterministic	Algebraic	Compares micro-simulation to indicators for robustness evaluation.
Jensen, Schmidt & Nielsen (2020)	Evaluation	Ex-ante	Deterministic	Algebraic	Defines timetable-free network capacity by maximizing feasible train sets.
Kroon et al. (2008)	Design	—	Stochastic	—	Stochastic improvement of cyclic timetables using probabilistic delays.
Polinder, Breugem, Dollevoet & Marti (2019)	Design	—	Deterministic	—	Adjustable robust optimization for periodic timetabling with uncertainty.
Polinder, Schmidt & Huisman (2021)	Design	—	Deterministic	—	Strategic passenger-focused timetabling formulation for SPOT.
Salido, Barber & Ingolotti (2008)	Evaluation	Ex-ante & Ex-post	Det. & Stoc.	Alg. & Sim.	Framework distinguishing ex-ante vs. ex-post robustness metrics
Salido, Barber & Ingolotti (2012)	Evaluation	Ex-ante & Ex-post	Det. & Stoc.	Alg. & Sim.	Analytical robustness indicators validated with microscopic simulation on a single-line.
Solinen, Nicholson & Peterson (2017)	Evaluation	Ex-post	Deterministic	Simulation	Microscopic evaluation of robustness and recovery in critical points.
Sparing & Goverde (2017)	Design	—	Deterministic	—	Optimizes periodic timetables through minimum cycle time.
Takeuchi, Tomii & Hirai (2007)	Evaluation	Ex-post	Stochastic	Alg. & Sim.	Probabilistic indicators based on estimated delay distributions.

Methodology

This chapter describes the framework used to develop a new performance indicator for temporary timetables under PTC conditions. Section 3.1 describes the methodology used to develop a data-driven multi-criteria performance quality indicator to quantify and evaluate the effectiveness of temporary railway timetables. The definition of timetable effectiveness, associated indices, and integrated indicator are developed and elaborated in Section 3.2.

3.1. Methodological Approach

The methodology to develop and evaluate a timetable effectiveness indicator is a multi-method approach, combining conceptual development, interviews, data-analysis, and microscopic simulation. A high-level overview of the entire methodological process is shown in Figure 3.1. This diagram distinguishes between the methods (shown in blue) and their inputs and outputs (shown in yellow). The methodological approach follows a three-stage method: (i) identifying relevant assessment characteristics, (ii) defining a new definition and conceptual framework with design requirements, and (iii) translating this definition into measurable indices.

3.1.1. Identification of Relevant Characteristics

The first stage consists of identifying which characteristics are essential to design and evaluate temporary timetables. These insights are gained through a literature review and a series of semi-structured interviews with a diverse set of experts, including planners, operational specialists, performance analysts, and capacity management advisors. The semi-structured interview format allows for a consistent set of guiding questions while also having flexibility to elaborate on issues most relevant to their expertise. Therefore, both comparability across interviews and role-specific insights are gained. Recurring concepts, concerns, and decision-making criteria mentioned by different experts are defined. Specific attention is paid to aspects emphasized by planners directly involved in designing temporary timetables. The characteristics obtained form a set of potentially relevant timetable attributes which serve as input for the contextualization and definition of timetable effectiveness.

3.1.2. Defining Effectiveness of Temporary Timetables

Building on the set of potentially relevant characteristics, a new definition of timetable effectiveness is constructed. Temporary timetables operate under fundamentally different conditions than nominal schedules, such that the classical definitions are insufficient. Identifying which characteristics are relevant for evaluating temporary timetables and how their meaning changes under PTC conditions is crucial. For each characteristic, it is examined how it behaves specifically under constrained conditions. Through this process, the nominal characteristics are re-framed. Furthermore, the analysis examines whether the set of timetable attributes is sufficient to assess temporary timetables. Certain performance aspects of temporary timetables are not adequately captured, which highlights the need of introducing an additional timetable attribute. The reformulated existing attributes and the new attribute are integrated into a definition of timetable effectiveness. This definition forms the basis for developing a set of design criteria that specify what an effectiveness indicator must measure to be meaningful in practice.

3.1.3. Translating Characteristics into Indices

The final step involves translating the definition of effectiveness and its associated design criteria into measurable indices. Where the previous step defines what aspects of timetable performance are

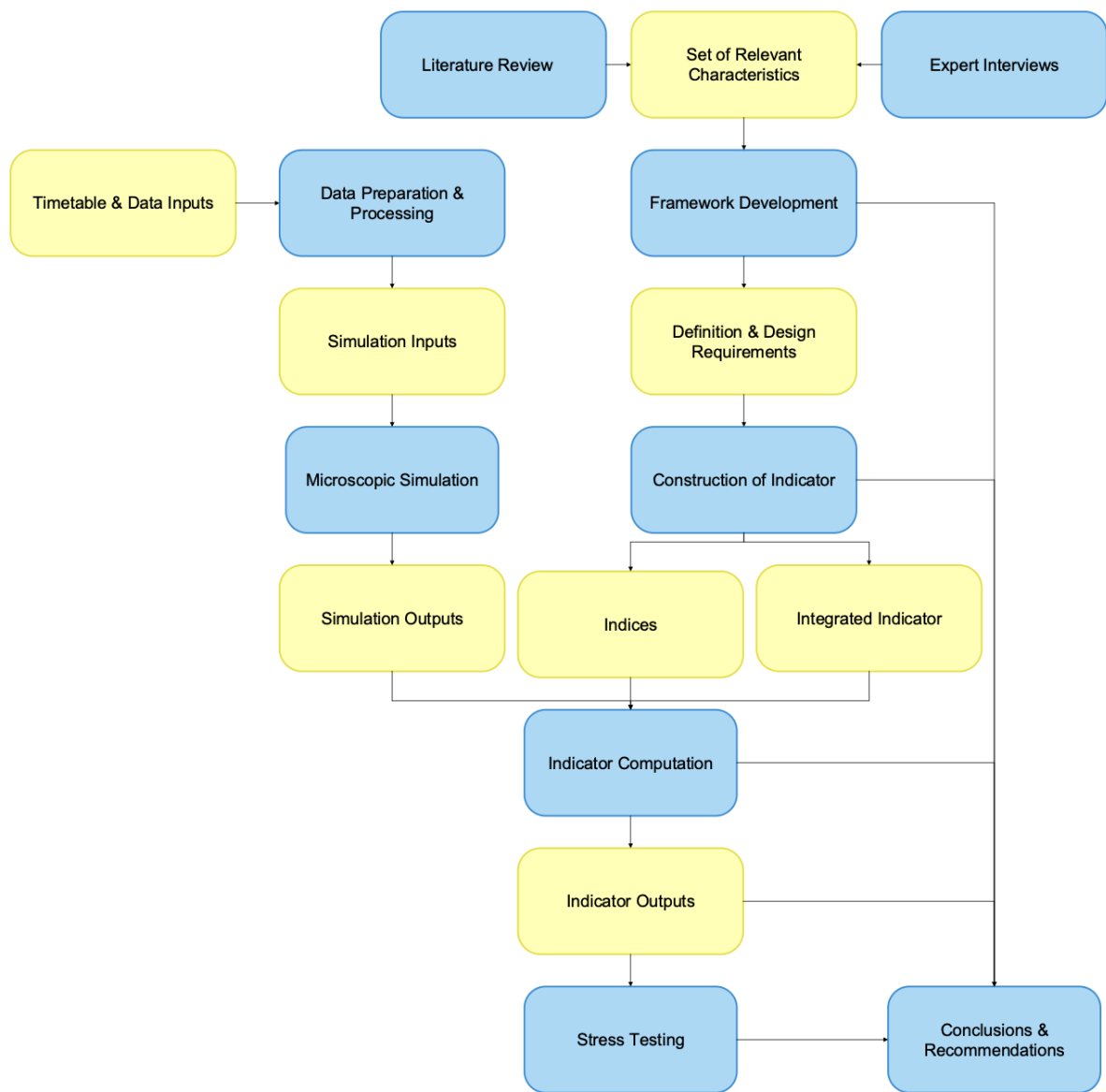


Figure 3.1: Overview of the methodological process, including methods (blue), and inputs and outputs (yellow).

relevant for temporary timetables, this step focuses on how these aspects are quantified. For each attribute, the associate design criteria are interpreted in terms of the underlying system behavior they represent during constrained conditions (e.g., capacity-related behavior, delay absorption, or operational executability). Potential indices are identified by reviewing KPIs in the literature. Existing metrics are adapted to reflect the constrained conditions of temporary timetable contexts. For attributes for which no existing KPI could be adapted, new indices are developed based on theoretical reasoning. These potential indices are assessed on their practical feasibility by evaluating whether the required input can be derived from the available simulation output and operational data. Some indices are theoretically suitable but cannot be operationalized within the available data and thus are simplified. Finally, the indices are evaluated against the primary objective to enable a systematic comparison and quantification of alternative temporary timetables. This iterative process of theoretical formulation, practical feasibility testing, and comparative validation results in a set of indices that quantify timetable effectiveness under constrained operational conditions.

3.2. Developing a Temporary Timetable Effectiveness Indicator

This section outlines the relevant timetable characteristics, develops the formal definition of timetable effectiveness, and translates it into measurable indices and an integrated indicator.

3.2.1. Relevant Characteristics

The identification of relevant characteristics for temporary timetables is based on the theoretical insights and practical observations gained in [Chapter 2](#). Attributes of timetable quality under nominal conditions include feasibility, stability, robustness, resilience, efficiency, and capacity-related measures. Feasibility is commonly defined as the absence of conflicts. Stability and robustness extend this by considering the ability of the timetable to absorb small disturbances and prevent delay propagation. Resilience captures the adaptive capacity through real-life traffic management interventions. The focus on optimal resource use is reflected through efficiency. Capacity-related concepts describe how intensively the available infrastructure is used. These attributes are strongly interrelated and various trade-offs exist between them. Because temporary timetables operate under conditions different from nominal operations, not all timetable characteristics that are commonly used for nominal timetable evaluation are relevant in the context of PTCs. Also, the interrelations and dependencies between the attributes change when considering temporary timetables. Insights from ProRail further emphasize that temporary timetables are designed and assessed differently from nominal timetables. An important insight is that temporary schedules may contain conflicts or tight margins but still remain operationally manageable.

The literature and practice indicate feasibility, stability, robustness, resilience, efficiency and capacity as important concepts for understanding timetable performance in general. The relative importance and meaning may differ when timetables are designed under constrained conditions. Therefore, these characteristics are treated as input for further contextualization and scoping when evaluated against the specific conditions and objectives of temporary timetables.

3.2.2. Design Requirements and Definition

The aspects of timetable performance relevant for temporary timetables are identified by formulating a set of design requirements. The first design requirement states that timetable quality should be grounded in microscopic detail but should produce aggregated performance outcomes at the corridor and timetable level. Margins such as headway buffers and running time supplements in microscopic detail are foundations of feasibility, stability, and robustness. However, individual train effects must be combined into higher level interpretable measures such that the overall effectiveness of a timetable can be assessed. Secondly, the evaluation is limited to small disturbances, defined as stochastic variations (operational spread) that do not require service cancellation. This excludes large-scale disruptions that cause capacity loss, cancellation, or require rescheduling of rolling stock, crew, or timetables. This requirement ensures that the proposed metric focuses on the variability that timetables can realistically absorb. Third, temporary schedules are characterized by reduced infrastructure availability, which reduces capacity. Therefore, temporary timetables must be evaluated in terms of relative capacity, the extent to which planned transport volumes can flow through the constrained infrastructure. Capacity is not considered an absolute benchmark but is considered only in relative terms to avoid conceptual ambiguity in absolute capacity definitions. Fourth, a temporary timetable must have a cyclic structure with a fixed reference period (e.g., one hour). Such cyclic timetables have a repetitive (hourly) pattern which provides a consistent temporal reference. Moreover, temporary timetables are designed within the tactical-operational planning horizon and must be computable within 18-weeks. Timetables should be quantifiable using a metric that provides quick and actionable feedback. This requirement ensures that the method is not only theoretically feasible, but also practically applicable within ProRail's planning environment, contributing to timetable feasibility. Finally, the indicator must be designed for ex-post evaluation. Ex-ante planning norms remain useful for initial timetable design, but these metrics do not capture the dynamic operational variations that determine real-world stochastic variability. Therefore, developing an ex-post effectiveness indicator that can be used for both simulated as real operational data is essential.

These six design requirements define how timetable performance should be understood in the context of temporary timetables. Characteristics that are primarily related to nominal optimization objectives (such as efficiency) or to large-scale recovery through traffic management (such as resilience) fall outside

the scope defined by the requirements. The core characteristics of timetable performance include feasibility, stability, robustness, and capacity. For feasibility, the constrained PTC environment often results in temporary schedules that are not conflict-free, yet still operationally workable. The relevant question becomes whether a conflicting schedule allows safe and executable train movements, reflected by the characteristic ‘operability’. To avoid propagation of small delays for timetable stability, it is important to absorb deterministic delays. For robustness, the focus is on preventing delay propagation under stochastic variability. Capacity relates to the level of planned transport volumes that the reduced infrastructure can accommodate, relative to the nominal situation.

In line with the design requirements and core characteristics, temporary timetable effectiveness in this research is defined as:

The ability of a temporary timetable operating under constrained conditions to maintain operable, stable, and robust train operations under deterministic and stochastic operational variability.

3.2.3. Defining Indices

The nominal characteristics of feasibility, stability, and robustness translate under PTC conditions into four fundamental performance indices:

- **Capacity Index:** The preserved throughput of timetables through a corridor.
- **Stability Index:** The ability of timetables to absorb input delays.
- **Robustness Index:** The ability of timetables to prevent delays under stochastic variability.
- **Operability Index:** The degree of executability of (in)feasible timetables.

All variables listed in the indices and mathematical equations are summarized in [Table 3.1](#). A distinction is made between sets, parameters, and variables. A set is defined as a collection of (indexed) elements over which mathematical expressions operate. A parameter is an exogenous input quantity whose value is fixed within a given simulation run. The parameters represent a constant that the simulation takes as given. A variable is an endogenous output whose value changes as a result of a simulation run. Variables represent outcomes or calculations based on simulations.

Table 3.1: Mathematical notation for the TTE framework

Symbol	Description	Domain
Sets and indices		
\mathcal{B}	Set of corridors	$b \in \mathcal{B}$
\mathcal{R}	Set of temporary timetables	$r \in \mathcal{R}$
\mathcal{M}	Set of train types	$m \in \mathcal{M}$
\mathcal{K}	Set of trains	$k \in \mathcal{K}$
\mathcal{P}	Set of timetable points	$p \in \mathcal{P}$
\mathcal{D}	Set of disturbance events	$d \in \mathcal{D}$
\mathcal{N}	Set of replications	$i \in \mathcal{N}$
$p^{exit}(k)$	Exit timetable point of train k	$p^{exit}(k) \subseteq \mathcal{P}$
\mathcal{A}	Set of activities	$a = (k, p) \in \mathcal{A} = \mathcal{K} \times \mathcal{P}$
\mathcal{C}	Set of blocking-time conflicts	$c \in \mathcal{C}$
$k_{f(c)}$	Following train associated with blocking-time conflict c	$k_{f(c)} \in \mathcal{K}$
Parameters		
N_b^{nom}	Number of scheduled trains passing corridor b in nominal timetable	\mathbb{N}
$N_{b,r}^{tmp}$	Number of scheduled trains passing corridor b in temporary timetable r	\mathbb{N}
$H_{b,r}$	Heterogeneity index of traffic through corridor b of temporary timetable r	$[0, 1]$
$p_{m,b,r}$	Share of scheduled trains belonging to class m at corridor b in timetable r	$[0, 1]$

Symbol	Description	Domain
δ_d^{inj}	Injected deterministic delay of disturbance event d	$\mathbb{R}_{\geq 0}$
$t_a^{sched,(r)}$	Scheduled time of activity a in temporary timetable r	$\mathbb{R}_{\geq 0}$
λ	Allowed schedule deviation (tolerance)	$\mathbb{R}_{\geq 0}$
$T^{cycle,ref}$	Reference operational time period	$\mathbb{R}_{> 0}$
T^{inj}	Time block in which input disturbance is injected	\mathbb{N}
T^{abs}	First time block after which no output delay remains	\mathbb{N}
$\tau_{c_{b,r}}$	Blocking-time overlap duration of conflict $c_{b,r}$	$\mathbb{R}_{> 0}$
$\alpha, \beta, \gamma, \epsilon$	Weights for component indices in TTE_r	$[0, 1]$
T_c	Compressed reference period used in stress testing	$\mathbb{R}_{> 0}$
T_{min}	Minimum reference period considered in stress testing	$\mathbb{R}_{> 0}$
T_{nom}	Nominal reference period of the timetable	$\mathbb{R}_{> 0}$
Variables		
$C_{b,r}^{pres}$	Preserved capacity ratio at corridor b in temporary timetable r	$[0, 1]$
$C_{b,r}^{mix}$	Capacity index for corridor b in temporary timetable r	$[0, 1]$
$\delta_{k,p}^{disturb,(r)}$	Realized delay of train k at timetable point p in disturbed run per temporary timetable r	$\mathbb{R}_{\geq 0}$
$\delta_{k,p}^{base,(r)}$	Realized delay of train k at timetable point p in baseline run per temporary timetable r	$\mathbb{R}_{\geq 0}$
$\delta_{k,p}^{add,(r)}$	Additional delay of train k at timetable point p per temporary timetable r	$\mathbb{R}_{\geq 0}$
$\delta_k^{add,exit,(r)}$	Additional delay of train k at exit point p^{exit} per temporary timetable r	$\mathbb{R}_{\geq 0}$
$\Delta^{in,(r)}$	Total injected delay per temporary timetable r	$\mathbb{R}_{\geq 0}$
$\Delta^{out,(r)}$	Total output delay at exit points per temporary timetable r	$\mathbb{R}_{\geq 0}$
S_r	Stability index of temporary timetable r	$[0, 1]$
$t_a^{real,(r,i)}$	Realized activity time of train k at timetable point p per temporary timetable r per replication i	$\mathbb{R}_{\geq 0}$
$q_a^{(r,i)}$	Binary variable: 1 if $t_a^{real,(r,i)} \leq t_a^{sched,(r)} + \lambda$	$\{0, 1\}$
$R_{r,i}$	Robustness index of temporary timetable r per replication i	$[0, 1]$
R_r	Robustness index of temporary timetable r	$[0, 1]$
$T_{b,r}^{cycle}$	Cycle time for temporary timetable r of corridor b	$\mathbb{R}_{> 0}$
$T_{b,r}^{cycle,feasible}$	Cycle time for a feasible temporary timetable r of corridor b	$\mathbb{R}_{> 0}$
$T_{b,r}^{cycle,infeasible}$	Cycle time for a infeasible temporary timetable r of corridor b	$\mathbb{R}_{> 0}$
$\Delta T_{k,b,r}^{accum}$	Accumulated minimum time shift required to resolve all conflicts for train k in timetable r for corridor b	$\mathbb{R}_{\geq 0}$
$\Delta T_{b,r}^{conflict}$	Minimum additional time shift required to eliminate all conflicts in timetable r for corridor b	$\mathbb{R}_{\geq 0}$
$O_{b,r}$	Operability index of temporary timetable r for corridor b	$[0, 1]$
TTE_r	Integrated timetable effectiveness indicator of temporary timetable r	$[0, 1]$
$AUEC_r$	Area Under the Effectiveness Curve for temporary timetable r	$\mathbb{R}_{\geq 0}$

Capacity Index

The effects of reduced capacity on temporary timetables are captured by the capacity index $C_{b,r}^{mix}$. Let \mathcal{B} denote the set of corridors, and let \mathcal{R} denote the set of timetables. This index reflects how much capacity is maintained under constrained infrastructure conditions and includes the complexity induced by the heterogeneity of traffic flows. Equation 3.1 defines the mathematical formulation.

$$C_{b,r}^{mix} = C_{b,r}^{pres} \cdot H_{b,r} \quad (3.1)$$

High values of $C_{b,r}^{mix}$ indicate effective capacity utilization under constrained and heterogeneous operating conditions. Lower values of the capacity index indicate reduced effective capacity.

Let N_b^{nom} denote the number of scheduled trains that pass through corridor b in one cycle of the nominal timetable. Similarly, let $N_{b,r}^{tmp}$ denote the number of scheduled trains passing through the same corridor b in temporary timetable r . The ratio of scheduled trains in the nominal and temporary timetable ($\frac{N_{b,r}^{tmp}}{N_b^{nom}}$) indicates the share of capacity preserved under PTC conditions. This is shown in Equation 3.2. Due to the decrease in infrastructure availability under PTC conditions, it is assumed that temporary timetables never increase capacity ($\frac{N_{b,r}^{tmp}}{N_b^{nom}} \leq 1$).

$$C_{b,r}^{pres} = \frac{N_{b,r}^{tmp}}{N_b^{nom}} \quad (3.2)$$

If $C_{b,r}^{pres} = 1$, this implies that temporary timetable r maintains the full planned corridor throughput of the nominal schedule. In contrast, if $C_{b,r}^{pres} < 1$ there is a loss of capacity as less trains are planned in the temporary timetable compared to the nominal timetable.

However, $C_{b,r}^{pres}$ does not capture the effects of heterogeneity on capacity. The operation of a heterogeneous train mix is more complex than a homogeneous traffic flow and decreases capacity (Salido et al., 2012). Therefore, a schedule that maintains the same capacity utilization under a heterogeneous mix indicates higher effectiveness than a timetable under homogeneity. Heterogeneity is included by rewarding heterogeneous traffic flows through the corridor. This is done using the heterogeneity index $H_{b,r}$, shown in Equation 3.3

$$H_{b,r} = \frac{1 - \sum_{m=1}^{\mathcal{M}} p_{m,b,r}^2}{1 - \frac{1}{\mathcal{M}}} \quad (3.3)$$

\mathcal{M} denotes the number of train classes (e.g., IC, SPR, GDR) and $p_{m,b,r}$ is the share of planned trains that belong to class m that run through corridor b in timetable r . If $H_{b,r} = 0$, this indicates a homogeneous flow (i.e., all trains are of one type). In contrast, $H_{b,r} = 1$ corresponds to all m train classes occurring in equal proportions, representing the maximum possible diversity in traffic composition. For example, if the train mix consists of 1/3 IC, 1/3 SPR, and 1/3 GDR, then $H_{b,r} = 1$. $H_{b,r}$ is only captured in terms of traffic composition and does not account for sequencing and bundle effects of different train services.

Stability Index

Effectiveness also concerns the ability of a schedule to limit the impact of delays. Stability measures how effectively a timetable can absorb input delays, to limit the propagation and realization of output delays spreading across the network. In this study, input and output delays are defined relative to the injected disturbances. An input delay is defined as the initial injected disturbance. All additional delay that arises after this initial disturbance, whether on the same train or on other trains, is classified as output delay. Under multiple simultaneous disturbances, delays cannot be attributed to a single initially disturbed train. Therefore, only the injected disturbance is defined as input delay.

For each timetable $r \in \mathcal{R}$, the stability index S_r quantifies how effectively injected input delays ($\Delta^{in,(r)}$) are absorbed to avoid propagation into output delays ($\Delta^{out,(r)}$). This is shown in Equation 3.4.

$$S_r = \max \left(0, \min \left(1, 0.5 - 0.5 \left(\frac{\Delta^{out,(r)} - \Delta^{in,(r)}}{\Delta^{in,(r)}} \right) \right) \right) \quad (3.4)$$

Here, the stability index S_r reflects the normalized difference between the input and output delays $\left(\frac{\Delta^{out,(r)} - \Delta^{in,(r)}}{\Delta^{in,(r)}} \right)$ which gives insight into how initial disturbances evolve. The stability index S_r quantifies how the timetable absorbs initial disturbances without generating additional output delay. Equation 3.4 maps this normalized difference to a bounded scale between 0 and 1. A value of $S_r = 0.5$ indicates neutral stability, meaning that the output delay equals the input delay. High values of S_r indicate stability, as the output delay is smaller than the input delay due to delay absorption. If the normalized difference between output and input delays increases for $\Delta^{out,(r)} > \Delta^{in,(r)}$, this indicates an increase in delays in the network, lowering S_r .

S_r is computed relative to a reference scenario to isolate the causal effect of the input delays. An undisturbed baseline simulation represents the operation without any injected disturbances. However, these timetables may contain delays that arise from structural conflicts in the timetable itself. These baseline delays are not part of the stability assessment and therefore must be removed before identifying input and output delay events. Let \mathcal{K} denote the set of trains and \mathcal{P} denote the set of timetable points. Then $\delta_{k,p}^{disturb,(r)}$ is the simulated realized delay of timetable r of train k at timetable point p in an experiment of disturbed operations and let $\delta_{k,p}^{base}$ denote the corresponding delay in the undisturbed baseline. The additional delay for train k at timetable point p for timetable r is defined by Equation 3.5.

$$\delta_{k,p}^{add,(r)} = \delta_{k,p}^{disturb,(r)} - \delta_{k,p}^{base,(r)} \quad (3.5)$$

Let \mathcal{D} denote the set of injected disturbances, in which each disturbance d affects a specific train k at a specific timetable point p . As stability is assessed under deterministic disturbance scenarios, a fixed set of input delays is specified and applied. For each disturbance $d \in \mathcal{D}$, let δ_d^{inj} represent the fixed injected input delay. The total input delay for temporary timetable r is given by Equation 3.6.

$$\Delta^{in,(r)} = \sum_{d \in \mathcal{D}} \delta_d^{inj} \quad (3.6)$$

All additional delay, whether on the same train further downstream or on other trains, is classified as output delay. However, summing $\delta_{k,p}^{add,(r)}$ over all trains k and all timetable points p will inflate the total additional delay, because the same additional delays reappear at multiple downstream points along the same train path. The additional delay is evaluated only at the moment a train leaves the study area or arrives at its final destination, indicating how much delay a specific train k remains after traveling through the network. This reflects stability as the delay absorption over the train path. For each train $k \in \mathcal{K}$, let $p^{exit}(k) \subseteq \mathcal{P}$ denote the last timetable point within the study area, where after train k leaves the network or arrives at its final destination. The additional delay of train k at exit point p^{exit} is given by $\delta_k^{add,exit,(r)}$, mathematically shown in Equation 3.7. The total output delay is defined as the sum of all additional delays at the exit point of each train path, given by Equation 3.8.

$$\delta_k^{add,exit,(r)} = \delta_{k,p^{exit}(k)}^{add,(r)} \quad (3.7)$$

$$\Delta^{out,(r)} = \sum_{k \in \mathcal{K}} \delta_k^{add,exit,(r)} \quad (3.8)$$

Stability also requires that input delays are absorbed within a limited time frame. Therefore, output delays that persist over multiple cycles indicate instability. Let the simulation horizon be divided into time blocks, corresponding to timetable cycles. Let T^{inj} denote the time block in which the initial disturbance is injected and let T^{abs} denote the first time block after which no remaining output delay is observed. A timetable is considered stable if all output delay is absorbed within two timetable cycles

($T^{abs} - T^{inj} \leq 2$) (Pachl, 2014). If the delay persists beyond this threshold, the timetable is classified unstable ($S_r = 0$), regardless of the relative magnitude of $\Delta^{out,(r)}$.

The stability index S_r only reflects the propagation delay caused by the input delay. In the undisturbed baseline experiment, no initial disturbance is injected. Consequently, the baseline does not contain any input delay ($\Delta^{in,(r)} = \text{undefined}$). Because input and output delays are defined relative to the injected disturbances, the baseline by definition does not contain any output delay ($\Delta^{out,(r)} = 0$). This indicates that S_r would be undefined in the baseline. Instead, the baseline is assigned a normalized difference value of 0 ($\frac{\Delta^{out,(r)} - \Delta^{in,(r)}}{\Delta^{in,(r)}} = 0$), with a corresponding value of $S_r^{base} = 1$. This represents the absence of any input and output delay, indicating perfect delay containment, and forms an upper bound against which all disturbed experiments are compared.

Robustness Index

Robustness is the ability of a timetable to mitigate the propagation of stochastic delays and withstand stochastic variability in operational processes. The robustness index R_r is given in Equation 3.9.

$$R_r = \frac{1}{N} \sum_{i=1}^N R_{r,i} \quad (3.9)$$

High values of R_r indicate the ability of a timetable to absorb random fluctuations and maintain operational performance without generating delays. Lower robustness values indicate that small variations in day-to-day operational behavior lead to schedule deviations.

For each timetable $r \in \mathcal{R}$, a set of N replications is generated. For each $i \in N$, random delays are drawn from empirically fitted distributions, representing realistic operational variability. A robustness score $R_{r,i}$ is calculated for each replication $i \in N$, based on the proportion of activities that remain within a specified range from their scheduled times. Let \mathcal{A} denote the set of activities for which $\mathcal{A} = \mathcal{K} \times \mathcal{P}$. For each activity $a = (k, p) \in \mathcal{A}$, let $t_a^{real,(r,i)}$ denote the realized time of activity a in timetable r and replication i and let $t_a^{sched,(r)}$ be the corresponding scheduled time. A tolerance parameter $\lambda \geq 0$ reflects the level of variation that can be absorbed without operational degradation. The robustness score $R_{r,i}$ is defined by the sum over binary variable $q_a^{(r,i)}$, shown in Equation 3.10.

$$q_a^{(r,i)} = \begin{cases} 1, & \text{if } t_a^{real,(r,i)} \leq t_a^{sched,(r)} + \lambda, \\ 0, & \text{otherwise.} \end{cases} \quad (3.10)$$

Equation 3.11 shows the robustness score $R_{r,i}$ of timetable r for replication i as fraction of activities that remain within the allowed tolerance.

$$R_{r,i} = \frac{1}{|\mathcal{A}|} \sum_{a \in \mathcal{A}} q_a^{(r,i)} \quad (3.11)$$

$R_{r,i} = 1$ indicates that all activities ($a \in \mathcal{A}$) remained within the allowed margin λ . Timetable r thus absorbs the stochastic variability for that specific replication i . Lower values of $R_{r,i}$ indicate that a larger proportion of activities exceed the allowed deviation.

Operability Index

Feasibility forms the structural foundation of an effective timetable and should always be included in a timetable analysis. In its formal definition, a timetable is feasible if and only if no blocking-time conflicts occur. Any overlap between blocking-times implies infeasibility, for which stability and robustness cannot be ensured. In practice, schedules are not always conflict-free. Although such timetables are theoretically infeasible, they are still operationally executable. Nevertheless, a timetable can become operationally in-executable when the total amount, duration, or severity of overlaps becomes too large. In this context, the degree of operability is quantified rather than taking (in)feasibility into account as a strict binary condition.

Operability reflects whether a timetable can be executed within a given reference operational cycle period ($T^{cycle,ref}$), despite the presence of blocking-time conflicts. The operability of a timetable is related to the additional cycle time that would be required to eliminate all blocking-time conflicts of the former cycle ($\Delta T_{b,r}^{conflict}$) for a specific corridor. Equation 3.12 defines the operability index $O_{b,r}$.

$$O_{b,r} = 1 - \min\left(1, \frac{\Delta T_{b,r}^{conflict}}{T^{cycle,ref} - T_{b,r}^{cycle,infeasible}}\right) \quad (3.12)$$

Let $T_{b,r}^{cycle}$ denote the time required to run one timetable cycle of timetable r for the corridor b . If timetable r in corridor b contains conflicts, the time to execute one full timetable cycle is denoted by $T_{b,r}^{cycle,infeasible}$. By de-conflicting the schedule (i.e., shifting following trains until their blocking-times become adjacent and not overlapping), a conflict-free version of the operational timetable can be obtained with cycle time $T_{b,r}^{cycle,feasible}$. By definition, as conflicts are resolved by shifting trains while preserving the train order and routing, $T_{b,r}^{cycle,feasible} \geq T_{b,r}^{cycle,infeasible}$. Furthermore, let $T^{cycle,ref}$ denote the time window for which the timetable is intended to operate a given number of trains, for example one hour. The ratio $\frac{\Delta T_{b,r}^{conflict}}{T^{cycle,ref} - T_{b,r}^{cycle,infeasible}}$ measures how much of the available operational slack, the unused time within the reference period ($T^{cycle,ref} - T_{b,r}^{cycle,feasible}$), is consumed by blocking-time conflicts. If there are no conflicts ($\Delta T_{b,r}^{conflict} = 0$) and $T_{b,r}^{cycle,infeasible} = T_{b,r}^{cycle,feasible} \leq T^{ref}$, a timetable attains maximum operability. For $O_{b,r} = 1$ the schedule is conflict-free and can be executed within the reference operational period. If there are delays present, the additional time to carry out the timetable in a conflict-free manner increases with $\Delta T_{b,r}^{conflict}$, such that $T_{b,r}^{cycle,feasible} = T_{b,r}^{cycle,infeasible} + \Delta T_{b,r}^{conflict}$. A larger share of the available slack is required to resolve conflicts, such that $0 \leq O_{b,r} \leq 1$. For schedules where the additional time exceeds the available slack, the timetable can no longer be executed within the reference period ($T_{b,r}^{cycle,feasible} \geq T^{cycle,ref}$). In such cases, the timetable is considered non-operable with $O_{b,r} = 0$.

Let C denote the set of blocking-time conflicts for a corridor b in timetable r . For each conflict $c_{b,r} \in C$, the blocking-time overlap is given by $\tau_{c_{b,r}}$. The following train involved in $c_{b,r}$ is given by $k_{f(c)}$, where $f(c)$ maps each conflict to its following train k_f . Resolving a conflict requires shifting the departure times of the following train by at least $\tau_{c_{b,r}}$. When multiple conflicts involve the same following train, the required time shift accumulates. For each train $k \in \mathcal{K}$, the total time shift required to resolve all conflicts in which train k is the following train $k_{f(c)}$ is given by Equation 3.13. The accumulated time shift represents the minimum shift in departure times that must be applied to train k to eliminate all its conflicts.

$$\Delta T_{k,b,r}^{accum} = \sum_{\substack{c_{b,r} \in C \\ k_{f(c)} = k}} \tau_{c_{b,r}} \quad (3.13)$$

Because shifting the departure of one train will possibly create conflicts with other subsequent trains in the timetable, conflict resolution is carried out on all trains k in corridor b for timetable r . However, the secondary *network* effects of shifting one train on the other parts of the network are not taken into account. Therefore, the minimum additional cycle time required to obtain a conflict-free timetable is determined by train k with the largest accumulated shift $\Delta T_{k,b,r}^{accum}$. This results in a conservative lower bound on the cycle time required for a conflict-free timetable r for corridor b .

$$\Delta T_{b,r}^{conflict} = \max_{k \in \mathcal{K}} \Delta T_{k,b,r}^{accum} \quad (3.14)$$

3.2.4. Towards an Integrated Effectiveness Indicator

In the context of temporary timetables designed under PTCs, effectiveness requires: (i) maintaining an adequate level of traffic flow through temporary constrained infrastructure, (ii) containing input delays, (iii) absorbing the stochastic variability that characterizes day-to-day operations, and (iv) ensuring that timetables remain operationally executable, even when they are infeasible. Based on these requirements, an ex-post performance indicator is proposed that evaluates timetable effectiveness using realized or simulated traffic data. The Temporary Timetable Effectiveness indicator (TTE_r) integrates capacity, stability, robustness, and operability into a single indicator. TTE_r is developed as a multi-criteria data-driven performance quality indicator that assesses the gap between expert-based evaluation and quantitative assessment of temporary timetable effectiveness. It is defined by Equation 3.15. TTE_r is defined at the level of an entire temporary timetable, yet not all indices can be quantified at this level. Therefore, $C_{b,r}^{mix}$ and $O_{b,r}$ are evaluated at the corridor level.

$$TTE_r = \alpha \cdot C_{b,r}^{mix} + \beta \cdot S_r + \gamma \cdot R_r + \epsilon \cdot O_{b,r} \quad (3.15)$$

$C_{b,r}^{mix}$ captures the ability of a timetable to maintain traffic flow through the critical corridor, under heterogeneous traffic conditions. The stability index S_r represents the ability of the schedule to contain delay propagation. The robustness index R_r reflects the degree to which operational variability is absorbed without generating delays. Finally, $O_{b,r}$ evaluates the operability of a temporary timetable, given both feasible and infeasible schedules. The weights α , β , γ , and ϵ determine the relative contribution of each dimension to reflect preferences and priorities. All indices are bounded between [0-1]. A higher TTE_r value corresponds to a more effective temporary timetable, whereas lower values indicate lower timetable effectiveness. TTE_r is defined as a weighted sum, such that each dimension contributes to the overall assessment. This allows for a nuanced comparison between alternative temporary timetables and highlights the strengths and weaknesses across the four performance characteristics.

3.2.5. Stress Testing

A stress test is conducted to evaluate the performance of the TTE_r indicator under increasing operational pressure. Timetable compression progressively removes slack from the schedule and simultaneously affects the reference period of the timetable and the planned running times of trains. Dwell times and other process times are kept unchanged. First, the reference period is reduced, forcing the schedule to fit within a shorter time window and compressing headway buffers and slack. Second, the planned running times are proportionally compressed. Trains are allocated less planned time to travel between successive timetable points, which increases infrastructure utilization. For small compression levels, this primarily removes running time supplements. For larger compression levels, the planned running times may become shorter than the minimum technical running times. This does not imply that trains actually run with shorter running times than technically possible. Instead, the compressed schedule represents an increasingly unrealistic plan. Once the planned running times fall below technical minima, trains are no longer able to realize the schedule and systematically accumulate delay. With increasing operational pressure, these deviations propagate through the system and may lead to deadlocks. For each compression level, there is scaled by a factor equal to the ratio between the constrained reference period and the nominal reference period. By keeping dwell and process times constant and compressing only the reference period and planned running times, the stress test isolates the effect of diminishing headway buffers, reduced running time margins, and increased infrastructure utilization.

In the stress test, TTE_r is examined across a range of compression levels using the Area Under the Effectiveness Curve ($AUEC$). Let $TTE_r(T_c)$ denote the timetable effectiveness corresponding to the compressed reference period T_c . The $AUEC_r$ is defined by Equation 3.16.

$$AUEC_r = \int_{T_{min}}^{T_{nom}} TTE_r(T_c) dT_c \quad (3.16)$$

Here, T_{min} represents the most compressed reference period and T_{nom} represents the nominal reference period. A larger $AUEC_r$ indicates that a timetable maintains higher effectiveness under constrained conditions.

Timetable Design Analysis

This chapter provides empirical insights into the factors that influence the effectiveness of temporary timetables used during constrained infrastructure conditions. [Section 4.1](#) outlines the general framework for the analyzes in this chapter. The qualitative factor analyzes based on planning norms, expert interviews and performance evaluations of real temporary timetables is given in [Section 4.2](#). In [Section 4.3](#), data-driven probability distributions are derived from realized operational data, which describe the realistic process variability for microscopic simulations.

4.1. Timetable Performance Analysis Method

The timetable design analysis consists of two components, namely a qualitative and quantitative analysis. The qualitative analysis uses planning norms, interviews and Prestatie Analyse Bureau (PAB) evaluation reports (i.e., timetable performance analysis reports from the corresponding PAB-department at the rail infrastructure manager ProRail). These sources are used to identify which factors influence timetable effectiveness and how these factors interact in real-life operations. The planning norms provide information on the nominal situation, while interviews and PAB reports reveal which factors are unique to temporary timetables under PTC conditions. Realized operational data are used to analyze and estimate stochastic activity time distributions in a quantitative manner. These distributions are used as input for a realistic microscopic simulation model used in the case study. An overview of the methodology implemented to perform such a timetable performance analysis is illustrated in [Figure 4.1](#).

