Quantifying the operational benefits of a railway Driver Advisory System

Julia Michels

Delft University of Technology

TU Delft ProRail

Quantifying the operational benefits of a railway Driver Advisory System

by

Julia Michels

to obtain the degree of Master of Science in Transport, Infrastructure and Logistics
at the Delft University of Technology,
to be defended publicly on Monday November 24, 2025 at 14:00.

Student number:	4996569	
Project duration:	March 24, 2025 – November 24, 2025	
Thesis committee:	Prof. dr. ir. R.M.P. Goverde,	TU Delft, Committee chairman
	Ir. Z. Wang,	TU Delft, Daily Supervisor CiTG
	Dr. ir. J.A. Annema,	TU Delft, Daily Supervisor TPM
	Ir. J.J.W. van Luipen,	ProRail, Company supervisor

TU Delft ProRail

Acknowledgements

This thesis marks the completion of my Master's in Transport, Infrastructure and Logistics at TU Delft, carried out in collaboration with ProRail. The past years have been educational, inspiring, and memorable. I am deeply grateful to the university for giving me the opportunity to learn and grow throughout my studies.

Over the past months, I have had the pleasure of working under the daily supervision of Ziyulong Wang. He supported me throughout every stage of this thesis and continuously inspired me with his knowledge and ideas. After every meeting, I left feeling motivated, confident, and full of new perspectives. His openness and reliability have been invaluable to me. I would also like to express my sincere appreciation to Rob Goverde and Jan Anne Annema for their constructive feedback and guidance during the process, and to Jelle van Luipen for giving me the opportunity to collaborate with ProRail.

Finally, I would like to thank my family and friends for their constant support and encouragement. Their love has meant the world and kept me going through every stage of this journey.

Summary

Background and Motivation

Freight transport is an essential component of Europe's rail network, yet it often faces capacity constraints and reliability challenges due to mixed traffic with frequent passenger services. In the Netherlands, these pressures are particularly evident on key freight corridors such as the Brabantroute, where freight trains share limited infrastructure with dense regional traffic. To maintain punctuality and efficiency without expensive infrastructure upgrades, the rail sector has increasingly turned to Driver Advisory Systems (DAS). These systems provide real-time or pre-computed driving advice, helping train drivers optimise speed and timing in response to operational conditions. Despite their promise, most studies on DAS performance have relied on simulation or laboratory environments, providing limited insight into their real-world impact on train operations. The RouteLint (RTL) system, developed by ProRail, is a connected DAS (C-DAS) that delivers timekeeping and situational information directly to freight drivers. Since late 2024, RouteLint has been made freely available to freight operators as part of the Betuweroute maintenance period, creating a unique opportunity to assess its effectiveness under real operational conditions. This thesis bridges the gap between theoretical research and real-world railway performance by empirically evaluating how RouteLint influences freight train operations on the Brabantroute corridor.

Therefore main research question of this study is:

How can the operational performance of freight trains be evaluated using empirical data from Driver Advisory Systems such as RouteLint?

To main question is divided into five sub-questions: What is the current state-of-the-art and state-of-practice of DAS? Which methods and Key Performance Indicators (KPIs) can quantify DAS effects in real-world freight operations? How do operations differ between trains with and without RouteLint? How do these impacts vary across operational scales (macro vs. micro) and over space and time? What policy implications arise from the findings?

Data and Methodology

The analysis focused on the Brabantroute between Eindhoven and Tilburg, a 68 km corridor serving as the main diversionary route for the Betuweroute closure. The study period covered August 2024 – July 2025, encompassing the first year of increased freight intensity, up to 120 above normal levels.

Figure 1 shows an overview of the methodology.

Three datasets formed the analytical foundation:

- Verkeersleiding Ondersteunend Systeem (VOS), containing timetable and passing-point records.
- Trein Observatie en Tracking Systeem (TROTS), providing event times and aspects of signals.
- RouteLint usage logs, identifying whether each train used the system.

After cleaning and matching, 23 573 individual freight trains were included in the final dataset. Four Key Performance Indicators (KPIs) were defined to capture different aspects of train operation:

- Transport volume
- Deviation from timetable
- Unplanned yellow signals
- Train interaction

Operational data were mapped onto two aggregated event graphs: a macroscopic graph connecting timetable points, and a microscopic graph linking signal-level events. To compare trains with and without RTL, the continuous KPIs (deviations from the planned timetable and average running times)

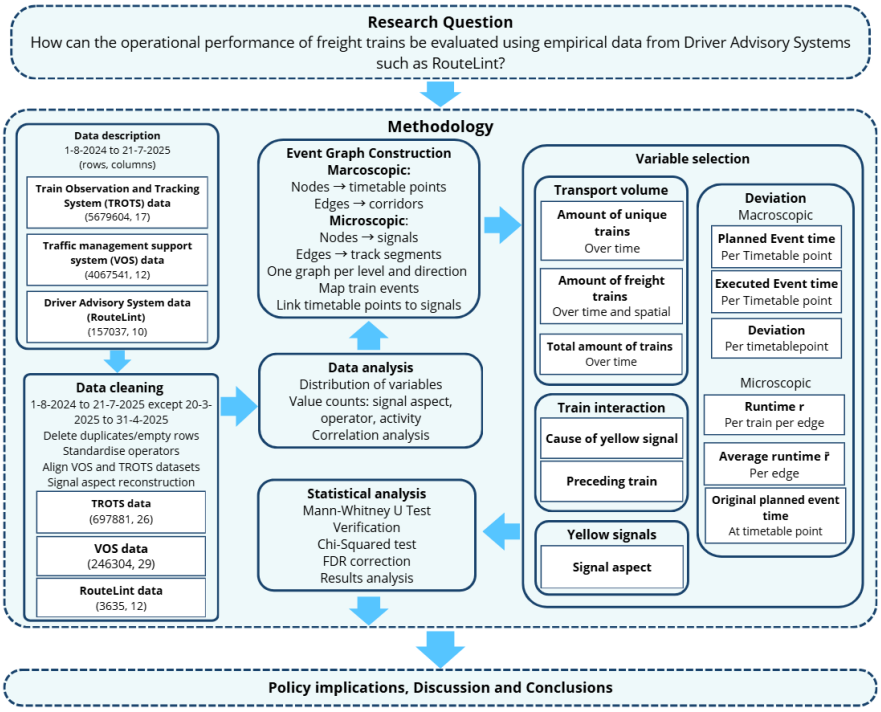


Figure 1: Flow diagram of the methodology.

were tested using the Mann–Whitney U non-parametric test, while categorical outcomes (yellow signal passages and train interactions) were analysed using Chi-squared statistics. Given the difficulty of guaranteeing complete independence between repeated drivers, random subsampling (30 % of data) was applied as a robustness check. To ensure reliable significance levels across multiple tests, all p-values were corrected using the Benjamini–Hochberg False Discovery Rate (FDR) procedure.

Results

At the timetable-point scale (macroscopic level), RouteLint users consistently ran closer to the planned schedule than non-users. Average deviations at most stations were lower for RTL-equipped trains, with the largest difference of approximately 66 seconds at Tilburg Industrie, and typical differences of 10–30 seconds around major nodes such as Eindhoven and Tilburg. Over time, as system usage grew, the average absolute deviation for RTL users remained below that of non-users, indicating improved consistency and driver adaptation to the system.

At the signal level (microscopic level), effects were more localised. Significant improvements appeared on several track segments, particularly at locations around stations where trains are most sensitive to preceding movements. This pattern is most evident in the eastbound direction around timetable point Ehv and in the westbound direction around Tb. Overall, the eastbound graph exhibits higher deviations and a greater number of significant edges than the westbound graph. For example, eastbound trains between AT$1346 $\rightarrow$ AT$1378 improved by roughly 117 seconds, while other links exhibited smaller differences of 1–5 seconds.

The proportion of trains encountering unplanned yellow signals was around 6 % lower among RTL users, corresponding to fewer restrictive aspects and reduced need for braking. When accounting for preceding-train types, the reduction was most notable when freight trains followed sprinter services or other freight trains, with around 15 % lower likelihood of unplanned yellow signal passages with RTL usage.

Despite a large rise in traffic volume during the diversion, RTL users maintained more stable performance levels. The system's benefits were strongest in stable flow conditions, while in complex merging or crossing scenarios involving delayed passenger trains, RTL users sometimes performed slightly worse, likely reflecting RTL's current limitation of not modelling real-time route conflicts in the case of a preceding merging passenger train, it caused 31 % more unplanned yellow aspects when

using RTL.

Conclusions and discussion

Overall, the results show that RTL yields modest but consistent improvements in freight train performance. The gains, typically on the order of tens of seconds per train and a few percentage points in conflict reduction, accumulate across the network, supporting smoother and more predictable operations.

The context-dependence of the results is a key finding: RTL proved most effective in at major stations, but less so during merging traffic interactions. This likely reflects both the nature of the advice (focused on timekeeping rather than conflict resolution) and the system's limited visibility of merging or crossing traffic in such situations.

Importantly, the analysis demonstrates that real-world DAS performance can be meaningfully quantified using operational datasets. By employing event-graph analysis and multiple KPIs, this study provides one of the first large-scale empirical assessments of a connected DAS under mixed-traffic conditions.

The findings suggest several practical recommendations, Together, these actions could help maintain punctuality and effective capacity on busy freight corridors, complementing traditional infrastructure expansion.

- Infrastructure managers should promote broader DAS adoption among operators. Since RouteLint already improves punctuality without physical upgrades, making the system widely available (e.g., free to freight operators) could yield network-wide benefits.
- IMs should institutionalise data-driven monitoring of DAS performance. The KPI framework used here can be integrated into regular operations, enabling continuous verification of where DAS deployment is most effective.
- System developers should enhance RouteLint's capabilities by adding predictive conflict alerts and driver feedback mechanisms to better manage merging and crossing scenarios.
- Railway undertakings and drivers can use RouteLint not only to adhere to schedules but also to reduce unnecessary braking and energy use, contributing to lower operational costs and environmental impact.

This research provides empirical evidence that a connected Driver Advisory System such as RTL can improve freight train performance in real-world conditions. By linking operational data with advisory usage, it bridges the gap between simulation-based studies and everyday railway practice. Although the improvements are moderate in size, they are operationally meaningful and demonstrate the potential of digital advisory support to enhance timetable adherence, stability, and sustainability. The methodological framework developed here, combining large-scale operational data, event-graph analysis, and statistical testing, offers a replicable approach for evaluating DAS effects across other corridors and European networks.

Contents

List of Figures

List of Tables

Nomenclature

Abbreviations

Abbreviation	Definition
AI	Artifical Intelligence
ATB	Automatische TreinBeïnvloeding
ATP	Automatic Train Protection
BH	Benjamini–Hochberg
CATO	Computer-Aided Train Operation
C-DAS	Connected Driver Advisory System
CPM	Critical Path Method
DAG	Directed Acyclic Graph
DAS	Driver Advisory System
DIL	Driver-In-the-Loop
ETCS	European Train Control System
EETO	Energy-efficient Train Operation
FDR	False Discovery Range
GPS	Global Positioning System
IM	Infrastructure Manager
HMI	Human-Machine Interface
KPIs	Key Performance Indicators
MTTC	Minimum-Time Train Control
N-DAS	Networked Driver Advisory System
OTR	Op Tijd Rijden
Q-Q	Quantile–Quantile
RITS	Rail Infra Toestand Service
RU	Railway Undertaking
RTI	Related Train Information
RTL	RouteLint
S-DAS	Standalone Driver Advisory System
TCC	Train Traffic Control Centre
TPS	Trein Positie Service
TROTS	Trein Observatie en Tracking Systeem
TSU	Track-Side Units
VOS	Verkeersleiding Ondersteunend Systeem

1 Introduction

1.1. Background

The rail network of the Netherlands is one of the busiest and most complex railway systems in the world (Verstappen et al., 2022). With increasing traffic demand and a largely saturated infrastructure, the system faces growing challenges in maintaining reliability and operational efficiency (Wang & Goverde, 2016). Rail freight transport is expected to grow by 50% in 2030 and 100% in 2050 compared to 2015, and for passenger transport, a growth of 30% is forecasted for 2030 (European Commission, 2011). As shown in Figure 1.1, several lines in the Dutch railway network are expected to face potential capacity shortages on a standard day in the 2026 timetable year, particularly in the Randstad, the southern corridors, and the routes towards Belgium and Germany.

Figure 1.1: Overview of the Dutch railway network on a standard day in the 2026 timetable year. *Source: Timetable Redesign: Capacity Model 2027, Netherlands (European Capacity Management Tool, ProRail)*

These capacity pressures are particularly evident on shared-use lines, where dense passenger services leave limited time slots for freight trains. Although freight trains operate according to pre-designed

timetables, their services are often less punctual (Wang et al., 2019). Compared with passenger trains, freight services differ substantially in power, mass, braking performance, and acceleration behaviour (Sogin et al., 2013). Their greater mass leads to longer acceleration and braking curves, higher energy consumption during speed changes (George, 2022), and lower operational speeds on some tracks to reduce rail wear. Consequently, freight train operations should be as conflict-free as possible to minimise delay and its propagation.

A promising approach to increasing operational efficiency without extensive infrastructure investment is to assist train drivers through real-time guidance provided by a Driver Advisory System (DAS) (Wang et al., 2019). DAS is an onboard tool that helps drivers follow an optimal reference trajectory (i.e., speed profile), thereby avoiding potential red or yellow signals indicating track occupation conflicts, while also promoting energy efficiency and achieving other operational objectives. To understand how such systems contribute to operational performance, it is important to consider the different types of DAS and the principles on which they are based.

1.1.1. Driver advisory system

DAS can be grouped into three types: Standalone-DAS (S-DAS), Networked-DAS (N-DAS), and Connected-DAS (C-DAS) (Panou et al., 2013; Wang et al., 2019).

- S-DAS operates independently by pre-loading all necessary data onto the train before or at the start of the journey. It functions without real-time communication with the traffic control centres and does not adapt to changes in the operational plan during the train run.
- N-DAS, on the other hand, is capable of communicating with one or more traffic control centres, allowing it to receive updates regarding schedules or routing information. While more complex than S-DAS, the updates in N-DAS are typically periodic rather than continuous, and thus not truly real-time.
- C-DAS represents the most advanced form, featuring real-time bi-directional communication with traffic control centres in the areas where the train is operating. C-DAS can dynamically receive updated timetable targets and generate a corresponding updated speed profile to avoid conflicts, providing advice to the driver to adhere to the updated schedule.

The design of C-DAS is based on distributing intelligence between the traffic control centre and the train, offering a practical solution for managing real-time traffic and schedule disturbances. This architecture can be crucial for addressing the connection between existing DAS approaches and future advancements, as it provides a framework for evolving solutions to network capacity issues (Panou et al., 2013). To understand the distribution of C-DAS intelligence, three key system clusters have been identified:

- DAS-Onboard, where the onboard unit computes the optimal speed profile, derives driving advice, and displays it to the driver.
- DAS-Distributed, where the computation of the optimal speed profile is off-train at Track-Side Units (TSUs) and transmitted onboard. The onboard functionality includes the advice derivation and display.
- DAS-Central, where the optimal speed profile computation and advice derivation are performed in the TSUs, leaving no onboard intelligence.

All C-DAS clusters require real-time tracking of the train's location to provide accurate and timely advice to the driver. A critical aspect of C-DAS design is how the advice is integrated with the driver. There are several approaches to delivering advice to the driver (Panou et al., 2013). One key distinction is between follow-me and trade-off advice. In the follow-me mode, advice is delivered in small, frequent increments, requiring the driver to make continuous adjustments to the train's speed. This mode demands higher attention from the driver, but offers the advantage of precise control over the train's trajectory (Panou et al., 2013). On the other hand, trade-off advice provides the driver with discrete, less frequent changes in speed targets. While this reduces the need for constant adjustments, it may require the driver to minimise the difference between the target and the current speed over a larger period. The research of Panou et al. (2013) describes types of advice updates. The different contexts of advice are:

- Distance or time interval-based advice: This form of advice is valid for specific distances or time intervals. While it provides clear and simple targets with little computer processing, it may not always align with the train's actual speed or trajectory.
- Event-driven advice: This type of advice is updated when a significant event occurs. Events in the train operation are defined as departure, arrival, and passing events (Spanninger et al., 2022). Event-driven advice usually fulfils the system's goals, as it adapts dynamically to the changing operational environment. However, this requires a filtering mechanism to avoid frequent, disruptive changes in advice.
- Updated speed profile advice: The DAS continually generates and displays an updated speed profile that reflects the train's real-time movements. This approach allows for fine-tuned adjustments to the driving strategy based on the train's actual progress over the designated route. Nevertheless, a significant amount of data has to be transmitted to the train if DAS-Onboard is not chosen.
- Contextual advice: No specific driving targets are presented to the driver, but instead contextual information, for instance, background information, leaving the driver to find an optimal trajectory on their own.

Moreover, Panou et al. (2013) also describes the different forms of advice:

- Suggested speed: This is the most straightforward form of advice, providing optimal speed instructions for the driver to follow. It is particularly useful when specific schedule targets need to be met.
- Timekeeping: This advice shows the deviation between the actual time at a specific point on the track and the scheduled time on the timetable, helping the driver manage delays.
- Control action advice: Including suggestions for coasting, tractive power, or brake pressure to achieve the most efficient operation.
- Energy savings: Besides speed or time suggestions, DAS can also advise the driver with the objective of minimising energy usage.

Among the DAS implementations in Europe, RouteLint (RTL) is the system developed by the Dutch Infrastructure Manager (IM), ProRail, and serves as the primary case study in this research. It is a real-world example of a C-DAS, in which all advice is centrally generated and transmitted to the train. The onboard functionality is limited to displaying the prepared advice to the driver (Panou et al., 2013). Nowadays, RTL is widely used by several European freight companies that deliver goods between the Netherlands and its neighbouring countries. It provides timekeeping advice and gives additional information on the upcoming route, scheduled event times at critical timetable points, and the real-time positions of nearby trains. Figure 1.2 shows the dashboard of the RTL app and its available advice and information.