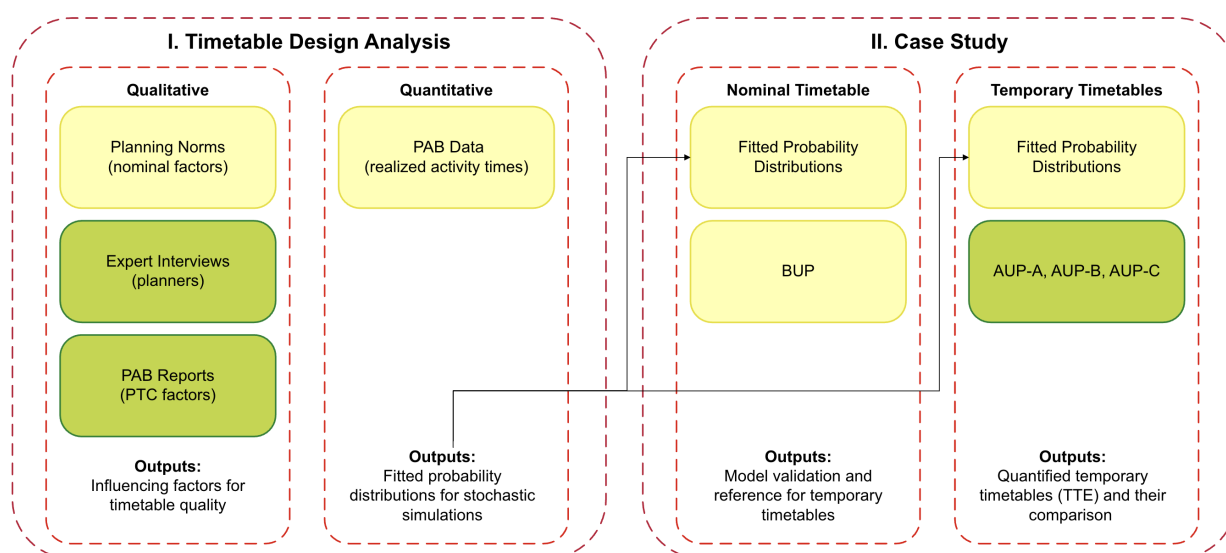


Figure 4.1: Overview of data sources and their outputs.

Note: Yellow represents nominal data and green represents PTC-specific data.

4.1.1. Data Sources

To define which factors and their correlations affect timetable effectiveness in real-life operations, to define probability distributions, and to validate the proposed multi-criteria performance quality indicator *TTE*, various data sources are needed. These data sources enable both qualitative understanding of factors affecting timetable quality and quantitative modeling of stochastic variability for simulation. Firstly, a nominal timetable for a specific corridor is needed. This timetable represents regular operations without infrastructure limitations. Secondly, different alternative temporary timetables for the same corridor are needed. These AUPs, represent a different planning solution for the same constrained infrastructure condition. Having alternative AUPs makes it possible to identify how different design choices affect capacity, feasibility, stability, and robustness, and thus to evaluate which temporary timetable performs best according to *TTE*. For all timetables, data on the infrastructure is necessary. Furthermore, realized operational data of the corridor are necessary to provide detailed process time measurements. Approximately one month of data is sufficient to capture typical variability patterns while remaining manageable for (pre)processing and analysis. This data is used to fit probability distributions that represent the normal stochastic variation on this corridor, defining the input for microscopic stochastic simulations. This ensures that the modeled disturbance behavior reflects realistic operational spread. Fourth, planning norm documentation is necessary. These norms are derived from large-scale empirical analyzes of realized performance under nominal conditions. They represent the baseline principles of good planning practice and thus reveal which key structural and operational factors are believed to determine the quality of the schedule. Although they do not specifically focus on temporary timetables, they provide valuable information and serve as a reference framework against which PTC-specific factors can be interpreted. Lastly, post-hoc analyzes of specific PTCs contain a structured description of the situation, expectations, performance realized, and conclusions and recommendations. By comparing multiple analyzes, factors are identified that specifically influence the quality of temporary timetables, complementing the insights gained for the nominal schedules. Therefore, these reports help bridge the gap between nominal and temporary operational contexts.

4.1.2. Data Processing

Different process steps transfer raw input data into usable datasets. Figure 4.2 provides an overview of the complete data pipeline. Various software applications are used, namely DONNA (Designer Of Network schedules for National use), the national planning system of NS and ProRail used for capacity allocation and timetable management of the Dutch railway network, DONS (Designer Of Network Schedules), the timetable planning system, and FRISO (Flexible Rail Infrastructure Simulation of the Operation), ProRail’s micro-simulation tool for analyzing and assessing the quality of infrastructure, timetables, and/or traffic control options.

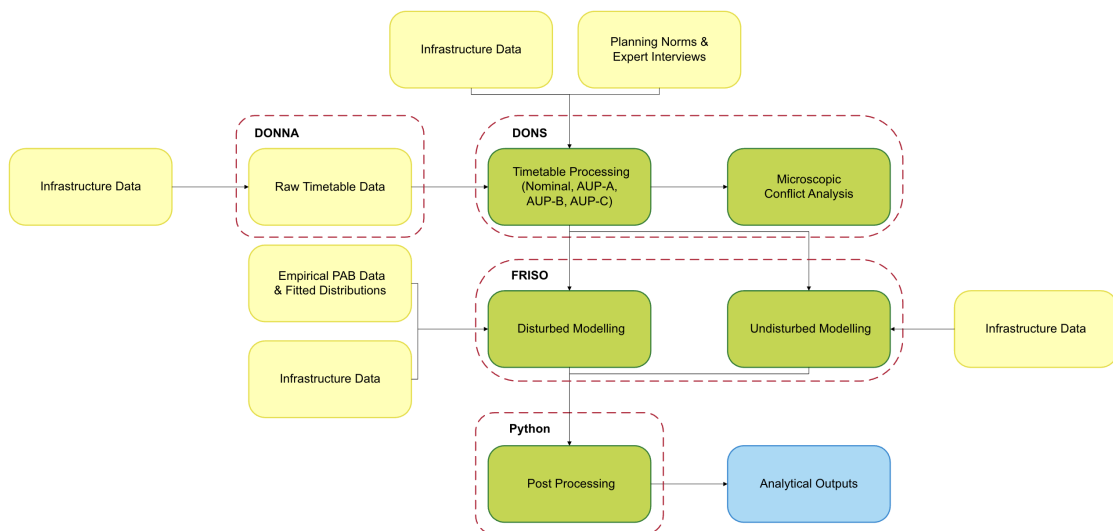


Figure 4.2: Overview of data processing steps.

Note: Yellow elements represent inputs, green elements represent intermediate processing steps, and blue represents outputs.

The raw DONNA timetables are imported into DONS, where overlapping train patterns, often resulting from duplicated patterns at different times of the day (e.g. week/weekend, off-peak/peak), are identified and resolved. This removal of patterns should be done in consultation with experienced planners to preserve the representative service structure. Empty rolling stock and shunting movements can be removed from the timetable as they may cause station-level conflicts. These conflicts are of a too detailed level and have negligible influence on the network-level delay propagation. However, pattern-based shunting movements with operational relevance should be kept and validated by planners. These timetable processing steps ensure that the remaining train paths realistically represent operational services, while focusing on network-scale dynamics. After these processing steps, conflicts are identified and quantified. For each timetable, the number, duration, and location of conflicts is extracted. The processed schedules are exported to FRISO and PAB realization data is imported and linked to timetable plans to generate stochastic disturbance distributions. These fitted distributions are used to introduce realistic disturbances, and thus stochasticity, in the simulations. For each timetable, two scenarios are simulated. An undisturbed scenario to verify the consistency of the input data and to provide baseline performance outputs of each timetable. From these first simulations, data are retrieved (i.e. delay monitoring and realized departure and arrival times), which serve as inputs for further experimental analysis considering fitted delay distributions. Second, simulations of delayed operational conditions are conducted to evaluate timetable performance under disturbed conditions and stochastic variability.

4.1.3. Data Verification

Verification of data ensures that the raw data sources are consistent and correctly linked. Several consistency checks between the data sources and the simulation environment exist. First, variables such as train numbers and schedule times were cross-checked between DONNA, DONS, and FRISO to ensure identical data structures after import and export. Furthermore, temporary timetables are verified against the infrastructure layout defined for the PTC, ensuring that no train paths use track sections, switches, or other infrastructure that are unavailable. Lastly, experiments under undisturbed operational conditions are simulated for all timetables to confirm that the simulation does not freeze or fill up (i.e., deadlocks). In addition, several checks are performed to assess whether the processed empirical data accurately represent the operational reality. The selected month is examined for the absence of PTCs and unusual demand patterns and operational conditions. This ensures that the fitted distributions reflect normal variability rather than exceptional patterns. Moreover, expert confirmation and reviews of cleaned timetables ensure that the remaining train paths, frequencies, and routing choices are realistic. Together this ensures that empirical input data are reliable and representative for deriving stochastic distributions and further quantitative analysis.

4.2. Qualitative Factor Analysis

This section analyzes how real-life operational and planning factors influence the effectiveness of temporary timetables. These insights are based on ProRail's planning norms, post-hoc PAB evaluations of temporary timetables, and expert interviews.

4.2.1. Factors Based on Planning Norms

The effectiveness of temporary timetables is influenced by structural and operational factors. Based on internal documentation on planning norms, the following key factors of effectiveness of the schedule have been determined, particularly in terms of feasibility, stability, and robustness.

The first key factor is the degree and distribution of headway buffers. These planned time gaps between successive train movements on shared infrastructure reduce the risk of conflicts and limit delay propagation. Standard norms prescribe a 30-second buffer behind passenger trains, while freight trains are typically planned without a buffer. This buffer is included on top of the technical infrastructure occupation. In simulation, this planned buffer is not automatically enforced. Therefore, a deterministic norm of 60 seconds ("kielzog") is applied. This includes the 30 second buffer and an additional 30 seconds to assess robustness. The kielzog prolongs the infrastructure occupation behind a train, representing a situation in which a train experiences a small delay and releases the infrastructure later than planned. In practice, there are exceptions to these buffer norms, such as platform succession, where two trains consecutively use the same platform in the same direction. In all cases, technical minimum progress must be respected and, for any violation, the feasibility must be guaranteed.

Secondly, running time supplements and their distributions over a train path influence effectiveness. These supplements are added to the technical minimum running time to absorb variability in train driving performances over its own route (due to, e.g. weather conditions, leaves on tracks, driving style of the driver). This technical minimum is based on speed limits, rolling stock characteristics, maximum speed of the specified timetable, and traffic heterogeneity. Freight services are scheduled using conservative worst-case-scenario running times and receive no additional supplement. For passenger services, supplements vary between 4-8%, depending on operational complexity. Shunting movements have longer running time supplements (10%), as they are carried out over shorter distances and periods and thus have high variability. To resolve conflicts, it is permitted to in/out bend running times, to a maximum of 0.5 minute and 1.0 minute for passenger and freight services, respectively.

For dwell and turnaround times, ProRail has defined dwell times norms, depending on factors such as the type of service and rolling stock. One-person operations require 0.5 minutes dwell, regional services with a conductor in the departure procedure require 0.7 minutes dwell, and all other rolling stock have a minimum of 0.9 minutes dwell. Turnaround times vary between the 4-6 minutes depending on switching drivers and train lengths. Combining always requires 3 minutes, while splitting has a norm of 2 minutes. A long line (i.e., lines with more than 60 minutes of travel time from start to destination) can have long-lasting and propagating effects, such that additional buffer time is required for dwelling.

Fourth, logical and consistent infrastructure use avoids unnecessary movements, keeps operations simple, and increases safety and feasibility. Plan norms to maintain directional consistency are default to right-hand running, avoid corridor mixing, one-direction use of tracks, minimize cross-flows when entering and leaving nodes, and choosing logical halting tracks for the chosen rolling stock. Local infrastructure features (e.g., tunnels, bridges, platform length) impose additional planning constraints.

Finally, temporal and operational context affects how these factors interact. During peak hours or high traffic density, buffers and supplements are used close to their limits, making the timetable more sensitive to small disturbances and increasing the variability of operational process times. The planning phase and the type of temporary schedule determine how strictly the norms are applied.

4.2.2. Factors Specific to Temporary Timetables

Plan norms are based on nominal schedules, but applied to both nominal and temporary timetables, where the trade-off between efficiency and effectiveness shifts under constrained conditions. Temporary timetables can be based on stricter norms, as the network operates under more constrained conditions. However, these schedules are temporary, so the operational risk associated with delays is limited, while capacity and operational services remain critical. To balance these viewpoints, the same plan norms are applied in temporary timetable design as under nominal conditions. However, the context of PTC introduces additional factors that further determine how effective a temporary schedule is.

First, in addition to planning norms, temporary timetables are guided by design principles. These principles are used as qualitative guidelines to ensure that deviations from standard norms are made consciously and transparently. These principles can be followed or deliberately relaxed if required. Design principles include the consistent use of platform tracks for one direction, as this reduces conflicting movements and complex platform changes. Furthermore, corridors should be kept as separate as possible, and infrastructure is used as intended. This includes avoiding mixing traffic from multiple lines and unnecessary left-hand running and turning at non-standard stations. Limiting interactions with unreliable services (e.g., High Speed Lines (HSL) or international freight services) decreases interdependencies and sensitivity to accumulating delays. Another design principle takes into account temporary speed and infrastructure restrictions, as these affect running times and their operational realism and accuracy. Design principles give room to exceptions and therefore it is necessary to check for accumulated exceptions. Individual deviations may be acceptable, while combinations may significantly reduce feasibility, stability, or robustness.

Secondly, analysis of PAB evaluation reports reveals additional factors that affect timetable effectiveness. The incomplete or outdated incorporation of Temporary Speed Restrictions (TSB) causes realized running times to systematically exceed planned running times, causing severe delay propagation despite compliance with norms. Secondly, temporary timetables often use dwell times from the nominal timetable, while changes in passenger boarding, platform usage, or passenger intensity ask for different

norms. Ensuring location-specific dwell times can mitigate these risks. Another influencing factor is long or coupled line operations. Extended line lengths decrease stability and robustness as delays accumulate over longer distances. Splitting lines and introducing additional turn-backs or layovers have proven to increase robustness. Similarly, short and complex reversals on atypical tracks or single-track configurations are critical for delay propagation and should be accompanied with larger buffers.

The intensity, dependent on the type of PTC, is a critical factor. A distinction exists between blocking and non-blocking closures. Blocking PTCs entirely obstruct train movements between two timetable points A and B and require either turn-back operations or rerouting. In turn-back operations, trains terminate at (one of the) boundary stations and reverse direction. This often requires left-hand running on short segments near the work zone. *Rerouting*, where trains bypass the closed section entirely via an alternative path (e.g., A->C->B), maintains services but often increases travel times and possible conflicts on the rerouted route. Non-blocking PTCs allow trains to continue through the work area, but with reduced capacity, which requires timetable adjustments to maintain feasibility, stability, and robustness. The level of impact and adjustments required are dependent on the intensity and spatial extent of the closure. A graphical representation of these PTC categories is shown in [Figure 4.3](#).

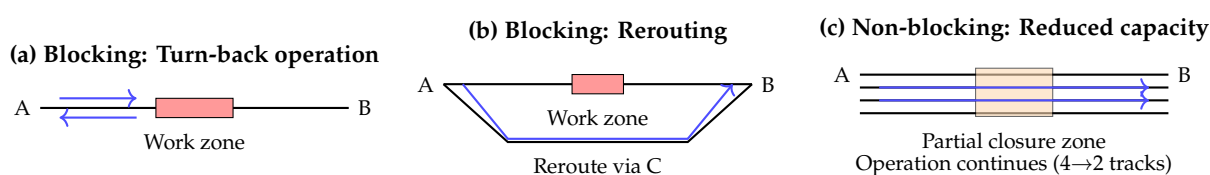


Figure 4.3: Schematic overview of different types of Planned Track Closures (PTCs).

Lastly, the duration of a PTC influences both timetable design and evaluation. Short-term closures, typically lasting a weekend, prioritize continuity and efficiency, accepting tighter margins and reduced robustness. Medium-term closures require more coordination across departments and partial timetable redesign, while long-term closures (≥ 18 weeks), are integrated into the national planning framework and allow for buffer adjustments and passenger service adaptations.

[Table 4.1](#) summarizes all identified factors and indicates relevance in nominal and temporary contexts.

Table 4.1: Comparison of key factors influencing timetable effectiveness under nominal and PTC conditions.

Factor	Nominal timetables	Temporary timetables
Headway buffers	Prevent delay propagation. Standard norm of 30 or 60s.	Same norms applied, sometimes reduced locally due to capacity loss.
Running time supplements	Absorb small operational variability. Standard norm of 4–7% for passenger and 0% for freight.	Same norms applied, but running times dependent on TSBs.
Dwell times	Ensure reliable boarding. Standard norms fixed per rolling stock and operation type.	Same norms applied, but higher variability due to irregular passenger flows and changed transfer patterns.
Turnaround time	Ensure reliable reversals. Standard norm of 4–6 min.	More critical due to reduced margins and complex reversals.
Infrastructure usage	Enable directional consistency.	Design principles applied to ensure safety and feasibility.
Time of day	Peak hours require stricter respect of buffers and supplements.	Same norms applied, mostly carried out during off-peak hours and weekends.
TSBs	Incorporated in planning if permanent.	Underestimated running times.
Long line lengths	Additional norm of extra buffers.	Splitting lines improves robustness.
Crossing movements	Managed by node design and buffers.	Riskier under reduced capacity where minor conflicts can propagate quickly.
Type of PTC	—	Blocking vs. non-blocking PTC determines operational adjustments.
Duration PTC	—	Trade-off efficiency and effectiveness.

4.2.3. Inter-dependencies and Combined Effects

Each of the discussed individual key factors strongly influences timetable quality, but their inter-dependencies largely determine actual performance. Temporary timetables operate under constrained conditions, where trade-offs between efficiency and effectiveness become increasingly important. Headway buffers and running time supplements determine the timetable’s ability to absorb stochastic variability. Headway buffers prevent delay propagation, whereas running time supplements enable an individual train to make up for a small delay. An imbalance between these factors can decrease effectiveness. Therefore, it is important to maintain sufficient recovery time within train paths and enough separation between them, simultaneously. Secondly, dwell and turnaround times interact and depend on headway buffers and running time supplements. Short dwell times improve throughput and efficiency, but increase sensitivity to operational and passenger-related variability. In contrast, longer dwell or turnaround times can serve as a local buffer. Infrastructure usage further sharpens these interactions as unconventional operations increases the need for buffers and supplements. The advantages of well-distributed buffers can be decreased by additional complexity, such as coupled line operations and directional inconsistencies. The operational context also influences how disturbances propagate and how factors should be prioritized. Short-term PTCs place more emphasis on efficiency, while longer-term PTCs often include longer recovery times and less complex line structures, such that timetable effectiveness increases. The time of day and traffic intensity influence how all these factors behave and interact. As a result, the same planning norms can lead to different effectiveness outcomes depending on the overall traffic load and contextual circumstances on the network.

4.2.4. Implications for Timetable Effectiveness

Headway buffers, running time supplements, and dwell and turnaround times are essential for timetable effectiveness. However, under PTC conditions, the operational circumstances become more constrained and less predictable. TSBs, reduced track availability, complex operations, and modified routing patterns increase the sensitivity of timetables to factors that are otherwise stable under nominal conditions. This requires not only a PTC-specific interpretation of existing factors, but also the definition of additional factors unique for temporary timetables. The effectiveness of the temporary timetables depends both on quantitative buffers and supplements and on qualitative coordination and design choices.

4.3. Quantitative Data-Analysis

The quantitative analysis derives disturbance distributions which represent operational variability. [Subsection 4.3.1](#) dives into detail on the realized PAB data, [Subsection 4.3.2](#) reveals the main insights, and [Subsection 4.3.3](#) presents the distributions.

4.3.1. Data Description

The basis for this analysis are the realized operation data extracted from the PAB database. This database stores detailed event-level information on train characteristics (e.g. series, numbers, type, etc.), network characteristics (e.g. timetable control points), and train movements and plans (e.g., arrival (A), departure (V), drive through (D)). The raw PAB data set contains timestamped observations of train movements, structured according to the control point of the timetable, the type of activity, and the train number. For each event, the observed and scheduled activity times are compared to the second. A snippet of the data set is shown in [Figure 4.4](#). The variables from left to right are: operating date, train series, train number, service type, timetabling point, activity type, planned timestamp, actual timestamp, original planned time from traffic control, and timestamp of a safe departure/entry signal aspect.

nvgb.verkeersdatum	basic.treinr_treinserie	basic.treinr	basic.treinr_rijkskarakter	basic.drp	basic.drp_act	basic.plan	basic.uitvoer	vklo_plan	trento_av.verteknijsein_seinveilig
2025-04-01	120NB	123	ICE	Btl	D	2025-04-01 11:46:00	2025-04-01 11:48:05		
2025-04-01	120NB	123	ICE	Lpe	D	2025-04-01 11:49:00	2025-04-01 11:49:56		
2025-04-01	120NB	123	ICE	Beto	D	2025-04-01 11:52:00	2025-04-01 11:51:48		
2025-04-01	120NB	123	ICE	Bet	D	2025-04-01 11:53:00	2025-04-01 11:52:34		
2025-04-01	120NB	123	ICE	At	D	2025-04-01 11:56:00	2025-04-01 11:54:26		
2025-04-01	120NB	123	ICE	Ehs	D	2025-04-01 11:58:00	2025-04-01 11:56:42		
2025-04-01	120NB	123	ICE	Ehv	A	2025-04-01 12:00:00	2025-04-01 11:59:21		2025-04-01 11:56:43
2025-04-01	120NB	123	ICE	Ehv	V	2025-04-01 12:01:00	2025-04-01 12:00:38		2025-04-01 12:00:10
2025-04-01	120NB	125	ICE	Btl	D	2025-04-01 13:46:00	2025-04-01 13:47:00		

Figure 4.4: Overview of PAB data set, indicating all variables (columns).

The reference period for this analysis is April 2025. According to planners and data performance analysts, one month of reference data is sufficient to derive representative disturbance distributions. April 2025 is a stable reference, as it was a month without PTCs on and around this corridor, and it lies outside the peak holiday season. These factors ensure that the traffic mix, the service intensity, and the demand for passengers are reflected according to nominal operations.

4.3.2. PAB Data-Analysis

Insights are gained regarding the statistical characteristics and operational variability on the Boxtel-Eindhoven corridor using various analyzes. Delays longer than 5 minutes (300 seconds) are excluded from this analysis, such that the focus is on small, frequent operational variations. Larger disturbances and disruptions that cause capacity loss, cancellation, or require rescheduling of rolling stock, crew, or timetables are not taken into account. Therefore, the derived delay distributions represent the realistic stochastic variability that temporary timetables should be able to absorb. This is consistent with planning practice, where timetable performance concerns the absorption of day-to-day variability.

Figure 4.5 and Figure 4.6 show the delay percentiles with 95% confidence. The lower percentiles (10th-40th) lie close to zero, indicating that a large share of trains arrive almost exactly on time. Negative delays indicate early arrivals, which result from unused buffer times and running time supplements. From the 50th to the 80th percentile, delay values increase and the confidence interval widens. The 90th percentiles reach roughly 1:45 minutes and 2:25 minutes with the widest confidence interval, for arrival and drive through delays, respectively. Thus, the majority of train movements is stable, but a small subset of trains experiences substantially larger delays. Jumps from the 80th to the 90th percentile indicate right-tailed distributions.

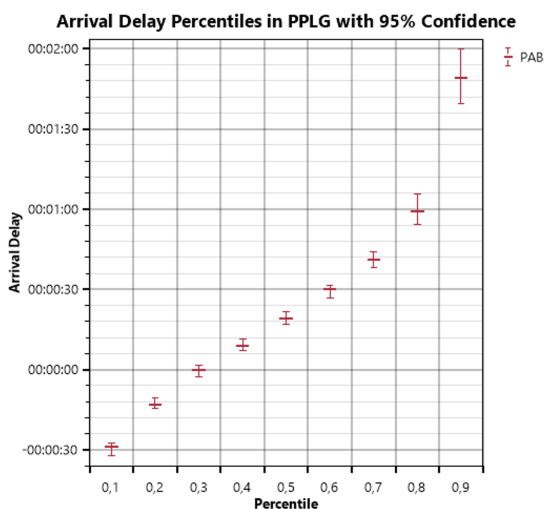


Figure 4.5: Arrival delay percentiles of PAB data.

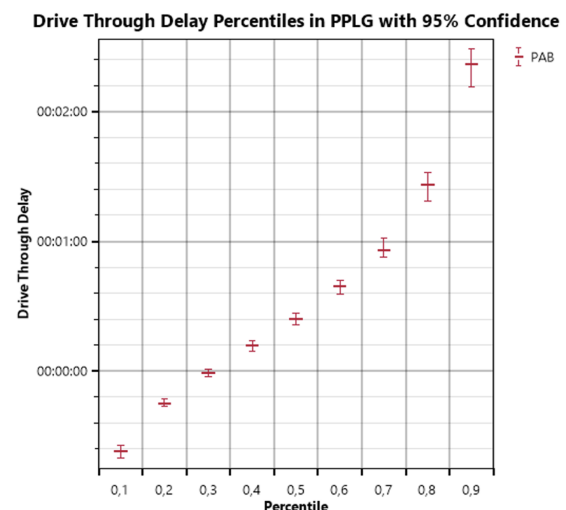


Figure 4.6: Drive through delay percentiles of PAB data.

Arrival and drive through delay distributions are visualized with a histogram, shown in Figure 4.7 and Figure 4.8. These histograms have a high density peak between -60 and 60 seconds. For arrival delays, the plot shows a strong decline beyond a one minute delay, yet outliers up to 5 minutes remain present. This indicates the right-skewed nature of operational delays, where most delays occur close to the scheduled time, but larger delays occasionally occur. The drive through delay distribution has a longer left tail, indicating a higher occurrence of early passages. This can be attributed to the absence of dwell times at drive through points, which allows trains to pass earlier than scheduled when upstream slack remains unused. Moreover, the right tail is thicker for the drive through distribution as the decline is more gradual. This indicates that drive through delays are more sensitive to delay accumulation, as there are no dwell times of local buffers to absorb delays.

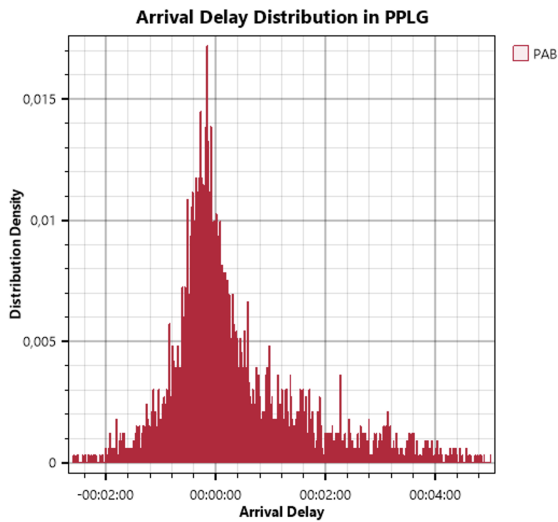


Figure 4.7: Arrival delay distribution of PAB data.

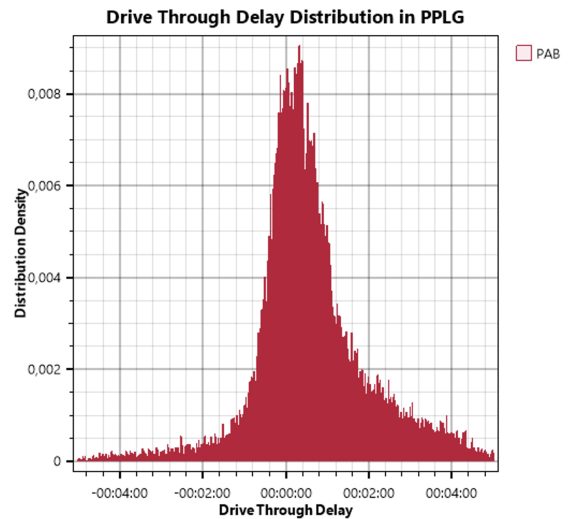


Figure 4.8: Drive through delay distribution of PAB data.

The moving average arrival delay, using windows of 900, 1800, and 3600 seconds, is presented in Figure 4.9. During the first hours, the delay shows strong fluctuations, which are progressively smoothed out when longer windows are applied. The moving average delay level stabilizes at approximately 30 seconds, independent of the chosen window length. The absence of an increasing trend indicates that the corridor does not experience accumulation of delays throughout the day.

Figure 4.10 compares delays across four train services at different timetable points. Intercity (IC) and Sprinter (SPR) services have the highest average delays at most points. The variability is limited and reflects predictable patterns. Freight (GDR) services show lower average delays because these services are planned with conservative margins. However, these delays have greater variability, which can be attributed to the lower priority in traffic operations. HSL services show the lowest delays, but with the highest variability. This is possibly influenced by the number of observations which is substantially lower for the HSL than for other services. Furthermore, HSL delays are unevenly distributed in both directions. A large share of data concerns services departing from the Netherlands towards international destinations, where delays are limited due to the short distance traveled. Observations in the opposite directions are more prone to accumulated delays due to the long distances traveled, but are less frequent in this dataset. Across the corridor, a slight increase in delay levels appears when approaching Eindhoven, indicating the increasing complexity and potential bottleneck effects.

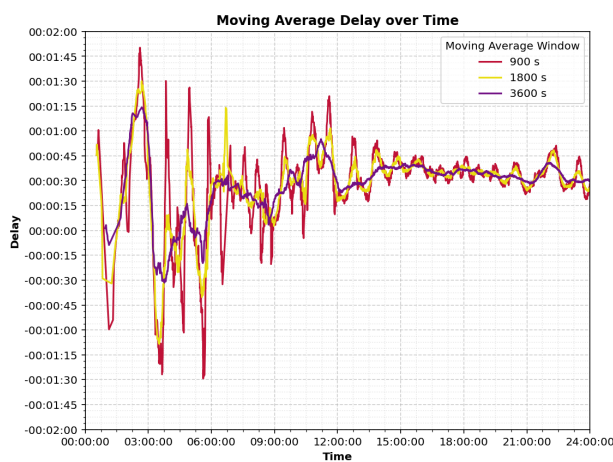


Figure 4.9: Moving average delays over time of PAB data.

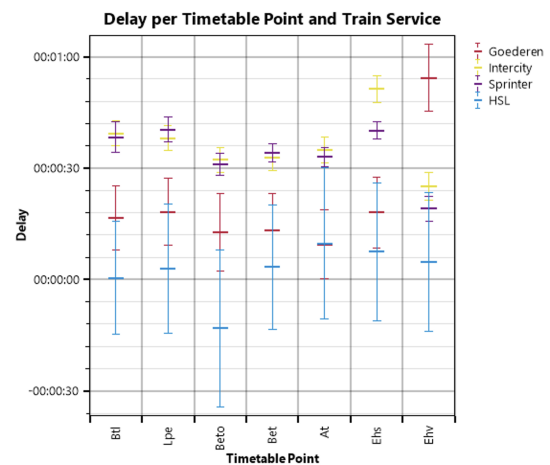


Figure 4.10: Delays per timetable point and service.

These analyzes reveal that most delays follow a right-skewed distribution. The majority of trains operate close to schedule, but a small subset of trains experiences larger delays. Despite these occasional greater delays, the Boxtel-Eindhoven corridor has a stable average operating system. The moving-average delay stabilizes around 30 seconds, indicating that delay propagation does not escalate throughout the timetable. Furthermore, delay variability is not uniform across train types. Freight trains tend to pass timetables points earlier than scheduled, with a larger spread. Intercity and sprinter services on average have larger delays, but have lower operational variability.

4.3.3. Fitted Disturbance Distributions

Disturbance distributions are derived and fitted that provide a quantitative representation of the stochastic variability observed in realized operational data. From the realized PAB data delays up to 300 seconds are included. Furthermore, distributions are estimated for train services with sufficient operations ($N > 100$). Four candidate distributions are fitted, namely Normal, Gamma, LogNormal, and Negative Exponential. For all non-Normal distribution types the data are shifted such that the minimum delays equal zero to ensure correct parameter estimation.

Each fitted distribution is visualized through a histogram with theoretical Probability Density Function (PDF) and a Cumulative Density Function (CDF). The histogram with theoretical PDF provides a comparison between empirical delay frequencies and predicted probability density. The CDF represents the probability that a delay is smaller than or equal to a certain value. A steeper CDF indicates that most observed delays are concentrated within a narrow range, whereas higher degrees of variability are observed for flatter curves. The quality of the fit is evaluated using the Quantile-Quantile (Q-Q) plot and the Root Mean Square Error (RMS). The Q-Q plot provides a visual measure of fit quality by comparing empirical quantiles of data with theoretical quantiles of the fitted distribution. Deviations of the observed data points from the theoretical diagonal indicate a bias in fit. S-shapes indicate heavier or lighter tails than the reference distribution, and concave distributions indicate skewness. The quality of fit is quantitatively evaluated using the Root Mean Square (RMS) on the CDF. The average difference in probability between the empirical and theoretical distributions is measured. For each observed delay, the probability calculated by the theoretical distribution is compared with the probability observed in the data. This RMS is bounded between 0 and 1, where smaller RMS values indicate better fits. A visual overview of all fitted distributions is shown in the histogram with theoretical PDF, CDF plot, and Q-Q plot in [Appendix A](#). The fitted parameters and distributions are summarized in [Table 4.2](#).

In general, the RMS values and Q-Q plots indicate that most fitted distributions achieve a good to moderate fit. For all departure processes, a strong peak close to 0 seconds is present. In the CDF plot, this can be seen by a steep theoretical CDF for the first 60 seconds. This reflects the dominant pattern where the majority of trains depart with no or little delay. For the central delay range (up to 150 seconds), the fitted distribution typically captures the empirical shape well. However, in the upper tails (delays > 250 seconds), the distribution fit is more poor. Empirical delays tend to be more extreme and variable than the theoretical distributions. This mismatch in distribution fit is visible in the Q-Q plots, where the highest quantiles systematically deviate from the theoretical distribution. A large share of the departure activities is fitted with a LogNormal distribution. The LogNormal distribution is a right-skewed distribution that is clustered around small values. This matches the empirical structure of departure delays. Normal distributions are able to capture the central mass adequately, yet struggle to correctly reflect the skewness and heavier positive delays correctly. Negative Exponential distributions perform well at locations with highly concentrated delays and limited tail behavior and delay variability.

Table 4.2: Overview of fitted disturbance distributions, derived from PAB analysis.

Process	Location	Distribution	RMS	Fit Quality
Depart A6400	Bet	NegExp(0.009940) -60	0.192	Good
Depart A6400	Btl	LogNormal(107.550, 71.046) -60	0.159	Good
Depart A6400	Ehs	LogNormal(129.227, 90.320) -0	0.374	Poor
Depart A6400	Ehv	LogNormal(491.683, 62.772) -480	0.158	Good
Depart B6400	Bet	NegExp(0.010587) -60	0.212	Moderate
Depart B6400	Btl	LogNormal(99.488, 62.200) -120	0.171	Good
Depart B6400	Ehs	LogNormal(155.870, 65.665) -120	0.192	Good
Depart B6400	Ehv	LogNormal(502.739, 67.400) -480	0.185	Good
Depart E1100	Ehv	Normal(16.185, 96.735) -60	0.232	Moderate
Depart E3500	Ehv	Normal(19.661, 85.782) -60	0.263	Moderate
Depart F3500	Ehv	LogNormal(144.869, 58.483) -120	0.209	Moderate
Depart G1100	Ehv	Normal(22.468, 97.535) -60	0.227	Moderate
Depart H3500	Ehv	NegExp(0.009693) -60	0.189	Good
Depart H6400	Bet	NegExp(0.010246) -60	0.206	Moderate
Depart H6400	Btl	LogNormal(229.503, 77.172) -240	0.165	Good
Depart H6400	Ehs	NegExp(0.009973) -60	0.198	Good
Depart H6400	Ehv	LogNormal(569.059, 74.664) -540	0.194	Good
Depart I2700	Ehv	LogNormal(106.469, 76.582) - 60	0.203	Moderate
Depart I6500	Bet	Normal(23.490, 90.553) - 60	0.262	Moderate
Depart I6500	Btl	Normal(27.812, 98.353) - 60	0.236	Moderate
Depart I6500	Ehs	Normal(30.729, 88.031) - 60	0.276	Moderate
Depart I6500	Ehv	Normal(4.479, 76.742) - 60	0.284	Moderate
Depart J2700	Ehv	LogNormal(142.957, 127.393) - 60	0.214	Moderate
Depart J6500	Bet	LogNormal(139.840, 79.223) - 120	0.121	Excellent
Depart J6500	Btl	LogNormal(111.615, 71.225) - 60	0.134	Excellent
Depart J6500	Ehs	LogNormal(129.639, 74.766) - 120	0.119	Excellent
Depart J6500	Ehv	LogNormal(4.479, 76.742) - 60	0.135	Excellent
Depart K2700	Ehv	Normal(24.653, 97.179) - 60	0.224	Moderate
Depart K6500	Bet	LogNormal(136.730, 61.082) - 120	0.170	Good
Depart K6500	Btl	LogNormal(138.991, 70.486) - 120	0.147	Excellent
Depart K6500	Ehs	Normal(29.237, 78.767) - 60	0.290	Moderate
Depart K6500	Ehv	LogNormal(487.451, 58.370) - 480	0.182	Good
Depart L2700	Ehv	LogNormal(168.611, 165.591) - 60	0.173	Good
Depart L6500	Bet	LogNormal(149.059, 86.954) - 120	0.125	Excellent
Depart L6500	Btl	LogNormal(1113.436, 75.588) - 60	0.138	Excellent
Depart L6500	Ehs	LogNormal(380.015, 85.994) - 360	0.134	Excellent
Depart L6500	Ehv	LogNormal(91.449, 56.768) - 60	0.300	Poor

5

Case Study

This chapter applies the methodological framework developed in [Chapter 3](#) and the empirical insights of [Chapter 4](#) to a case study. The objective of this case study is to validate the proposed model performance framework against real-world data and interpret how the design choices of the timetable affect the effectiveness. [Section 5.1](#) introduces the context of the case study and [Section 5.2](#) describes the simulation setup and input parameters. [Section 5.3](#) validates and verifies the simulation model using realized operational data from the same corridor during the PTC period. The results of the simulated scenarios are presented in [Section 5.4](#).

5.1. Case Definition

This case study aims to evaluate the performance of temporary timetables under realistic operating conditions, for the same PTC, by comparing the indices and the TTE_r indicator.

5.1.1. Study Area

The selected corridor for this case study is the Boxtel-Eindhoven Centraal (Btl-Ehv) corridor, located in the southern part of the Dutch railway network, see [Figure 5.1](#). This section is chosen because the case study considers a non-blocking PTC on the selected corridor. Non-blocking refers to a situation in which the infrastructure remains partially available for train services, but the capacity is reduced by closing a part of the infrastructure. Such cases are particularly interesting from a timetabling perspective, as trade-offs between maintaining service frequency, minimizing delays, and ensuring operational feasibility have to be taken into account. Because the PTC associated with this corridor reduced capacity without completely interrupting operations, it creates a realistic context in which different temporary timetable designs (AUPs) can be evaluated and compared.



(a) Study area shown within the Netherlands.



(b) Zoomed-in view of the southeast of the Netherlands.

Figure 5.1: Geographical representation of study area in the southeast Netherlands.

Note: Major nodes are marked in yellow, while all other operational points are indicated in red. The green lines represent the corridor Btl-Ehv.

The Btl-Ehv corridor includes the intermediate timetable points Liempde (Lpe), Best (Bet), Acht (At), Eindhoven Strijp-S (Ehs), and Eindhoven Stadion (Ehst) (which is not structurally operational, only during events). While the Btl-Ehv corridor forms the main focus, a broader spatial area is selected. This is necessary to determine the network effects of possible delay propagation. The study area, as shown in Figure 5.1b, includes the stations Breda (Bd), Zaltbommel (Zbm), Nijmegen (Nm), Venlo (VI), Roermond (Rm), and all intermediate minor stations. These additional nodes ensure that delay interactions between train services outside of the Btl-Ehv corridor (i.e., cross-series dependencies) are captured. Further extension of the study area is limited to avoid excessive complexity of additional freight corridors (e.g., the Betuwe Route) and large passenger stations (e.g., Utrecht Centraal, Rotterdam Centraal). Including other nodes would make it more difficult to distinguish and isolate the effects caused by the PTC and which are due to other external factors. Moreover, maintaining this bounded study area keeps the case-study computationally efficient for both data processing and simulations.

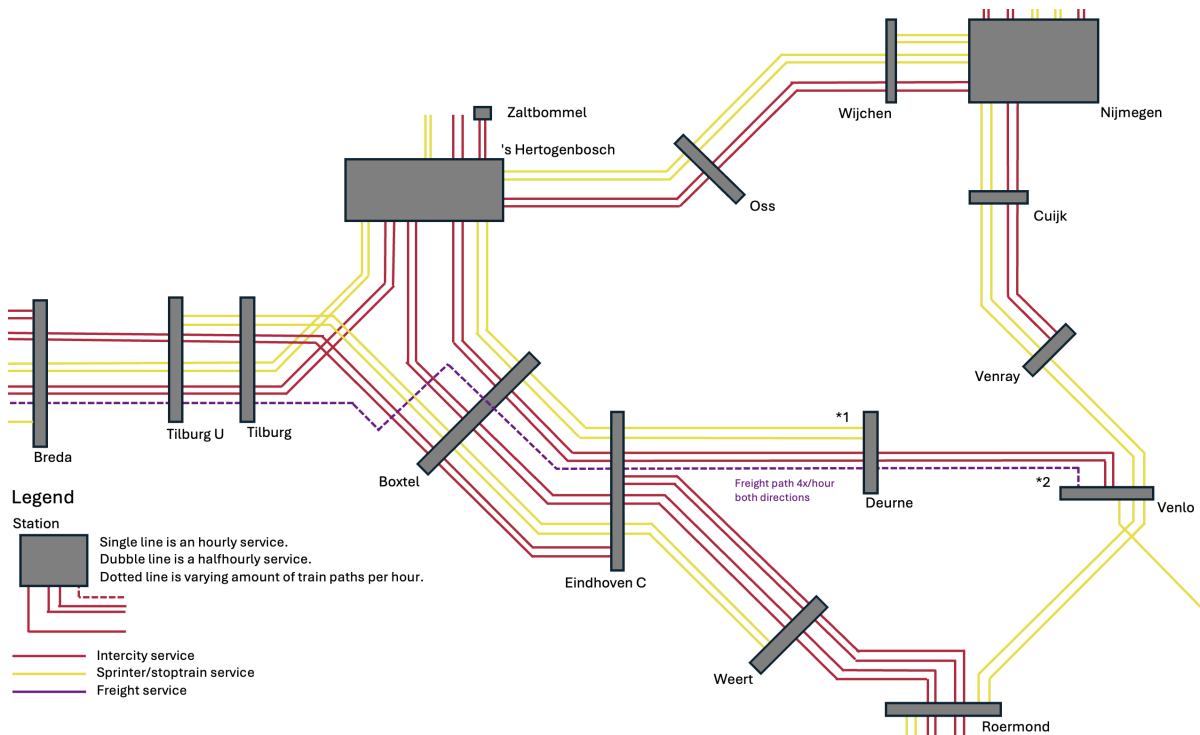


Figure 5.2: Schematic representation of passenger and freight service plan for the study area.

Note: Schematic representation is equal to the nominal timetable and temporary timetable AUP-C. *1: For temporary timetable AUP-A, this sprinter service is eliminated. *2: For temporary timetable AUP-B, the freight service is decreased to two paths/hour.

5.1.2. Corridor Characteristics

The Btl-Ehv corridor forms part of a multi-track corridor that connects several major nodes. A schematic representation of the passenger and freight services is illustrated in Figure 5.2 and a schematic overview of the infrastructure is shown in Figure 5.3. North of Boxtel, trains approach either from 's-Hertogenbosch or Tilburg, while south of Eindhoven, trains diverge towards Venlo or Roermond. This corridor is a combination of open line sections and junctions. Boxtel serves as a key convergence point for passenger and freight trains and Eindhoven Centraal functions as a major passenger interchange hub. The corridor has a mixed traffic composition, including IC, SPR, and GDR services, which have distinct origin-destination locations outside the corridor. For passenger trains, according to the nominal weekend schedule, three IC services (i.e., 1100, 2700, 3500) and two SPR services (i.e., 6400, 6500) operate on this corridor. All passenger services operate in a half-hourly pattern (i.e. twice per hour), per direction. Freight services operate on predefined train paths. The corridor is part of the Brabant Route, the southern freight corridor that connects the port of Rotterdam to Germany, via Tilburg, Eindhoven, and Venlo. In contrast to the Betuwe Route, this is a mixed-traffic line where both freight and passenger

services share the same infrastructure. This results in heterogeneous traffic dynamics and complex inter-train dependencies that influence timetable effectiveness. The time-distance diagram for this corridor is visually represented in Figure 5.4.

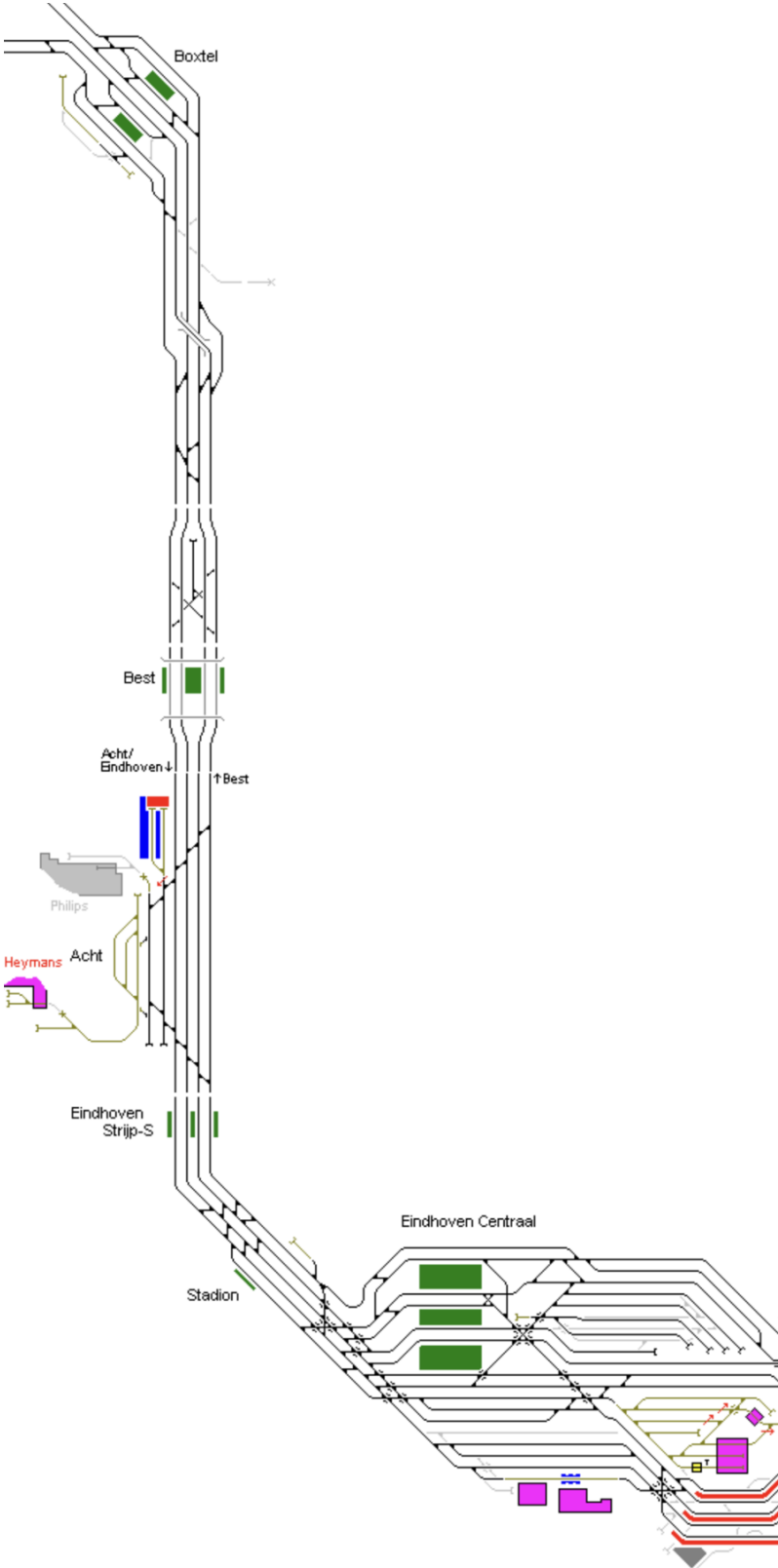
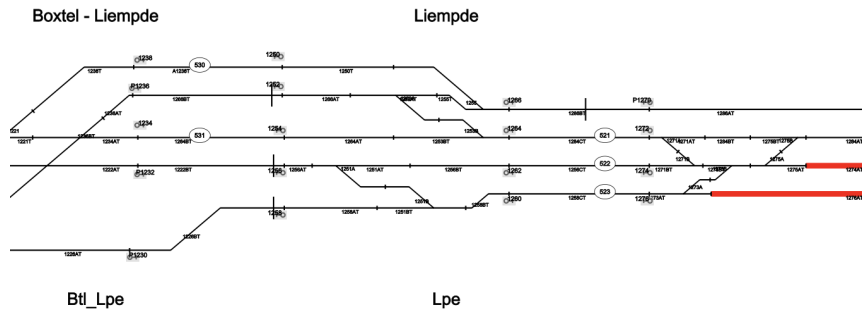
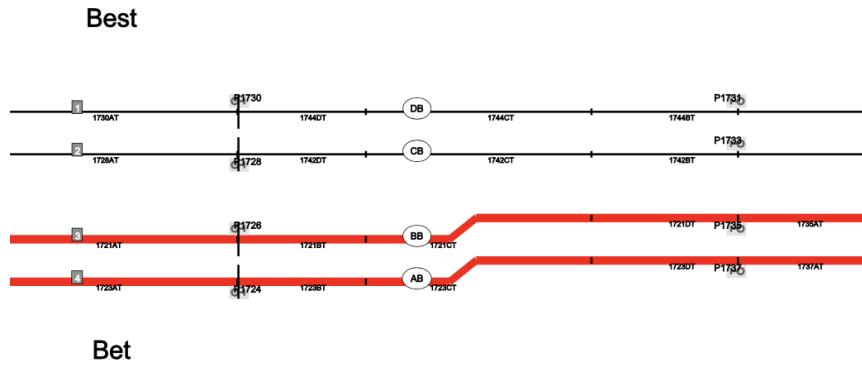


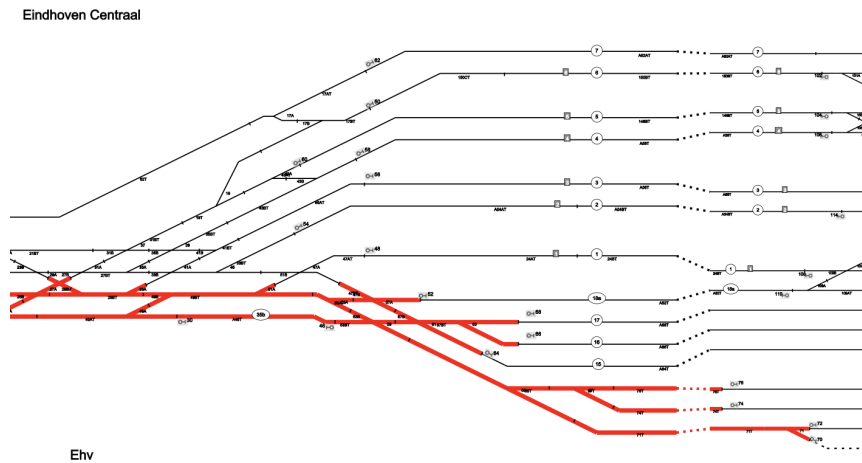
Figure 5.3: Schematic representation of infrastructure layout for the corridor Btl-Ehv (Sporenplan.nl, 2025).



(a) Schematic representation of part of the infrastructure at Boxtel-Liempde.



(b) Schematic representation of part of the infrastructure at Best.



(c) Schematic representation of part of the infrastructure at Eindhoven Centraal.

Figure 5.5: Schematic representation of parts of the infrastructure between Btl-Ehv during the PTC.

Note: Black lines represent available tracks, red lines represent unavailable tracks.

5.2.1. Defining Scenarios and Experiments

It is essential to distinguish between the different types of delays that arise within this case study. Initial disturbances refer to exogenous delays that are injected into the simulation as inputs. An input delay is defined as the delay carried forward by the disturbed train as a direct consequence of its initial disturbance. Output delays refer to all other additional delays that emerge endogenously by the disturbed train carrying its delay forward, train interactions, and shared infrastructure usage.

Four variants of the schedule are made. First, the nominal timetable which has no restrictions on the available infrastructure, and all services are maintained. Secondly, the AUP-A schedule has limited infrastructure availability, as two of the four tracks are unavailable. Consequently, the number of train

services decreases. Where all freight paths are maintained, passenger supply is reduced by eliminating a SPR service. AUP-B is similar, only the passenger service is fully maintained, while the freight service is reduced to two paths per hour. AUP-C maintains all services, both freight and passenger.

For each timetable variant, simulation scenarios are defined that are classified into undisturbed and disturbed operational conditions. The scenario of undisturbed operational conditions serves as a baseline in which no initial disturbances are injected. This scenario allows for the identification of structural conflicts and their impact under idealized operating conditions. This undisturbed scenario consists of a single experiment, denoted as Experiment U1 (Baseline). Simulation experiments of disturbed operational conditions represent realistic and degraded conditions. Two sets of experiments are conducted to assess the performance of temporary timetables in terms of stability and robustness under operational uncertainty. For stability, a set of experiments (S1 - S3) is conducted to assess the timetable's ability to absorb deterministic disturbances. A limited number of trains are subject to a single departure delay at a single infrastructure location. The experiments S1 - S3 differ in how the initial disturbances are distributed across trains, while the total deterministic delay is kept constant at 600 seconds. Experiment S1 considers a delay of 300 seconds to two trains (E1100 at EHV and F3500 at Btl), Experiment S2 considers a delay of 150 seconds to four trains (E1100 at EHV, F3500 at Btl, G6400 at EHV, and H6400 at Btl), and Experiment S3 considers a delay of 100 seconds to six trains (E1100 at EHV, F3500 at Btl, G6400 at EHV, H6400 at Btl, I2700 at EHV, and L2700 at Btl). To assess robustness, fitted delay distributions are applied in experiment R1, to evaluate timetable performance under stochastic variability. In this experiment, all fitted stochastic disturbance distributions are applied throughout the whole simulation horizon. The disturbances are continuously realized, reflecting normal day-to-day operational variability, and capture the average ability of the timetable to absorb stochasticity. To account for the randomness of stochastic disturbances, 25 replications of Experiment R1 are executed. Each replication has a different random realization of the disturbance processes, drawn from the fitted probability distributions.

5.2.2. Driving Behavior and Operational Assumptions

The driving behavior is varied because it determines how trains utilize available running time margins and react to disturbances. Without additional restrictions on driving behavior, trains operate at their maximum permitted speed and use the scheduled running time supplements. This leads to early arrivals, which reduces the simulation realism. A baseline static driving behavior "Stick to the plan" is applied, restricting all trains from operating at maximum speed to prevent unrealistic early running. In this case, trains are limited to a predefined fraction of their maximum speed. Various parameter values are used for different types of trains (e.g., passenger, regional, and freight) to reflect their operational characteristics. In addition, an adaptive driving behavior "Make up for delay" is introduced for passenger trains (i.e., IC, SPR, and HSL services). Under this behavior, trains operate at the same limited speed as long as they remain close to the scheduled activity times. If the delay stays below 60 seconds, the reduced speed is imposed. Once a train exceeds a delay of 60 seconds, the driving behavior changes and the train is allowed to operate at its maximum permitted speed. This represents a situation in which the driver attempts to recover delays. To avoid excessive and unrealistic delay recovery, the driving behavior switches back to the reduced-speed when delays decrease below 20 seconds. Very small delays (< 20 seconds) are often not noticed and corrected by drivers, such that continuously trying to recover from minor delays might lead to early arrivals. This adaptive driving behavior introduces heterogeneity in driving behavior within a single simulation, reflecting more realistic responses to delays.

5.2.3. Time Horizon and Warm-up Phase

A 30 minute warm-up phase is applied to ensure that sufficient trains have entered the network and that interactions between trains are established. Without warm-up time, a limited number of services are present on the network, which may lead to an underestimation of delay propagation effects. Performance measurement starts after the warm-up period and continues over a duration of 6 hours. A sufficient time horizon is needed to observe how deterministic and stochastic disturbances evolve over time and space, allowing for delay propagation and absorption. Shorter time horizons have the risk of truncating these dynamics, especially for disturbances introduced at the end of a timetable cycle.

5.3. Model Verification and Validation

Before analyzing the performance of temporary timetables, it is essential to ensure that the model produces technically correct and operationally reliable output. Verification dives into the correctness of the model, while validation concerns the reproducibility of observed traffic conditions from the simulation model given the same set of inputs. Together, verification and validation ensure that the simulation framework is technically correct, that stochastic processes behave as expected, and that the resulting effectiveness indicator can be calculated. Model verification focuses on the correctness and logical consistency of the simulation setup. All timetables are loaded into FRISO and checked for completeness. The 'test timetable' functionality is used to ensure network connectivity and data completeness (i.e., verifying if all trains can fulfill their planned paths from origin to destination). For alternative timetables, the routing patterns are verified against the PTC infrastructure layout. Furthermore, a baseline scenario of undisturbed operating conditions is executed to verify if the simulation runs without deadlocks or unrealistic standstill times. The validation steps focus on the nominal timetable and assess whether the simulation results align with real-world operational behavior. The simulated fitted distributions of the disturbed scenario are compared to the realized data. To assess whether the simulation model reproduces real-world operational behavior, the baseline experiment (U1) and the experiment of disturbed operational conditions with fitted distributions (R1) for the nominal timetable are compared against realized PAB data. The comparison is performed using the arrival delay distribution, delay percentiles, and the moving-average. Figure 5.6 - Figure 5.10 are the visual results.

The arrival delay histogram, Figure 5.6, shows a broader and flatter peak around -30 to +30 seconds for the PAB data. The day-to-day operational variation therefore results in a wider distribution of minor delays than the stochastic distributions generate. The experiments show sharper and more concentrated peaks, meaning that the simulation arrivals cluster more tightly. Both PAB and the disturbed simulation have a long right tail, confirming that the fitted delay distributions realistically reproduce the asymmetric structure and rare large delay events observed in practice. However, for longer delays (> 120 seconds), the disturbed simulation overestimates the probability of delay compared to the PAB data.

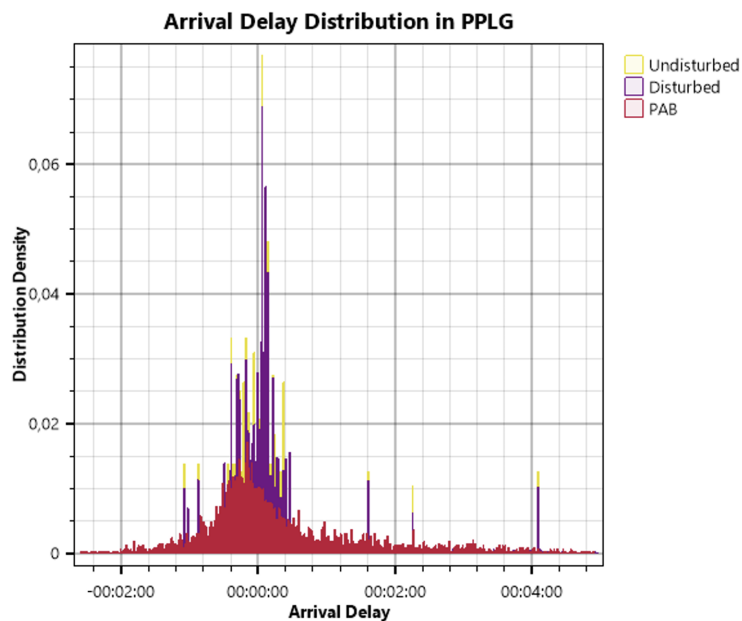


Figure 5.6: Arrival delay distribution of 'Undisturbed', 'Disturbed', and 'PAB'.

Figure 5.7 and Figure 5.8 present the arrival and departure delay percentiles. The PAB data and disturbed scenario include a 95% confidence interval to reflect statistical variability. The undisturbed experiment has small delays across all percentiles, indicating that the timetable contains structural conflicts. The disturbed arrival delays align closely with the PAB data for the lower and mid-range percentiles. For higher percentiles, arrival delays are underestimated. Infrequent but more severe delays are not fully

captured by the distributions. The disturbed experiment shows nearly zero departure delays up to the 50th percentile. The fitted disturbance expressions are truncated at zero, which creates a point mass at exactly zero. In contrast, the PAB data show a gradual increase in delays up to the 80th percentile. For the highest percentile, the delays increase, reflecting the accumulation of delays.

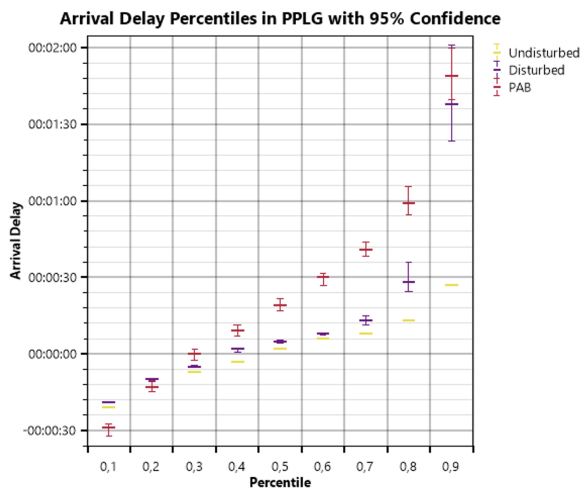


Figure 5.7: Arrival delays percentiles.

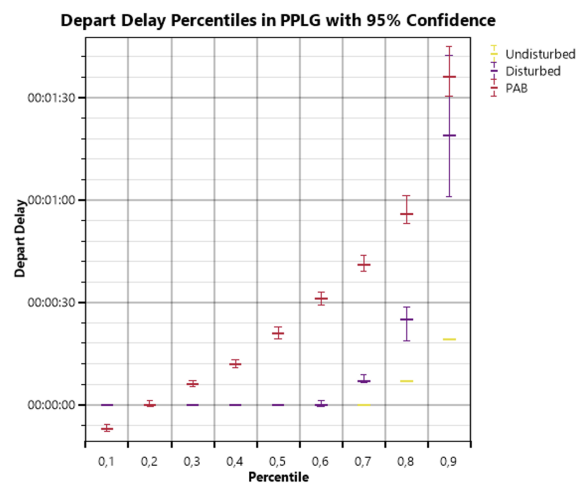


Figure 5.8: Departure delays percentiles.

The moving averages are shown in Figure 5.9 and Figure 5.10. For the undisturbed scenario, no stochasticity is introduced, which means that the present delays come from structural delays. The delay stabilizes at a minor delay level. For the disturbed experiment, the delay level is less stable, reflecting the stochastic operational behavior. For the first three hours, the delay level increases due to an accumulation of delays. The delay stabilizes at approximately 20 seconds. In contrast to the undisturbed scenario, the delay eventually decreases. This decrease may reflect partial recovery. The disturbed experiment has a more variable delay process due to continuous stochastic variability, while the undisturbed scenario converges to a deterministic, recurring pattern with a stable delay level over time.

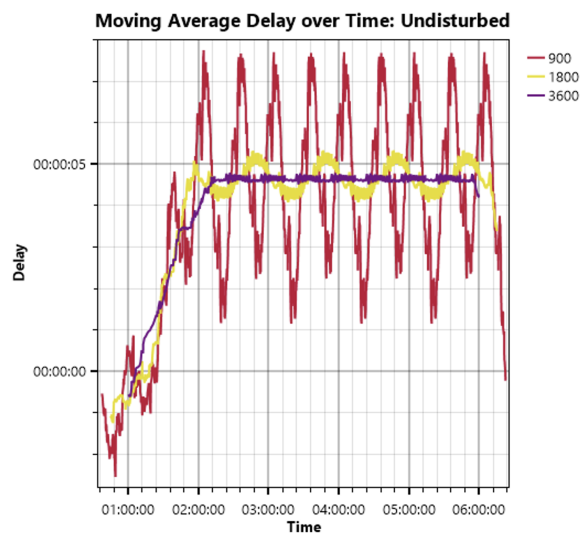


Figure 5.9: Moving average delays over times 'Undisturbed'.

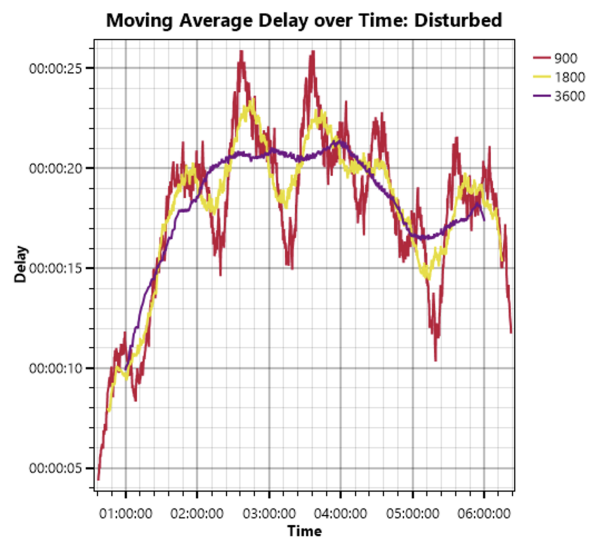


Figure 5.10: Moving average delays over time 'Disturbed'.

In conclusion, the non-zero delays in the undisturbed scenario confirm that the nominal timetable contains structural conflicts that generate delays, even without injected disturbances. The fitted

distributions capture everyday variability well, while extreme delay propagation fits more poorly. Despite this limitation in tail fitting, the simulation framework reproduces the essential structure of real-world patterns and therefore the fitted distributions are valid for the comparative analysis of timetable effectiveness.

5.4. Results

The results of the case study and the comparison of the timetables are presented. First, the results are given for each performance index, providing insight into the mechanisms underlying the effectiveness score. Secondly, the indices are aggregated into the *TTE* indicator. Additional sensitivity analyzes are conducted to assess the impact of modeling assumptions on the results. Lastly, a stress test is presented to evaluate the behavior of the *TTE* indicator under increasing operational pressure.

5.4.1. Timetable Assessment per Index

Capacity Index

Table 5.1 summarizes the capacity-related metrics. AUP-C preserves the largest share of nominal capacity, followed by AUP-A and then AUP-B. This is reflected in higher values of $C_{b,r}^{pres}$. The traffic heterogeneity for AUP-B is slightly lower than for AUP-A and AUP-C. Temporary schedule AUP-C maintains the same level of heterogeneity as the nominal timetable. AUP-C maintains more traffic through the bottleneck, despite the higher heterogeneity of the traffic compared to AUP-A and AUP-B. The combined capacity index $C_{b,r}^{mix}$ has the highest value for AUP-C, followed by AUP-A and AUP-B.

Table 5.1: Capacity-related metrics for nominal and alternative temporary timetables.

Metric	Nominal	AUP-A	AUP-B	AUP-C
Scheduled number of trains (nominal) N_b^{nom}	28	-	-	-
Scheduled number of trains (temporary) $N_{b,r}^{tmp}$	-	26	24	28
Preserved capacity ratio $C_{b,r}^{pres}$	1.000	0.929	0.857	1.000
Traffic heterogeneity $H_{b,r}$	0.980	0.959	0.917	0.980
Capacity index $C_{b,r}^{mix}$	0.980	0.890	0.786	0.980

Stability Index

The stability behavior of all timetables for the three deterministic disturbance experiments is illustrated in Figure 5.11. The output delay reflects the extent to which the schedule is able to absorb the initial disturbances over timetable cycles. These stability results are obtained under static driving behavior.

For AUP-A higher output delays are present for more concentrated disturbances, as the output delays for S1 are larger than for S2 as for S3. The input delay does not generate persistent delay propagation across all experiments, indicating delay absorption. AUP-B shows a more varying absorption pattern. For S1 and S2 the output delay is absorbed immediately, without first causing delay propagation. In contrast, for experiment S3 AUP-B experiences delay propagation which is visible by the increase in output delay over timetable cycles 1 to 2. This timetable absorbs the input delay within two timetable cycles for all experiments, indicating stability. For AUP-C, the input delay leads to substantial output delays that persist over multiple timetable cycles. For S2 and S3 the output delay initially decreases after the first timetable cycle, but increases again in later cycles. This reflects delay propagation effects. Timetable AUP-C shows limited ability to absorb disturbances.

Timetable AUP-B shows lower output delays for the same initial disturbances compared to AUP-A and AUP-C. Across most intermediate output delays, AUP-B is better able to absorb disturbances within the first timetable cycle. Both timetables AUP-A and AUP-B prevent long-term delay propagation, but AUP-B shows a more stable short-term response. AUP-B results in a schedule that is less sensitive to deterministic shocks, indicating better delay absorption across different disturbance distributions. Therefore, AUP-B has an overall higher stability score of $S_{AUPB} = 0.785$, compared to $S_{AUPA} = 0.736$ for AUP-A. Following the stability definition adopted in this research, a timetable is considered stable if a deterministic input delay is absorbed within two timetable cycles (Pachl, 2014). Timetable AUP-C

does not absorb the deterministic input delay within this horizon for all experiments. Figure 5.11d illustrates that, even when allowing for more than two timetable cycles, the output delay of AUP-C is not absorbed. Therefore, AUP-C fails to satisfy the hard constraint and is assigned an overall stability score of $S_{AUPC} = 0$.

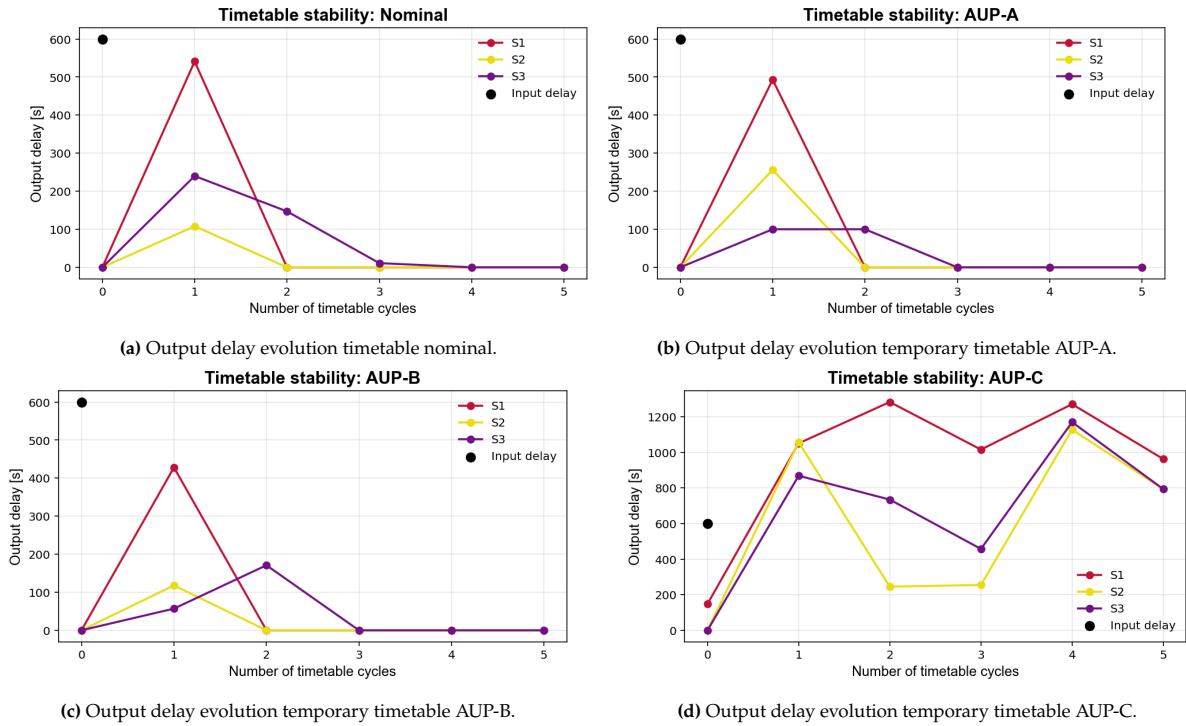


Figure 5.11: Output delay dynamics under deterministic disturbance scenarios for timetables nominal, AUP-A, AUP-B, and AUP-C.

Robustness Index

The box plots of robustness scores are shown in Figure 5.12. This index is evaluated using a tolerance parameter of $\lambda = 30$ seconds. As train activities are planned at tenths of minutes (i.e., 6 seconds), a tolerance larger than this resolution is required to exclude deviations that arise from measurement accuracy. These robustness results are obtained under static driving behavior.

AUP-A consistently attains higher robustness scores $R_{r,i}$ than AUP-B and AUP-C, but lower than the nominal timetable variant. A larger fraction of activities remain within the allowed tolerance under stochastic disturbances for AUP-A compared to the other temporary timetables. The interquartile range of AUP-A is comparable to that of the nominal timetable, but narrower than that of AUP-B and AUP-C. On average, the robustness for AUP-A is higher and more consistent across different replications i than that of AUP-B, while AUP-B in turn outperforms AUP-C in both level and consistency. The lower-tail of the interquartile range of robustness scores $R_{r,i}$ is lower for AUP-B and AUP-C than AUP-A. The presence of these lower outliers for AUP-B and AUP-C reflect higher sensitivity to stochastic delays.

The empirical cumulative distribution functions (ECDFs), visualized in Figure 5.13, are derived from Monte Carlo simulation experiments. For each timetable variant, 25 stochastic replications are performed in which stochastic disturbances are sampled from the empirically fitted delay distributions. This results in a set of 25 robustness scores ($R_{r,i}$) per timetable.

The ECDF of the nominal timetable and AUP-A is shifted to the right, compared to AUP-B and AUP-C, indicating a higher probability of attaining a given robustness level. For any robustness threshold, the probability that AUP-A attains this level of robustness is higher than for AUP-B and AUP-C. Temporary timetable AUP-C has substantially lower robustness scores $R_{r,i}$ across all replications compared to the other variants. This indicates a structurally lower ability to absorb stochastic variation. The markers in the ECDF indicate the average robustness score (R_r) per timetable. These resulting overall robustness scores are $R_{Nominal} = 0.850$, $R_{AUPA} = 0.844$, $R_{AUPB} = 0.812$, and $R_{AUPC} = 0.630$.

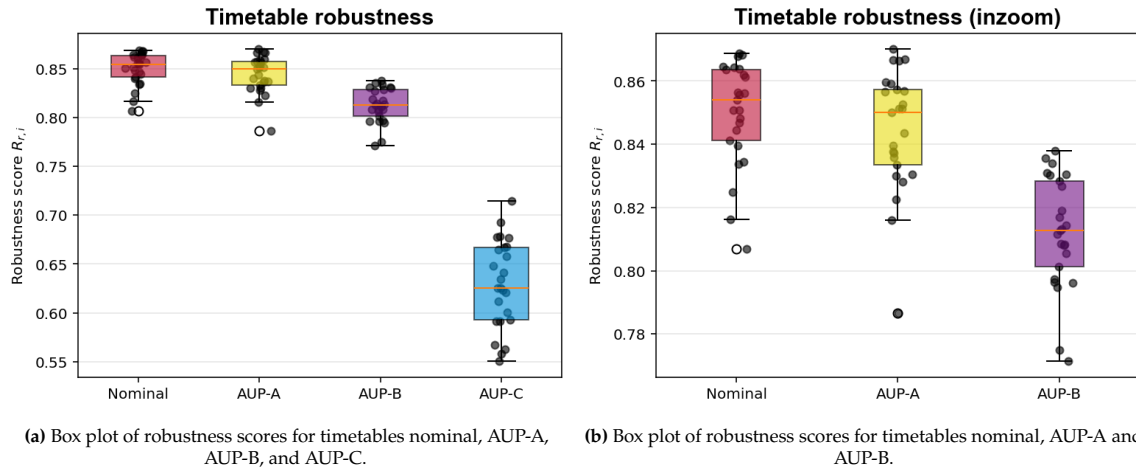
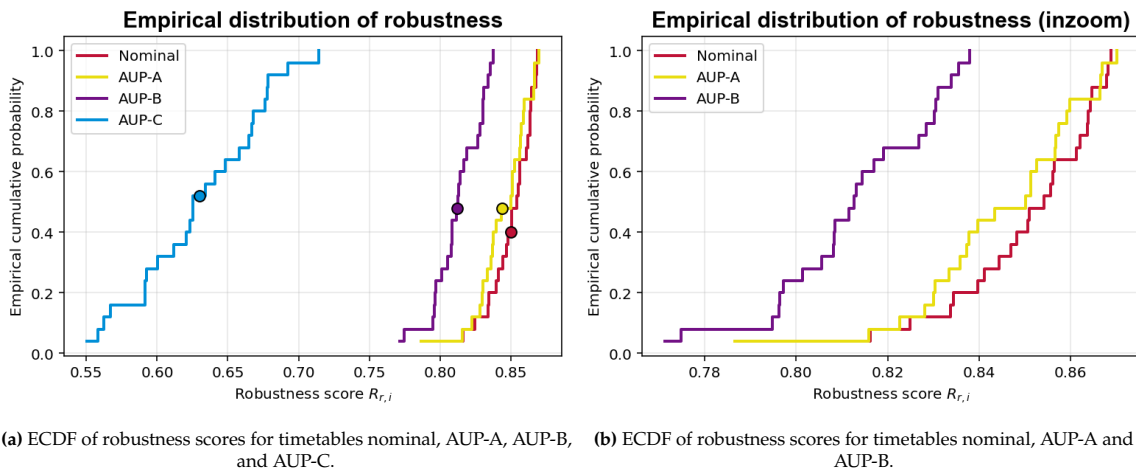


Figure 5.12: Box plot of robustness scores under stochastic disturbances for timetables nominal, AUP-A, AUP-B, and AUP-C.



Note: Markers denote the average robustness score per timetable.

Figure 5.13: ECDF of robustness scores under stochastic disturbances for timetables nominal, AUP-A, AUP-B, and AUP-C.

These box plots and ECDFs indicate that AUP-A is more robust to continuous stochastic variability than AUP-B and AUP-C. While AUP-A and AUP-B prevent extreme degradation of performance, AUP-A more effectively absorbs the day-to-day operational variation. In contrast, AUP-C exhibits lower robustness performance compared to AUP-A and AUP-B, with higher sensitivity to stochastic disturbances.

Operability Index

The time-distance diagrams (TWDs) of AUP-A, AUP-B, and AUP-C for both directions of the corridor (see Figure 5.14 - Figure 5.19) illustrate how train paths are distributed along the corridor before and after timetable compression and conflict resolution. The original timetable depicts the scheduled train paths and may contain blocking-time overlaps (i.e., conflicts). The compressed timetable is obtained using the UIC 406 timetable compression method (International Union of Railways (UIC), 2004). In this compression, any possible conflicts are not resolved. This compressed timetables shows the minimum cycle time in which the planned timetable, including potential conflicts, can be executed. The compressed feasible timetable illustrates the minimum cycle time of a conflict-free (feasible) timetable. The compressed timetable is re-expanded by resolving blocking-time conflicts. Here, the smallest necessary time extension, the maximum blocking-time overlap between conflicting train pairs within the corridor, is determined. Every train in the compressed timetable is shifted with this time extension to resolve all blocking-time conflicts on the corridor.

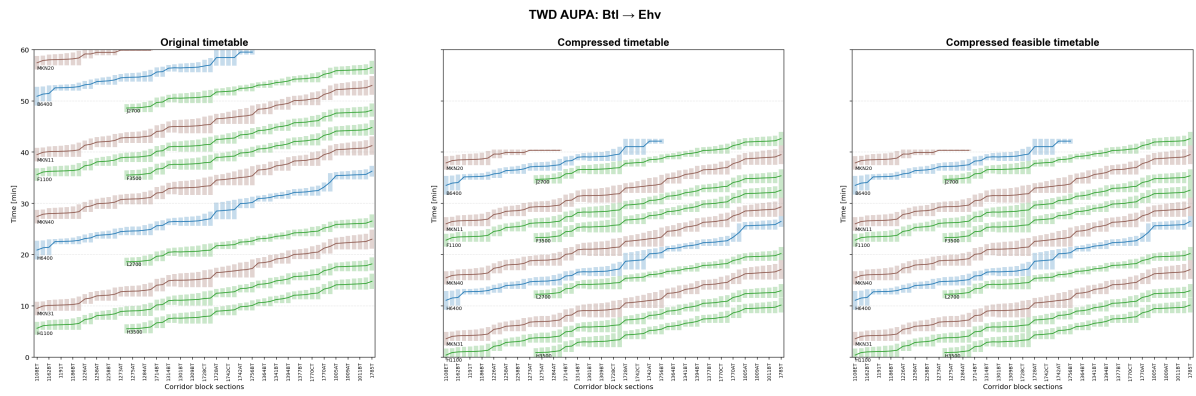


Figure 5.14: Time-distance diagram of the corridor Boxtel (Btl) - Eindhoven Centraal (Ehv) for temporary timetable AUP-A.