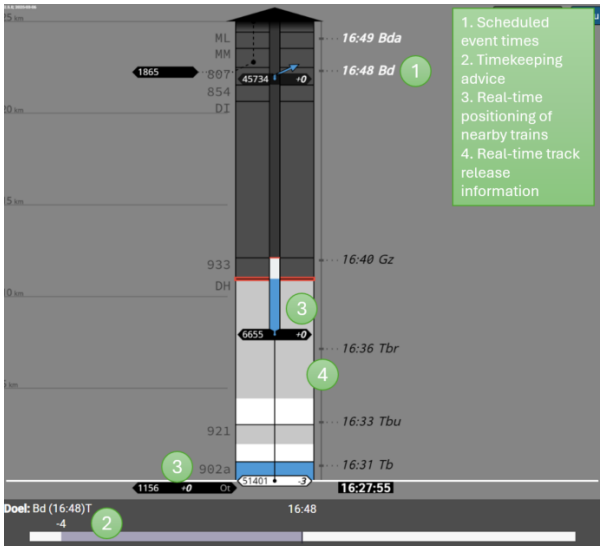


Figure 1.2: RTL dashboard showing scheduled times, timekeeping advice, nearby trains, and track release status.

However, due to the vertical separation between the Railway Undertaking (RU) companies and the IM, RTL cannot present speed advice, as speed advice would be considered as directing of operation, a task explicitly for RU (Abbott, 2017).

For DAS to be effective, the quality of the information it provides is key. This includes accurate information about the driver's surroundings, such as nearby trains and accurate signalling and infrastructure information (Panou et al., 2013). Therefore, understanding the signalling system in which DAS operates is crucial.

1.1.2. Dutch Railway Signalling System

This section describes the Dutch Railway Signalling system NS'54, which provides information that is the basis of the Dutch DAS RTL. To ensure both safe and efficient spacing between trains, signalling systems were developed and initially operated by humans. From 1830, Railway Policemen were replaced by fixed signals (Clark, 2012). Since 1954, the signalling system NS'54 has been implemented in the Netherlands (Goverde et al., 2013). NS'54 is a three-aspect, two-block system, as shown in figure 1.3. The main aspects, green and yellow, are extended with target speed indicators, allowing for braking over multiple blocks by progressively reducing speed from one block to the next Wang (2017). It uses a fixed block section with lengths that are at least equal to the maximum braking distance of the worst braking train, plus a margin for the reaction time of the driver and the Automatic Train Protection (ATP) system. There is a difference between automatic and controlled signals. Automatic signals function automatically when a train passes through track sections. They can only be utilised on tracks without switches or points where conflicts with movements on conflicting routes may occur (Pachl, 2020). Their default setting is green, and they turn red automatically when a train enters the block section, yellow when the train leaves the block and green when the train leaves the following block. In the Netherlands, aspects of automatic blocks are not saved, but can be derived from train position data. The aspect of a controlled signal, on the other hand, is by default red and has to be set to green by a traffic control centre or an automatic route setting system. This type of signal is implemented with a section that contains movable track elements or points where movement conflicts may occur (Pachl, 2020).

Goverde et al. (2013) describe the two types of blocks considered in the Netherlands: normal and short blocks. A normal block has a length corresponding to the line speed and no movable track elements, in which case NS'54 is an automatic signal with only 3 aspects: green, continue with maximum track speed; yellow, reduce speed to 40 km/h and prepare to stop before the next red sign; red, stop before the signal.

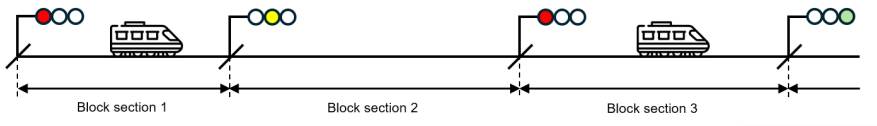


Figure 1.3: Schematic drawing of a three-aspect two-block signalling.

Short blocks are sections of track typically located near stations. Their length is shorter than the maximum braking distance required at line speed, allowing trains to operate at shorter headways. To facilitate safe operation under these constraints, the NS'54 already communicates speed restrictions at one or more signals in advance. Consequently, train drivers receive timely indications to adjust their speed to prescribed limits before reaching the following signals, ensuring the train enters the short block at a speed consistent with the block's reduced length. These speed restrictions are displayed by a yellow aspect combined with a white numeral, indicating the maximum permitted speed at the next signal. For instance, a signal showing a yellow aspect with a white '8' instructs drivers to reduce speed to 80 km/h before the next signal.

Since NS'54 is a lineside signalling system that relies on driver compliance, it is complemented with an ATP system to guard against errors of train drivers (Goverde et al., 2013). Currently, three main ATP systems are operational in the Netherlands: ATB-Eerste Generatie (ATB-EG), ATB-Next Generation (ATB-NG) and European Train Control System (ETCS). Figure 1.4 shows the different systems on

the map. The first ATB version, ATB-EG, is the most commonly used ATP system as it is deployed on the core mainline railway network. It continuously monitors train speed and intervenes based on predefined speed steps, enforcing braking actions as instructed by the signalling system. The speed steps are transmitted to the train via coded track circuits. After a speed reduction order, the driver has to reduce speed until the permitted speed is reached. If a driver fails to brake sufficiently after a speed reduction order, then ATB warns the driver. If still no action is taken by the driver, the ATB will intervene with an emergency brake to bring the train to a standstill. Speeds below 40 km/h are not supervised because it is assumed that the driver runs on sight and shall stop in time before a red signal. The supervised speeds correspond to the nearest ATB speed step above the permitted speed. To improve low-speed protection, an enhancement known as ATB-Vv (Verbeterde versie) has been added at selected high-risk locations, such as switches or intersections.

The second system, ATB-NG, is mainly on regional lines that did not have ATB at the time of the ATB-EG implementation. It is a completely different intermittent ATP system.

Figure 1.4: Coverage of Automatic Train Protection systems on the Dutch railway network.

The third and most advanced operating ATP system, ETCS Level 2, is completely different from ATB, as it has integrated cab signalling, continuous supervision, and radio-based communication. It operates as a fixed-block system using track vacancy detection and interlockings to manage route setting.

The combination of dense mixed-traffic operations and growing demand places high requirements on both traffic management and train driving. While RTL and other C-DAS implementations aim to support drivers by improving anticipation and coordination with network control, their real-world operational benefits have not yet been quantified. Previous research has demonstrated the theoretical potential of DAS to enhance punctuality, energy efficiency, and workload reduction (Dong et al., 2018; Panou et al., 2013; Verstappen et al., 2022). However, most studies have relied on simulations, controlled environments, or small-scale trials, offering only partial insight into the system's day-to-day effectiveness under realistic signalling and timetable conditions. As a result, there remains limited empirical evidence of how DAS performs in regular freight operations. Therefore, a data-driven assessment of

RTL's operational performance within the Dutch signalling environment is required to better understand where and when such systems deliver measurable improvements.

1.2. Research questions

Building on the identified gap in the quantification of DAS benefits using real-world operational data, this research aims to address this aspect through a data-driven approach. By utilising established railway datasets and developing a methodological framework for empirical analysis, the study demonstrates how the operational performance of a DAS can be evaluated under real-life conditions. By comparing operations with and without RTL and analysing performance in train operations such as merging, diverging, and short-following scenarios (where freight trains operate closely behind preceding services), the study seeks to reveal measurable improvements in operational efficiency. The outcomes are expected to support infrastructure managers and freight operators in assessing the effectiveness of DAS and improving traffic flow on capacity-constrained corridors.

The main research question is:

How can the operational performance of freight trains be evaluated using empirical data from Driver Advisory Systems such as RouteLint?

Next, the sub-questions are addressed to answer the main research question:

1. What is the state-of-the-art and the state-of-practice of Driver Advisory Systems?
2. Which methods and KPIs can be used to quantify the effect of Driver Advisory Systems on freight train operations using empirical data?
3. How do freight train operations with and without a Driver Advisory System differ?
4. How do the performance impacts of a Driver Advisory System vary across scales of operation and over space and time?
5. What policy implications can be drawn from the data-driven analysis?

1.3. Thesis Structure

The remainder of this thesis is structured in the following manner: In Chapter 2, the state-of-the-art of DASs and the state-of-practice of the Dutch DAS, RTL, together with its international counterparts, are reviewed to answer sub-question 1. Chapter 2 investigates the current methods that are used in the literature to identify the effects of DAS in train operation, and the Key Performance Indicators (KPIs) applied in the analyses. In Chapter 3, the methodology is described, and the results are presented in Chapter 4 to answer sub-questions 3 and 4, whether freight train performance differs with or without the use of RTL. Subsequently, Chapter 5 discusses the policy implications to address sub-question 5, based on the findings of Chapters 3 and 4, and identifies potential improvements for RTL. In addition, this chapter outlines the study's limitations and provides suggestions for future research. Finally, Chapter 6 concludes the thesis by summarising the answers to the research questions and presenting the main takeaways of the study.