Note: Blue lines represent SPR-services, green lines represent IC-services and brown lines represent GDR-services.

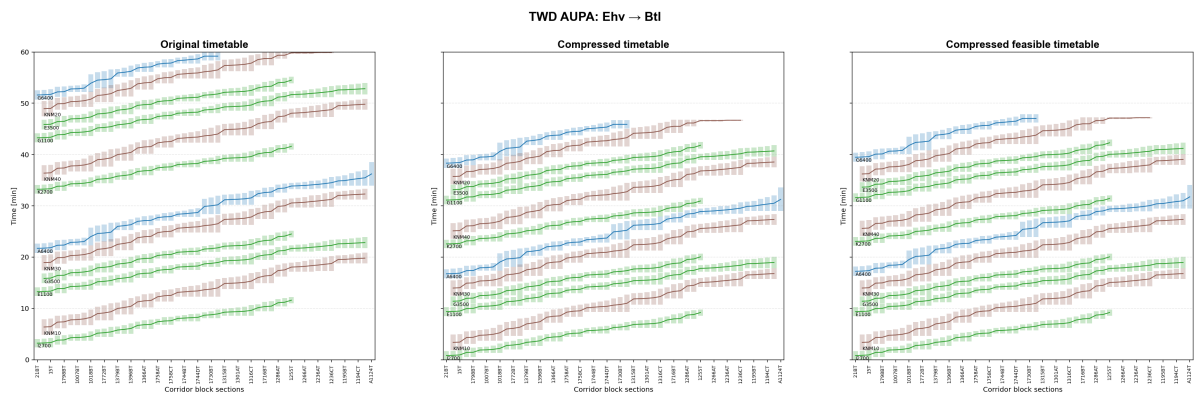


Figure 5.15: Time-distance diagram of the corridor Eindhoven Centraal (Ehv) - Boxtel (Btl) for temporary timetable AUP-A.

Note: Blue lines represent SPR-services, green lines represent IC-services and brown lines represent GDR-services.

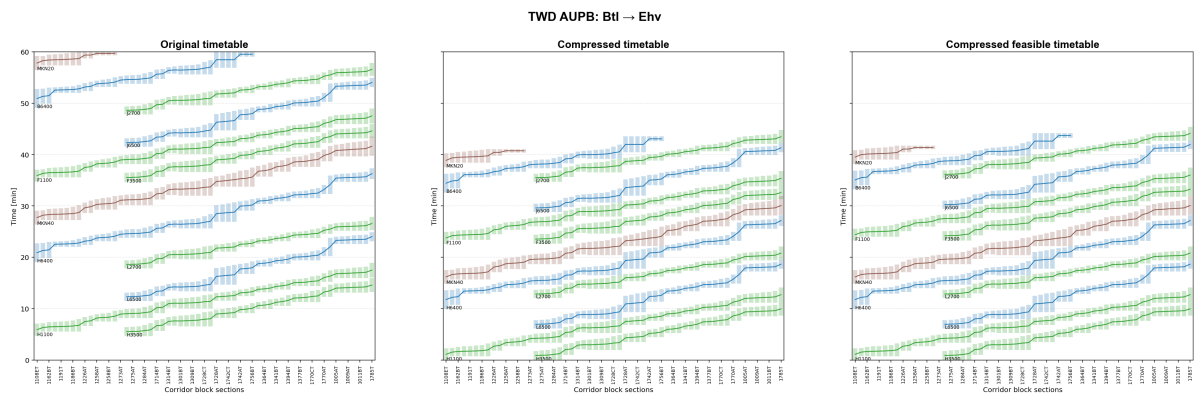


Figure 5.16: Time-distance diagram of the corridor Boxtel (Btl) - Eindhoven Centraal (Ehv) for temporary timetable AUP-B.

Note: Blue lines represent SPR-services, green lines represent IC-services and brown lines represent GDR-services.

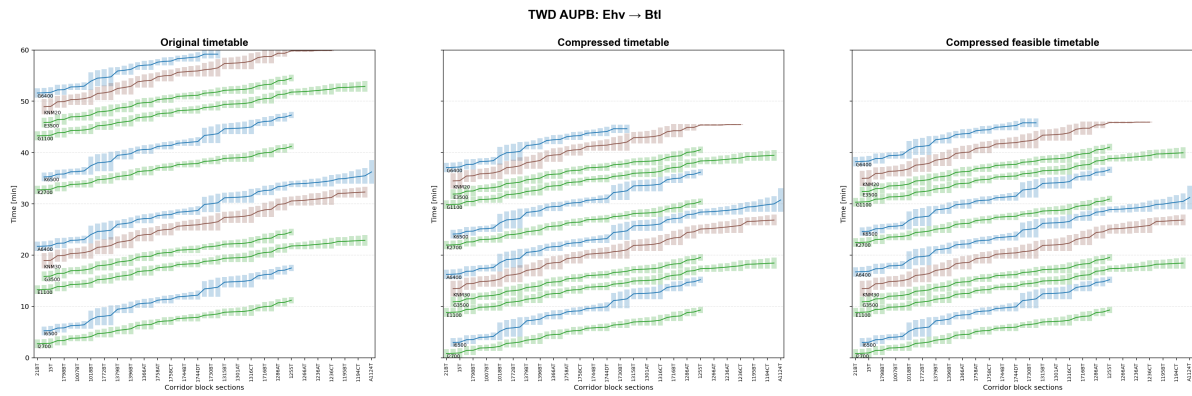


Figure 5.17: Time-distance diagram of the corridor Eindhoven Centraal (Ehv) - Boxtel (Btl) for temporary timetable AUP-B.

Note: Blue lines represent SPR-services, green lines represent IC-services and brown lines represent GDR-services.

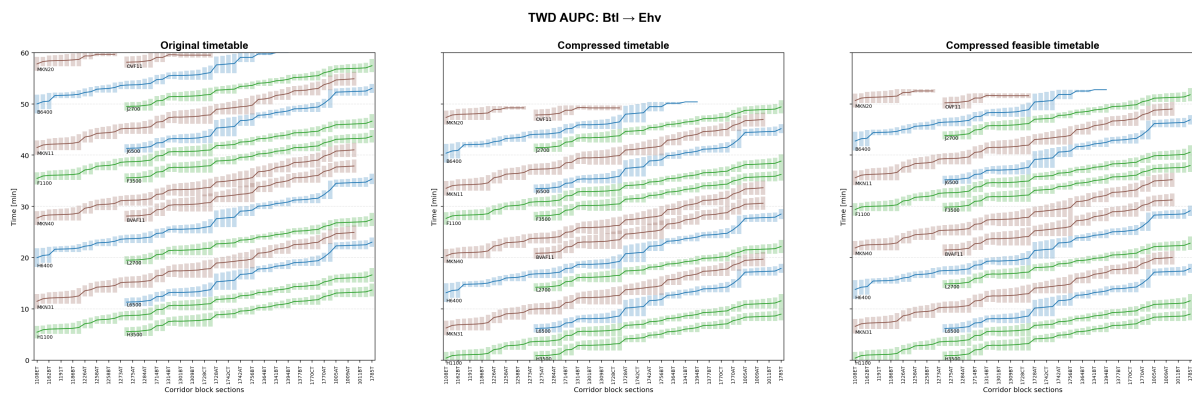


Figure 5.18: Time-distance diagram of the corridor Boxtel (Btl) - Eindhoven Centraal (Ehv) for temporary timetable AUP-C.

Note: Blue lines represent SPR-services, green lines represent IC-services and brown lines represent GDR-services.

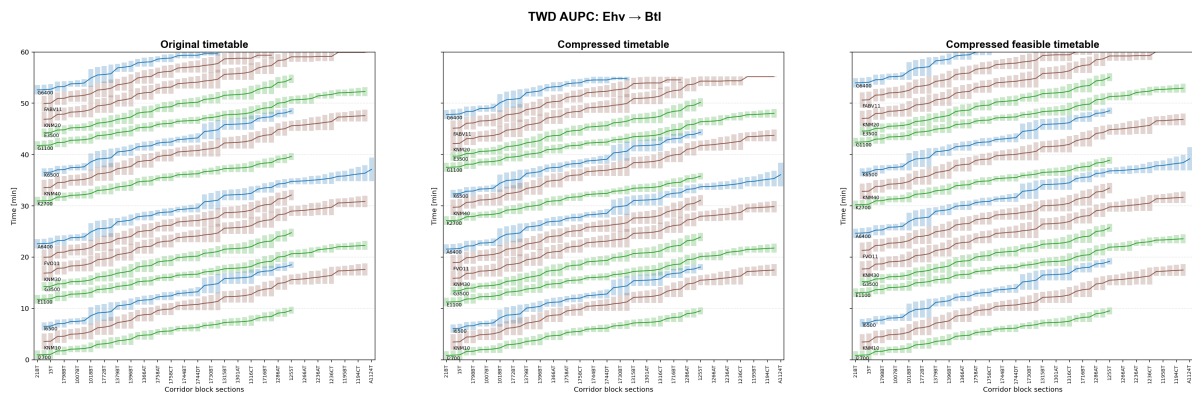
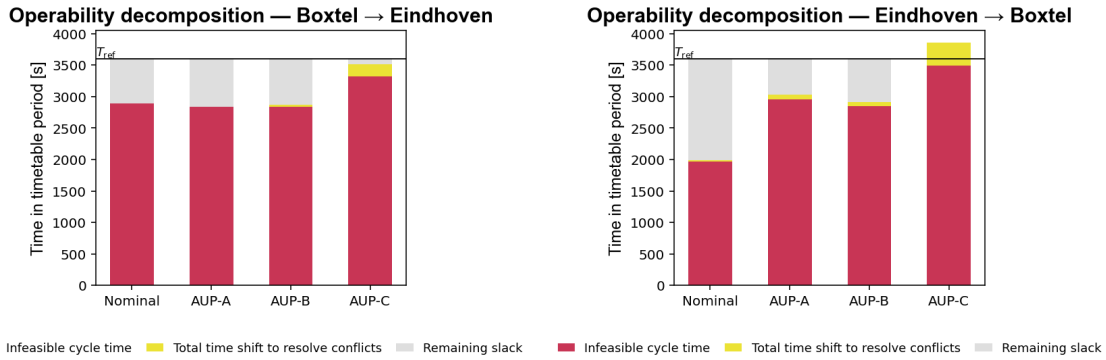


Figure 5.19: Time-distance diagram of the corridor Eindhoven Centraal (Ehv) - Boxtel (Btl) for temporary timetable AUP-C.

Note: Blue lines represent SPR-services, green lines represent IC-services and brown lines represent GDR-services.

The operability is assessed using the infeasible cycle time, the total time shift required to resolve all conflicts, and the remaining slack. The infeasible cycle time represents the compressed timetable duration under minimum headway constraints, allowing for conflicts. By applying the total time shift required to resolve all conflicts, a feasible cycle time is obtained. This feasible cycle time is compared

against the reference time period to assess operability. The decomposition is illustrated in Figure 5.20. For all timetables, the infeasible cycle time constitutes the largest share of the reference period. For the direction Btl-Ehv, the infeasible cycle times of the nominal timetable, AUP-A and AUP-B are similar. Although the nominal schedule and AUP-C have the same scheduled number of trains, AUP-C exhibits a larger infeasible cycle time. This can be attributed to the fact that AUP-C operates on two tracks, while the nominal timetable does not take into account the constrained conditions and operates on four tracks. For this direction, AUP-A is conflict-free, while AUP-B contains minor conflicts. All timetables are executable within the reference period. For the direction Ehv-Btl, larger deviations in the infeasible cycles times are observed and there are more pronounced conflicts. The nominal timetable has a low infeasible cycle time, indicating a substantial amount of slack. AUP-B has a lower infeasible cycle time than AUP-A, while the total amount of retiming to make the timetable conflict-free is identical. AUP-C has many conflicts, such that the total time shift required to resolve all conflicts exceeds the operational slack. In contrast, AUP-A and AUP-B are operable in one hour.



(a) Decomposition of operability direction Boxtel-Eindhoven Centraal for timetables nominal, AUP-A, AUP-B, and AUP-C. (b) Decomposition of operability direction Eindhoven Centraal-Boxtel for timetables nominal, AUP-A, AUP-B, and AUP-C.

Figure 5.20: Decomposition of operability into infeasible cycle time, total time shift required to solve all conflicts, and remaining slack for timetables nominal, AUP-A, AUP-B, and AUP-C per corridor direction.

Operability is evaluated separately for both directions, as structural conflicts and minimum feasible cycle times differ per direction. Table 5.2 shows the main results. Although AUP-A and AUP-B are structurally executable for both directions, the operability for AUP-B is slightly higher. AUP-C has the least remaining slack in the BE-direction and is not executable within the reference period for the EB-direction. The operability at the corridor-level is defined as the minimum operability in both directions. This reflects the fact that executability is defined as a lower bound. Based on this conservative aggregation, AUP-B attains the highest corridor-level operability score of $O_{AUPB} = 0.908$ compared to the other temporary timetables with operability scores of $O_{AUPA} = 0.892$ and $O_{AUPC} = 0$, respectively.

Table 5.2: Operability-related metrics for nominal and alternative temporary timetables.

Metric	Nominal	AUP-A	AUP-B	AUP-C
Direction Boxtel → Eindhoven				
Infeasible cycle time ($T_{BE}^{cycle, infeasible}$)	2885.01	2834.27	2834.54	3317.49
Total time shift required to resolve all conflicts ($T_{BE}^{conflict}$)	0.000	0.000	38.06	197.99
Operability (O_{BE})	1.000	1.000	0.950	0.299
Direction Eindhoven → Boxtel				
Infeasible cycle time ($T_{EB}^{cycle, infeasible}$)	1971.30	2957.51	2846.75	3489.15
Total time shift required to resolve all conflicts ($T_{EB}^{conflict}$)	14.50	69.23	69.23	371.14
Operability (O_{EB})	0.991	0.892	0.908	0.000
Minimum corridor operability				
Operability ($\min(O_{b,r})$)	0.991	0.892	0.908	0.000

5.4.2. Effectiveness Indicator Results

The integrated timetable effectiveness indicator (TTE_r) combines the four performance indices (i.e., capacity utilization, stability, robustness, and operability) into a single score. The objective of this aggregation is to provide a transparent synthesis that supports the comparison of alternative temporary timetables under constrained conditions. Table 5.3 summarizes the effectiveness results for all variants of the timetable for different weight compositions. These sets are considered to reflect alternative operational priorities that may arise during PTCs. Since no weighting scheme is optimal, multiple compositions are evaluated to illustrate the sensitivity of TTE_r to varying weights. The operability-constrained balanced reference reflects a perspective in which operability is considered a prerequisite for assessing timetable performance, while the remaining performance indices are weighted in a relatively balanced manner. An operational reliability-focused composition assigns a higher importance to stability and robustness, often referred to as a passenger-oriented set. A capacity-preservation focused composition emphasizes maintaining throughput. Finally, a conservative and risk-averse weight set assigns a dominant weight to operability, representing contexts where executability and conflict avoidance are critical.

Table 5.3: Integrated timetable effectiveness results under different weight compositions.

Metric	Nominal	AUP-A	AUP-B	AUP-C
Capacity index $C_{b,r}^{mix}$	0.980	0.890	0.786	0.980
Mean stability S_r	0.486	0.736	0.785	0.000
Robustness R_r	0.850	0.844	0.812	0.630
Operability (conservative) $O_{b,r}$	0.991	0.892	0.908	0.000
Operability-constrained balanced reference: $\alpha=0.30, \beta=0.15, \gamma=0.15, \epsilon=0.40$				
Timetable Effectiveness TTE_r	0.891	0.861	0.839	0.388
Operational reliability-focused: $\alpha=0.20, \beta=0.30, \gamma=0.30, \epsilon=0.20$				
Timetable Effectiveness TTE_r	0.795	0.831	0.818	0.385
Capacity preservation-focused: $\alpha=0.45, \beta=0.15, \gamma=0.10, \epsilon=0.30$				
Timetable Effectiveness TTE_r	0.896	0.863	0.825	0.504
Conservative and risk-averse: $\alpha=0.20, \beta=0.15, \gamma=0.15, \epsilon=0.50$				
Timetable Effectiveness TTE_r	0.892	0.861	0.851	0.290

Note: Green values indicate the highest value over temporary timetable alternatives AUP-A, AUP-B, and AUP-C.

The results in Table 5.3 show that AUP-A achieves higher overall effectiveness scores than AUP-B and AUP-C, in all weight compositions. AUP-B performs comparably well and attains TTE_r scores close to those of AUP-A. AUP-C consistently under-performs when all indices are considered simultaneously. Important to emphasize is that these results are conditional on the selected weight compositions and reflect plausible varying stakeholder preferences. When extreme weight sets are considered, such that a single index receives full weight while the other are assigned a zero-weight, the ranking of the timetable alternatives may change. For example, under a purely capacity-driven evaluation, AUP-C becomes the preferred alternative due to its high capacity index. Under strict stability- or operability-focused evaluations, AUP-B performs best, while in a solely robustness-driven evaluation, AUP-A performs best.

5.4.3. Sensitivity Analysis

Sensitivity to Parameter λ

Figure 5.21 shows the effect of λ on robustness scores. Larger values of λ lead to higher R_r values, as a greater share of activities remain within the tolerance margin. The marginal gains are greatest for small values of λ . Furthermore, the difference in robustness scores between the temporary timetables AUP-A, AUP-B, and AUP-C increases up to $\lambda = 30$ seconds. Beyond this, the difference gradually decreases because most activities of both timetables fall within the acceptable tolerance margin. Thus, the difference in R_r between these timetables is primarily driven by small to moderate delays.

The average sum of differences between the robustness score of the nominal timetable and each temporary timetable ($\frac{(R_{Nominal}-R_{AUPA})+(R_{Nominal}-R_{AUPB})+(R_{Nominal}-R_{AUPC})}{3}$) is visualized in Figure 5.22. The average differences in robustness scores R_r decrease when increasing the tolerance margin λ . This relationship is relatively linear, indicating that increasing the tolerance margin reduces the differences in robustness scores between the nominal and temporary timetables but does not alter the comparative assessment between the timetable alternatives.

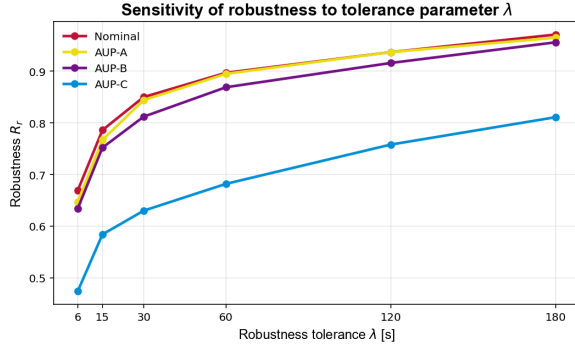


Figure 5.21: Sensitivity of robustness scores to λ for timetables nominal, AUP-A, AUP-B, and AUP-C.

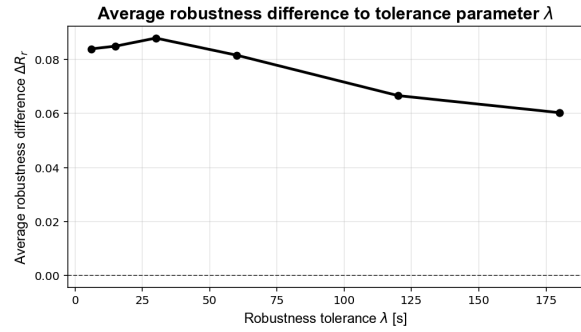


Figure 5.22: Sum of differences of R_r between nominal and temporary timetables AUP-A, AUP-B, and AUP-C to λ .

Sensitivity to Driving Behavior

To assess the sensitivity to the driving behavior, all disturbance experiments are evaluated under the static driving behavior "Stick to the plan" (MOP: Mik Op Pad) and the adaptive driving behavior "Make up for delay" (IOV: Inspelen Op Vertragenen). The capacity and operability indices are unaffected by this distinction in behavior, as they are derived from timetable characteristics and do not depend on realized train activities. The MOP driving behavior corresponds to the static driving behavior used in the stability and robustness results presented in Subsection 5.4.1. The adaptive driving behavior IOV allows trains to recover delays.

Figure 5.23 illustrates the effects of dynamic driving behavior on the stability and output delays. For all timetables, the output delays decrease in size and the output delays are absorbed within fewer timetable cycles. This is reflected by an increase in stability score for all timetables. For the nominal timetable the stability score increases from $S_{Nominal}^{MOP} = 0.486$ to $S_{Nominal}^{IOV} = 0.940$. For the temporary timetables the increases in stability scores are $S_{AUPA}^{MOP} = 0.736$ to $S_{AUPA}^{IOV} = 0.980$, $S_{AUPB}^{MOP} = 0.785$ to $S_{AUPB}^{IOV} = 0.964$, and $S_{AUPC}^{MOP} = 0.000$ to $S_{AUPC}^{IOV} = 0.506$, respectively.

This highlights two important insights. First, adaptive driving behavior affects both the absolute level of stability, as well as the relative ranking of the timetable alternatives. Under MOP, AUP-B outperforms AUP-A. However, when adaptive driving behavior is applied, this ranking reverses, and AUP-A attains a higher stability score. This indicates that timetable stability is sensitive to driving behavior in a structural manner, where difference in timetable design are amplified or mitigated depending on how slack and recovery margins are used in operation. Consequently, stability is determined by the interaction between timetable design and driving behavior. Secondly, the stability improvement for AUP-C is the largest among all timetable variants. Under static driving behavior, AUP-C fails to absorb output delays within two timetable cycles for all experiments, resulting in a stability score of zero. Under adaptive driving behavior, this hard constraint is satisfied for two experiments, increasing the stability to a non-zero value. This highlights that adaptive driving behavior is beneficial for timetables operating close to the boundary of instability, where small changes in delay absorption dynamics can lead to large improvements in stability.

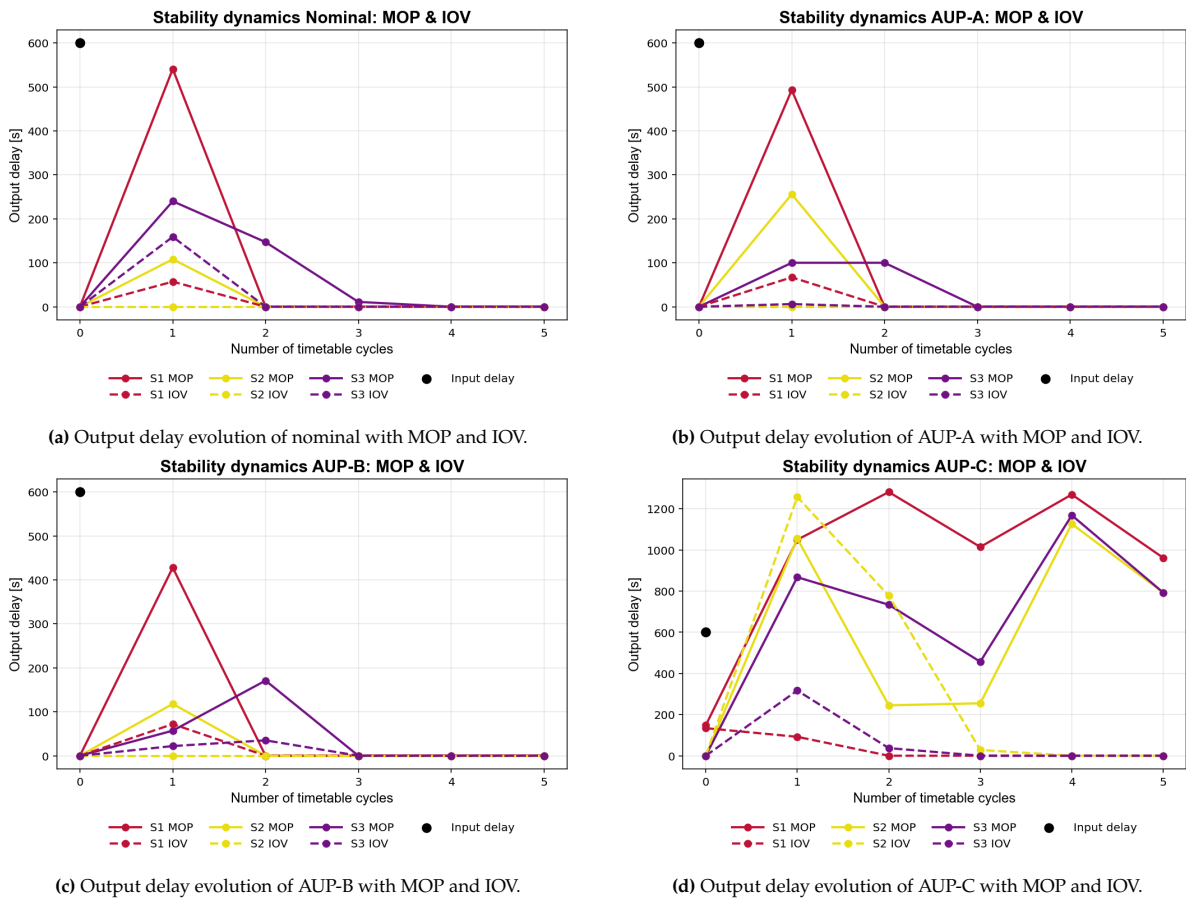


Figure 5.23: Output delay dynamics under deterministic disturbance scenarios for timetables nominal, AUP-A, AUP-B, and AUP-C with varying driving behavior MOP and IOV.

Figure 5.24 shows how under IOV driving behavior the robustness score distribution shifts towards higher values with reduced spread. This indicates that with adaptive driving behavior, a larger share of activities remain within the tolerance margin as trains can recover delays. Therefore, stochastic disturbances are absorbed more frequently before propagating. This dampening effect of the variability is reflected by the reduced interquartile ranges. An important observation is the increase of R_{AUPC} . Under MOP, AUP-C clearly under-performs, both in median value and variability. However, under IOV, R_{AUPC} increases substantially, where the difference with other timetables becomes considerably smaller. This indicates that part of the robustness is attributable to the driving behavior applied.

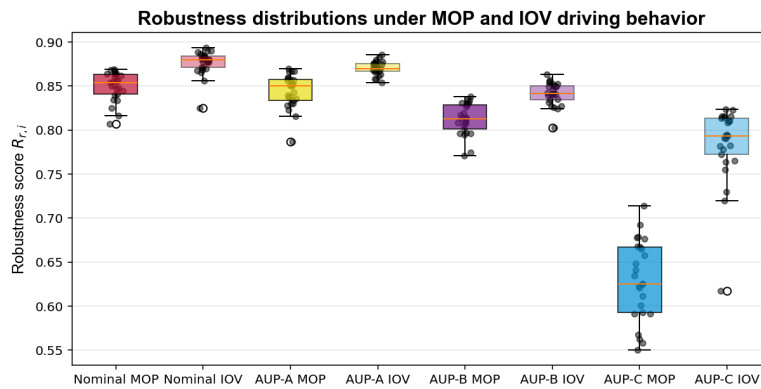


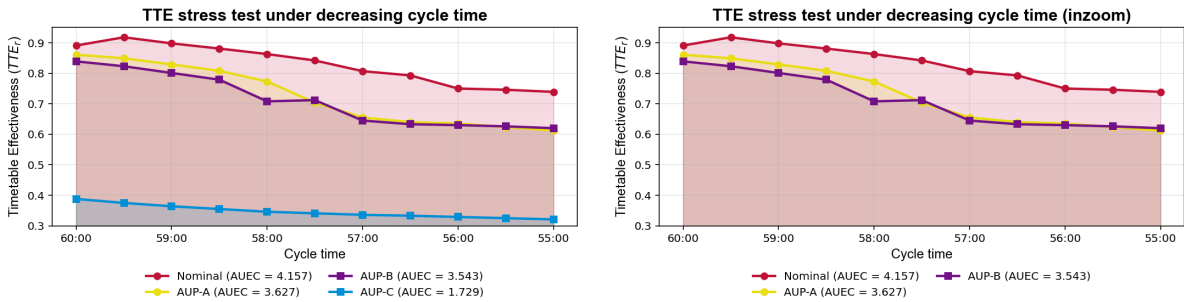
Figure 5.24: Distribution of robustness scores under stochastic disturbances for timetables nominal, AUP-A, AUP-B, and AUP-C with varying driving behavior MOP and IOV

This sensitivity analysis shows how adaptive driving behavior decreases short-term delay propagation and improves both stability and robustness. Adaptive driving behavior increases absolute performance levels and may change the relative differences between timetable alternatives. As shown for stability, the relative ranking of timetables can change when adaptive recovery is allowed. For robustness, adaptive driving behavior reduces the performance gap between alternatives. These results indicate that stability and robustness, and thus timetable effectiveness, is sensitive to the assumed driving behavior.

5.4.4. Stress Testing

The results of the stress testing experiment are presented. This stress test is performed with static driving behavior (MOP), $\lambda = 30$ seconds, and the operability-constrained balanced reference weight composition for the aggregation of the performance indices. Details on the stress test and the definition of $AUEC$ are described in [Subsection 3.2.5](#).

[Figure 5.25](#) illustrates how TTE_r evolves as the reference period is progressively reduced. For all timetable variants, effectiveness decreases as operational pressure increases. This reflects the gradual removal of slack and running time margins. For low levels of compression, TTE_r remains relatively high for the nominal schedule and temporary timetables AUP-A and AUP-B. A limited reduction in the operating time window can be absorbed without severe performance degradation. However, both AUP-A and AUP-B exhibit a strong decline in effectiveness once the reference period is reduced below 58 minutes. For compression levels below 57 minutes, the effectiveness scores of these temporary timetables converge to a lower bound of approximately $TTE_{AUPA/B} = 0.625$. Similarly, the nominal schedule and AUP-C converge to stable lower bounds of approximately $TTE_{Nominal} = 0.75$ and $TTE_{AUPC} = 0.325$, respectively. These lower performance bounds, observed in the effectiveness curves, indicate that the timetable has become so compressed that it is no longer operationally realistic and becomes structurally infeasible. At these compression levels, planned running times are no longer achievable, leading to persistent delay accumulation and saturation of the simulation outcomes (i.e. deadlocks).



(a) Timetable effectiveness dynamics for nominal, AUP-A, AUP-B, and AUP-C. (b) Timetable effectiveness dynamics zoomed-in for nominal, AUP-A, and AUP-B.

Figure 5.25: Timetable effectiveness dynamics under decreasing reference periods for timetables nominal, AUP-A, AUP-B and AUP-C.

The cumulative effect of increasing operational pressure is summarized by the $AUEC_r$ values. AUP-A has a slightly higher score ($AUEC_{AUPA} = 3.627$) than AUP-B ($AUEC_{AUPB} = 3.543$), indicating that AUP-A preserves timetable effectiveness marginally better across the tested compression range. The nominal timetable has the highest score with $AUEC_{Nominal} = 4.157$. AUP-C has the lowest score with $AUEC_{AUPC} = 1.729$, reflecting its limited ability to maintain effectiveness under increasing operational pressure.

6

Discussion

The previous chapters demonstrate that the proposed temporary timetable effectiveness framework can systematically quantify and compare timetables. This chapter discusses the interpretation, relevance, and limitations of the framework. Section 6.1 interprets the results from both a methodological and a practical perspective, addressing whether the framework works as intended and discussing its practical usability. Section 6.2 reflects on the contributions and the research gaps identified. Finally, Section 6.3 discusses the main limitations of the research.

6.1. Interpretation of Results

The results demonstrate that the proposed effectiveness framework is capable of systematically quantifying and comparing temporary railway timetables. By translating timetable characteristics and simulated operational performance into a consistent set of indices, the framework results in numerical outcomes that reflect the capacity utilization, stability, robustness, and operability of temporary timetables. Aggregating these outcomes leads to an integrated TTE_r score that is interpretable and comparable across different timetable alternatives. The case study also shows that the method is capable of distinguishing both small and large differences between alternative temporary timetables. Subtle differences between similar timetables are reflected in marginal but consistent differences in index scores. More pronounced timetable changes result in clearly different effectiveness outcomes. The decomposition of effectiveness into four indices allows differences between timetables to be interpreted in terms of their underlying performance mechanisms, in addition to their aggregated effectiveness score.

The proposed framework thus works conceptually and methodologically. Discussions with planners and simulation experts indicate that the framework is particularly valuable in situations where multiple temporary timetable alternatives are available and where trade-offs between different performance indices are not immediately clear. In such cases, the framework can support decision-making by making trade-offs between capacity, stability, robustness, and operability explicit and transparent. In current planning practice, often only a single temporary timetable is designed. In this context, the framework remains useful as an ex-ante and ex-post assessment tool, enabling prediction of expected performance and the identification of potential weaknesses within a given timetable. The added value of timetable quantification is expected to be highest in complex planning situations, either where multiple feasible timetable alternatives exist or when the effects of design choices cannot be easily overseen.

6.2. Reflection on Contributions

From a scientific perspective, this research contributes to the literature on railway timetable evaluation by addressing the gap in the quantification of temporary timetables. Existing studies focus on nominal timetables and have developed a range of methods to define and measure feasibility, stability, and robustness. This research focuses on temporary schedules used for PTCs and positions temporary timetables as a distinct research domain within railway timetabling. By defining temporary timetables as a separate category, this research provides a foundation for the development of performance indices and evaluation methods specified to constrained operating conditions. At the same time, the insights gained highlight that nominal schedules may also contain minor infeasibilities while remaining executable. This motivates a broader interpretation of timetable evaluation that extends beyond temporary timetables only. A key contribution lies in the development of timetable effectiveness as a multi-criteria performance concept where effectiveness is defined as the ability of a timetable to preserve

capacity while remaining operable, stable, and robust under deterministic and stochastic variability. From this new effectiveness definition, a quantitative framework is developed that combines timetable characteristics with realized operational performance into a set of performance indices and an integrated indicator TTE_r . Existing concepts are reformulated to ensure applicability under PTC conditions. In addition, a new operability index is introduced to define and quantify the executability of timetables that contain minor conflicts, but remain operationally feasible. By linking timetable compression, conflict resolution, and remaining slack, operability extends existing feasibility-based approaches. Although this methodological framework is developed and demonstrated in the context of temporary timetables, the concept of timetable effectiveness and the proposed performance indices are applicable to railway timetables more generally. In different operational contexts, definitions and modeling assumptions can be adapted to reflect the relevant conditions without changing the underlying principles of the method.

Beyond these scientific contributions, the proposed framework also provides a decision-support tool for planning practice. It enables planners to systematically assess expected timetable performance, compare alternative designs when available, and identify potential weaknesses within a given timetable. Furthermore, stochastic approaches are well established in nominal timetable research and practice, where application to temporary timetables remains limited. This research contributes to this gap by embedding stochastic variability into timetable evaluation by fitting empirical delay distributions to historical performance data and evaluating robustness in a probabilistic matter.

6.3. Limitations

The TTE_r framework allows stakeholders to express different priorities and objectives through alternative weight compositions, which is a key strength of the proposed method. However, the selection of such weights proved to be challenging in practice. Discussion with planners and other experts revealed that assigning weights to abstract performance indices is difficult. This highlights a mismatch between the analytical advantages of a weighted multi-criteria framework and the way users assign weights. Although the inclusion of stakeholder and their preferences in the framework is essential for decision support, future applications may benefit from alternative ways of capturing preferences.

A second limitation relates to the scope of the case study. The framework is applied to a single PTC context with a small corridor. This provides a controlled and realistic setting to test the framework, but the direct applicability to large-scale PTCs may be limited without further extension. The underlying concepts and indices are not limited to small-scale contexts, but applying the framework to multiple and structurally different PTC contexts may require methodological extensions. In particular, the capacity and operability indices may need to be adjusted to account for multiple corridors and secondary network effects.

Finally, a limitation lies in the scoping of the framework to temporary timetables. This specific focus is necessary to define effectiveness and develop performance indices and a quantification method for constrained operating conditions. Despite this limited scope, the underlying framework has a high degree of generalizability. Although operational conditions differ, temporary and nominal timetables can be evaluated using a comparable effectiveness-based approach. Therefore, the proposed method and concepts are applicable to railway timetables more generally. Extending the framework beyond temporary timetables would require reformulating definitions and mathematical formulations to reflect the relevant operational context (e.g. idealized conditions for nominal timetables, the absence of cyclic structures for non-cyclic timetables, approaching of yellow signals for operability, or additional attributes for financial-oriented decision-making), while preserving the underlying methodological principles.

Conclusion and Future Research

This chapter presents the practical recommendations, conclusions, and fields of future research of this study. Section 7.1 provides practical implications and guidelines. Section 7.2 answers the main research question, addresses the research objective, and points out directions for future research.

7.1. Practical Recommendations and Guidelines

Several recommendations are provided to support the quantification of effective temporary railway timetables, particularly within the context of ProRail's planning and timetable design processes. First, the modeling of stochastic disturbances in microscopic simulation can be improved by allowing a broader set of delay distributions and selecting the best-fitting distribution using objective criteria. This leads to more representative assessments of operational variability and thus robustness. Second, driving behavior has a substantial influence on both absolute performance levels and the relative ranking of timetable alternatives. Therefore, driving behavior should be treated as a source of uncertainty. It is recommended to evaluate timetable effectiveness under multiple representative driving behaviors, such that the framework allows for the simulation of alternative driving behavior models and the aggregation of their outcomes. Third, current deterministic robustness assessments can be complemented with stochastic simulation-based assessment using empirically fitted delay distributions derived from historical performance data. By continuously sampling delays from these distributions, robustness is evaluated in a probabilistic manner that better reflects day-to-day operational uncertainty. Finally, it is recommended to improve the technical integration and organizational alignment between DONNA, DONS, and FRISO. Ensuring cross-system expertise of these tools by planning and simulation experts and increasing the software interoperability support a more coherent and transparent process for temporary timetable design and evaluation.

7.2. Conclusion and Future Research

7.2.1. Conclusion

The main question of this research is to determine how the effectiveness of temporary railway timetables can be systematically quantified in the context of PTCs. This study demonstrates that effectiveness can be assessed through an integrated quantification and evaluation framework that combines timetable characteristics with simulated operation performance under both deterministic and stochastic variability.

Temporary railway timetables are revised schedules implemented during periods of interrupted operations, such as maintenance activities or unforeseen disruptions. These schedules operate under constrained infrastructure conditions and differ fundamentally from nominal schedules. Temporary timetables often have reduced slack, altered services patterns, and minor theoretical infeasibilities, yet remain operationally executable. To capture these characteristics, this research introduces timetable effectiveness as a multi-dimensional performance concept, defined as the ability of a temporary timetable operating under constrained conditions to maintain operable, stable, and robust train operations under deterministic and stochastic operational variability. Based on this definition, a quantitative framework is developed that translates timetable performance into capacity utilization, stability, robustness, and operability. By decomposing effectiveness into distinct indices, the framework reveals both overall results and the underlying trade-offs. Capacity relates to the preserved throughput through a corridor. Stability captures the ability to absorb deterministic delays. Robustness focuses on preventing delays

under stochastic variability. Operability represents a new index that quantifies the executability of timetables that may contain minor conflicts. These four indices are aggregated through a weighted sum into an integrated effectiveness indicator TTE_r . Here, alternative weight compositions allow for different stakeholder priorities to be represented. This allows timetable effectiveness to be assessed in absolute terms and relative to operational objectives.

The application to a real-world case study on the Boxtel-Eindhoven corridor shows that the framework is able to systematically quantify timetable effectiveness and enables the comparison between alternative temporary schedules. Under the operability-constrained balanced reference weight composition, the nominal timetable achieves a $TTE_{Nominal}$ score of 0.891, while temporary timetables AUP-A and AUP-B attain $TTE_{AUPA} = 0.861$ and $TTE_{AUPB} = 0.839$, respectively. AUP-C performs substantially worse with $TTE_{AUPC} = 0.388$, reflecting its poor stability, robustness, and operability despite its high preserved capacity. These results indicate that the framework is capable to distinguish both subtle and pronounced differences. Small design changes result in marginal differences, such as between AUP-A and AUP-B. More substantial different timetable alternatives, such as the nominal schedule and AUP-C, lead to clearly different performance results. Stress testing further demonstrates the ability of the framework to distinguish between timetable alternatives under increasing operational pressure. When progressively compressing the timetable, AUP-A and AUP-B converge to a lower effectiveness bound of approximately $TTE_{AUPA/B} = 0.625$, while the nominal timetable stabilizes around $TTE_{Nominal} = 0.75$ and AUP-C around $TTE_{AUPC} = 0.325$. The cumulative impact of compression is captured by $AUEC$, where the nominal timetable achieves the highest score ($AUEC_{Nominal} = 4.157$), followed by AUP-A ($AUEC_{AUPA} = 3.627$), AUP-B ($AUEC_{AUPB} = 3.543$), and AUP-C ($AUEC_{AUPC} = 1.729$).

In conclusion, this research shows that temporary railway timetables, in the context of PTCs, can be systematically quantified through a data-driven multi-criteria performance quality indicator, Temporary Timetable Effectiveness TTE . This indicator provides both a single interpretable performance score and insight into the underlying operational mechanisms and trade-offs between capacity utilization, stability, robustness, and operability. In this way, the method offers a structured decision-support tool for planning practice, supporting transparent comparison of timetable alternatives and systematic assessment of expected performance. Although the framework is developed and demonstrated for temporary schedules, the underlying concept of timetable effectiveness and the proposed methodology of indices and an integrated indicator are not limited to PTC contexts. With adaptation of the definitions and modeling assumptions to the specific operational conditions, the framework can be applied to railway timetables more generally.

7.2.2. Future Research

Several directions for future research follow from this study. First, timetable effectiveness is shown to be sensitive to driving behavior. However, only two driving behaviors were considered. Future research is required to study how driving behavior varies across train types, corridors, and operational contexts, and how drivers respond to delays under different conditions. Incorporating multiple driving behaviors will allow to account for behavioral uncertainty. Second, the framework could be extended to a network-level perspective. While this research focuses on a single corridor, PTCs often involve multiple interacting corridors. Exploring how timetable effectiveness can be aggregated across multiple corridors or how local effectiveness relates to network-wide performance would increase the applicability of the framework for larger scale PTCs. This is particularly relevant for the operability index, which is derived from a corridor-based timetable compression approach. This method captures the minimum addition shifts to resolve local conflicts, but does not account for secondary network effects of blocking-time conflict resolution. Also, future work could refine the modeling of stochastic variability under PTC conditions. In this research, delay distributions are derived from nominal operations and applied to temporary timetables. This provides a stable and consistent baseline for day-to-day operational variability, but does not include the additional variability that arises during PTCs. Further research could explore the possibility of piecewise fitting approaches in which multiple theoretical distributions are fitted to different parts of the empirical distribution. Such an approach could improve the representation of extreme tail behavior while maintaining a realistic description of central variability behavior. Lastly, empirical research should explore the affects and applicability of the proposed quantification framework for timetable effectiveness on real-world decision-making. The role of the framework as decision-support tool is promising, but requires further validation in real-life planning practice.

References

- Andersson, E., Peterson, A., & Krasemann, J. T. (2013a). Robustness in swedish railway traffic timetables. *Proceedings of the 5th International Seminar on Railway Operations Modelling and Analysis (Rail-Copenhagen)*, 1–20. <https://www.diva-portal.org/smash/get/diva2:539662/FULLTEXT01.pdf>
- Andersson, E., Peterson, A., & Krasemann, J. T. (2013b). Quantifying railway timetable robustness in critical points [Robust Rescheduling and Capacity Use]. *Journal of Rail Transport Planning Management*, 3(3), 95–110. <https://doi.org/https://doi.org/10.1016/j.jrtpm.2013.12.002>
- Besinovic, N., & Goverde, R. M. P. (2018). Capacity assessment in railway networks. In R. Borndorfer, T. Klug, L. Lamorgese, C. Mannino, M. Reuther, & T. Schlechte (Eds.), *Handbook of optimization in the railway industry* (pp. 25–45). pringer International Publishing. https://doi.org/10.1007/978-3-319-72153-8_2
- Besinovic, N., Goverde, R. M., Quaglietta, E., & Roberti, R. (2016). An integrated micro-macro approach to robust railway timetabling. *Transportation Research Part B: Methodological*, 87, 14–32. <https://doi.org/https://doi.org/10.1016/j.trb.2016.02.004>
- Cacchiani, V., & Toth, P. (2012). Nominal and robust train timetabling problems [Feature Clusters]. *European Journal of Operational Research*, 219(3), 727–737. <https://doi.org/https://doi.org/10.1016/j.ejor.2011.11.003>
- Chung, S., Sardak, D., Kitsak, M., Jin, A., & Linkov, I. (2025). Contested logistics: Resilience of strategic highways and railways. *Transportation Research Interdisciplinary Perspectives*, 32, 101507. <https://doi.org/https://doi.org/10.1016/j.trip.2025.101507>
- CORDIS: Publications Office of the European Union. (n.d.). Final report summary - on-time (optimal networks for train integration management across europe) [Accessed: 2025-09-12].
- Correia Duarte, P. J., Dauzere-Peres, S., Huisman, D., Mannino, C., Sartor, G., Weik, N., & Widmann, P. (2025). 50 years of or in railway timetabling and rolling stock planning. *EURO Journal on Transportation and Logistics*, 14, 100155. <https://doi.org/https://doi.org/10.1016/j.ejtl.2025.100155>
- Davenport, A., & Beck, J. (2001). Slack-based techniques for robust schedules. *Proceedings of the Sixth European Conference on Planning (ECP-2001)*.
- Dewilde, T., Sels, P., Cattrysse, D., & Vansteenwegen, P. (2014). Improving the robustness in railway station areas. *European Journal of Operational Research*, 235(1), 276–286. <https://doi.org/https://doi.org/10.1016/j.ejor.2013.10.062>
- Drummond, M., Bresina, J., & Swanson, K. (1994). Just-in-case scheduling. *Proceedings of the National Conference on Artificial Intelligence*, 2, 1098–1104.
- Fischetti, M., & Monaci, M. (2009). Light robustness. In R. K. Ahuja, R. H. Mohring, & C. D. Zaroliagis (Eds.), *Robust and online large-scale optimization: Models and techniques for transportation systems* (pp. 61–84). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-05465-5_3
- Goverde, R. M. (2007). Railway timetable stability analysis using max-plus system theory [Advanced Modelling of Train Operations in Stations and Networks]. *Transportation Research Part B: Methodological*, 41(2), 179–201. <https://doi.org/https://doi.org/10.1016/j.trb.2006.02.003>
- Goverde, R. M. (2010). A delay propagation algorithm for large-scale railway traffic networks [11th IFAC Symposium: The Role of Control]. *Transportation Research Part C: Emerging Technologies*, 18(3), 269–287. <https://doi.org/https://doi.org/10.1016/j.trc.2010.01.002>
- Goverde, R. M., Corman, F., & D'Ariano, A. (2013). Railway line capacity consumption of different railway signalling systems under scheduled and disturbed conditions [Robust Rescheduling and Capacity Use]. *Journal of Rail Transport Planning Management*, 3(3), 78–94. <https://doi.org/https://doi.org/10.1016/j.jrtpm.2013.12.001>
- Goverde, R. M., & Hansen, I. A. (2013). Performance indicators for railway timetables. *2013 IEEE International Conference on Intelligent Rail Transportation Proceedings*, 301–306. <https://doi.org/10.1109/ICIRT.2013.6696312>

- Hansen, I. A. (2009). Railway network timetabling and dynamic traffic management [Discusses deterministic running, dwell, and headway times; use of supplements buffer times based on rules of thumb; compares combinatorial optimisation and simulation for timetable optimisation and traffic management.]. *2nd International Conference on Recent Advances in Railway Engineering (ICRARE-2009)*. https://repository.tudelft.nl/file/File_f6e58db0-56df-4d89-87b0-067a23d2aad3?preview=1
- Hansen, I., & Pachl, J. (Eds.). (2008). *Railway timetable & traffic: Analysis, modelling, simulation*. Eurailpress.
- Hoos, H. H., & Stutzle, T. (2005). 6 - propositional satisfiability and constraint satisfaction. In H. H. Hoos & T. Stutzle (Eds.), *Stochastic local search* (pp. 257–312). Morgan Kaufmann. <https://doi.org/https://doi.org/10.1016/B978-155860872-6/50023-8>
- International Union of Railways (UIC). (2004). *Uic leaflet 406: Capacity* (Leaflet). International Union of Railways. Paris, France.
- Janic, M. (2018). Modelling the resilience of rail passenger transport networks affected by large-scale disruptive events: The case of hsr (high speed rail). *Transportation*, 45(4), 1101–1137. <https://doi.org/10.1007/s11116-018-9875-6>
- Jensen, L. W., Landex, A., & Nielsen, O. A. (2014). Evaluation of robustness indicators using railway operation simulation. In C. A. Brebbia, N. Tomii, P. Tzieropoulos, & J. M. Mera (Eds.), *Computers in railways xiv* (pp. 329–339). WIT Press.
- Jensen, L. W., Schmidt, M., & Nielsen, O. A. (2020). Determination of infrastructure capacity in railway networks without the need for a fixed timetable. *Transportation Research Part C: Emerging Technologies*, 119, 102751. <https://doi.org/https://doi.org/10.1016/j.trc.2020.102751>
- Kroon, L., Maroti, G., Helmrich, M. R., Vromans, M., & Dekker, R. (2008). Stochastic improvement of cyclic railway timetables. *Transportation Research Part B: Methodological*, 42(6), 553–570. <https://doi.org/https://doi.org/10.1016/j.trb.2007.11.002>
- Lee, Y., Lu, L.-S., Wu, M.-L., & Lin, D.-Y. (2017). Balance of efficiency and robustness in passenger railway timetables. *Transportation Research Part B: Methodological*, 97, 142–156. <https://doi.org/https://doi.org/10.1016/j.trb.2016.12.004>
- Leon, V. J., Wu, S., & Storer, R. (1994). Robustness measures and robust scheduling for job shops. *Iie Transactions*, 26, 32–43. <https://doi.org/10.1080/07408179408966626>
- Lindfeldt, O. (2008). Evaluation of punctuality on a heavily utilised railway line with mixed traffic. *WIT Transactions on The Built Environment*, 103, 545–553.
- Lusby, R. M., Larsen, J., & Bull, S. H. (2018). A survey on robustness in railway planning. *European Journal of Operational Research*, 266(1), 1–15. <https://doi.org/10.1016/j.ejor.2017.07.044>
- NS. (2025). Annual report 2024 [Accessed: 17 March 2025]. <https://www.nsjaarverslag.nl/external/asset/download/project/f9822a0a-03ed-0000-2e8a-0cb47fbeb7e3/name/NS-Jaarverslag-2024.pdf>
- NS. (n.d.). Responsibilities – about ns [Accessed: 2025-09-24].
- OpenAI. (2025). Chatgpt [Version 5.2 [Large Language Model] Used as a supporting tool for structuring and refining academic text and assisting in the formulation, spelling, grammar and clarification of written content during this research].
- Pachl, J. (2014). *Railway operation and control* (2nd) [Chapter 6: Stability and delay propagation]. VTD Rail Publishing. <https://www.vtd.net/VTDPUB/ROC2/ROC2.html>
- Polinder, G.-J., Breugem, T., Dollevoet, T., & Maroti, G. (2019). An adjustable robust optimization approach for periodic timetabling. *Transportation Research Part B: Methodological*, 128, 50–68. <https://doi.org/https://doi.org/10.1016/j.trb.2019.07.011>
- Polinder, G.-J., Schmidt, M., & Huisman, D. (2021). Timetabling for strategic passenger railway planning. *Transportation Research Part B: Methodological*, 146, 111–135. <https://doi.org/10.1016/j.trb.2021.02.006>
- ProRail. (2017, June). *Lagere kosten, minder hinder bij werkzaamheden op het spoor* [Accessed: 29 september 2025]. <https://www.prorail.nl/nieuws/lagere-kosten-minder-hinder-bij-werkzaamheden-op-het-spoor>
- ProRail. (2024a). Jaarverslag 2024 [Accessed: 6 October 2025]. <https://jaarverslagprorail.nl/external/asset/download/project/f970f10a-03e8-0000-2e8a-0cb47fbeb7e3/name/ProRail%20Jaarverslag%202024.pdf>
- ProRail. (2024b, December). *Netverklaring 2026* (Official network statement) (Definitieve versie d.d. 13 december 2024). ProRail. Retrieved January 5, 2026, from <https://www.prorail.nl/siteassets/>

[homepage / samenwerken / vervoerders / documenten / 2026-netverklaring / netverklaring-2026---definitieve-versie-d.d.-13-december-2024.pdf](#)

- ProRail. (2025). Onverminderd veel werk op het spoor in 2025 [Accessed: 2025-09-24].
- Rijksoverheid. (n.d.). Afspraken met prorail — openbaar vervoer [Accessed: 2025-09-24].
- Salido, M., Barber, F., & Ingolotti, L. (2012). Robustness for a single railway line: Analytical and simulation methods. *Expert Systems with Applications*, 39(18), 13305–13327. <https://doi.org/https://doi.org/10.1016/j.eswa.2012.05.071>
- Salido, M., Barber, F., & Ingolotti, L. (2008). Robustness in railway transportation scheduling. *Proceedings of the 7th World Congress on Intelligent Control and Automation*, 2880–2885. <https://doi.org/10.1109/WCICA.2008.4594481>
- Scheepmaker, G. M., & Goverde, R. M. (2015). Running time supplements: Energy-efficient train control versus robust timetables [Accessed: 2025-09-D17]. *Proceedings of the 6th International Conference on Railway Operations Modelling and Analysis (RailTokyo 2015)*. https://repository.tudelft.nl/file/File_02a115ec-061e-4558-9c5c-1e6bd7a3ea37
- Schobel, A., & Kratz, A. (2009). A bicriteria approach for robust timetabling. In R. K. Ahuja, R. H. Möhring, & C. D. Zaroliagis (Eds.), *Robust and online large-scale optimization* (pp. 119–144, Vol. 5868). Springer. https://doi.org/10.1007/978-3-642-05465-5_6
- Serafini, P., & Ukovich, W. (1989). A mathematical model for periodic scheduling problems. *SIAM Journal on Discrete Mathematics*, 2(4), 550–581. <https://doi.org/10.1137/0402049>
- Solinen, E. (2025). *Improving railway timetable robustness: Development and application of robustness indicators* (Publication No. 2415) [Doctoral dissertation, Linköping University]. <https://liu.diva-portal.org/smash/get/diva2:1923967/FULLTEXT01.pdf>
- Solinen, E., Nicholson, G., & Peterson, A. (2017). A microscopic evaluation of railway timetable robustness and critical points. *Journal of Rail Transport Planning Management*, 7(4), 207–223. <https://doi.org/https://doi.org/10.1016/j.jrtpm.2017.08.005>
- Sparing, D., & Goverde, R. M. (2017). A cycle time optimization model for generating stable periodic railway timetables. *Transportation Research Part B: Methodological*, 98, 198–223. <https://doi.org/https://doi.org/10.1016/j.trb.2016.12.020>
- SpoorData ProRail. (n.d.). Wat betekent btd? [Accessed: 2025-09-16].
- Sporenplan.nl. (2025). *Sporenplan: Geldermalsen – 's-hertogenbosch – boxtel – tilburg & eindhoven – venlo – roermond* [Accessed: 20 January 2026]. <https://sporenplan.nl/>
- Takeuchi, Y., Tomii, N., & Hirai, C. (2007). Evaluation method of robustness for train schedules. *Quarterly Report of RTRI*, 48(4), 197–201. <https://doi.org/10.2219/rtriqr.48.197>
- Tielman, W. (2017, October 6). *Validatie – minimale rijtijd* (tech. rep.) (Internal report). ProRail RailwayLab.
- Vromans, M. J. C. M. (2005). *Reliability of railway systems* [PhD thesis]. EPS-2005-062-LIS, Erasmus Research Institute of Management. https://repub.eur.nl/pub/6773/EPS2005062LIS_9058920895_VROMANS.pdf
- Wolniewicz, L. (2024). Railway transport system modelling approach for robustness analysis. *Transport*, 39(4), 287–301. <https://doi.org/10.3846/transport.2024.22877>

A

Fitted Distributions

This appendix provides a visual overview of all fitted distributions with a histogram with theoretical PDF, a CDF plot, and a Q-Q plot.

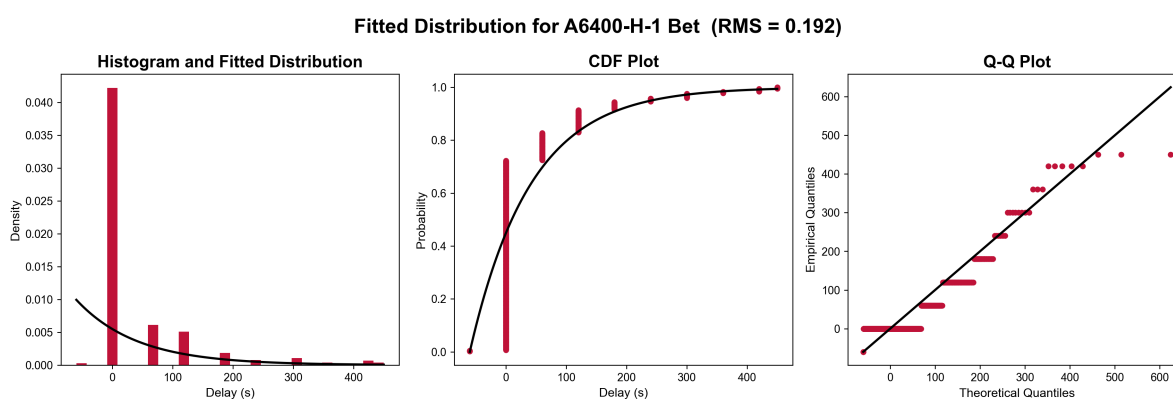


Figure A.1: Histogram with theoretical PDF, CDF and Q-Q plot of A6400 train at Best (Bet).

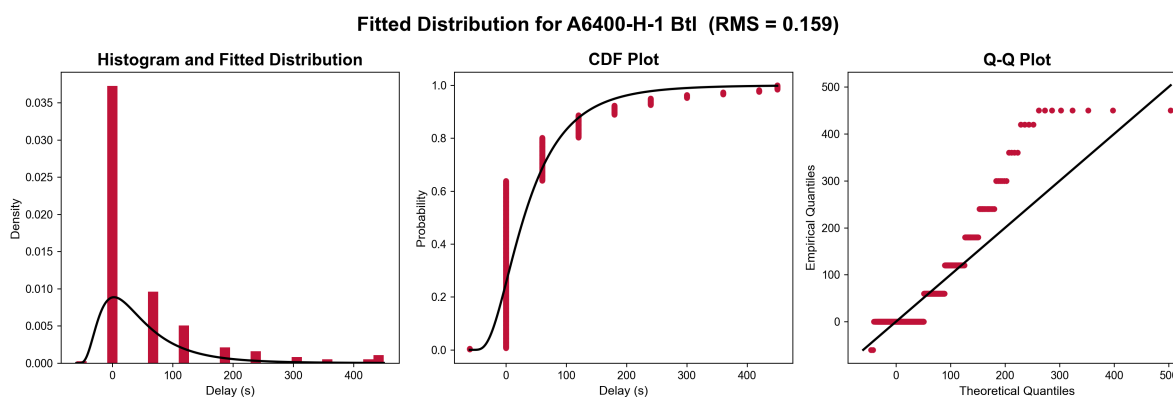


Figure A.2: Histogram with theoretical PDF, CDF and Q-Q plot of A6400 train at Bostel (Btl).

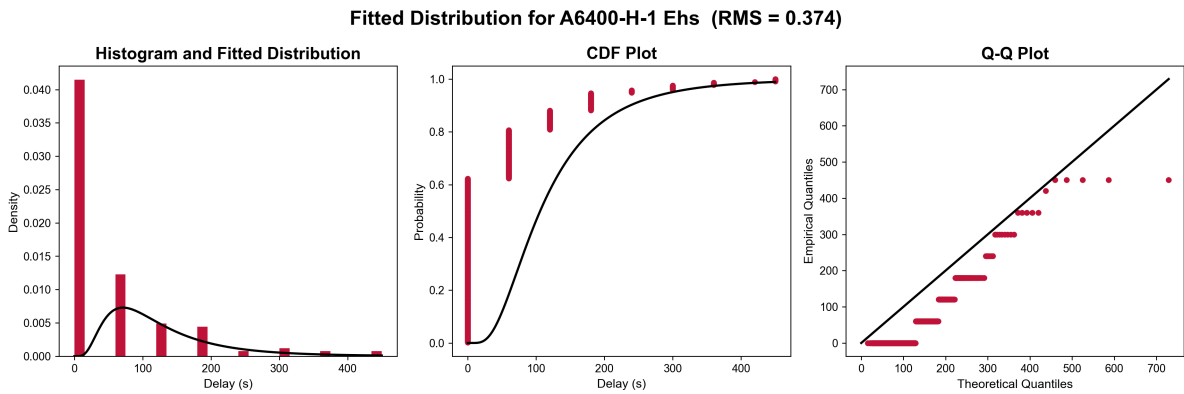


Figure A.3: Histogram with theoretical PDF, CDF and Q-Q plot of A6400 train at Eindhoven Strijp-S (Ehs).

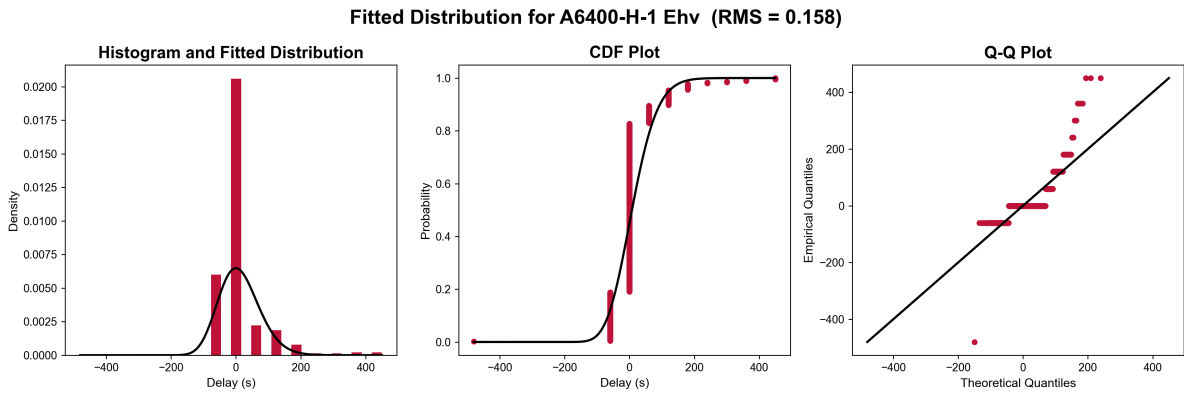


Figure A.4: Histogram with theoretical PDF, CDF and Q-Q plot of A6400 train at Eindhoven Centraal (Ehv).

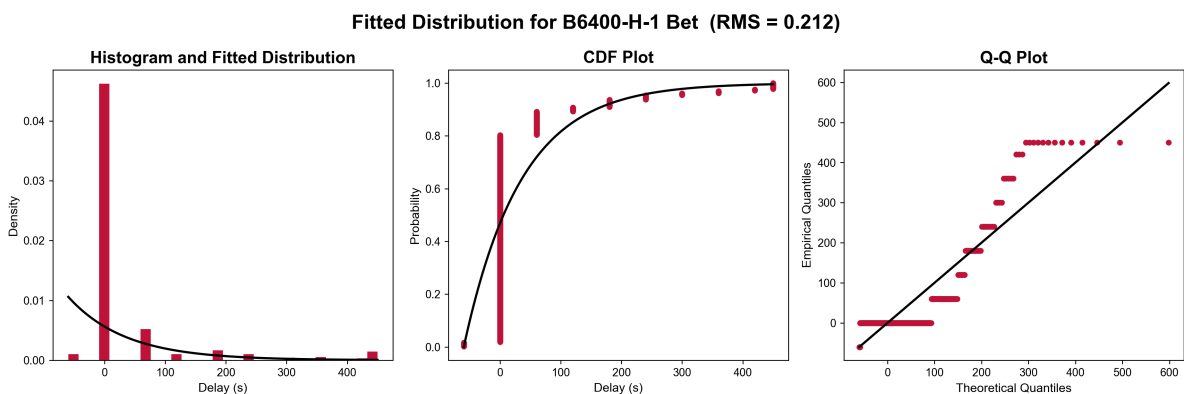


Figure A.5: Histogram with theoretical PDF, CDF and Q-Q plot of B6400 train at Best (Bet).

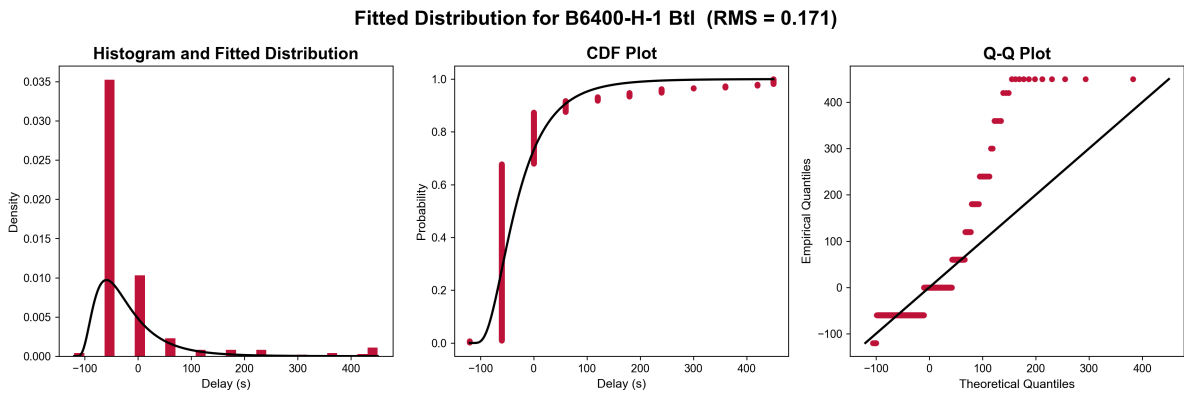


Figure A.6: Histogram with theoretical PDF, CDF and Q-Q plot of B6400 train at Bostel (Btl).

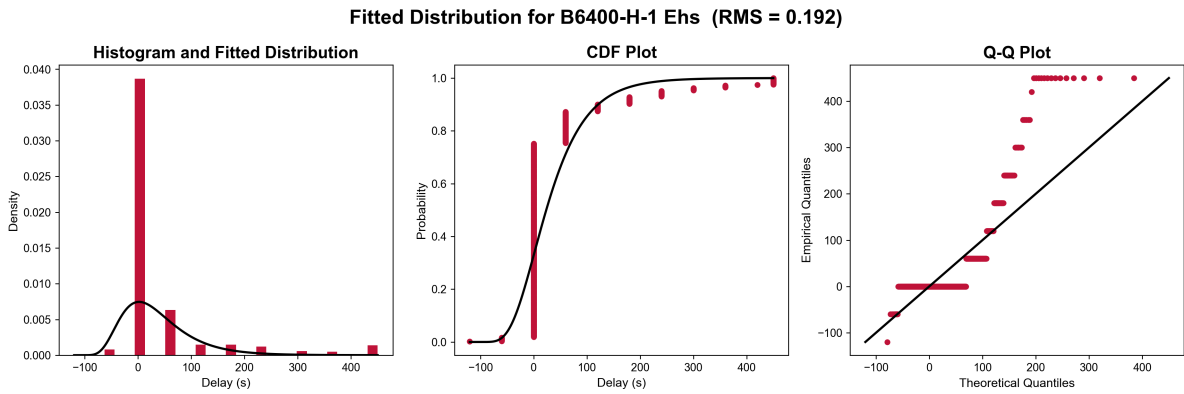


Figure A.7: Histogram with theoretical PDF, CDF and Q-Q plot of B6400 train at Eindhoven Strijp-S (Ehs).

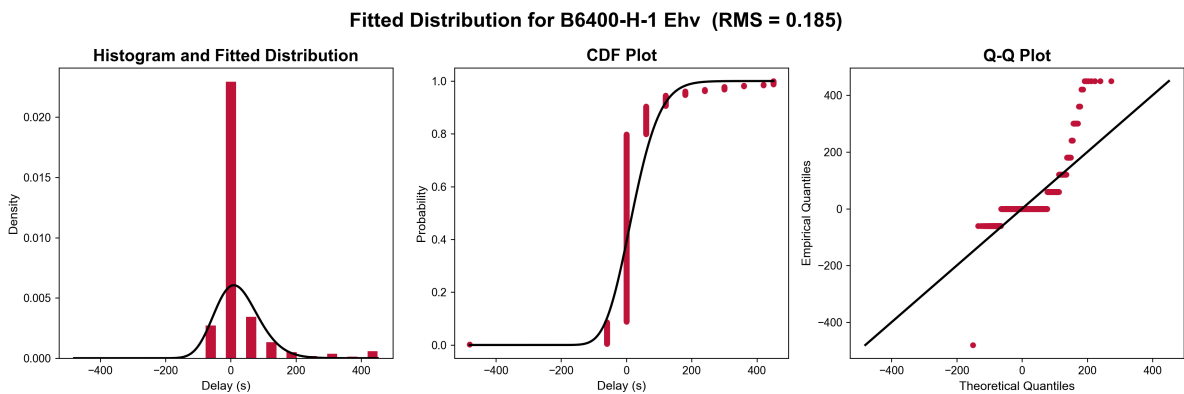


Figure A.8: Histogram with theoretical PDF, CDF and Q-Q plot of B6400 train at Eindhoven Centraal (Ehv).

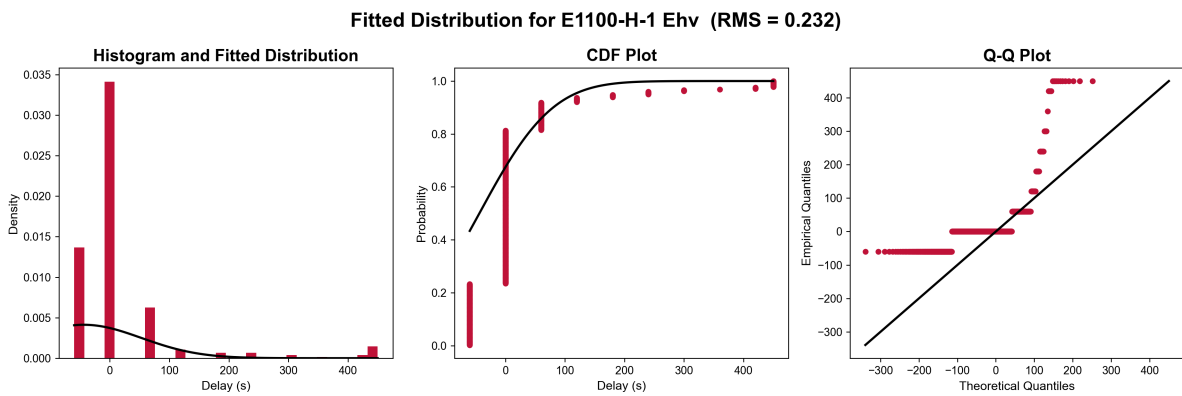


Figure A.9: Histogram with theoretical PDF, CDF and Q-Q plot of E1100 train at Eindhoven Centraal (Ehv).

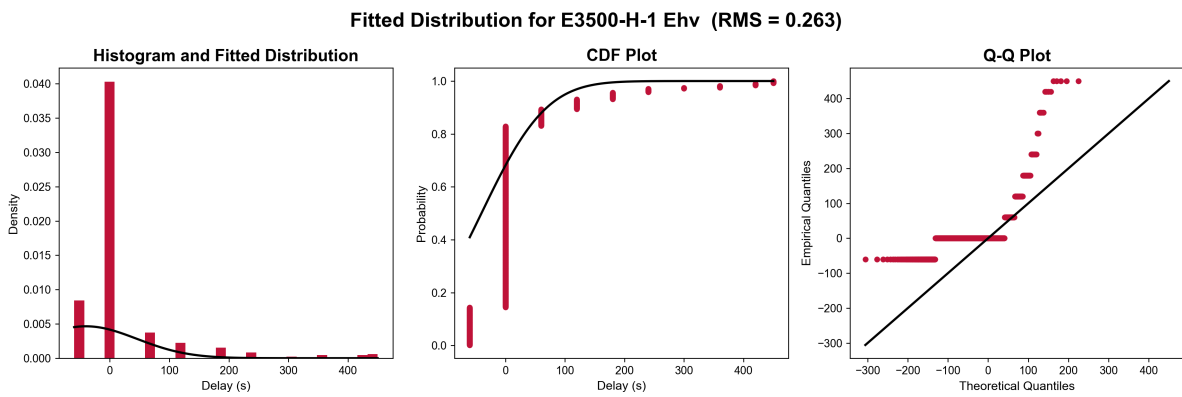


Figure A.10: Histogram with theoretical PDF, CDF and Q-Q plot of E3500 train at Eindhoven Centraal (Ehv).

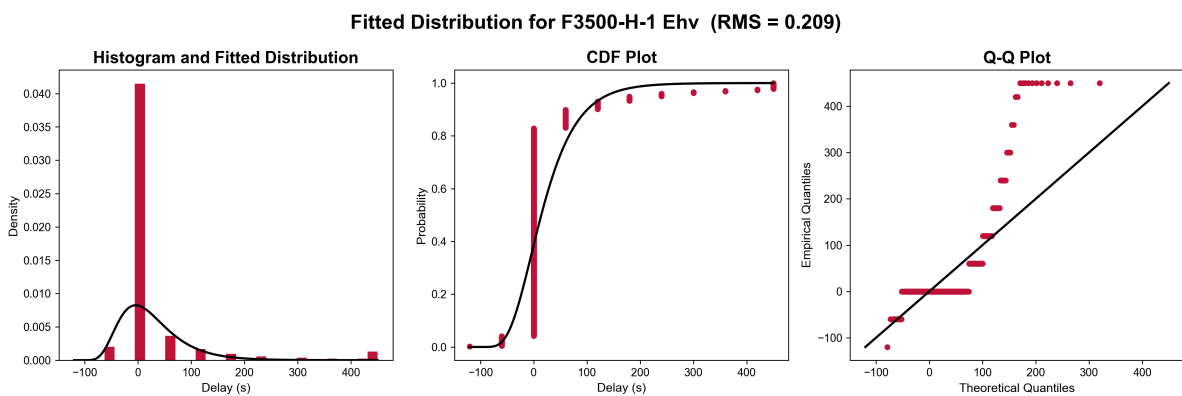


Figure A.11: Histogram with theoretical PDF, CDF and Q-Q plot of F3500 train at Eindhoven Centraal (Ehv).

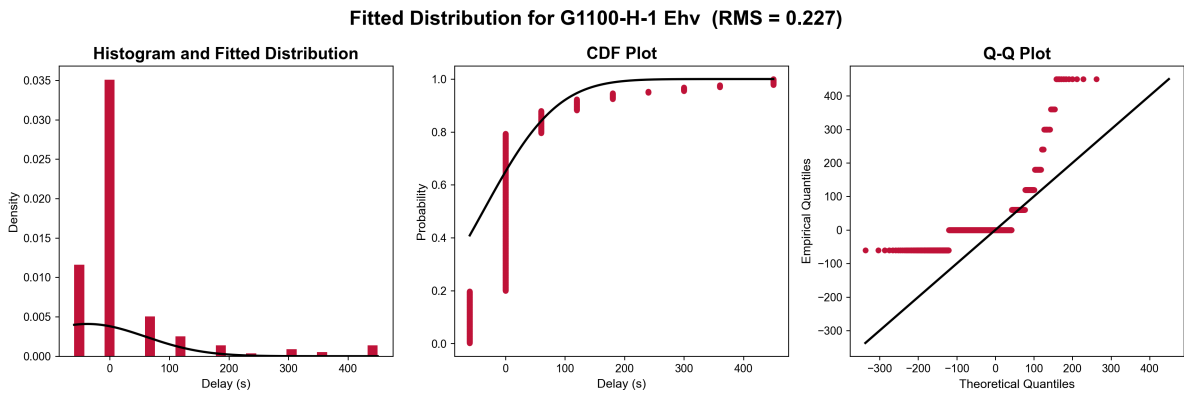


Figure A.12: Histogram with theoretical PDF, CDF and Q-Q plot of G1100 train at Eindhoven Centraal (Ehv).

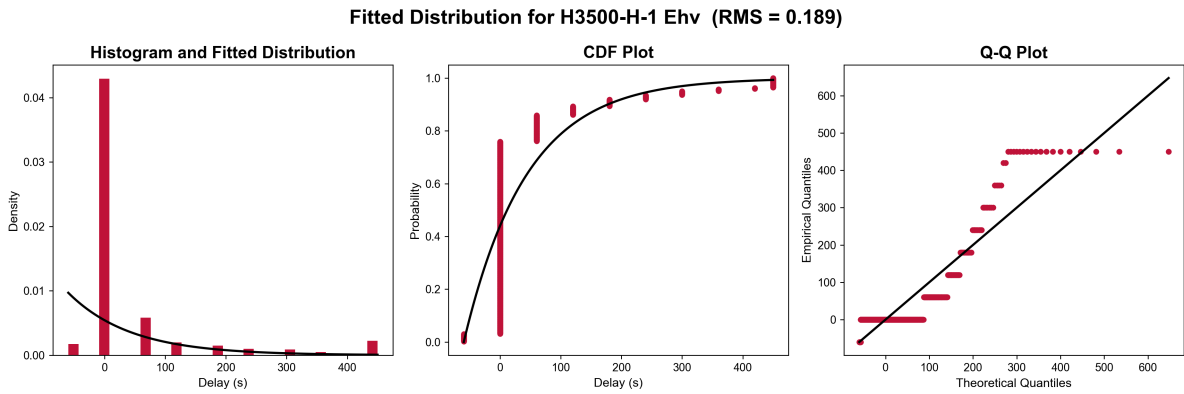


Figure A.13: Histogram with theoretical PDF, CDF and Q-Q plot of H3500 train at Eindhoven Centraal (Ehv).

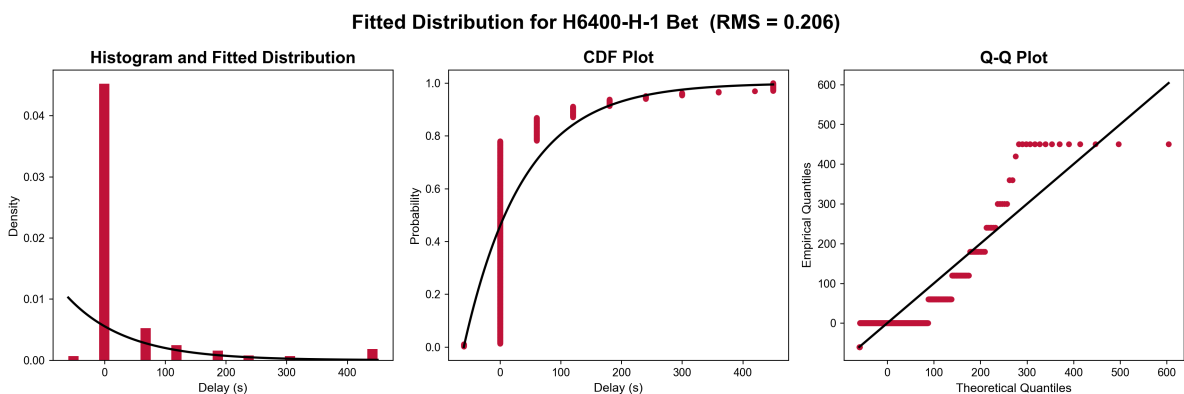


Figure A.14: Histogram with theoretical PDF, CDF and Q-Q plot of H6400 train at Best (Bet).

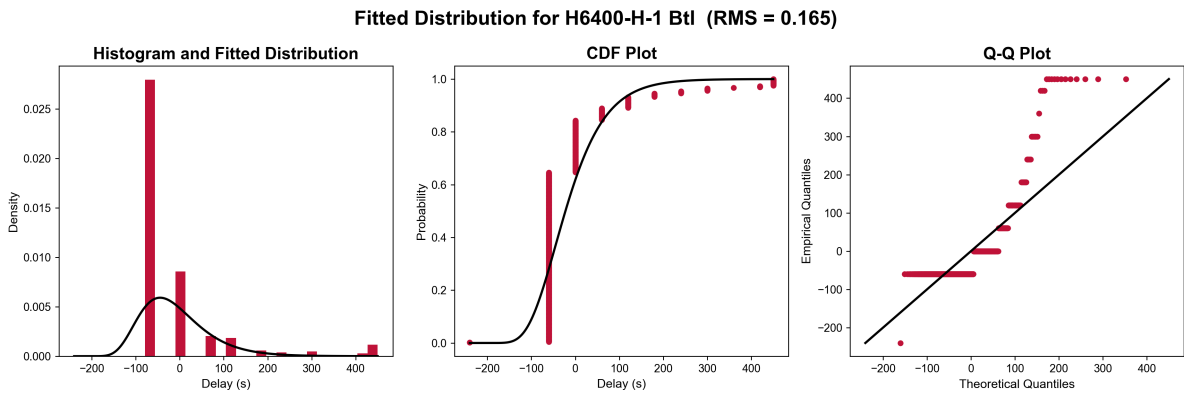


Figure A.15: Histogram with theoretical PDF, CDF and Q-Q plot of H6400 train at Boxtel (Btl).