2

Literature review

Although DAS has been widely tested in simulations and prototype tests, little is known about its real-world operational impact in freight transport. This study aims to fill that gap through a data-driven evaluation of RouteLint (RTL).

This chapter provides an overview of the existing literature and current practices related to DAS. It starts by discussing the state-of-the-art developments in DAS technology and research, followed by a review of the current applications and implementations in European rail networks. Also, the chapter outlines commonly used KPIs for evaluating DAS performance and concludes with an overview of the limitations found in existing DAS studies.

2.1. State-of-the-art of DAS

As rail transport evolves to meet demands for improved efficiency, sustainability, and safety, DAS has become an important tool for optimising train operations (Verstappen et al., 2022). This section presents the state of the art in DAS, highlighting recent developments, innovations, and scientific research.

Studies like Dong et al. (2018), Wang et al. (2019), and Xiao et al. (2021) highlight the potential of DAS for decreasing delay impacts by improving energy efficiency and punctuality under both nominal and disturbed conditions. DAS can support Energy-efficient Train Operation (EETO) by advising optimal driving strategies. The literature of Scheepmaker et al. (2017, 2020) identifies several such strategies, including:

- Maximal Coasting strategy, which applies maximum speed (if reached) until coasting.
- Reduced Maximum Speed strategy, which applies an optimal cruising speed without coasting.
- Energy-Efficient Train Control strategy, which combines optimal coasting and cruising.

Another driving strategy discussed in the literature is Minimum-Time Train Control (MTTC). Unlike energy-efficient approaches, this strategy does not aim to minimise energy consumption but focuses on compensating for delays, instructing the train to drive as fast as possible. The causes of delays can be categorised into disturbances and disruptions (Cacchiani et al., 2014). Disturbances are generally seen as minor perturbations to the railway system that can be managed by adjusting the timetable without altering rolling stock or crew assignments. In contrast, disruptions are more severe incidents that necessitate changes not only to the timetable but also to the deployment of rolling stock and crew (Cacchiani et al., 2014). Effectively rescheduling the timetable can enhance service reliability, improve working conditions, reduce unnecessary travel times, and decrease both operational costs and investments in rolling stock and personnel (Cheng, 1996).

Research by Xiao et al. (2021) and Dong et al. (2018) demonstrates the real-time applicability of DAS in responding to delays and temporary speed restrictions. Xiao et al. (2021) find that the operational efficiency of a single high-speed train is improved for punctuality and energy efficiency with the help of DAS in disturbed conditions. Dong et al. (2018) show through a case study and analysis how the train trajectory is optimised for ETO in nominal conditions and punctuality-prioritised driving during disturbed conditions. Similarly, Wang et al. (2019) show how DAS under nominal conditions can maintain an optimal speed profile and minimise delays.

The technology of a DAS consists of five layers: Task distribution, data processing, driver interface, train positioning, and communication (Panou et al., 2013). Albrecht (2014) describes the enabling inputs and interfaces of a typical DAS system as four main components: a positioning or speed measurement device, a clock, a database containing the timetable, infrastructure data, and roller stock characteristics, and a Human-Machine Interface (HMI) to present advice to the drivers. Wang and Goverde (2017) present an example of a prototype DAS, the ETO system, which includes five main functions: data processor, Train State Monitor, Trajectory Calculator, Trajectory Processor, and Advice Generator. In this system, real-time monitoring and trajectory recalculation ensure that the driving advice remains adaptive to deviations from the planned timetable, thereby optimising operational performance. Successful integration of a DAS requires careful coordination between these layers to ensure the advice provided to drivers is both accurate and operationally useful (Panou et al., 2013).

Recent developments of DAS explore the potential of Artificial-Intelligence (AI) driven DAS, such as the study of Luo et al. (2025), where an Intelligent DAS based on Large Language Model framework was proposed to compute context-specific advice for high-speed trains to prevent accidents. Although promising, such AI-driven systems remain largely experimental and lack robust validation to ensure reliability in real-world operations.

Further research has also focused on improving traditional DAS designs. Ghaviha et al. (2017) developed an S-DAS for Android devices, including dynamic loss with EETC instead of constant efficiency for a more realistic energy consumption. Z. Li et al. (2018) proposed a dynamic optimisation method to generate optimised train trajectories for DAS using a combination of online and offline optimisation under varying driving styles.

Overall, current research demonstrates that DAS technology has matured from conceptual frameworks to advanced prototypes integrating AI and optimisation techniques. Nevertheless, the majority of these studies focus on developing new algorithms or improving optimisation models, for example, trajectory recalculation (Z. Li et al., 2018), adaptive coasting control (Scheepmaker et al., 2017), or AI-driven decision logic (Luo et al., 2025). Many of these approaches have been validated primarily through simulation-based experiments (e.g. Dong et al. (2018), Wang et al. (2019), and Xiao et al. (2021)), providing valuable insights but limited evidence of operational effectiveness in daily freight operations. The next subsection describes the state-of-practice, highlighting how DAS technologies are currently implemented in the Netherlands and across other European railways.

2.2. State-of-practice of DAS

In this study, the Dutch DAS RTL is analysed as a case study. RTL is currently provided by ProRail to freight companies. It is a C-DAS that provides timekeeping advice, information about the position of surrounding trains, the upcoming route, and the up-to-date schedule. Figure 2.1 illustrates how various data sources interact within RTL to generate advice for the driver. The timekeeping advice shows the expected deviation from the event time at the next scheduled timetable point in minutes, with + values indicating delays and - values for early arrivals. To compute this deviation, the expected time of the upcoming event is calculated by estimating the train's current speed from the ratio of travelled distance to elapsed time, and by considering the remaining distance to the timetable point. Since July 9, the calculation has also incorporated the speed limits. The train's position, as well as that of nearby trains, is determined in real time using either Global Positioning System (GPS)-based data or track occupancy data. The default is via GPS signals, provided by GEO data, which contains all geographical information of the Dutch railway and its surroundings. However, poor reception caused by steel freight locomotives and tunnels often disrupts GPS signals. In such cases, the system switches to the Trein Positie Service (TPS), which estimates train locations based on trackside sensor data.

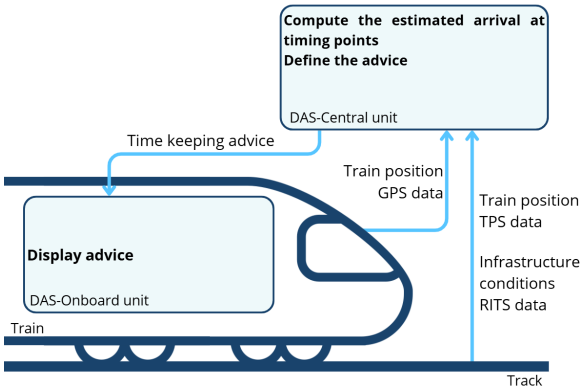


Figure 2.1: Current data flow in the RouteLint system.

The information about the upcoming route includes the cleared route for the train, which is derived from the Rail Infra Toestand Service (RITS). This database contains all information about the positions of switches and signals, and together with the positions of trains from TPS, it clears routes for the train. In RTL, the upcoming track can be displayed in blue, dark grey, light grey, or white, as shown in Figure 2.2. The blue line represents the real-time position of the train, while the light grey and white sections indicate that the upcoming route has been cleared. Dark grey indicates that no route has been cleared and that the train does not have permission to enter that section. A section is shown in light grey when it lies on the open line, denoted by automatic signals. A section appears in white when it lies within a controlled area, where manually operated signals are used, and the system has more precise information about signal aspects.

The displayed schedule reflects the planning generated by planners in DONNA. DONNA is the planning system used for railway capacity allocation in the Netherlands (Meijer et al., 2012). It can be updated in real-time by Train Traffic Control Centre (TCC) in response to driver-submitted requests for alternative paths or when disruptions occur.

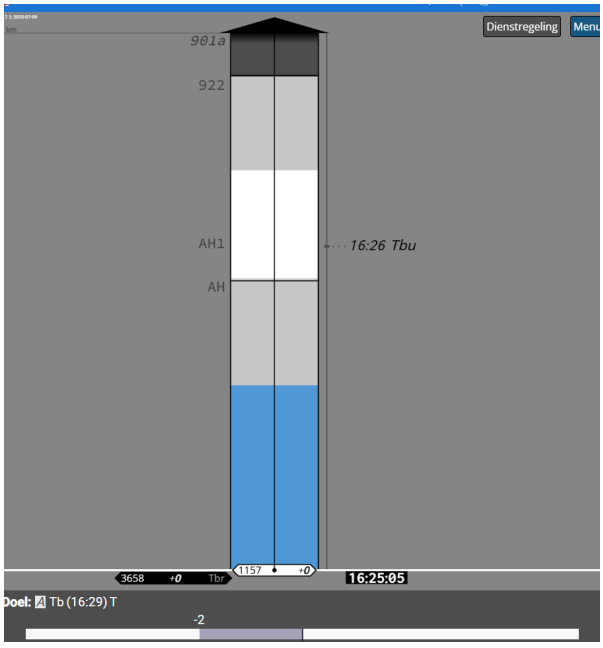


Figure 2.2: A screenshot of the RTL dashboard, showing the difference in cleared routes.

Despite the advanced functionalities implemented in RTL, some limitations remain:

- A lack of consideration for the delay status of surrounding trains when generating advice.

- Limited granularity in feedback to drivers (only timekeeping advice, not speed).
- Occasional mismatches between GEO-data and track-side data, which may lead to errors in released track sections or misinterpreted signals.

DAS are developed either by IMs, RUs, system suppliers or train manufacturers. When an IM develops a DAS, it is usually provided free of charge to the RUs operating on that network, such as infraDOAS in Austria (ÖBB Infra) or the Infrabel DAS in Belgium. An exception is RTL, which is offered as a paid service to railway companies. When system suppliers like Siemens or KeTech develop DAS solutions, they act as commercial vendors selling their tools to RUs. In contrast, NS (the largest Dutch RU) chose to develop its own DAS, TIMTIM, to incorporate coasting advice specifically aimed at reducing energy consumption. In many other European countries, the structure of information provision and services is comparable to NS with TIMTIM, where Related Train Information (RTI), provided by the IM ProRail, is utilised to generate and present advice to train drivers (Gueudar-Delahaye et al., 2022; Salden et al., 2024).

RTL is currently used by freight operating companies, while TIMTIM, developed by NS for passenger services, functions as a C-DAS that integrates GPS-based train positioning and RTI data from RTL. The central DAS unit calculates coasting advice when a train runs on schedule and provides timekeeping recommendations for arrivals and departures at stations. This information is transmitted to the onboard DAS subsystem, which presents the advice to the driver. The timing for initiating coasting is based on the remaining running time and speed limits between stops (Cunillera et al., 2023; Scheepmaker et al., 2020). The interface also displays surrounding trains, temporary speed restrictions, station information, train characteristics, and track conditions. To achieve energy efficiency, NS employs the MC driving strategy, which maintains maximum speed up to the limit before coasting, the most energy-efficient and low-workload strategy, particularly for Sprinter services.

In most European countries, the structure of information provision is similar to the Dutch approach: the IM supplies real-time traffic information, which the RU uses to generate driver advice. In Sweden, the Computer-Aided Train Operation (CATO) system calculates an energy-efficient speed profile while maintaining punctuality (Dong et al., 2018; Sandblad et al., 2015). In Germany, LEADER (by Knorr-Bremse) provides multi-level optimisation combining timetable, infrastructure, and live traffic data to improve flow and energy use (Knorr-Bremse, n.d.).

Overall, the comparison (Table 2.1) shows a clear trend toward hybrid architectures: Systems with centralised intelligence for timetable management and onboard subsystems for real-time driver interaction and local optimisation.

Table 2.1: Current implementation and developments of DAS in Europe.

Name	Developer	IM	RU	Intelligence	Features	Source
LEADER	Knorr-Bremse	DB InfraGo	DB Cargo	Central	Control action advice, related trains, up-to-date planning	Knorr-Bremse
Infrabel DAS	Infrabel	Infrabel	Multiple RUs	Central	Suggested speed advice	Infrabel
TIMTIM	NS	ProRail	NS	Central	Coasting advice, related trains, infrastructure characteristics, up-to-date planning	(Salden et al., 2024)
infraDOAS	ÖBB Infra	ÖBB Infra	Multiple RUs	Central	Suggested speed advice, up-to-date planning	InfraStruktur (Gueudar-Delahaye et al., 2022)
RouteLint	ProRail	ProRail	Multiple RUs	Central	Timekeeping advice, related trains, track occupation, up-to-date planning	(Gueudar-Delahaye et al., 2022)

Continued on next page

Name	Developer	IM	RU	Type	Features	Source
Opti-Conduite	SNCF	SNCF Réseau	SNCF	Onboard	Control action driving advice, related trains, track occupation, infrastructure characteristics, up-to-date planning	OptiDriving
DASEM	CAF	–	–	Onboard	Energy-efficient driving advice, up-to-date planning	CAF
Siemens C-DAS	Siemens	Network Rail	Multiple RUs	Central	Speed and energy-efficient driving advice, up-to-date planning	Siemens
Ketech C-DAS	Ketech	Network Rail	West Midlands Train	Onboard	Speed and energy-efficient driving advice, up-to-date planning	KeTecch
GreenSpeed	CUBRIS	DSB Infra	Multiple RUs	Onboard	Energy-efficient driving advice, related trains, track condition, up-to-date planning	CUBRIS

2.3. KPIs for evaluating DAS

To assess the performance and risks of railway systems, the European standard EN 50126 has been established. This standard defines the RAMS aspects: Reliability, Availability, Maintainability, and Safety (Stenström et al., 2012). Within this framework, specific KPIs are defined to provide quantitative insights for decision-makers in railway planning and operations.

According to Goya et al., 2018, KPIs such as accuracy, protection level (limiting the uncertainty of the real position), availability, operational availability, integrity, and continuity are defined to support the integration of onboard position systems as GNSS within the standardisation of EN 50126. Nicholson et al. (2015) developed an algorithm for timetabling and real-time disturbance management using indicators including transport volume, journey time, connectivity, punctuality, resilience, energy consumption, and resource usage. Similarly, Goverde and Hansen (2013) and Goverde et al. (2016) identified KPIs such as infrastructure occupation, timetable stability, feasibility, robustness, and efficiency to evaluate and construct effective timetables.

Other studies focus more on reducing energy consumption in train operation. González-Gil et al. (2015) aim to analyse the actual energy performance of the railway system, and describe the KPIs related to energy consumption: specific CO_2 emissions, energy consumption, share of renewable energy, waste heat recovery, traction power supply efficiency, in-service traction energy consumption, in-service auxiliaries' energy consumption, braking energy recovery, energy consumption in depots, and energy consumption in stations. Building on this, Scheepmaker and Goverde (2021) introduce a multi-objective optimisation framework for energy-efficient timetabling, balancing travel time, total traction energy consumption, and buffer time.

Overall, railway performance evaluation encompasses both operational efficiency and punctuality. Commonly used indicators include average delay per train, on-time arrival rates, delay propagation metrics, and dwell time deviations (Kristoffersson & Palmqvist, 2021; Scheepmaker et al., 2020). Together, these KPIs form a comprehensive framework for assessing both the technical efficiency and the service reliability of railway operations, providing the analytical foundation for evaluating the effectiveness of DAS.

In the Dutch railway context, the national rail operator NS applies a dedicated KPI known as Op Tijd Rijden (OTR), which is used on all unhindered train runs, i.e. services that were not affected by external disturbances. If the delay was caused by preceding traffic or infrastructure constraints, delay propagation was not included in the OTR metric. This indicator evaluates whether a journey adhered to a strict schedule window defined in the timetable. A ride is considered "on time" if the train departed within a tolerance of −10 to +30 seconds of the scheduled departure time and arrived within −30 to +10 seconds of the scheduled arrival time. Alternatively, suppose the train departed late but did not accumulate additional delay (i.e., more than 10 seconds before the scheduled arrival), the ride is still considered a good performance (Weeda & van Onna, 2019). At a higher policy level, ProRail and the

Dutch government define train punctuality as arrival within a 3-minute tolerance window, meaning trains arriving up to 3 minutes early or 3 minutes late are considered on time. This threshold applies to both passenger and freight services van Infrastructuur en Waterstaat (2020). In this study, the focus will be on timeliness: the deviations from the planned time of passage at scheduled timetable points.

RTL has been developed to improve train punctuality and, ultimately, the overall capacity of the railway network. However, at present, no functionality related to energy efficiency has been integrated into the system. Besides, in current operational practice, energy efficiency is considered secondary to punctuality. When the scheduled journey time cannot be achieved, trains typically adopt the MTTC driving strategy to recover delays as much as possible. Therefore, when quantifying the effects of the DAS, the focus will be on the punctuality of train services, evaluated through the following KPIs: Transport volume, timetable deviation, number of unplanned yellow signals, and train interaction effects. Table 2.2 provides a detailed description of these KPIs and indicates the levels at which they are assessed. Chapter 3 provides a detailed explanation of the methods used to derive the KPIs.

Table 2.2: Key Performance Indicators applied to evaluate the effectiveness of RTL.