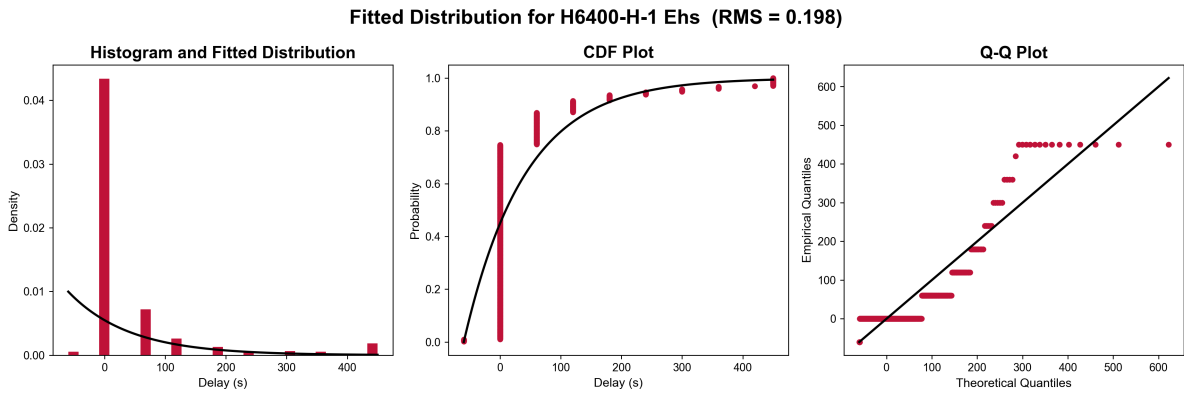


Figure A.16: Histogram with theoretical PDF, CDF and Q-Q plot of H6400 train at Eindhoven Strijp-S (Ehs).

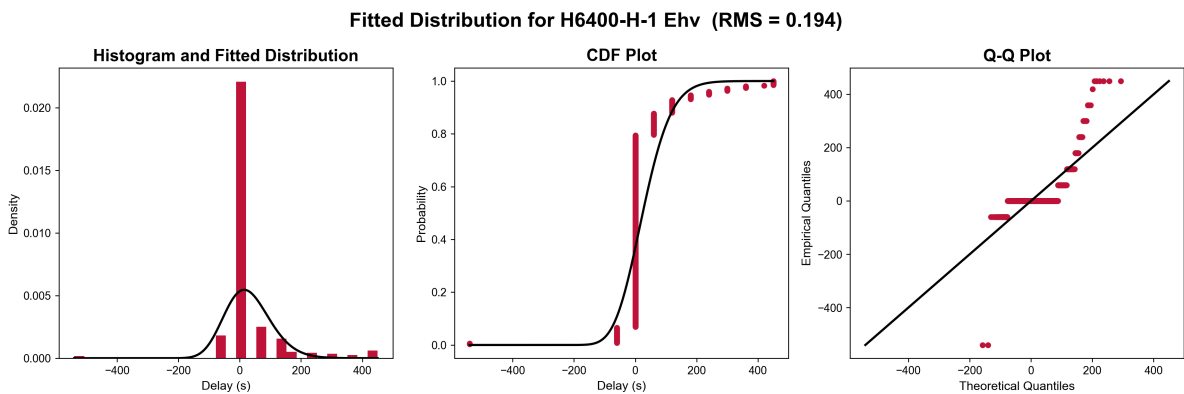


Figure A.17: Histogram with theoretical PDF, CDF and Q-Q plot of H6400 train at Eindhoven Centraal (Ehv).

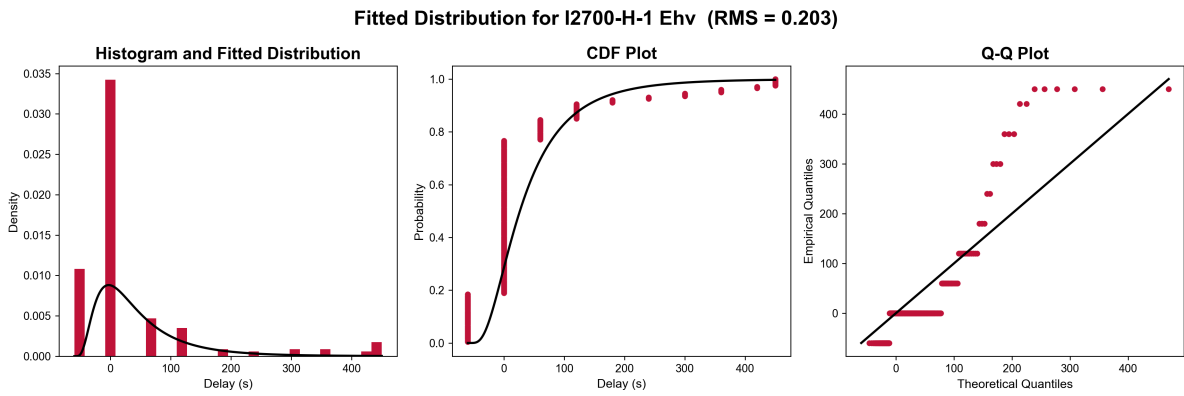


Figure A.18: Histogram with theoretical PDF, CDF and Q-Q plot of I2700 train at Eindhoven Centraal (Ehv).

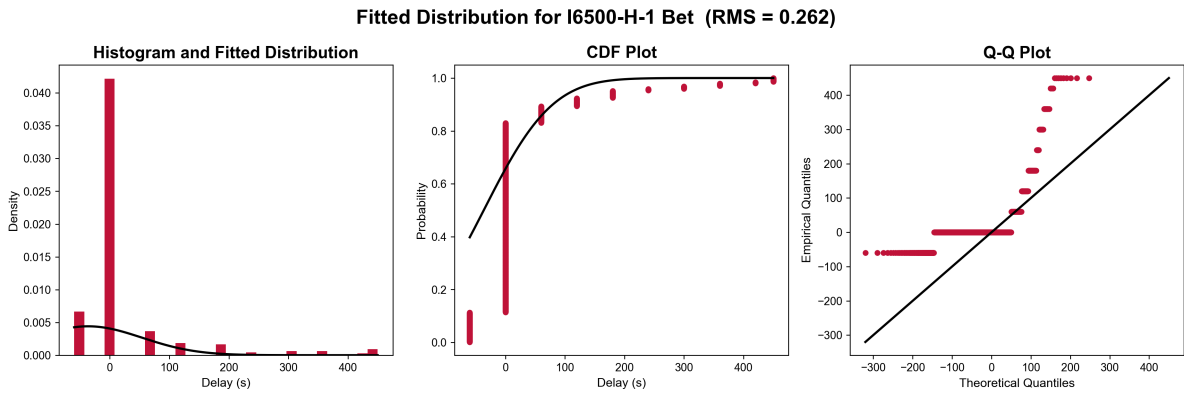


Figure A.19: Histogram with theoretical PDF, CDF and Q-Q plot of I6500 train at Best (Bet).

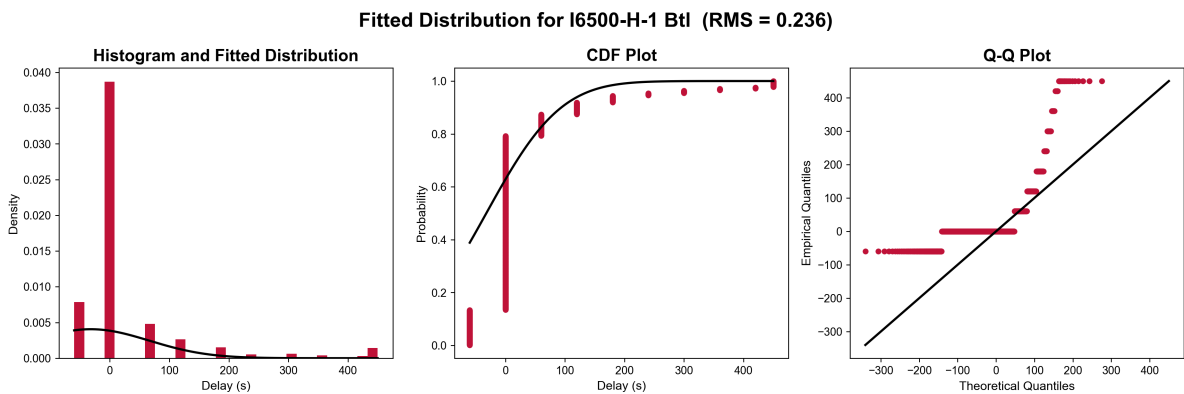


Figure A.20: Histogram with theoretical PDF, CDF and Q-Q plot of I6500 train at Boxtel (Btl).

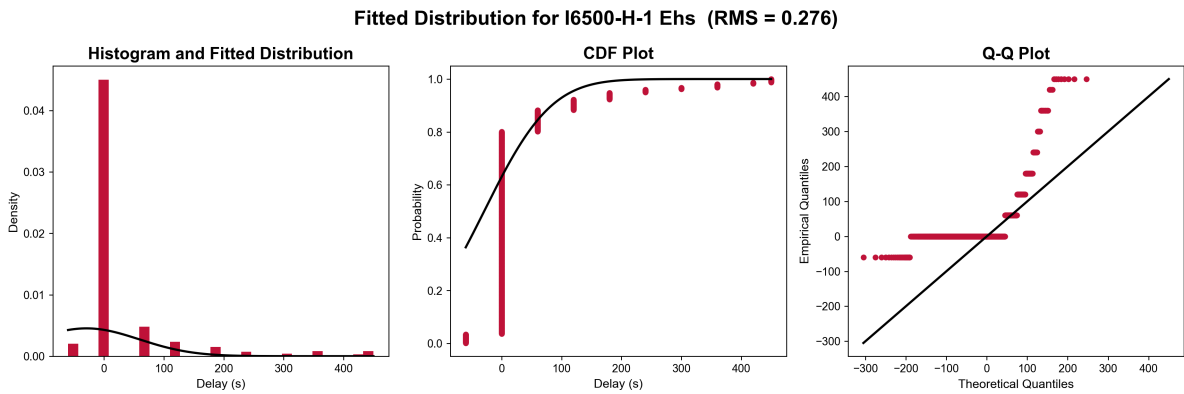


Figure A.21: Histogram with theoretical PDF, CDF and Q-Q plot of I6500 train at Eindhoven Strijp-S (Ehs).

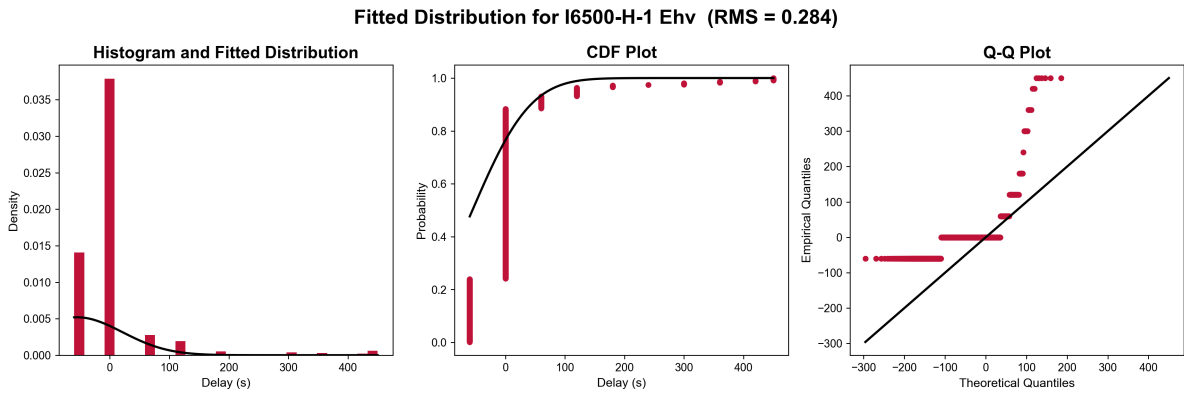


Figure A.22: Histogram with theoretical PDF, CDF and Q-Q plot of I6500 train at Eindhoven Centraal (Ehv).

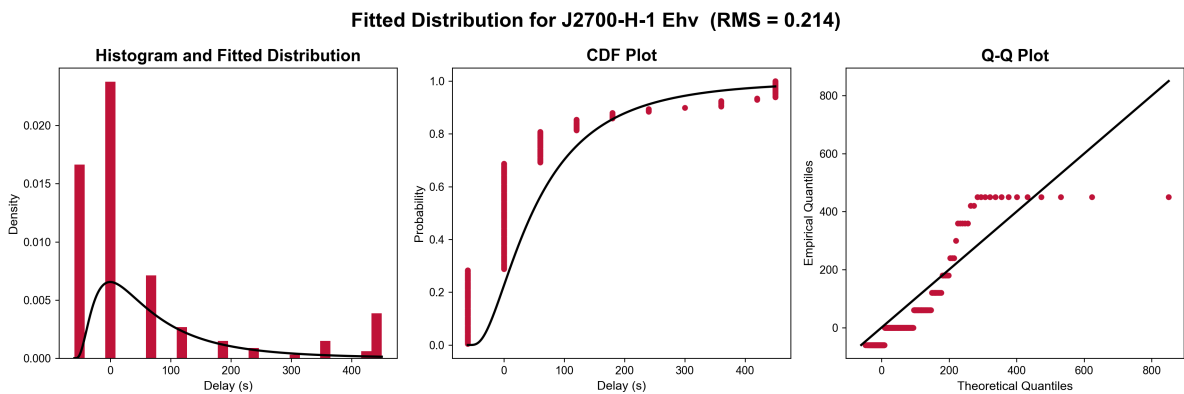


Figure A.23: Histogram with theoretical PDF, CDF and Q-Q plot of J2700 train at Eindhoven Centraal (Ehv).

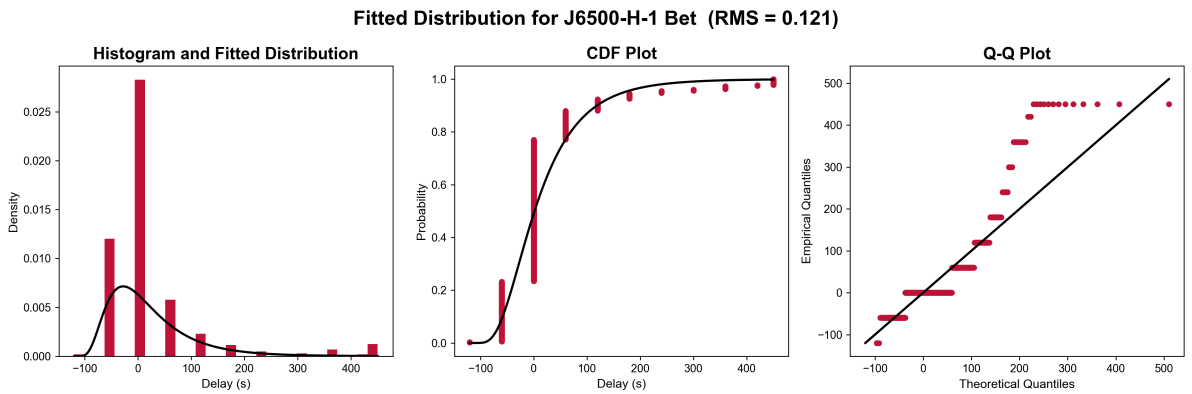


Figure A.24: Histogram with theoretical PDF, CDF and Q-Q plot of J6500 train at Best (Bet).

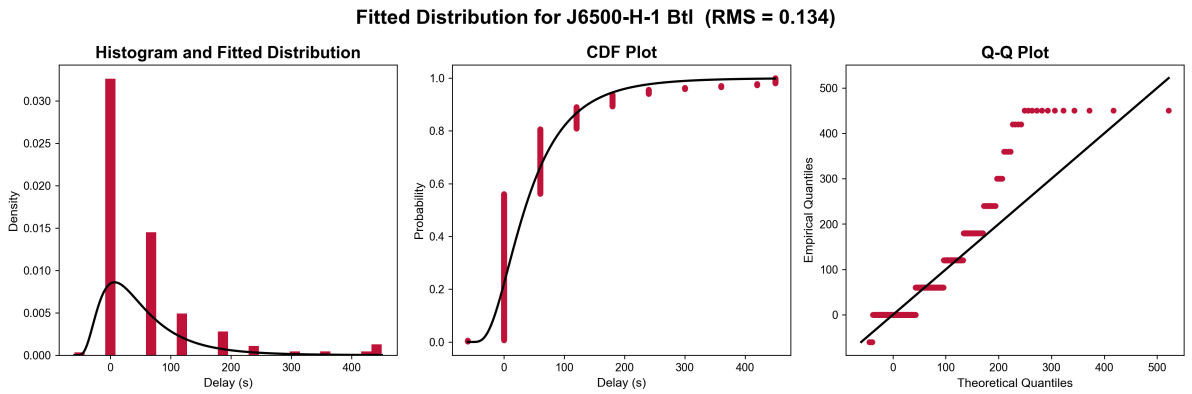


Figure A.25: Histogram with theoretical PDF, CDF and Q-Q plot of J6500 train at Boxtel (Btl).

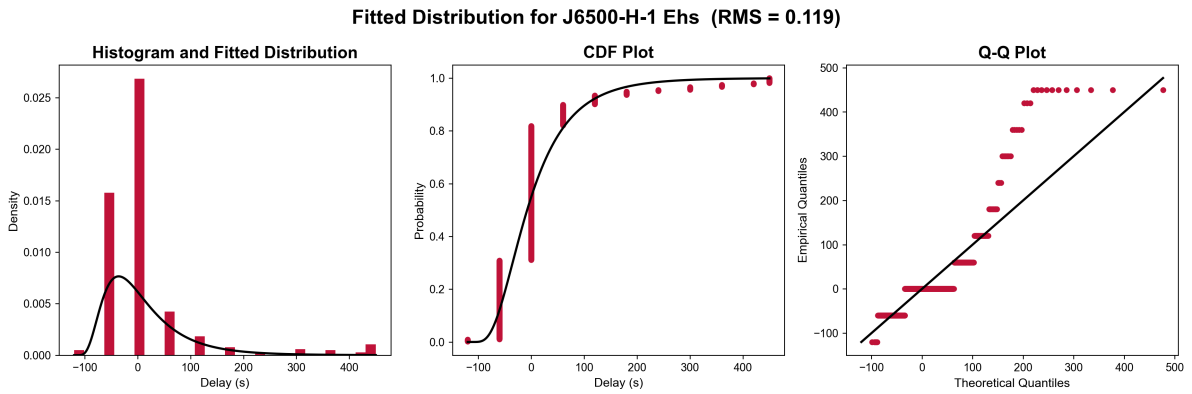


Figure A.26: Histogram with theoretical PDF, CDF and Q-Q plot of J6500 train at Eindhoven Strijp-S (Ehs).

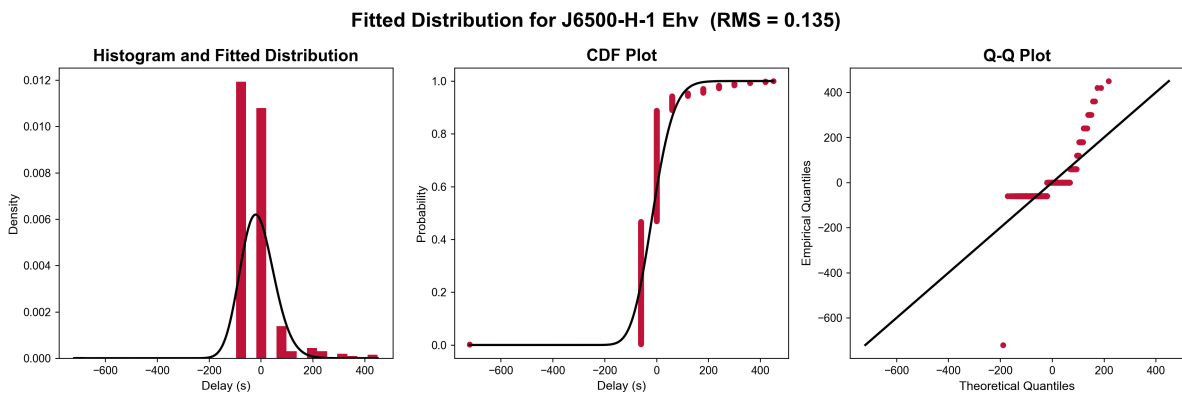


Figure A.27: Histogram with theoretical PDF, CDF and Q-Q plot of J6500 train at Eindhoven Centraal (Ehv).

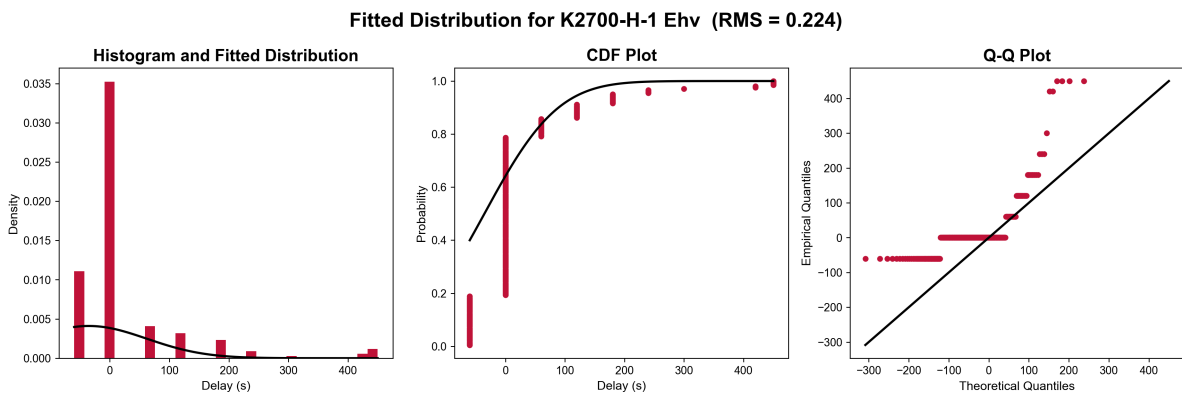


Figure A.28: Histogram with theoretical PDF, CDF and Q-Q plot of K2700 train at Eindhoven Centraal (Ehv).

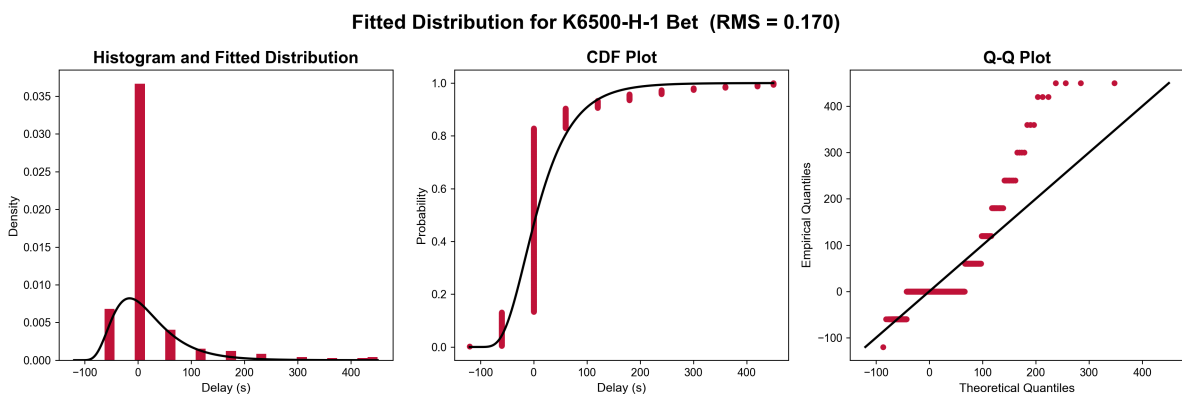


Figure A.29: Histogram with theoretical PDF, CDF and Q-Q plot of K6500 train at Best (Bet).

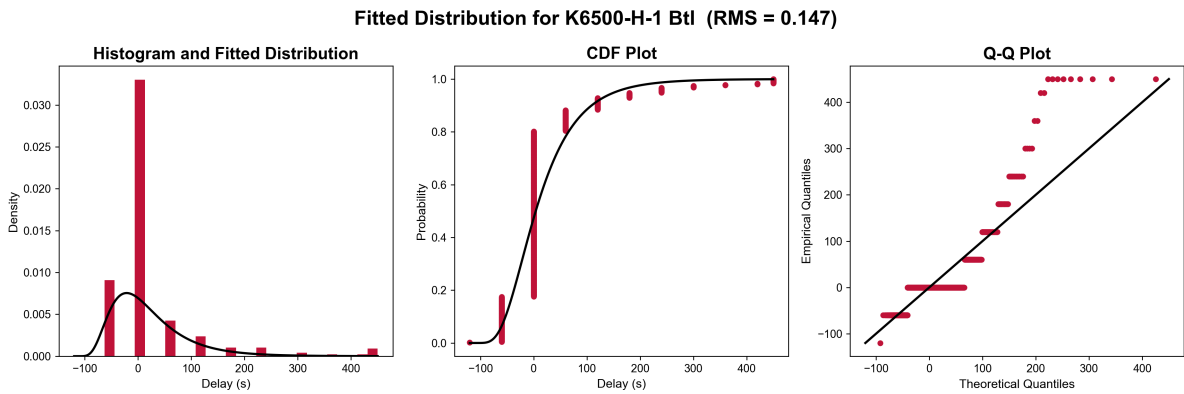


Figure A.30: Histogram with theoretical PDF, CDF and Q-Q plot of K6500 train at Boxtel (Btl).

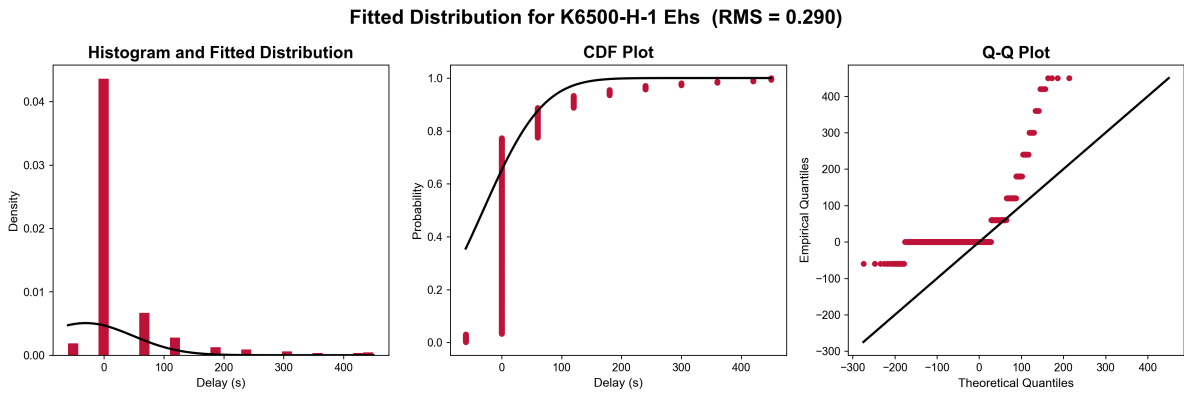


Figure A.31: Histogram with theoretical PDF, CDF and Q-Q plot of K6500 train at Eindhoven Strijp-S (Ehs).

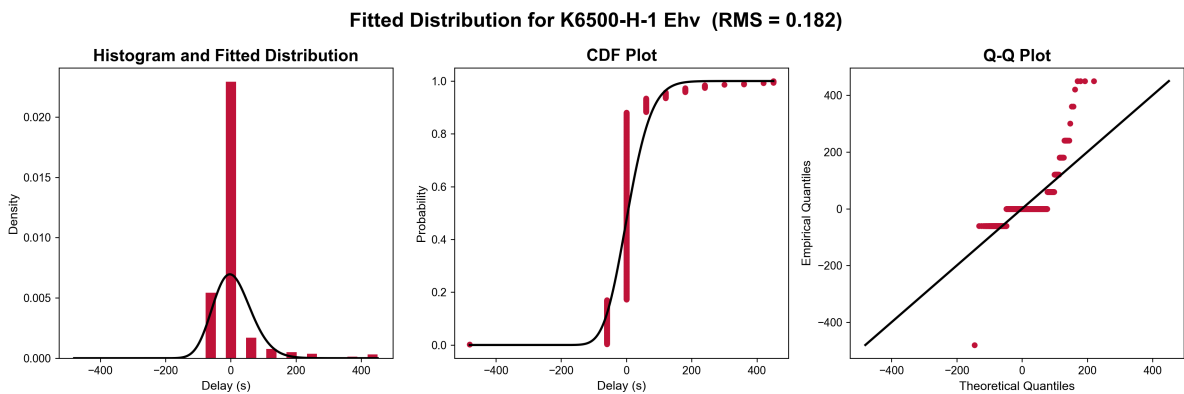


Figure A.32: Histogram with theoretical PDF, CDF and Q-Q plot of K6500 train at Eindhoven Centraal (Ehv).

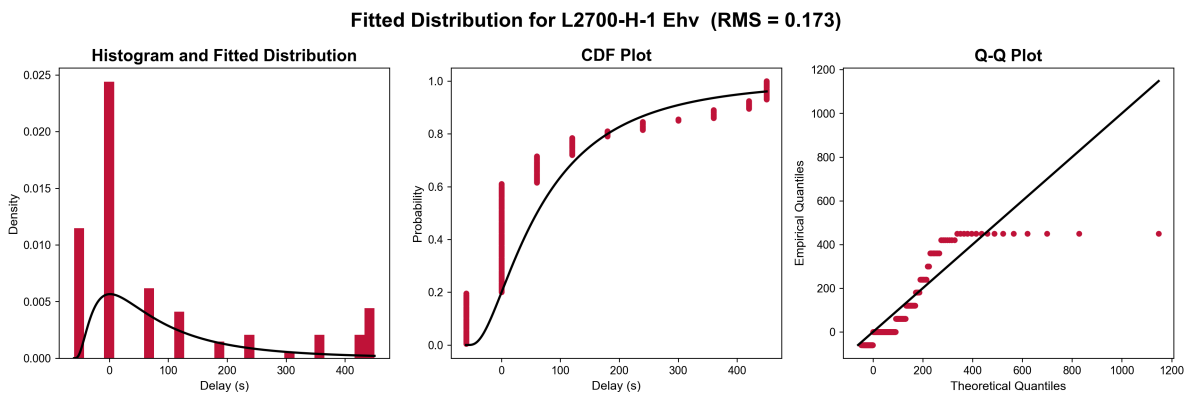


Figure A.33: Histogram with theoretical PDF, CDF and Q-Q plot of L2700 train at Eindhoven Centraal (Ehv).

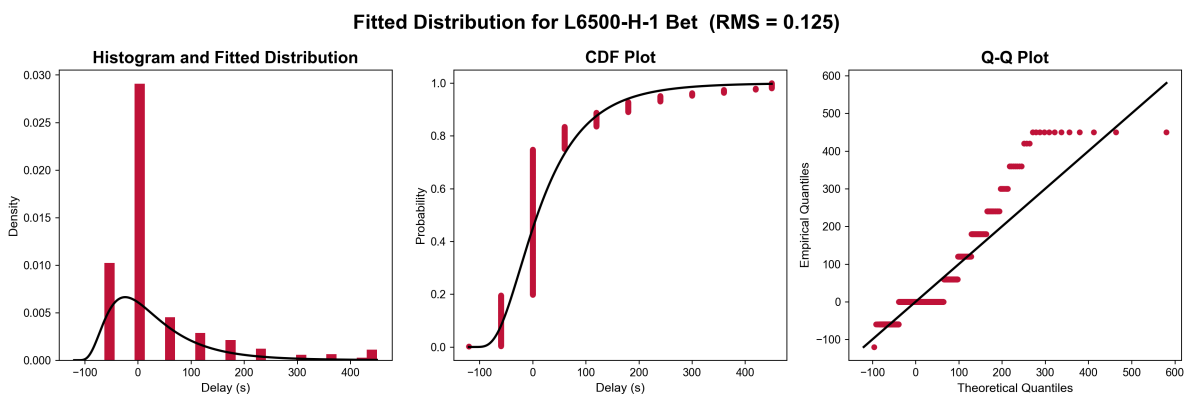


Figure A.34: Histogram with theoretical PDF, CDF and Q-Q plot of L6500 train at Best (Bet).

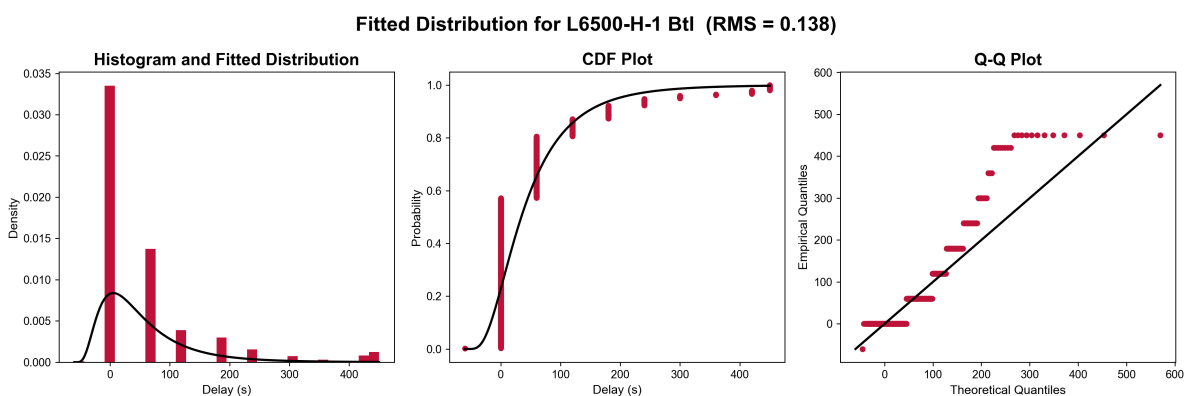


Figure A.35: Histogram with theoretical PDF, CDF and Q-Q plot of L6500 train at Boxtel (Btl).

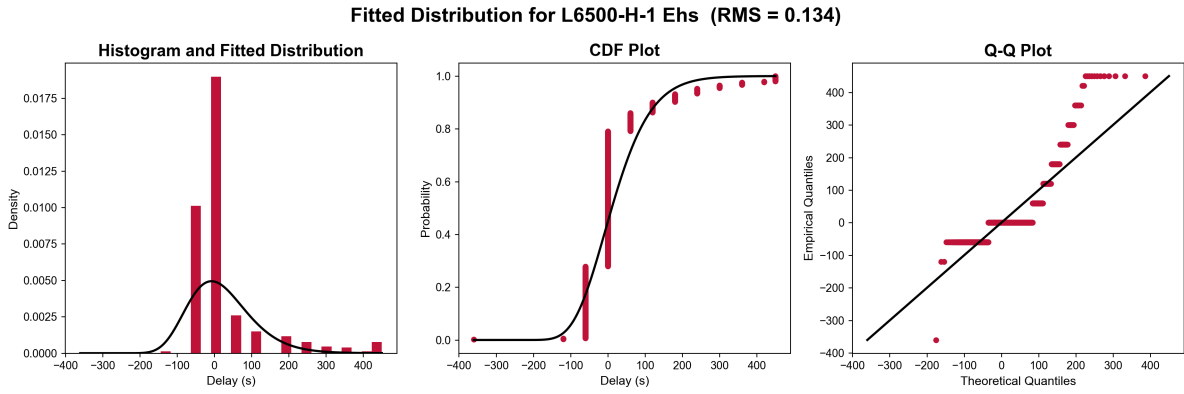


Figure A.36: Histogram with theoretical PDF, CDF and Q-Q plot of L6500 train at Eindhoven Strijp-S (Ehs).

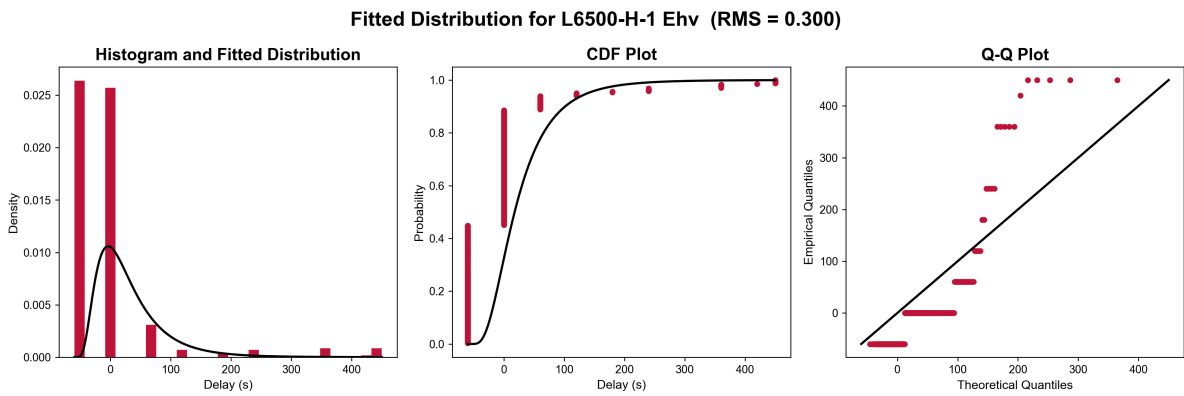


Figure A.37: Histogram with theoretical PDF, CDF and Q-Q plot of L6500 train at Eindhoven Centraal (Ehv).

Quantifying the Effectiveness of Temporary Railway Timetables under Planned Track Closures

Femke T. Tamsma,¹*

¹*Delft University of Technology, The Netherlands*

10 February 2026

ABSTRACT

Temporary railway timetables are implemented during periods of disrupted operations, such as planned track closures, and operate under constrained conditions. Temporary timetables lack quantitative methods to assess their overall performance and systematic comparison. This paper introduces an approach to quantify the effectiveness of temporary railway timetables. Timetable effectiveness is defined as the ability of a temporary timetable operating under constrained conditions to maintain operable, stable, and robust train operations under deterministic and stochastic variability. Based on this definition, a multi-criteria evaluation framework is proposed that integrates four performance indices: capacity utilization, stability, robustness, and operability. Existing timetable attributes are reformulated to reflect constrained operating conditions. In addition, a new timetable attribute, operability, is introduced to capture the executability of infeasible timetables. The framework aggregates the indices into a single Temporary Timetable Effectiveness (*TTE*) indicator. The proposed methodology is demonstrated through a real-world case study of a planned track closure on the Dutch railway network. Multiple temporary timetables are quantified and compared using microscopic simulation. The results show that the framework is capable of distinguishing both marginal and substantial differences between alternative timetable designs and making trade-offs between performance indices explicit. The proposed framework provides a systematic quantitative assessment and comparison of temporary railway timetables.

Key words: Temporary railway timetables – Effectiveness – Quantification – Operability

1 INTRODUCTION

Rail transportation is essential in modern transportation systems. Trains offer a mode of transport that combines high capacity, energy efficiency, and reliability. Across Europe, the demand for rail transport continues to grow due to population growth, increasing mobility demand, and expanding freight traffic, although current demand remains below levels before COVID (ProRail (2024)). Having a high-functional railway system is crucial to meet environmental, economic, and social goals. At the same time, many railway systems are under continuous pressure as increasing demand for capacity must be accommodated on infrastructure that requires regular maintenance and upgrades.

Railway timetabling plays a central role in the planning and operation of rail transport systems. Schedules coordinate the use of limited infrastructure capacity while balancing the efficiency and stability and robustness of operations. The increasing demand for rail transport, mixed traffic operations, and intensive use of infrastructure have made timetabling more complex. The quality of timetables has become a key determinant of operational performance. Much of the scientific literature has focused on railway timetabling problems and evaluating timetable quality. A wide range of concepts and methods are developed to assess aspects such as feasibility, stability, robustness, resilience, and efficiency.