	Macro-level	Micro-level
Transport volume	- Transport volume over time - Freight volume between timetable points - Freight volume per timetable point over time - Freight volume per timetable point	- Freight volume between signals
Timetable deviation	- Average absolute deviation from planned timetable per timetable point - Average deviation from planned timetable over time	- Absolute deviation from estimated average running time per train
Yellow signals		- Number of encountered unplanned yellow signals - Number of unplanned yellow signals over time
Train interaction		- Operational causes associated with unplanned yellow signals

2.4. Limitations of existing DAS studies

Although the theoretical potential of DAS has been explored in the literature (Panou et al., 2013; Quaglietta et al., 2016), empirical studies on their real-world performance remain scarce, particularly those based on actual operational data. While many simulation studies, simulator tests, and DAS development research exist, most of these studies focus on prototype systems, small-scale trials, or controlled environments, providing qualitative insights into the potential benefit of DAS. For example, in the research of Wang and Goverde (2017), ETO was introduced as a new prototype of DAS, demonstrated in a laboratory environment with real-world instances. Similarly, Large et al. (2017) used simulation software to compare the performance of experienced drivers with that of novice drivers. Additionally, Verstappen et al. (2022) assessed the impact of DAS on train driver workload, attention allocation, and safety performance through a simulator study. It contributes to the impact of practical DAS in the railway industry, as the research concludes that DAS has no negative impact on attention allocation or safety performance and can decrease workload at several events. Furthermore, Dong et al. (2018) investigated an approach to improve energy efficiency and punctuality of trains equipped with DAS. A hardware-in-the-loop experiment is used to simulate the driving processes of passenger trains with and without DAS, and a prototype DAS is developed to deliver advice to the driver. On the other hand, in the research of Xiao et al. (2021), a Driver-In-the-Loop (DIL) test was conducted in combination with a driving simulator and an energy-efficient DAS to help validate the system's real-time response to delays and temporary speed limitations. Despite these advances, the majority of existing studies

still rely on controlled conditions and lack large-scale empirical data. While DIL tests, such as the one conducted by Xiao et al. (2021), are valuable for validating the real-time responsiveness of DAS, they do not provide a comprehensive analysis of DAS's performance under varied operational conditions over extended periods. Furthermore, despite the valuable insights from these simulations and test-based studies, a gap exists in quantifying the actual benefits of DAS, which limits the understanding of its broader implications in real-world railway systems. Simulation and test-based studies can provide valuable insights, but the lack of quantifiable benefits limits the understanding of how effectively DAS can mitigate real-world challenges such as delays and speed restrictions, which impact railway performance (Z. C. Li et al., 2021).

In summary, the literature establishes the technological promise of DAS but offers limited empirical evidence of operational benefits; the next chapter (Chapter 3) outlines the methods used to derive KPIs to address this gap.

3

Methodology

This chapter outlines the methodological framework used to assess the operational performance of freight trains, both with and without a DAS. It first provides a detailed description of the datasets used in the analysis, followed by a comprehensive overview of the data cleaning and merging procedures. Next, the methods applied for data exploration and event graph construction are presented. Finally, the analytical approaches employed to evaluate train performance are discussed. Figure 3.1 shows how the data is used for the analysis.

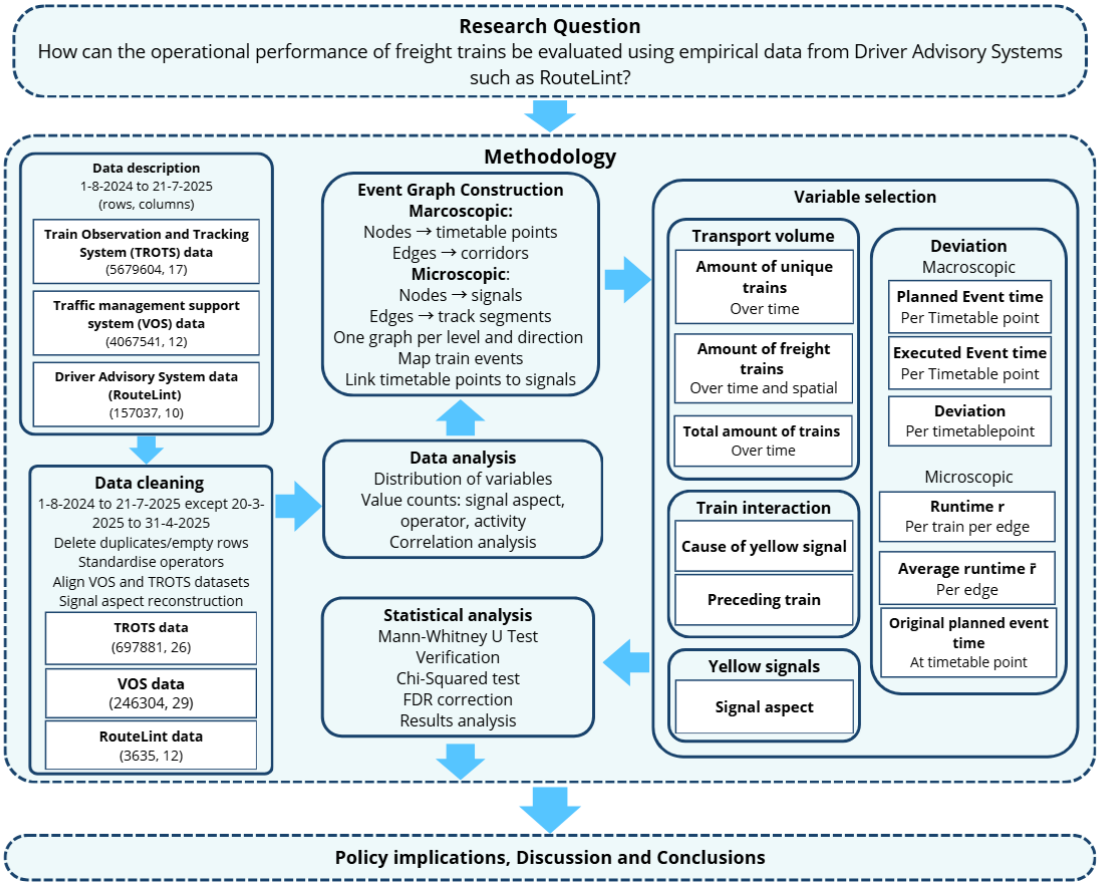


Figure 3.1: Flow diagram of the methodology.

3.1. Data description

The analysis uses two types of historical datasets: Verkeersleiding Ondersteunend Systeem (VOS) data and Trein Observatie en Tracking Systeem (TROTS) data. These datasets differ in the level of operational detail they provide, representing two distinct modelling scales: macroscopic and microscopic.

In railway research, the level of detail chosen in the modelling determines how infrastructure, train movements, and interactions are represented. A train movement is an authorised movement of a train within a defined section of the track Pachl (2020). Macroscopic models provide a high-level view of the railway network, abstracting it into stations or junctions and corridor-level track sections. Such models are suitable for traffic forecasting, routing, or long-term capacity studies, for example. Still, they lack the resolution required to describe operational details such as signalling or train conflicts. Microscopic models, on the other hand, represent every operational element in detail, including individual signals, switches, and block sections. This level of precision enables the simulation of train dynamics, conflict detection and resolution, and accurate computation of running and blocking times. However, their computational cost and data requirements often make them impractical for large-scale or long-term analyses (Radtke, 2014).

The Betuweroute is the main freight corridor in the Netherlands; however, due to major maintenance work in Germany, it was partially closed for a period of 80 weeks starting in November 2024. As a result, freight trains will be rerouted via alternative routes, with the Brabantroute serving as the primary alternate route for the Betuweroute. The Brabant route is a shared-use rail that connects Kijfhoek to Germany through Venlo. During the maintenance period, an additional 5 to 75 freight trains per day are expected to operate on the Brabantroute (Figure 3.2), representing an increase of 120%. To ensure conflict-free train operations on the Brabant Route despite the increase in traffic, the DAS RouteLint (RTL) system is being offered free of charge to all freight operators. This creates a valuable opportunity to analyse the impact of RTL using real-time operational data. For the case study, the analysis focuses on the corridor between Tilburg and Eindhoven, chosen for its operational diversity. It includes both inflows and outflows to freight marshalling yards, as well as connections to passenger lines, such as those near Tilburg and Boxtel. This mix of traffic types makes the corridor representative of complex railway dynamics in the Dutch network. To establish a consistent baseline for comparison between operations with and without RTL, the data analysis covers a one-year period from August 1, 2024, to the end of July 2025. This time window captures several months before the start of the 80-week closure and continues till the end of the period, enabling a balanced assessment of the system's operational effects.

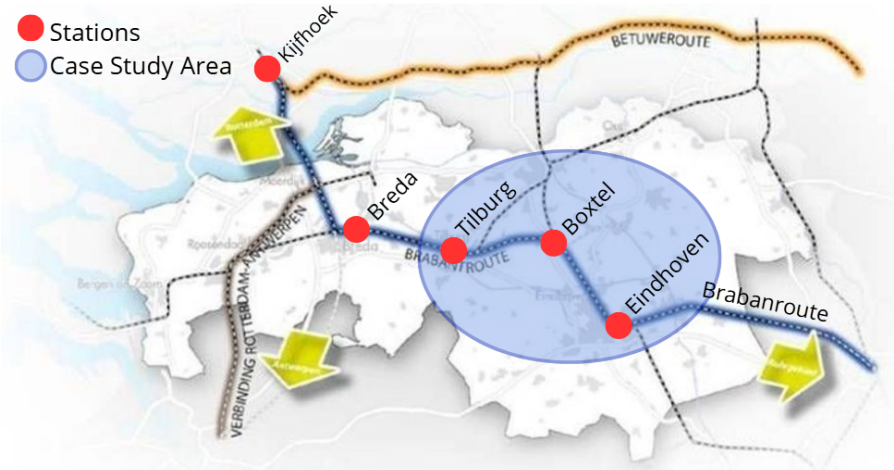


Figure 3.2: The Brabantroute on the Dutch railway mainline network with highlighted case study area. *Source: NOS*

In this study, the VOS dataset provides macroscopic-level data, including the planned timetable with scheduled event times at timetable points along the corridor; the adjusted timetable with up-to-date adjustments, such as real-time event time modifications due to disturbances or rerouting; and the actual event times at timetable points. Timetable points are locations in a train's schedule, with an associated planned or executed time labelled as an arrival, departure or passing time (Goverde et al., 2016).

In contrast, the TROTS dataset offers microscopic-level detail derived from dynamic infrastructure information, which includes information about signal passage times, signal clearance time, and signal aspects. The integration of both levels allows for a comprehensive analysis that balances high-level network behaviour with sufficient operational detail to study train performance and deviations.

The first step in preprocessing was selecting the appropriate time period: August 1, 2024, to July 21, 2025, for all trains driving over the Brabantroute within the corridor between Tilburg and Eindhoven. The corridor contains approximately 280 signals and 413 track sections connecting these signals, providing a rich network structure for constructing an event graph and analysing running time behaviour at a detailed level. The period from 17 to 24 April 2025 was excluded due to a technical malfunction in the RTL application, which failed to provide usable driving advice during that time. The datasets were cleaned by retaining only trains with the driving characteristic "freight" and removing empty columns.

As described in Chapter 1, the Dutch network uses a two-block, three-aspect signalling system (NS'54): a signal s shows red when the block ahead is occupied, yellow when the next signal $s{+}1$ is at red, and green when $s{+}1$ is not red (i.e., yellow or green). Passing a signal that is not cleared constitutes a red signal violation.

In the TROTS dataset, approximately 5% of signal aspects are missing. To determine the aspect shown at the moment of passage, we compare the time the train t passes signal s ($m_{t,s}^{\text{pass}}$) with the time the next signal $s{+}1$ clears from red to yellow ($m_{t,s+1}^{\text{clear}}$). If $s{+}1$ had already cleared before t passed s ($m_{t,s+1}^{\text{clear}} \leq m_{t,s}^{\text{pass}}$), then s must have displayed Green (GR). Conversely, if t passed s while $s{+}1$ was still red ($m_{t,s+1}^{\text{clear}} > m_{t,s}^{\text{pass}}$), then s must have displayed Yellow (GL).

We define the Signal Release Delay as the time between passing s and the subsequent clearing of $s{+}1$:

$$\text{Signal Release Delay} \;=\; \max\!\big(0,\; m_{t,s+1}^{\text{clear}} - m_{t,s}^{\text{pass}}\big) \qquad \forall\, t \in T,\, s \in S. \tag{3.1}$$

Where:

- $m_{t,s}^{\text{pass}}$ is the timestamp when train t passes s,
- $m_{t,s+1}^{\text{clear}}$ is when $s{+}1$ first shows a green of yellow aspect,
- $\max(0, \cdot)$ ensures that negative values are set to zero, which correspond to s being green upon approach.

Classification rule:

- Signal Release Delay $= 0$ Green (GR) at s;
- Signal Release Delay > 0 Yellow (GL) at s.

To ensure consistent merging of the TROTS and VOS datasets, train number and date combinations were used as unique identifiers, while duplicate train numbers per day were removed. Both datasets were then aligned to ensure that they contained exactly the same set of trains by checking the combinations of traffic date and train number. Trains without a corresponding match in the other dataset were excluded. Finally, a directional attribute was assigned to each train based on its sequence of event times and signal order along the corridor. This attribute identifies whether a train was operating in the eastbound or westbound direction and ensures that the event graphs correctly represent the directional flow of train movements.

The cleaned and harmonised datasets provide the temporal and spatial structure for the analysis. The main timetable points and their abbreviations are listed in Table 3.1.

Table 3.1: List of timetable points on the Brabantroute and their abbreviations and short descriptions.

Timetable point	Abbreviation	Description
Tilburg Reeshof	Tbr	Minor passenger Station
Tilburg Universiteit	Tbu	Minor passenger Station
Tilburg	Tb	Major passenger Station of Tilburg
Tilburg aansluiting	Tba	Junction at Tilburg
Tilburg industrie	Tbi	Industrial siding at Tilburg
Oisterwijk	Ot	MInor passenger station
Boxtel	Btl	Minor passenger Station
Vught aansluiting	Vga	Junction at Vught
Vught	Vg	Minor passenger station
Liempde	Lpe	Siding area at Liempde
Best	Bet	Minor passenger station
Acht	At	Siding area at Acht
Eindhoven Strijp-S	Ehs	Minor passenger station
Eindhoven	Ehv	Major passenger station of Eindhoven
Tongelre aansluiting	Tgra	Junction at Tongelre/Eindhoven

3.2. Aggregated event graph construction

To answer the second sub-question of how freight train performance differs between operations with and without a DAS, a network representation of the corridor is provided through an aggregated event graph. An event graph is a Directed Acyclic Graph (DAG) $G(V, E)$, consisting of vertices/nodes (V) and arcs/edges (E). A DAG graph is a graph whose every edge has a direction, and is acyclic because it contains no directed cycles (Vo, 1988). In this study, the event graph is aggregated because each node represents the combined event information of all trains at a given geographical location, rather than a single train-specific event. At every node, relevant performance indicators (e.g., planned and actual passage times, deviations, or signal aspects) are aggregated statistically across trains.

In the macroscopic graph $G_{macro}(V, E)$, the nodes represent the timetable points and the edges are the corridors in between. The nodes contain the information per train of the original planned event times, the real-time planned event times, the executed event times, and the deviation from the planned and original timetable. In contrast, in the microscopic graph $G_{micro}(V, E)$, the nodes represent signals, both automatic and controlled, and contain information about the signal aspect, passage time, clearance time, release delay, and the cause of the signal aspect for each train. The edges represent the track sections between the signals and include the running times for each train on those tracks.

Both graphs were implemented using the NetworkX[1] Python package, which allows for the visualisation of large directed acyclic graphs. The nodes and edges were defined based on the infrastructure topology, using the ordered sequence of timetable points or signals to establish directional connectivity. Train-specific events were then mapped onto these edges/nodes, and the aggregated values form the basis for the statistical analysis presented in the next section. From the aggregated event graphs, four KPIs are derived to evaluate the operational performance of freight trains with and without a DAS. These include transport volume, deviation, yellow signals, and train interaction. The event graphs provide both temporal and spatial context for these indicators: transport volume is examined through the distribution of train movements across time and along the corridor, while deviation, yellow signals, and train interactions are directly linked to the aggregated events at the macro- and microscopic levels. The following section outlines the statistical approach applied to analyse these indicators.

3.3. Statistical analysis

RTL usage was identified using the RTL dataset, which records driving sessions where the advisory system was active. Each train in the combined dataset was referenced with this information and labelled accordingly as either a RTL user or a non-user. This grouping forms the basis for all analyses.

[1] NetworkX

3.3.1. Mann-Whitney U test

To determine the appropriate statistical approach for analysing the deviation data, the distribution of the variable was first examined in the data description section. Since the choice between parametric and non-parametric tests depends on whether the data follow a specific distribution (usually normal), a visual assessment was performed using a Quantile–Quantile (Q–Q) plot. A Q–Q plot compares the quantiles of the observed sample with those of a theoretical normal distribution, providing a graphical method to detect deviations from normality (Shapiro & Wilk, 1965). The deviation data were visually inspected for normality using a Q–Q plot, as formal tests such as the Shapiro–Wilk are overly sensitive for large datasets (Razali & Wah, 2011). The distribution of deviation was visually assessed using a Q–Q plot (see Figure 4.15b in the results, section 4).

In the case of normally distributed data with approximately equal variances, an independent two-sample *t*-test would be the preferred method for comparing the performance of RTL users and non-users. The *t*-test directly evaluates differences in group means and provides high statistical power when its assumption of normality is satisfied. However, when these assumptions are violated, the *t*-test can yield unreliable results because the mean is highly sensitive to extreme values.

The data showed a right-skewed pattern, indicating non-normality and supporting the use of non-parametric tests. Therefore, the non-parametric equivalent of the *t*-test, the Mann–Whitney U test, was selected instead. Unlike the *t*-test, the Mann–Whitney U test compares the relative rank distributions of two independent groups without assuming normality or equal variances (McKnight & Najab, 2010). It therefore provides a more robust comparison under real-world operational conditions, where train performance data often display heteroscedasticity and heavy-tailed distributions.

The Mann-Whitney U test tests whether the distribution of deviations for RTL users systematically differs from that of non-users. The test assumes that the two groups are independent. Although this assumption cannot be formally verified, a robustness check based on random subsampling is conducted after each test to assess the stability of the results. In this procedure, 30% of the data are randomly sampled to confirm that the conclusions remain consistent across subsamples. This step, therefore, serves as a robustness validation of the statistical findings.