Traditionally, railway timetabling research has focused on nominal timetables, designed under stable and predefined conditions. However, real-world railway operations frequently operate under conditions that deviate from these nominal assumptions. Unforeseen disruptions and planned capacity restriction such as maintenance and construction works require the use of temporary timetables. Despite their operational importance, temporary timetables have received little explicit attention in the scientific literature, as they are treated as minor deviations from nominal schedules rather than as a distinct timetabling context. Existing approaches to timetable evaluation are not directly applicable to assess the performance of temporary timetables. Consequently, the overall performance of temporary timetables is difficult to evaluate systematically, and the comparison between multiple alternative temporary schedules remains challenging. In particular, there is a lack of integrated methods that combines multiple timetable attributes into a single performance measure that reflects the effectiveness of timetables as a whole.

To address this gap, this paper introduces the concept of timetable effectiveness as an overarching performance indicator that captures the ability of a timetable to support feasible, stable and robust train operations under constrained conditions. Timetable effectiveness extends beyond single key performance indices by including multiple dimensions into an integrated evaluation framework. Although the underlying indices and aspects of timetables have been studied extensively, explicit quantification of overall timetable effectiveness has received limited attention in railway research, and there is no

* E-mail: f.t.tamsma@student.tudelft.nl

dedicated framework for temporary timetables. This paper therefore proposes a multi-criteria indicator to quantify timetable effectiveness and demonstrate its application in the context of temporary railway timetables. This framework uses existing concepts such as capacity, stability, robustness, and feasibility while introducing new formulations specific to the context of contained operational conditions. In addition, a new timetable attribute, operability, is introduced to capture the degree to which a (infeasible) timetable can be executed in practice, despite minor conflicts.

The main contributions of this paper are:

- The conceptualization of temporary timetables as a distinct research domain within railway timetabling.
- A formal definition of timetable effectiveness, as an integrated perspective of timetable performance under constrained operational conditions.
- A generic methodological framework to quantify timetable effectiveness through a set of indices.
- New formulations of existing timetable attributes, specific to the context of temporary timetables.
- A new timetable attribute, 'operability', that captures the executability of infeasible timetables.

The remainder of this paper is structured as follows. [Section 2](#) reviews the literature on railway timetabling, focusing on timetable indices, evaluation methods, and the distinction between nominal and temporary timetables. The proposed timetable effectiveness framework and associated mathematical formulation are presented in [Section 3](#). [Section 4](#) applies the framework to a temporary timetable case study and discusses the resulting performance evaluation. In [Section 5](#) a reflection on the implication and limitations of the proposed framework is given. Finally, [Section 6](#) concludes and outlines directions for future research on timetable evaluation.

2 THEORETICAL BACKGROUND & LITERATURE REVIEW

Railway timetables can be assessed using a wide range of indices and evaluation methods. Most existing approaches have been developed in the context of nominal timetables operating under stable conditions. This section reviews the relevant timetable attributes, performance indicators, and evaluation approaches, and discusses how these concepts relate to both nominal and temporary timetables.

2.1 Railway Timetabling Fundamentals

Designing a timetable is about finding a conflict-free schedule and coordinating the use of limited infrastructure capacity ([Correia Duarte et al. \(2025\)](#)). Timetabling decisions are defined at three planning levels: strategic, tactical, and operational. Strategic planning addresses long-term network and service design, tactical planning focuses on timetable design, and the operational level concerns real-time traffic management and dispatching ([Correia Duarte et al. \(2025\)](#)).

2.1.1 Nominal versus Temporary Timetables

Nominal and temporary timetables serve different purposes within railway operations. Nominal timetables are designed under stable operating assumptions and aim to maximize efficiency while infrastructure capacity constraints are satisfied ([Cacchiani and Toth \(2012\)](#) & [Polinder et al. \(2021\)](#)). These timetables include deterministic slack times to absorb operational variations. Temporary timetables

are revised schedules implemented during periods of disrupted operations, such as maintenance activities. These schedules are applied for limited periods and operate under constrained conditions. Due to reduced infrastructure availability and adjusted arrival and departure times, routes, track allocations, and train frequencies, temporary timetables typically contain less slack. This makes these schedules more vulnerable to delay propagation. Temporary timetables represent a trade-off, namely ensuring sufficient service performance while working within limited infrastructure availability and avoiding excessive inefficiencies.

2.1.2 Deterministic and Stochastic Timetabling

Deterministic and stochastic approaches represent two ways of modeling timetabling problems. Deterministic approaches assume fixed input parameters, such as running times, dwell times, network availability, and headways, and aim to produce schedules that are optimized under ideal conditions ([Salido et al. \(2012\)](#)). Stochastic approaches explicitly incorporate uncertainty by representing operational processes through probability distributions ([Salido et al. \(2012\)](#)). These methods aim to design a schedule that performs well across a range of disturbance scenarios.

Both approaches can be used as strategies to manage operational variability. Where deterministic methods tackle variability by adding fixed buffers and supplements, stochastic methods embed uncertainty directly into the design process by representing probability distributions for sources of variability. Currently, deterministic approaches dominate in timetable design, while stochastic methods are increasingly used for robustness and resilience evaluation.

2.2 Timetable Attributes and Definitions

Timetables translate the long-term demand for train services into an operational plan that assigns to each train a conflict-free path and track allocation over the available infrastructure within a given period ([Hansen and Pahl \(2008\)](#), [Cacchiani and Toth \(2012\)](#), & [Correia Duarte et al. \(2025\)](#)). Several attributes are commonly used in literature to characterize timetable performance. Feasibility is defined as whether the timetable is conflict-free under ideal conditions ([Goverde and Hansen \(2013\)](#)). Stability refers to whether small delays can be absorbed without spreading or requiring rescheduling ([Goverde and Hansen \(2013\)](#) & [Goverde \(2010\)](#)). Resilience describes the ability of the system to recover after disruptions through system adaptation or rescheduling measures ([CORDIS: Publications Office of the European Union \(nd\)](#)). The ability of a timetable to prevent delay propagation and maintain scheduled train paths under small stochastic disturbances is referred to as robustness ([Lusby et al. \(2018\)](#) & [Solinen \(2025\)](#)). Efficiency relates to optimal or near-optimal use of resources under nominal conditions ([Cacchiani and Toth \(2012\)](#)).

These attributes are closely related to the concept of capacity, as feasibility, stability, resilience, robustness, and efficiency each depend on how the available infrastructure can accommodate train movements. Capacity in the rail context is generally defined as the number of trains per unit of time. In railway operations, different variants of capacity are distinguished:

- Theoretical capacity: The maximum number of trains (upper bound) in a network section, assuming a perfectly homogeneous traffic flow without any buffers and slack ([Besinovic and Goverde \(2018\)](#) & [Jensen et al. \(2020\)](#)).
- Maximum capacity: The maximum number of trains that can

be scheduled in a given network section, given the actual traffic heterogeneity, without headway buffers (Jensen et al. (2020)).

- Practical capacity: The effective capacity in a network section that can be scheduled in a robust and stable manner, including headway buffers (Besinovic and Goverde (2018) & Jensen et al. (2020)).
- Capacity occupation: The actual time that the infrastructure in a network section is occupied by a set of trains (Jensen et al. (2020)).
- Capacity occupation rate: The share of the time window in which the infrastructure is occupied in a network section, the ratio between capacity occupation and a time period (Jensen et al. (2020)).

2.3 Timetable Performance Evaluation

The evaluation of timetables involves performance indicators that quantify attributes and methodological approaches used to assess these indicators. Key Performance Indicators (KPIs) describe which aspects of timetable quality are measured, whereas evaluation methods determine how these indicators are assessed.

The KPIs can be evaluated ex-ante or ex-post. Ex-ante measures are embedded in the timetable itself and can be determined without knowledge of traffic performances (Andersson et al. (2013a)). They describe specific and quantifiable features of a schedule, such as planned running time supplements, headway buffer times, or the distribution of slack along a train path. Ex-post indicators, derived from traffic performances, evaluate timetables by simulation or observing their execution in real-world conditions. Commonly used ex-post indicators are punctuality, total delay, secondary delay, or cumulative delay propagation patterns across the network (Jensen et al. (2014)).

2.3.1 Key Performance Indicators

For feasibility, an ex-ante indicator is the scheduled process time. If the scheduled running, dwell, or turn-around times exceed the corresponding technical minimums, the timetable is considered feasible (Goverde and Hansen (2013)). A conflict is detected when, on a schedule, at least one train is forced to brake, stop, or deviate from its scheduled movements. The number of scheduled train conflicts should be zero. This can be achieved by ensuring sufficient (temporal) separation between successive departures. The minimum spacing required must be equal to the minimum headway time between successive trains (Hansen and Pacht (2008)). From an event-sequencing perspective, a KPI for timetable feasibility is the absence of logical inconsistencies, ensuring that trains cannot depart or arrive earlier than physically possible (Polinder et al. (2021)).

Ex-ante stability indices focus on the amount and distribution of slack, such as running time supplements and buffer times. A timetable is considered stable if the minimum cycle time (i.e., the closed sequence of events with the least mean time allowance (Goverde (2007) & Goverde (2010))) is shorter than the nominal timetable period. Here, the minimum cycle time of a schedule is an ex-ante indicator of timetable stability. Ex-post KPIs focus on delay absorption such as Weighted Travel Time Extension or heterogeneity indices, which quantify how uneven headways and traffic mixes impact the ability to absorb primary delays (Jensen et al. (2014) & Dewilde et al. (2014)).

An ex-ante efficiency indicator is the total time supplement, defined as the sum of running time supplements and buffer times (Scheepmaker and Goverde (2015)). Other indicators use capacity utilization (ex-ante), measured as the infrastructure occupation rate (Goverde and Hansen (2013)) and the average travel time (ex-post) (Lee et al. (2017)). The infrastructure occupation rate captures how much of the available track capacity is consumed by a schedule.

This indicator reflects the proportion of time that the infrastructure is blocked by train movements and thus indicates the utilization (Goverde and Hansen (2013)).

For robustness the most commonly used KPI is the timetable margins. There are two types of margins, the first can be added to ensure that delay propagation and tight dependencies between successive trains are avoided (i.e., buffer time) (Andersson et al. (2013b)). The other can be used to adjust the minimum running times to account for variations in actual driving behavior (i.e., running time supplements) (Andersson et al. (2013b)). Ex-post KPIs include punctuality (Goverde and Hansen (2013)), secondary delay (Dewilde et al. (2014)), percentile-based process times (Goverde and Hansen (2013)), and cumulative total delays (Salido et al. (2008)).

2.3.2 Analytical and Empirical Evaluation

There are two main categories for the methodological approaches to evaluate KPIs. First, analytical approaches use mathematical formulations without requiring full simulation of train operations. Classical methods include timetable compression, blocking time theory, and max-plus algebra (Hansen and Pacht (2008), Goverde et al. (2013) & Goverde (2007)). Secondly, empirical evaluation techniques commonly make use of simulation-based approaches that model train movements and disturbances to evaluate timetables under operational scenarios (Lindfeldt (2008) & Jensen et al. (2014)). Simulation techniques are especially useful in the context of timetable stability, resilience, and robustness due to the possibility of analyzing average delays caused by disturbances and the recovery time needed to return to the planned schedules (Salido et al. (2008) & Wolniewicz (2024)).

2.3.3 Timetable Performance Analysis Method

Where the reviewed literature distinguishes between individual performance indices and evaluation methods, the assessment of temporary timetables requires an integrated evaluation approach. A qualitative analysis identifies factors that influence the performance of the schedule under both nominal and constrained conditions and identifies how these factors interact in real-world operations. A quantitative analysis, using realized operational data, estimates stochastic probability distributions. These distributions reflect the operational day-to-day variability in stochastic simulations. Combining a qualitative factor identification with a quantitative simulation-based evaluation provides a systematic basis for assessing timetable performance.

2.4 Implications for Temporary Timetables

Quantification of timetable performance is based on both ex-ante schedule characteristics and ex-post performance measures. The choice of quantification methods depends strongly on the timetable context. Nominal timetables, designed under stable and ideal conditions, are assessed with standardized efficiency and capacity-oriented measures. On the other hand, temporary timetables require situation-specific evaluation under reduced infrastructure availability and higher operational variability. Quantification of temporary timetable effectiveness therefore must emphasize delay propagation and service continuity under constrained conditions. Limited studies have quantified performance for temporary schedules, thus to the best of our knowledge no research has so far provided measures to benchmark PTC timetables. This highlights an important research gap: the lack of systematic, data-driven quantification methods for temporary timetables.

3 QUANTIFYING TIMETABLE EFFECTIVENESS

The methodological framework used to quantify timetable effectiveness under constrained condition is presented. First, the conceptual foundations and definition of timetable effectiveness are introduced. Secondly, effectiveness is formalized through a set of performance indices that capture the key operational dimensions relevant for temporary timetables.

3.1 Conceptual Framework

Temporary timetables operate under fundamentally different conditions than nominal timetables. A performance indicator for temporary timetables must satisfy several design principles specific for constrained conditions. First, effectiveness must explicitly account for reduced capacity and altered traffic patterns. Secondly, effectiveness must capture delay absorption for both deterministic disturbances and stochastic operational variability. Third, infeasible yet operationally executable timetables should be evaluated using a graded notion of operability rather than a binary feasibility classification.

3.2 Definition of Timetable Effectiveness

Timetable effectiveness is a broad timetable performance indicator that is defined by the characteristics, feasibility, stability, and robustness. Because temporary timetables operate under constrained conditions, a new definition of timetable effectiveness is necessary.

Effectiveness is the ability of a temporary timetable operating under constrained conditions to maintain operable, stable, and robust train operations under deterministic and stochastic operational variability.

The nominal attributes of capacity, stability, robustness, and feasibility translate under constrained conditions into four fundamental performance dimensions. These dimensions form the basis of the proposed effectiveness framework:

- **Capacity Index:** The preserved throughput of timetables through a corridor.
- **Stability Index:** The ability of timetables to absorb input delays.
- **Robustness Index:** The ability of timetables to prevent delays under stochastic variability.
- **Operability Index:** The degree of executability of (in)feasible timetables.

3.3 Selected Indices for Temporary Timetables

Each of the indices captures a distinct operational dimension that is relevant for the evaluation of timetable effectiveness under constrained infrastructure conditions. All indices are normalized to a scale between 0 and 1, where higher values indicate better performance. The complete mathematical formulation, definitions, and implementation details are provided in [Section A](#).

For this study, it is essential to distinguish between the different types of delays that arise. Initial disturbances refer to exogenous delays that are injected into the simulation as inputs. An input delay is defined as the delay carried forward by the disturbed train as a direct consequence of its initial disturbance. Output delays refer to all other additional delays that emerge endogenously through train interactions, infrastructure conflicts, and delay propagation mechanisms after this initial disturbance, whether on the same train or on other trains. Under multiple simultaneous disturbances, delays cannot be uniquely attributed to a single initially disturbed train. Therefore, only the injected disturbance is defined as input delay.

3.3.1 Capacity Index

The effects of reduced capacity on temporary timetables are captured by the capacity index $C_{b,r}^{mix}$. Let \mathcal{B} denote the set of corridors, and let \mathcal{R} denote the set of timetables. [Equation 1](#) defines the formulation.

$$C_{b,r}^{mix} = C_{b,r}^{pres} \cdot H_{b,r} \quad (1)$$

Here, $C_{b,r}^{pres}$ denotes the ratio of trains preserved under constrained conditions relative to the nominal timetable ($\frac{N_{b,r}^{imp}}{N_{b,r}^{nom}}$). The heterogeneity of the traffic mix is captured by $H_{b,r}$. Maintaining traffic flow under heterogeneous conditions is operationally more demanding than homogeneous traffic, such that a schedule that maintains the same capacity utilization under a heterogeneous mix indicates higher effectiveness than a timetable under homogeneity.

High values of $C_{b,r}^{mix}$ indicate effective capacity utilization under constrained and heterogeneous operating conditions. If a temporary timetable preserves the nominal corridor capacity and operates under maximum heterogeneity, the index is equal to 1. Lower values of the capacity index indicate reduced effective capacity.

3.3.2 Stability Index

Stability measures how effectively a timetable can absorb deterministic input delays to limit the propagation and realization of output delays spreading across the network. For each timetable $r \in \mathcal{R}$, the stability index S_r quantifies how effectively injected input delays ($\Delta^{in,(r)}$) are absorbed to avoid propagation into output delays ($\Delta^{out,(r)}$). This is shown in [Equation 2](#).

$$S_r = \max \left(0, \min \left(1, 0.5 - 0.5 \left(\frac{\Delta^{out,(r)} - \Delta^{in,(r)}}{\Delta^{in,(r)}} \right) \right) \right) \quad (2)$$

Let \mathcal{D} denote the set of injected disturbances. The total input delay for temporary timetable r is given by $\Delta^{in,(r)} = \sum_{d \in \mathcal{D}} \delta_d^{inj}$.

Let \mathcal{P} denote the set of timetable points and let \mathcal{K} denote the set of trains, then $p^{exit}(k) \in \mathcal{P}$ denotes the last timetable point of train k within the study area. To isolate the causal effect of the injected disturbances, delays are evaluated relative to the structural conflicts in the timetable. Additional delays are defined as the difference between the realized delay in the simulation of disturbed operations and the corresponding delay in the baseline, undisturbed, simulation. The additional delay is given by $\delta_{k,p}^{add,(r)} = \delta_{k,p}^{disturb,(r)} - \delta_{k,p}^{base,(r)}$. The total output delay is defined as $\Delta^{out,(r)} = \sum_{k \in \mathcal{K}} \delta_k^{add,exit,(r)}$.

A value of $S_r = 0.5$ indicates neutral stability, where the output delay equals the input delay. High values of S_r indicate stability ($\Delta^{out,(r)} < \Delta^{in,(r)}$), while lower values indicate delay propagation. Furthermore, a timetable is considered unstable if the output delay persists beyond two timetable cycles, in which case the stability index is set to zero ($S_r = 0$).

3.3.3 Robustness Index

Robustness refers to the ability to withstand stochastic operational variability without generating excessive schedule deviations. For each timetable $r \in \mathcal{R}$, the robustness index R_r is defined as the average robustness score on a set of N stochastic replications, as shown in [Equation 3](#).

$$R_r = \frac{1}{N} \sum_{i=1}^N R_{r,i} \quad (3)$$

Let \mathcal{A} denote the set of activities for which $\mathcal{A} = \mathcal{K} \times \mathcal{P}$. For each activity $a = (k, p) \in \mathcal{A}$, $q_a^{(r,i)}$ denotes a binary variable equal to 1 if the activity time realized remains within the allowed tolerance and $q_a^{(r,i)} = 0$ otherwise. Each replication $i \in \mathcal{N}$ represents the realization of stochastic operational variability, generated by sampling random delays from empirically fitted distributions. For a given replication i , the robustness score $R_{r,i}$ is defined as the fraction of timetable activities that remain within an acceptable deviation from their scheduled times ($R_{r,i} = \frac{1}{|\mathcal{A}|} \sum_{a \in \mathcal{A}} q_a^{(r,i)}$).

High values of R_r indicate the ability of a timetable to absorb day-to-day operational fluctuations and maintain operational performance without generating delays. Lower robustness values indicate that small variations lead to schedule deviations.

3.3.4 Operability Index

Operability in this study reflects whether a timetable can be executed within a given reference operational time period ($T^{cycle,ref}$), despite the presence of blocking-time conflicts. While feasibility is traditionally defined as the complete absence of blocking-time conflicts, temporary timetables may remain operationally executable despite minor overlaps in blocking times. For each timetable $r \in \mathcal{R}$, the operability index $O_{b,r}$ is related to the additional cycle time required to eliminate all conflicts on corridor b . This is shown in Equation 4.

$$O_{b,r} = 1 - \min \left(1, \frac{\Delta T_{b,r}^{conflict}}{T^{cycle,ref} - T_{b,r}^{cycle,infeasible}} \right) \quad (4)$$

Let \mathcal{C} denote the set of blocking-time conflicts for a corridor b in timetable r . For each conflict $c_{b,r} \in \mathcal{C}$, the blocking-time overlap is given by $\tau_{c_{b,r}}$ and the following train involved in $c_{b,r}$ is given by $k_f(c)$. The minimum accumulated time shift required to resolve all conflicts in which train k is the following train $k_f(c)$ is given by $\Delta T_{k,b,r}^{accum} = \sum_{k_f(c)=k} \tau_{c_{b,r}}$. The shift in departure of one

train will propagate to all subsequent trains in the timetable, so that the minimum additional time shift required to obtain a conflict-free timetable is determined by train k with the largest accumulated shift: $\Delta T_{b,r}^{conflict} = \max_{k \in \mathcal{K}} \Delta T_{k,b,r}^{accum}$.

In $O_{b,r}$, $T^{cycle,ref} - T_{b,r}^{cycle,infeasible}$ represents the operational slack available. A timetable is considered non-operable ($O_{b,r} = 0$) if the cycle time is larger than the reference period. For all operable timetables ($T_{b,r}^{cycle,feasible} \leq T^{cycle,ref}$), the maximum operability score of $O_{b,r} = 1$ is defined for timetables without blocking-time conflicts ($\Delta T_{b,r}^{conflict} = 0$). As blocking-time conflicts increase, a larger share of the available slack is consumed, reducing $O_{b,r}$.

3.4 Aggregation into an Integrated Indicator

In the context of temporary timetables designed under constrained conditions, effectiveness requires four complementary qualities: (i) maintaining an adequate level of traffic flow through temporary constrained infrastructure, (ii) containing input delays, (iii) absorbing the stochastic variability that characterizes day-to-day operations, and (iv) ensuring that timetables remain operationally executable, even when they are infeasible. Based on these requirements, an ex-post performance indicator is proposed that evaluates timetable effectiveness. The Temporary Timetable Effectiveness indicator (TTE_r) integrates capacity, stability, robustness, and operability into a single indicator. It is defined by Equation 5.

$$TTE_r = \alpha \cdot C_{b,r}^{mix} + \beta \cdot S_r + \gamma \cdot R_r + \epsilon \cdot O_{b,r} \quad (5)$$

TTE_r is defined at the timetable network-level, yet not all indices can be quantified at this level. Therefore, $C_{b,r}^{mix}$ and $O_{b,r}$ are evaluated at the corridor level. The weights α , β , γ , and ϵ determine the relative contribution of each dimension and may be calibrated to reflect preferences and priorities. All indices are bounded to a scale [0-1]. A higher TTE_r value corresponds to a more effective temporary timetable, whereas lower values indicate lower timetable effectiveness. TTE_r is a multi-criteria data-driven performance quality indicator that assesses the research gap of the lack of quantitative evaluation of the effectiveness of temporary schedules.

4 CASE STUDY & RESULTS

The proposed timetable effectiveness framework is applied to a case study. The objective of the case study is to evaluate how temporary timetables perform under constrained infrastructure conditions and operational variability and to illustrate how the proposed TTE_r indicator can be used to compare alternative timetable designs.

4.1 Case Study Description

A case study is conducted on a temporary timetable applied during a planned track closure (PTC) on the Dutch railway network. The selected study area is the Boxtel-Eindhoven Centraal corridor, located in the southern part of the Netherlands, see Figure 1. Although the corridor between Boxtel (Btl) and Eindhoven Centraal (Ehv) forms the main focus of this case study, a broader spatial area is selected. This is necessary to determine the spread of possible output delays.

The Btl-Ehv corridor is part of a heavily used mixed-traffic route where Boxtel serves as a key convergence point for passenger and freight trains and Eindhoven Centraal functions as a major passenger interchange hub. Passenger services include intercity and regional trains, while the corridor also accommodates freight traffic. This corridor is part of the Brabant Route, which is the southern freight corridor that connects the port of Rotterdam to Germany.

The case study focuses on a non-blocking PTC, in which part of the infrastructure remains available while overall capacity is reduced. Such situations require temporary timetables that balance service continuity, operational feasibility, stability, and robustness. Three alternative temporary timetables are considered for the same PTC. All schedules operate under identical infrastructure constraints, but differ in their service composition. Alternative Hour Pattern A (AUP-A) prioritizes the preservation of freight services, AUP-B emphasizes passenger service continuity, and AUP-C maintains both passenger and freight services by maintaining the nominal service level.

By applying the TTE_r indicator to a realistic PTC scenario with different timetables, the case study illustrates how the effectiveness framework can discriminate between alternative timetable designs across multiple performance dimensions. The comparison focuses on understanding how design principles translate into capacity utilization, stability, robustness, and operability, and how timetable effectiveness is affected. The results should be interpreted as an illustration of how the framework can be used to quantify and compare temporary timetables.



(a) Study area shown within the Netherlands.



(b) Zoomed-in view of the southeast of the Netherlands.

Figure 1: Geographical representation of the study area.

Note: Major nodes are marked in yellow. Other operational points are indicated in red. The green lines represent the corridor Btl-Ehv.

4.2 Experimental Setup

4.2.1 Timetable Variants

Four timetable variants are considered. First, a nominal timetable represents regular operations without infrastructure limitations and serves as a reference. When assessing and evaluating the effectiveness of temporary timetables, their performance results are expressed relative to this nominal reference. Second, different alternative temporary timetables that are designed for the same PTC and operate under identical infrastructure constraints are analyzed. Here, infrastructure availability is reduced as two of the four tracks between Boxtel - Eindhoven Centraal are unavailable. Where for AUP-A all freight paths are maintained, passenger supply is reduced by eliminating an SPR service. For AUP-B, the passenger service is fully maintained, while the freight service is reduced to two paths per hour. For AUP-C, both passenger and freight services are maintained, accommodating the full set of scheduled train services within the constrained infrastructure.

4.2.2 Simulation Scenarios

To evaluate timetable effectiveness, various simulations are performed. Simulation of undisturbed operational conditions serves as a baseline in which no initial disturbances are injected. These allow for the identification of structural conflicts. Simulations of disrupted operations represent realistic and degraded conditions and aim to assess how each timetable variant performs under disturbances.

Two sets of experiments are conducted for the simulation of disturbed operational conditions. Deterministic disturbed traffic experiments aim to assess stability and thus evaluate how effectively a timetable absorbs disturbances and prevents delay propagation. A limited number of trains are subject to a single deterministic departure delay at a single location. Three disturbance scenarios, S1, S2, and S3, are defined. In all simulations, a total input delay of 600 s is injected into the simulation, only the distribution of this total delay across trains differs. S1 concentrates the disturbance on two trains with 300 s input delay each, S2 injects a 150 s delay on four trains, and S3 spreads the input disturbance on six trains with each a delay of 100 s. The simulations represent increasing levels of disturbance dispersion while maintaining the same total delay severity.

Simulating stochastic disturbances aims to evaluate timetable robustness under continuous stochastic variability. Delays are generated by sampling from empirically fitted probability distributions and are continuously realized, reflecting normal day-to-day operational variability. To account for the randomness of stochastic disturbances, multiple independent replications are executed. Each replication has a different random realization of the disturbance processes drawn from the fitted probability distributions. Robustness is assessed on the basis of the average performance across these replications.

4.2.3 Simulation Configuration

All simulations are executed in FRISO (Flexible Rail Infrastructure Simulation of the Operation), a microscopic simulation environment used by the Dutch infrastructure manager ProRail. The simulation setup is identical across all timetable variants to ensure fair comparison and isolate the effects of design choices on effectiveness.

Each timetable variant is simulated under undisturbed, deterministic disturbance conditions, and stochastic disturbance conditions. To prevent unrealistic early running, a driving behavior is applied that limits trains from operating at the maximum allowed speed. Freight trains are assumed to run at maximum speed, while Intercities and Sprinters are modeled with 93% of their maximum speed. Regional stopping trains operate at 96% of the maximum speed.

A 30 minute warm-up period is included at the beginning of each simulation to allow the network to reach a representative operational state. Without warm-up time, a limited number of services are present on the network, which may lead to underestimation of delay propagation effects. Performance measurement starts after the warm-up period and continues for 6 hours. A sufficient time horizon is needed to observe how deterministic and stochastic disturbances evolve, allowing for delay propagation and absorption. Shorter time horizons have the risk of truncating these dynamics, especially for disturbances introduced at the end of a timetable cycle.

4.2.4 Performance Evaluation Procedure

To derive the performance indices and the aggregated TTE_r indicator, the simulation outputs are post-processed. This analysis combines microscopic simulation results with timetable and infrastructure data to quantify capacity utilization, stability, robustness, and operability.

For each timetable variant and simulation run, the scheduled and realized train movements for each activity at each timetable point are logged. This offers a comparison between planned and realized operations. Based on the number of scheduled trains and the train composition of the timetable, the capacity index is determined.

Stability is evaluated by applying fixed input delays to selected trains at specific locations. For each activity, an added delay is computed as the difference between the disturbed and undisturbed realized simulation outcomes. To avoid double counting, only added delays at exit points (i.e., the last monitored activity before a train leaves the study area or reaches its final destination) are considered.

Robustness is assessed using stochastic disturbance scenarios. From real-world operational data on the corridor, empirically fitted probability distributions are derived. These distributions are used to generate random delays. For each replication, robustness is quantified as the fraction of realized timetable activities that remain within a predefined tolerance of their scheduled times.

The operability index is derived from timetable and infrastructure characteristics. Using section occupation data and blocking-time conflicts, the schedule is compressed by iteratively reducing slack until blocking times touch, following the UIC 406 timetable compression method (International Union of Railways (UIC) (2004)). The resulting compressed cycle time represents the minimum time required to operate the timetable. For infeasible timetables, the blocking-time conflicts are solved by re-expanding the timetable until all initial overlaps are eliminated. Operability is quantified by comparing the additional cycle time required to solve conflicts with the available slack within the reference cycle period.

4.3 Case Study Results

The results of the case study and the comparison of the temporary timetables are presented. First, the results of the timetable evaluation for each performance index are presented. This provides insight into the mechanisms underlying the overall effectiveness score. Secondly, the indices are aggregated into the TTE_r indicator to allow for comparison between temporary timetable variants.

4.3.1 Timetable Assessment per Index

Capacity Index

Table 1 summarizes the capacity-related metrics. AUP-C preserves the largest share of nominal capacity, followed by AUP-A and then AUP-B. This is reflected in higher values of $C_{b,r}^{pres}$. The traffic heterogeneity for AUP-B is slightly lower than for AUP-A and AUP-C. Temporary schedule AUP-C maintains the same level of heterogeneity as the nominal timetable. AUP-C maintains more traffic through the corridor, despite the higher heterogeneity of the traffic compared to AUP-A and AUP-B. The combined capacity index $C_{b,r}^{mix}$ has the highest value for AUP-C, followed by AUP-A and AUP-B.

Table 1: Capacity-related metrics for nominal and alternative temporary timetables.

Metric	Nominal	AUP-A	AUP-B	AUP-C
N_b^{nom}	28	-	-	-
$N_{b,r}^{tmp}$	-	26	24	28
$C_{b,r}^{pres}$	1.000	0.929	0.857	1.000
$H_{b,r}$	0.980	0.959	0.917	0.980
$C_{b,r}^{mix}$	0.980	0.890	0.786	0.980

Stability Index

The stability behavior for the three deterministic disturbance experiments is illustrated in Figure 2 - Figure 5. The output delay reflects the ability to absorb the initial disturbances over timetable cycles.

For AUP-A higher output delays are present for more concentrated disturbances, as the output delays for S1 are larger than for S2 as for S3. The input delay does not generate persistent delay propagation. AUP-B shows a more varying absorption pattern. For S1 and S2 the output delay is absorbed immediately, while experiment S3 experiences delay propagation. This is visible by the increase in output delay over timetable cycles 1 to 2. AUP-B absorbs the input delay within two timetable cycles for all experiments, indicating stability. AUP-C experiences substantial output delays that persist over multiple timetable cycles. For S2 and S3 the output delay initially decreases after the first timetable cycle, but increases again in later cycles. This reflects delay propagation. Timetable AUP-C shows limited ability to absorb disturbances.

Timetable AUP-B shows lower output delays compared to AUP-A and AUP-C. Across most intermediate output delays, AUP-B is better able to absorb disturbances. Both timetables AUP-A and AUP-B prevent long-term delay propagation, but AUP-B shows a more stable short-term response that is less sensitive to deterministic shocks. Therefore, AUP-B has a higher stability score of $S_{AUPB} = 0.785$, compared to $S_{AUPA} = 0.736$. In this research, a timetable is considered stable if deterministic input delay is absorbed within two timetable cycles (Pachl (2014)). Timetable AUP-C does not absorb the deterministic input delay within this horizon for all experiments. Even when allowing for more timetable cycles, the output delay of AUP-C is not absorbed. Therefore, AUP-C fails to satisfy the hard constraint and is assigned a stability score of $S_{AUPC} = 0$.

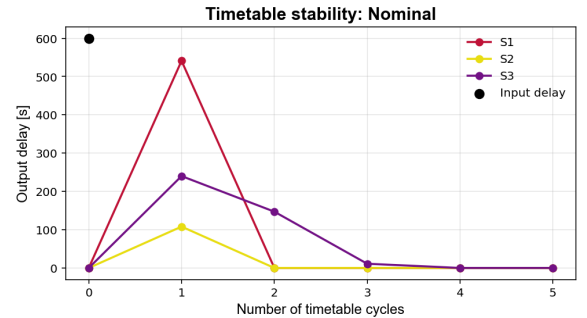


Figure 2: Output delay dynamics under deterministic disturbance scenarios for the nominal timetable.

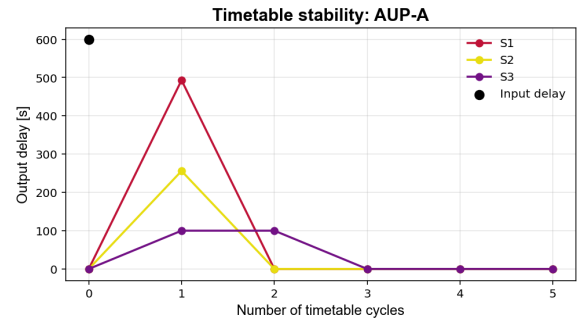


Figure 3: Output delay dynamics under deterministic disturbance scenarios for temporary timetable AUP-A.

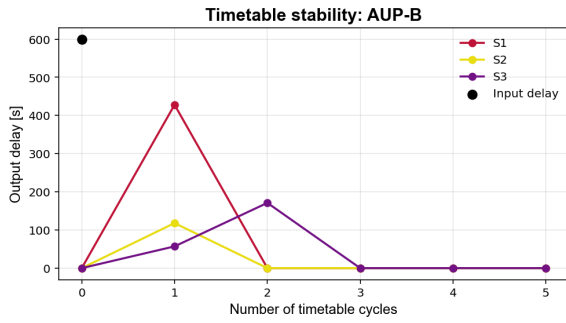


Figure 4: Output delay dynamics under deterministic disturbance scenarios for temporary timetable AUP-B.

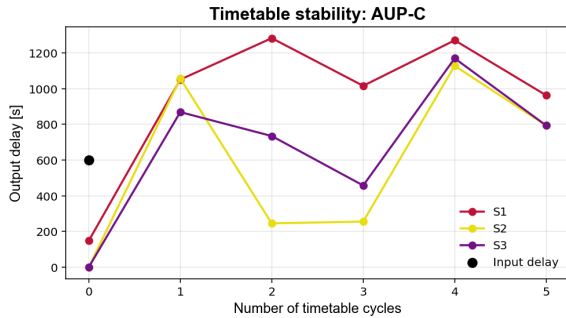


Figure 5: Output delay dynamics under deterministic disturbance scenarios for temporary timetable AUP-C.

Robustness Index

Figure 6 shows the robustness scores of 25 replications. Each data point represents the fraction of train events within a replication that remain below the predefined delay tolerance threshold of $\lambda = 30$ seconds. As train activities are planned at tenths of minutes (i.e., 6 seconds), a tolerance larger than this resolution is required to exclude deviations that arise from measurement accuracy.

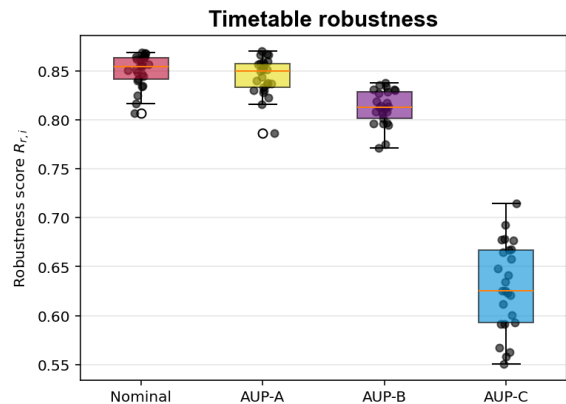


Figure 6: Box plot of robustness scores under stochastic disturbance scenario for timetables nominal, AUP-A, AUP-B, and AUP-C.

AUP-A consistently attains higher robustness scores $R_{r,i}$ than AUP-B and AUP-C, as a larger fraction of activities remain within the allowed tolerance under stochastic disturbances. The interquartile range of AUP-A is narrower than that of AUP-B and AUP-C.

On average, the robustness for AUP-A is higher and more consistent across different replications i than that of AUP-B. AUP-B in turn outperforms AUP-C in both level and consistency. The lower-tail of the interquartile range is lower for AUP-B and AUP-C than for AUP-A. The presence of these lower outliers for AUP-B and AUP-C reflect higher sensitivity to stochastic delays. Thus, AUP-A is most robust to continuous stochastic variability. Although AUP-A and AUP-B prevent extreme performance degradation, AUP-A more effectively absorbs day-to-day operational variation, resulting in higher overall robustness. In contrast, AUP-C exhibits lower robustness performance compared to the other temporary timetables with higher sensitivity to stochastic disturbances. The resulting overall robustness scores are $R_{AUPA} = 0.844$, $R_{AUPB} = 0.812$, and $R_{AUPC} = 0.630$.

Operability Index

The operability index is derived from timetable compression and blocking-time conflict resolution. The time-distance diagrams (TWDs) of AUP-A, AUP-B, and AUP-C for both directions of the corridor (see Figure B1 - Figure B6 in Section B) illustrate how train paths are distributed before and after timetable compression and blocking-time conflict resolution. The original timetable depicts scheduled train paths and may contain blocking-time overlaps. Figure 7 shows an original timetable, where slack between two trains is indicated by the red circle.

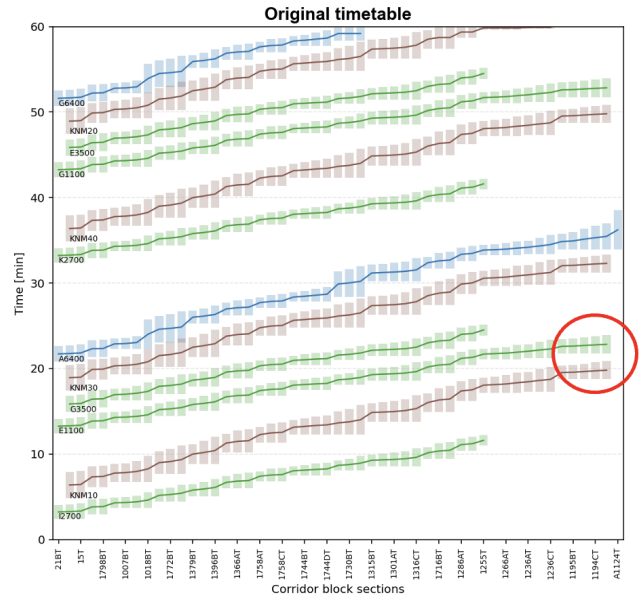


Figure 7: Time-distance diagram of an original timetable.

Note: Red circle indicates slack.

The compressed timetable is visualized in Figure 8. This schedule is obtained using the UIC 406 timetable compression method (International Union of Railways (UIC) (2004)). The red circled section shows how the slack between two trains is removed. The minimum headway time has been applied so that the blocking-times are critical and touch on at least one section. In this compression, any possible conflicts are not resolved. The blocking-times circled in black indicate an overlap. This compressed timetable shows the minimum cycle time in which the planned timetable can be executed. This infeasible cycle time represents the compressed timetable duration with minimum headway constraints, allowing for conflicts.

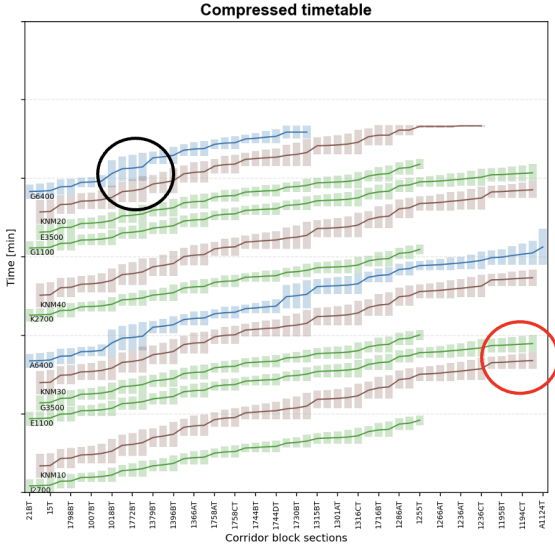


Figure 8: Time-distance diagram of a compressed timetable.

Note: Red circle indicates removed slack. Black circle indicates a conflict.

Lastly, the compressed feasible timetable illustrates the minimum cycle time of a conflict-free (feasible) timetable. The compressed timetable is re-expanded by applying the total time shift required to resolve all conflicts. This results in feasible cycle times which are compared against the reference time period to assess operability. The black circled blocking times in Figure 9 indicate how blocking times are shifted to resolve the conflict. Here, the smallest necessary time extension, the maximum blocking-time overlap between conflicting train pairs within the corridor, is determined. Every train in the compressed timetable is shifted with this time extension to resolve all blocking-time conflicts.

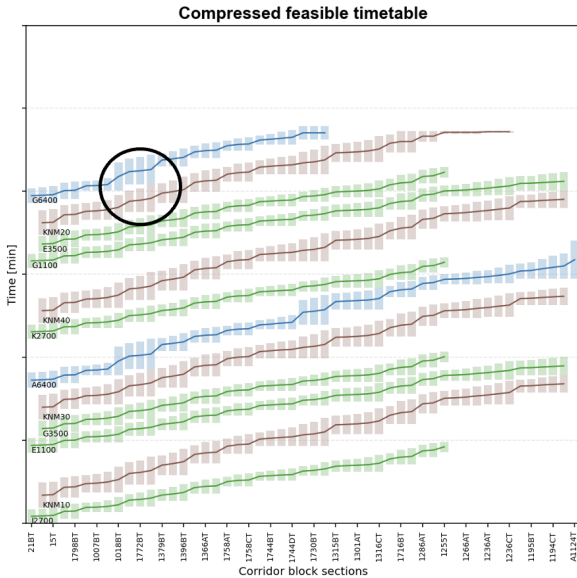


Figure 9: Time-distance diagram of a compressed and feasible timetable.

Note: Black circle indicates a resolved conflict.

The operability is assessed using the infeasible cycle time, the total time shift required to resolve all conflicts, and the remaining slack. The decomposition is illustrated in Figure 10 and Figure 11. Operability is evaluated separately for both directions, as structural conflicts and minimum feasible cycle times differ. For all timetables, the infeasible cycle time constitutes the largest share of the reference period. For the direction Btl-Ehv, the infeasible cycle times of the nominal timetable, AUP-A and AUP-B are similar. Although the nominal schedule and AUP-C have the same scheduled number of trains, AUP-C exhibits a larger infeasible cycle time. This can be attributed to the fact that AUP-C operates on two tracks, while the nominal timetable operates on four tracks. For this direction, AUP-A is conflict-free, while AUP-B contains minor conflicts. All timetables are executable within the reference period. For the direction Ehv-Btl, larger deviations in the infeasible cycles times are observed and there are more pronounced conflicts. The nominal timetable has a low infeasible cycle time, indicating a substantial amount of slack. AUP-B has a lower infeasible cycle time than AUP-A, while the total amount of retiming to make the timetable conflict-free is identical. AUP-C has many conflicts, such that the total time shift required to resolve all blocking-time conflicts exceeds the operational slack. In contrast, AUP-A and AUP-B are operable in one hour.

Operability decomposition — Boxtel → Eindhoven

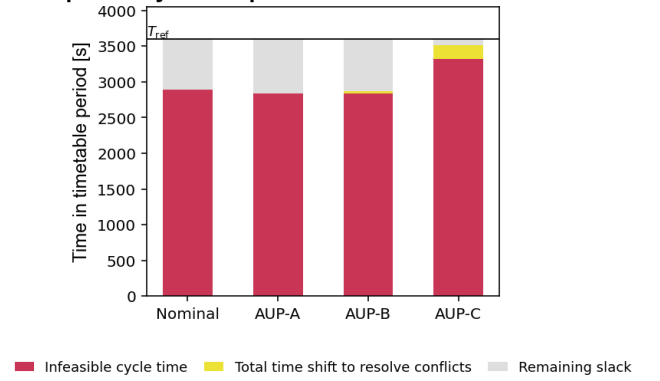


Figure 10: Decomposition of operability into infeasible cycle time, total time shift to solve conflicts, and remaining slack for Baxtel-Eindhoven for timetables nominal, AUP-A, AUP-B, and AUP-C.

Operability decomposition — Eindhoven → Baxtel

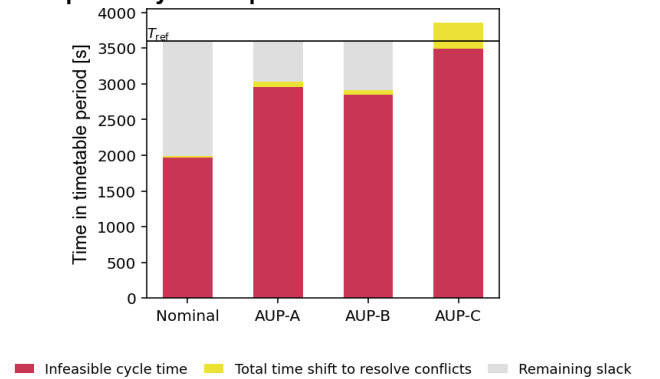


Figure 11: Decomposition of operability into infeasible cycle time, total time shift to solve conflicts, and remaining slack for Eindhoven-Baxtel for timetables nominal, AUP-A, AUP-B, and AUP-C.

Table 2 shows the main operability results. Although AUP-A and AUP-B are structurally executable for both directions, the operability for AUP-B is slightly higher. AUP-C has the least remaining slack in the BE-direction and is not executable within the reference period for the EB-direction. The operability at the corridor-level is defined as the minimum operability in both directions. This reflects the fact that executability is defined as a lower bound. Based on this conservative aggregation, AUP-B attains the highest corridor-level operability score of $O_{AUPB} = 0.908$ compared to the other temporary timetables with operability scores of $O_{AUPA} = 0.892$ and $O_{AUPC} = 0$.

Table 2: Operability-related metrics for nominal and alternative temporary timetables.

Metric	Nominal	AUP-A	AUP-B	AUP-C
Direction Boxtel → Eindhoven				
$T_{BE}^{cycle, infeasible}$	2885.01	2834.27	2834.54	3317.49
$T_{BE}^{conflict}$	0.000	0.000	38.06	197.99
O_{BE}	1.000	1.000	0.950	0.299
Direction Eindhoven → Boxtel				
$T_{EB}^{cycle, infeasible}$	1971.30	2957.51	2846.75	3489.15
$T_{EB}^{conflict}$	14.50	69.23	69.23	371.14
O_{EB}	0.991	0.892	0.908	0.000
Minimum corridor operability				
$\min(O_{b,r})$	0.991	0.892	0.908	0.000

4.3.2 Timetable Effectiveness Results

The integrated timetable effectiveness indicator (TTE_r) combines the four performance dimensions (i.e., capacity utilization, stability, robustness, and operability) into a single composite score. The objective of this aggregation is to provide a structured and transparent synthesis that supports the comparison of alternative temporary timetables under constrained conditions.

Table 3 summarizes the effectiveness results for all variants of the timetable for different weight compositions. These sets are considered to reflect alternative operational priorities that may arise during PTCs. Since no weighting scheme is optimal, multiple compositions are evaluated to illustrate the sensitivity of TTE_r to varying weights. The operability-constrained reference reflects a perspective in which operability is considered a prerequisite for assessing timetable performance, while the remaining performance indices are weighted in a relatively balanced manner. An operational reliability-focused composition assigns a higher importance to stability and robustness, often referred to as a passenger-oriented set. A capacity-preservation focused composition emphasizes maintaining throughput. Finally, a conservation and risk-averse weight set assigns a dominant weight to operability, representing contexts where executability and conflict avoidance are critical.

The results illustrate trade-offs between individual timetable attributes. A strength in one timetable attribute is often accompanied by weaknesses in others when the timetable is evaluated as a whole. The integrated TTE_r indicator therefore makes these trade-offs explicit and allows alternative timetable designs to be compared on a quantitative scale. AUP-A achieves higher overall effectiveness scores than AUP-B and AUP-C, in all weight compositions. AUP-B performs comparably well and attains TTE_r scores close to those of AUP-A. AUP-C consistently under-performs when all indices are considered simultaneously. Important to emphasize is that these results are conditional on the selected weight compositions and reflect

plausible varying stakeholder preferences. When extreme weight sets are considered, such that a single index receives full weight while the other are assigned a zero-weight, the ranking of the timetable alternatives may change. For example, under a purely capacity-driven evaluation, AUP-C becomes the preferred alternative due to its high capacity index. Under strict stability- or operability-focused evaluations, AUP-B performs best, while in a solely robustness-driven evaluation, AUP-A performs best.

Table 3: Integrated timetable effectiveness results under different weight compositions.

Metric	Nominal	AUP-A	AUP-B	AUP-C
$C_{b,r}^{mix}$	0.980	0.890	0.786	0.980
S_r	0.486	0.736	0.785	0.000
R_r	0.850	0.844	0.812	0.630
$O_{b,r}$	0.991	0.892	0.908	0.000
Operability-constrained reference: $\alpha=0.30, \beta=0.15, \gamma=0.15, \epsilon=0.40$				
TTE_r	0.891	0.861	0.839	0.388
Operational reliability-focused: $\alpha=0.20, \beta=0.30, \gamma=0.30, \epsilon=0.20$				
TTE_r	0.795	0.831	0.818	0.385
Capacity preservation-focused: $\alpha=0.45, \beta=0.15, \gamma=0.10, \epsilon=0.30$				
TTE_r	0.896	0.863	0.825	0.504
Conservative and risk-averse: $\alpha=0.20, \beta=0.15, \gamma=0.15, \epsilon=0.50$				
TTE_r	0.892	0.861	0.851	0.290

Note: Green values indicate the highest value over temporary timetable alternatives AUP-A, AUP-B, and AUP-C.

4.3.3 Sensitivity Analysis and Stress Testing

To test the sensitivity of the robustness index to the value of the tolerance margin, a sensitivity analysis with respect to λ is performed. Larger values of λ lead to higher R_r values, as a greater share of activities remain within the tolerance margin. The marginal gains are greatest for small values of λ . The average sum of differences between the robustness score of the nominal timetable and each temporary timetable ($(R_{Nominal} - R_{AUPA}) + (R_{Nominal} - R_{AUPB}) + (R_{Nominal} - R_{AUPC})$) is visualized in Figure 12. This average difference in robustness scores ΔR_r decreases when increasing the tolerance margin λ . This relationship is relatively linear, indicating that increasing the tolerance margin reduces the differences in robustness scores between the nominal and temporary timetables but does not alter the comparative assessment between the timetable alternatives.

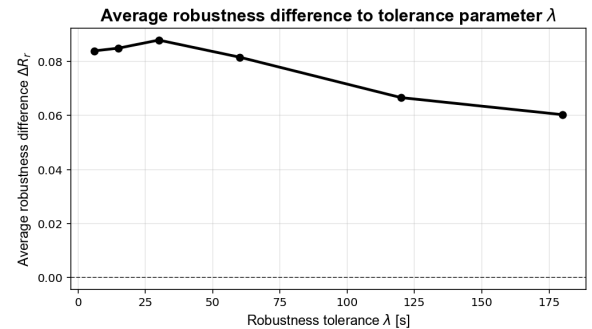


Figure 12: Sum of differences of R_r between nominal and temporary timetables AUP-A, AUP-B, and AUP-C to λ .

The behavior of the TTE_r indicator under increasing operational pressure is evaluated with a stress test. The timetable is progressively compressed by reducing the reference period and proportionally shortening planned running times, while keeping dwell times and other process times unchanged. The cumulative impact of increasing operational pressure is summarized using the Area Under the Effectiveness Curve ($AUEC_r = \int_{T_{min}}^{T_{nom}} TTE_r(T_c) dT_c$). Higher $AUEC_r$ values indicate better preservation of effectiveness across compression levels. Figure 13 shows the timetable effectiveness as the reference period is reduced. For all timetables, TTE_r decreases with increasing compression, reflecting the gradual loss of operational slack. For moderate compression levels, the nominal timetable and temporary schedules AUP-A and AUP-B maintain relatively high effectiveness. Beyond a critical compression threshold, effectiveness decreases and converges to lower bounds. This indicates that timetables become structurally infeasible and operationally unrealistic, resulting in delay accumulation and simulation deadlocks. The cumulative effect is summarized by the $AUEC_r$ values. AUP-A has a slightly higher score ($AUEC_{AUPA} = 3.627$) than AUP-B ($AUEC_{AUPB} = 3.543$), indicating that AUP-A preserves timetable effectiveness marginally better across the tested compression range. The nominal timetable has the highest score with $AUEC_{Nominal} = 4.157$. AUP-C has the lowest score with $AUEC_{AUPC} = 1.729$, reflecting its limited ability to maintain effectiveness under increasing operational pressure.

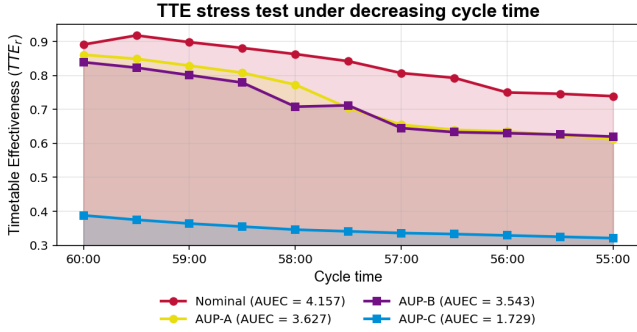


Figure 13: Timetable effectiveness stress test for nominal, AUP-A, AUP-B, and AUP-C timetables.

5 DISCUSSION

This paper examines whether temporary railway timetables, operating under constrained infrastructure conditions, can be assessed and compared in a systematic manner. By translating timetable characteristics and simulated operational performance into a consistent set of indices, the case study demonstrates that the framework results in numerical outcomes that reflect the capacity utilization, stability, robustness, and operability of temporary timetables. The framework makes trade-offs between these indices explicit and aggregates them into a data-driven multi-criteria performance quality indicator of temporary timetable effectiveness (TTE_r), together with a formal definition. The case study shows that both marginal and substantial differences can be captured between alternative schedule designs. Reformulating existing timetable attributes and defining a new attribute leads to a generic methodological approach that enables timetable effectiveness quantification. In this way, the framework provides a basis for the evaluation of temporary schedules as a distinct research domain within railway timetabling.

A key contribution of this study is the introduction of operability as a new timetable attribute. Where feasibility is traditionally defined as the absence of blocking-time conflicts, the results illustrate that temporary timetables may remain operationally executable, despite minor conflicts. The operability index quantifies the total time shift required to resolve blocking-time conflicts, relative to the available operational slack within a reference period. This approach extends beyond a binary feasible-infeasible classification and reflects how temporary timetables are used in practice.

The framework is demonstrated using temporary timetables operating under constrained conditions. However, the methodological applicability is not restricted to this specific context. The conceptual formulation of timetable effectiveness and the underlying performance indices can be adapted to reflect other relevant operational contexts, such as for nominal timetables. Consequently, the proposed framework should be interpreted as an evaluation methodology, rather than a case-specific assessment tool. Practical implementation requires context-specific adjustments, while the underlying mechanisms for quantification and aggregation of timetable performance remain unchanged. The framework proves a generalizable basis for timetable evaluation across different planning contexts, with temporary timetables representing a particularly relevant and previously underexplored application domain.

An important consideration concerns the assignment of weights in the integrated TTE_r indicator. The case study demonstrates that different weight compositions can lead to different absolute and relative timetable effectiveness scores. The objective of the framework is not to identify a single optimal timetable or optimal set of weights, but to enable quantification and comparison of alternative timetable designs. To support such a comparison, it is necessary to assign weights to reflect how stakeholders value trade-offs between capacity, stability, robustness, and operability. From a methodological perspective, the formulation of timetable effectiveness as a weighted sum is a key strength of the proposed method. However, translating stakeholder priorities into numerically weighed variables proved challenging in practice.

6 CONCLUSIONS & FUTURE RESEARCH

6.1 Main Conclusions

In this paper, the question of how effectiveness of temporary railway timetables can be systematically quantified in the context of PTCs is addressed. A formal definition of effectiveness is defined as the ability of a temporary timetable operating under constrained conditions to maintain operable, stable, and robust train operations under deterministic and stochastic operational variability. This study demonstrates that effectiveness can be quantified by combining the performance indices capacity utilization, stability, robustness, and operability into an evaluation framework. By reformulating existing timetable attributes and introducing operability as a new attribute, this paper provides a methodological approach that enables temporary timetables to be quantified and compared in a systematic manner. The aggregation of indices into a data-driven multi-criteria performance quality indicator of temporary timetable effectiveness (TTE_r) reflects both the trade-offs between the underlying indices and the overall effectiveness of each timetable alternative. In this way, the method offers a structured decision-support tool for planning practice, supporting transparent comparison of timetable alternatives and systematic assessment of expected performance. Although the

framework is developed and demonstrated for temporary schedules, the underlying concept of timetable effectiveness and the proposed methodology of indices and an integrated indicator are not limited to PTC contexts. With adaptation of the definitions and modeling assumptions to the specific operational conditions, the framework can be applied to railway timetables more generally.

6.2 Future Research Directions

Future research should focus on three directions. First, the case study results demonstrate the analytical potential of temporary timetable quantification. However, empirical research should explore the effects and applicability of the proposed framework on real-world decision-making. Second, further research is needed to develop systematic approaches for extracting and translating stakeholder preferences into numerical weights. Third, the framework should be extended and tested on larger network segments and other operational conditions, to assess its scalability and generalizability to network-scales and different contexts.

ACKNOWLEDGEMENTS

This research is supported by ProRail, the infrastructure manager in the Netherlands, by providing the required data.

DATA AVAILABILITY

The author does not have permission to share data.

REFERENCES

- Andersson, E., Peterson, A., and Krasemann, J. T. (2013a). Quantifying railway timetable robustness in critical points. *Journal of Rail Transport Planning Management*, 3(3):95–110. Robust Rescheduling and Capacity Use.
- Andersson, E., Peterson, A., and Krasemann, J. T. (2013b). Robustness in Swedish railway traffic timetables. *Proceedings of the 5th International Seminar on Railway Operations Modelling and Analysis (RailCopenhagen)*, pages 1–20.
- Besinovic, N. and Goverde, R. M. P. (2018). *Capacity Assessment in Railway Networks*, pages 25–45. Springer International Publishing.
- Cacchiani, V. and Toth, P. (2012). Nominal and robust train timetabling problems. *European Journal of Operational Research*, 219(3):727–737. Feature Clusters.
- CORDIS: Publications Office of the European Union (n.d.). Final report summary - on-time (optimal networks for train integration management across Europe). <https://cordis.europa.eu/project/id/285243/reporting>. Accessed: 2025-09-12.
- Correia Duarte, P. J., Dauzere-Peres, S., Huisman, D., Mannino, C., Sartor, G., Weik, N., and Widmann, P. (2025). 50 years of or in railway timetabling and rolling stock planning. *EURO Journal on Transportation and Logistics*, 14:100155.
- Dewilde, T., Sels, P., Cattrysse, D., and Vansteenwegen, P. (2014). Improving the robustness in railway station areas. *European Journal of Operational Research*, 235(1):276–286.
- Goverde, R. M. (2007). Railway timetable stability analysis using max-plus system theory. *Transportation Research Part B: Methodological*, 41(2):179–201. Advanced Modelling of Train Operations in Stations and Networks.
- Goverde, R. M. (2010). A delay propagation algorithm for large-scale railway traffic networks. *Transportation Research Part C: Emerging Technologies*, 18(3):269–287. 11th IFAC Symposium: The Role of Control.
- Goverde, R. M., Corman, F., and D’Ariano, A. (2013). Railway line capacity consumption of different railway signalling systems under scheduled and disturbed conditions. *Journal of Rail Transport Planning Management*, 3(3):78–94. Robust Rescheduling and Capacity Use.
- Goverde, R. M. and Hansen, I. A. (2013). Performance indicators for railway timetables. In *2013 IEEE International Conference on Intelligent Rail Transportation Proceedings*, pages 301–306.
- Hansen, I. and Pachl, J., editors (2008). *Railway Timetable & Traffic: Analysis, Modelling, Simulation*. Eurailpress, Hamburg.
- International Union of Railways (UIC) (2004). Uic leaflet 406: Capacity.
- Jensen, L. W., Landex, A., and Nielsen, O. A. (2014). Evaluation of robustness indicators using railway operation simulation. In Brebbia, C. A., Tomii, N., Tzieropoulos, P., and Mera, J. M., editors, *Computers in Railways XIV*, pages 329–339, Southampton. WIT Press.
- Jensen, L. W., Schmidt, M., and Nielsen, O. A. (2020). Determination of infrastructure capacity in railway networks without the need for a fixed timetable. *Transportation Research Part C: Emerging Technologies*, 119:102751.
- Lee, Y., Lu, L.-S., Wu, M.-L., and Lin, D.-Y. (2017). Balance of efficiency and robustness in passenger railway timetables. *Transportation Research Part B: Methodological*, 97:142–156.
- Lindfeldt, O. (2008). Evaluation of punctuality on a heavily utilised railway line with mixed traffic. *WIT Transactions on The Built Environment*, 103:545–553.
- Lusby, R. M., Larsen, J., and Bull, S. H. (2018). A survey on robustness in railway planning. *European Journal of Operational Research*, 266(1):1–15.
- Pachl, J. (2014). *Railway Operation and Control*. VTD Rail Publishing, 2nd edition. Chapter 6: Stability and delay propagation.
- Polinder, G., Schmidt, M., and Huisman, D. (2021). Timetabling for strategic passenger railway planning. *Transportation Research Part B: Methodological*, 146:111–135.
- ProRail (2024). Jaarverslag 2024. Accessed: 6 October 2025.
- Salido, M., Barber, F., and Ingolotti, L. (2008). Robustness in railway transportation scheduling. pages 2880 – 2885.
- Salido, M., Barber, F., and Ingolotti, L. (2012). Robustness for a single railway line: Analytical and simulation methods. *Expert Systems with Applications*, 39(18):13305–13327.
- Scheepmaker, G. M. and Goverde, R. M. (2015). Running time supplements: Energy-efficient train control versus robust timetables. In *Proceedings of the 6th International Conference on Railway Operations Modelling and Analysis (RailTokyo 2015)*, Tokyo, Japan. Accessed: 2025-09-17.
- Solinen, E. (2025). *Improving railway timetable robustness: Development and application of robustness indicators*. PhD thesis, Linköping University.
- Wolniewicz, L. (2024). Railway transport system modelling approach for robustness analysis. *Transport*, 39(4):287–301.

APPENDIX A: MATHEMATICAL NOTATION

All symbols listed in the equations are summarized in Table A1. A distinction is made between sets, parameters, and variables.

Table A1: Mathematical notation for the TTE framework

Symbol	Description	Domain
Sets and indices		
\mathcal{B}	Set of corridors	$b \in \mathcal{B}$
\mathcal{R}	Set of temporary timetables	$r \in \mathcal{R}$
\mathcal{M}	Set of train types	$m \in \mathcal{M}$
\mathcal{K}	Set of trains	$k \in \mathcal{K}$
\mathcal{P}	Set of timetable points	$p \in \mathcal{P}$
\mathcal{D}	Set of disturbance events	$d \in \mathcal{D}$
\mathcal{N}	Set of replications	$i \in \mathcal{N}$
$p^{exit}(k)$	Exit timetable point of train k	$p^{exit}(k) \subseteq \mathcal{P}$
\mathcal{A}	Set of activities	$a \in \mathcal{A}$
\mathcal{C}	Set of blocking-time conflicts	$c \in \mathcal{C}$
$k_{f(c)}$	Following train associated with blocking-time conflict c	$k_{f(c)} \in \mathcal{K}$
Parameters		
N_b^{nom}	Number of scheduled trains passing corridor b in nominal timetable	\mathbb{N}
N_b^{tmp}	Number of scheduled trains passing corridor b in temporary timetable r	\mathbb{N}
$H_{b,r}$	Heterogeneity index of traffic through corridor b of temporary timetable r	$[0, 1]$
$p_{m,b,r}$	Share of scheduled trains belonging to class m at corridor b in timetable r	$[0, 1]$
δ_d^{inj}	Injected deterministic delay of disturbance event d	$\mathbb{R}_{\geq 0}$
$t_a^{sched,(r)}$	Scheduled time of activity a in temporary timetable r	$\mathbb{R}_{\geq 0}$
λ	Allowed schedule deviation (tolerance)	$\mathbb{R}_{\geq 0}$
$T_{cycle,ref}$	Reference operational time period	$\mathbb{R}_{> 0}$
T^{inj}	Time block in which input disturbance is injected	\mathbb{N}
T^{abs}	First time block after which no output delay remains	\mathbb{N}
$\tau_{c_{b,r}}$	Blocking-time overlap duration of conflict $c_{b,r}$	$\mathbb{R}_{> 0}$
$\alpha, \beta, \gamma, \epsilon$	Weights for component indices in TTE_r	$[0, 1]$
T_c	Compressed reference period used in stress testing	$\mathbb{R}_{> 0}$
T_{min}	Minimum reference period considered in stress testing	$\mathbb{R}_{> 0}$
T_{nom}	Nominal reference period of the timetable	$\mathbb{R}_{> 0}$
Variables		
$C_{b,r}^{pres}$	Preserved capacity ratio at corridor b in temporary timetable r	$[0, 1]$
$C_{b,r}^{mix}$	Capacity index for corridor b in temporary timetable r	$[0, 1]$
$\delta_{k,p}^{disturb,(r)}$	Realized delay of train k at timetable point p in disturbed run per temporary timetable r	$\mathbb{R}_{\geq 0}$
$\delta_{k,p}^{base,(r)}$	Realized delay of train k at timetable point p in baseline run per temporary timetable r	$\mathbb{R}_{\geq 0}$
$\delta_{k,p}^{add,(r)}$	Additional delay of train k at timetable point p per temporary timetable r	$\mathbb{R}_{\geq 0}$
$\delta_k^{add,exit,(r)}$	Additional delay of train k at exit point p^{exit} per temporary timetable r	$\mathbb{R}_{\geq 0}$
$\Delta^{in,(r)}$	Total injected delay per temporary timetable r	$\mathbb{R}_{\geq 0}$
$\Delta^{out,(r)}$	Total output delay at exit points per temporary timetable r	$\mathbb{R}_{\geq 0}$
S_r	Stability index of temporary timetable r	$[0, 1]$
$t_a^{real,(r,i)}$	Realized activity time of train k at timetable point p per temporary timetable r per replication i	$\mathbb{R}_{\geq 0}$
$q_a^{(r,i)}$	Binary variable: 1 if $t_a^{real,(r,i)} \leq t_a^{sched,(r)} + \lambda$	$\{0, 1\}$
$R_{r,i}$	Robustness index of temporary timetable r per replication i	$[0, 1]$
R_r	Robustness index of temporary timetable r	$[0, 1]$
$T_{b,r}^{cycle,feasible}$	Cycle time for a feasible temporary timetable r of corridor b	$\mathbb{R}_{> 0}$
$T_{b,r}^{cycle,infeasible}$	Cycle time for an infeasible temporary timetable r of corridor b	$\mathbb{R}_{> 0}$
$\Delta T_{k,b,r}^{accum}$	Accumulated minimum time shift required to resolve all conflicts for train k in timetable r for corridor b	$\mathbb{R}_{\geq 0}$
$\Delta T_{b,r}^{conflict}$	Minimum additional time shift required to eliminate all conflicts in timetable r for corridor b	$\mathbb{R}_{\geq 0}$
$O_{b,r}$	Operability index of temporary timetable r for corridor b	$[0, 1]$
TTE_r	Integrated timetable effectiveness indicator of temporary timetable r	$[0, 1]$
$AUEC_r$	Area Under the Effectiveness Curve for temporary timetable r	$\mathbb{R}_{\geq 0}$

APPENDIX B: TIME-DISTANCE DIAGRAMS

This appendix presents the TWDs used to illustrate the timetable compression and conflict resolution steps underlying the operability analysis.

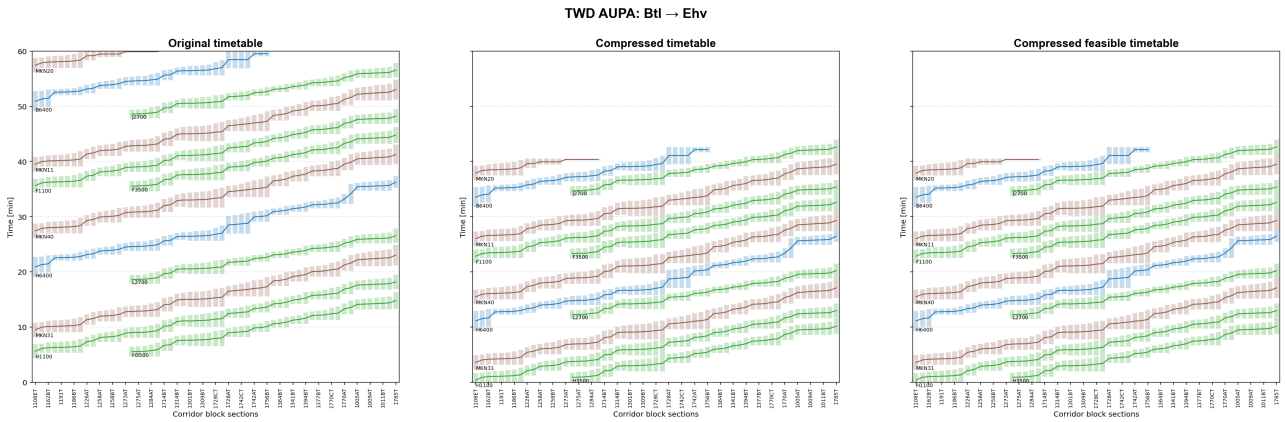


Figure B1: Time-distance diagram of the corridor Boxtel (Btl) - Eindhoven Centraal (Ehv) for temporary timetable AUP-A.

Note: Blue lines represent SPR-services, green lines represent IC-services and brown lines represent GDR-services.

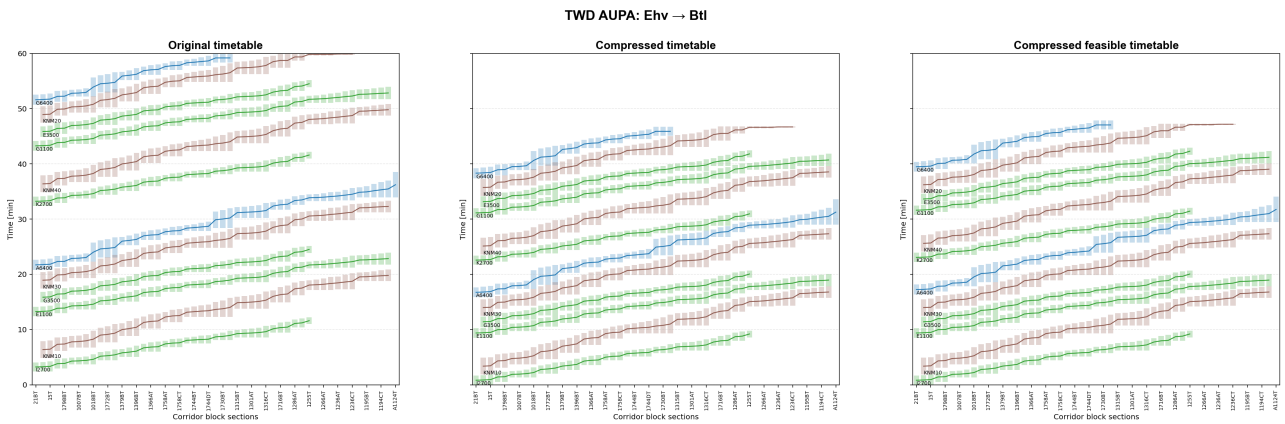


Figure B2: Time-distance diagram of the corridor Eindhoven Centraal (Ehv) - Boxtel (Btl) for temporary timetable AUP-A.

Note: Blue lines represent SPR-services, green lines represent IC-services and brown lines represent GDR-services.

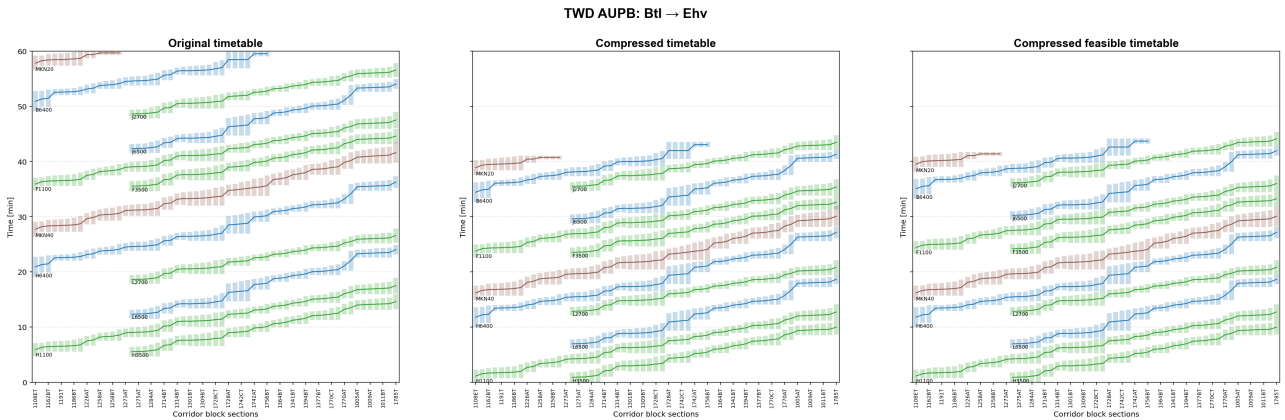


Figure B3: Time-distance diagram of the corridor Boxtel (Btl) - Eindhoven Centraal (Ehv) for temporary timetable AUP-B.

Note: Blue lines represent SPR-services, green lines represent IC-services and brown lines represent GDR-services.

TWD AUPB: Ehv → Btl

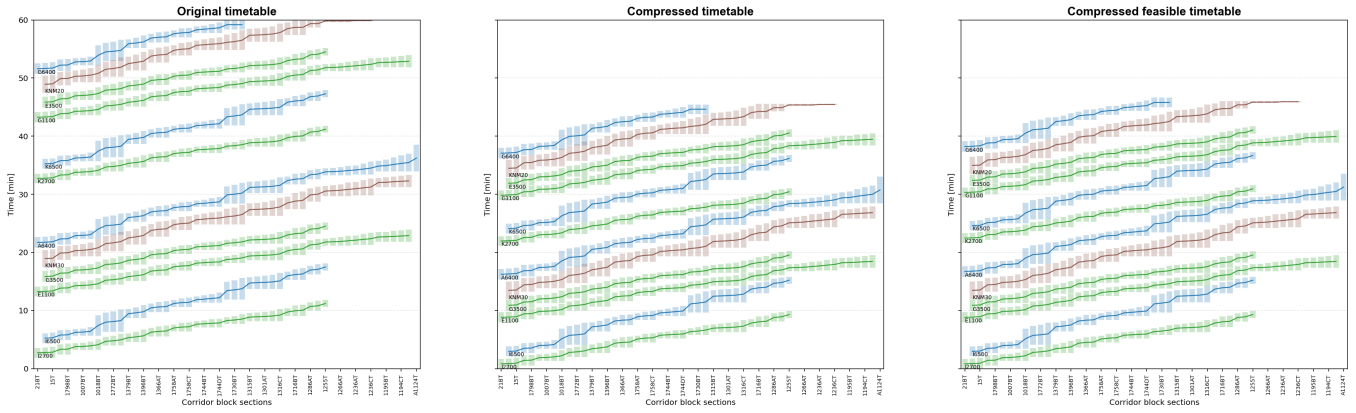


Figure B4: Time-distance diagram of the corridor Eindhoven Centraal (Ehv) - Boxtel (Btl) for temporary timetable AUP-B.

Note: Blue lines represent SPR-services, green lines represent IC-services and brown lines represent GDR-services.

TWD AUPC: Btl → Ehv

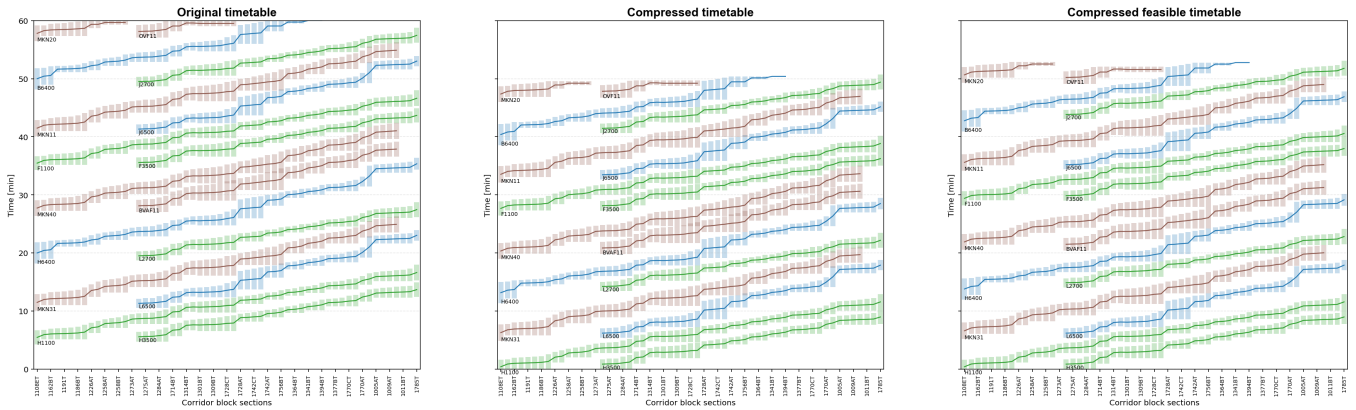


Figure B5: Time-distance diagram of the corridor Boxtel (Btl) - Eindhoven Centraal (Ehv) for temporary timetable AUP-C.

Note: Blue lines represent SPR-services, green lines represent IC-services and brown lines represent GDR-services.

TWD AUPB: Ehv → Btl

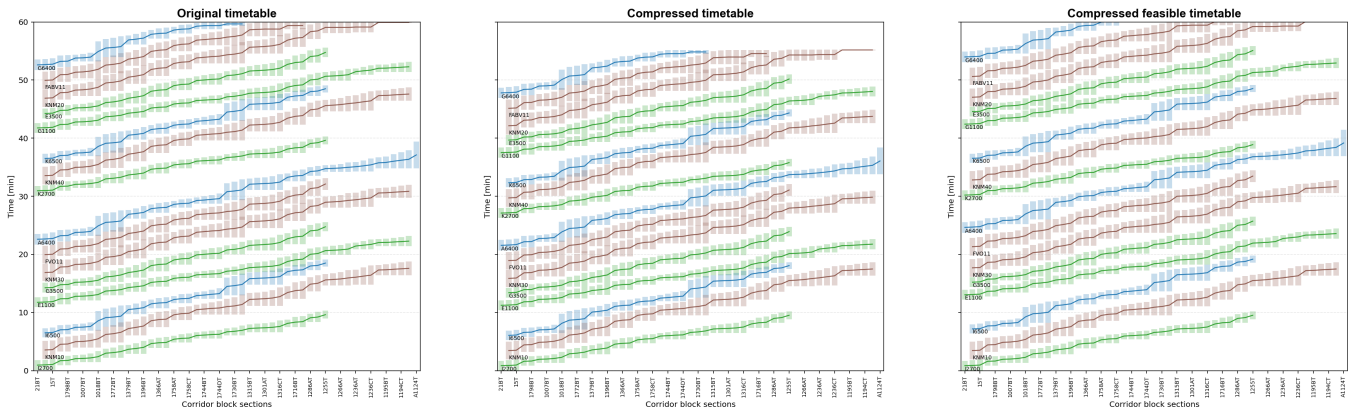


Figure B6: Time-distance diagram of the corridor Eindhoven Centraal (Ehv) - Boxtel (Btl) for temporary timetable AUP-B.

Note: Blue lines represent SPR-services, green lines represent IC-services and brown lines represent GDR-services.