The Mann-Whitney U test is applied per edge in the event graph, with the following hypotheses:

- Null hypothesis (H_0): There is no significant difference in driving performance between drivers with and without RTL.
- Alternative hypothesis (H_1): Drivers using RTL show significantly lower deviation from average running time.

The Mann–Whitney U statistic is calculated as:

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \tag{3.2}$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 \tag{3.3}$$

$$U = \min(U_1, U_2) \tag{3.4}$$

where n_1 and n_2 represent the sample sizes of the two groups, and R_1 and R_2 denote the sums of ranks for each group. The smaller of the two U values is used to determine the level of significance. The tests are evaluated at a significance level of $p < 0.05$, meaning that differences with a lower p-value are regarded as statistically significant.

In addition to statistical significance, effect sizes were computed to assess the magnitude of observed differences. For non-parametric comparisons performed using the Mann–Whitney U test, the effect size was derived from the standardized test statistic (Z) according to Cohen's definition:

$$\text{effect} = \frac{Z}{\sqrt{N}} \tag{3.5}$$

where N is the total number of observations included in the test. Following the conventions proposed by Cohen (1988), effect sizes of effect $= 0.10$, effect $= 0.30$, and effect $= 0.50$ were interpreted as small, medium, and large effects, respectively.

Macroscopic level

For the macroscopic evaluation, the deviation between the planned and executed event times is calculated at each timetable point for all trains. This comparison provides an overview of the punctuality performance across the corridor.

Microscopic level

On the microscopic level, the signals have no planned event times, only passing times. Using the Critical Path Method (CPM), as described by Cheng (1996), the sequence of events will be modelled to identify the critical path and quantify deviations from the ideal flow. CPM is selected for its ability to reveal timing dependencies and operational bottlenecks in the considered railway corridors. This allows for a comparative analysis of driver behaviour based on RTL usage. To examine the deviations on the track-based level, the average running time $\bar{r}_e$ is determined for $e \in E$ in the event graph. The reference running times are derived from the distributions of the original planned event times described in the data section. Based on these distributions, two operational categories are distinguished: heavy trains and regular trains. Heavy trains have longer running times due to their limited acceleration and speed restrictions on sections to reduce wear. To ensure a reliable reference running time, the average running time per category is calculated using only trains operating under nominal conditions. The VOS and TROTS datasets are combined to identify trains that fulfil these conditions. Some nodes are linked to timetable points based on geographical location, and only trains meeting the following criteria are included in the reference group:

- The train must have passed both consecutive timetable points with a deviation of no more than 180 seconds (3 minutes) from the scheduled time (van Infrastructuur en Waterstaat, 2020). This condition corresponds to equations 3.7 and 3.8.
- Only trains that did not pass yellow or red aspects at any of the signals are included, ensuring that the measured running times reflect conflict-free operating conditions unaffected by restrictive signalling (equation 3.9).
- The activity type a at both the current and next timetable point must be a through passage (DR), as expressed in equation 3.10.
- Each edge must contain at least 15 trains satisfying these criteria to ensure statistical reliability (equation 3.11).

The threshold of 15 trains per edge was selected after testing alternative cut-offs (10, 15, and 20), which showed that average running time estimates stabilised beyond 15 observations while preserving sufficient network coverage.

The average running time for each edge is computed as:

$$\bar{r}_e = \frac{\sum_{t \in T} z_{t,e} \cdot r_{t,e}}{\sum_{t \in T} z_{t,e}} \qquad \forall e \in E \tag{3.6}$$

Under the following conditions:

$$|p^{\text{plan}}_{t,d} - p^{\text{real}}_{t,d}| \leq 180 \qquad \forall t \in T, d \in D \tag{3.7}$$

$$|p^{\text{plan}}_{t,d+1} - p^{\text{real}}_{t,d+1}| \leq 180 \qquad \forall t \in T, d \in D \tag{3.8}$$

$$asp_{t,s} = GR \qquad \forall t \in T, \forall s \in S_t(d, d{+}1) \tag{3.9}$$

$$a_{t,d}, a_{t,d+1} = DR \qquad \forall t \in T, d \in D \tag{3.10}$$

$$\sum_{t \in T} z_{t,e} \geq 15 \qquad \forall e \in E \tag{3.11}$$

Where:

- E – set of all edges between consecutive signals $(s, s{+}1)$, $e \in E$
- T – set of trains considered in the analysis, $t \in T$
- S – set of signals along the corridor, $s \in S$

- D – set of timetable points along the corridor, $d \in D$
- $d(s)$ – timetable points located at the position of signal s
- $S(d, d{+}1)$ – set of all signals located between two consecutive timetable points d and $d{+}1$
- $S_t(d, d{+}1) \subset S(d, d{+}1)$ – subset of signals that were actually passed by train t between timetable points d and $d{+}1$
- $r_{t,e}$ – running time of train t over edge e [s]
- $a_{t,d}$ – activity of event of train t at timetable point d
- $p_{t,d}^{plan}$ – planned passing time of train t at timetable point d [s]
- $p_{t,d}^{real}$ – real (observed) passing time of train t at timetable point d [s]
- $asp_{t,s}$ – aspect shown by signal s for train t
- $z_{t,e}$ – binary variable, 1 if train t is included in the running time calculation for edge e, 0 otherwise

With the average running times per edge, the average absolute deviation from the average running time is computed as:

$$\bar{\Delta r}_e = \frac{1}{|T_e|} \sum_{t \in T_e} |r_{t,e} - \bar{r}_e| \qquad \forall e \in E \tag{3.12}$$

where r is the observed running time for a train over a given edge $e \in E$, and $\bar{r}_e$ is the average running time for that edge, as computed from a filtered reference group. This yields a distribution of absolute running time deviations per train and per edge. The absolute formulation is adopted because both early and late arrivals are considered deviations from the planned schedule and equally relevant in assessing punctuality. Once all deviations are determined, the aggregated event graphs are used to identify locations (edges) with the highest deviations. For each edge, the significance of performance differences between RTL users and non-users is then evaluated using the Mann–Whitney U test.

3.3.2. Chi-squared test

For the key performance indicators, yellow signals and train interaction, the categorical variables signal aspect and cause of aspect from the TROTS dataset are analysed to examine whether the occurrence of unplanned yellow signals differs between operations with and without RTL. During train operations, yellow signals are passed. In many cases, these are planned, for example, due to speed restrictions near switches or while passing through stations. However, yellow signals can also appear unplanned, such as when headways are reduced due to operational conflicts or when signalling systems experience failures.

To determine whether RTL usage influences the frequency of unplanned yellow signals, a Pearson's Chi-squared test of independence is applied. The Chi-squared test is suitable for this analysis because it evaluates the association between two categorical variables, in this case, RTL usage (user vs. non-user) and signal aspect (unplanned vs. planned yellow and green), without assuming any specific data distribution (McHugh, 2013; Pearson, 1900). It therefore provides an appropriate method for testing whether the proportion of unplanned yellow signals differs significantly between the two groups.

The test compares the observed frequencies O_i in each category to the frequencies E_i expected under the null hypothesis of independence.

The null and alternative hypotheses are formulated as:

- H_0: There is no association between RTL usage and the occurrence of unplanned yellow signals.
- H_1: RTL usage is associated with a different frequency of unplanned yellow signals.

The Chi-squared statistic χ^2 is calculated as:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}, \tag{3.13}$$

where k is the number of categories. Large deviations between observed and expected frequencies result in a higher χ^2 value, indicating a more substantial likelihood that the variables are not independent.

A significant result ($p < 0.05$) suggests that the distribution of yellow signal types differs between RTL users and non-users, indicating that RTL usage may influence how often trains passed unplanned yellow signals.

To better understand when and where unplanned yellow signals occur, the yellow signal events are also visualised over the aggregated event graphs. This spatial representation highlights the sections of the corridor where unplanned yellow aspects most frequently arise.

In addition to the signal-related analysis, the study also examines train interactions. Since RTL provides drivers with real-time information about preceding trains, this analysis focuses specifically on the interactions between freight trains and the trains operating directly in front of them. For each signal passage, the type of preceding train that had reserved the upcoming route section of the current freight train was identified.

The various categories of preceding trains considered in this analysis are presented in Table 4.11 in Chapter 4. These can be grouped into four main types: passenger trains, freight trains, empty freight trains, and other categories, which include international high-speed services (e.g., Thalys, ICE) and maintenance trains. The analysis evaluates whether different preceding-train scenarios are associated with variations in the causes of yellow signal passed and whether these effects differ between RTL users and non-users.

To account for multiple hypothesis testing, the False Discovery Rate (FDR) correction is applied using the Benjamini–Hochberg (BH) procedure (Benjamini & Hochberg, 1995). This correction is applied to both the Mann–Whitney U and Chi-squared test results whenever more than five independent tests are performed simultaneously. The BH method controls the expected proportion of falsely rejected null hypotheses (false positives) among all significant results, thus maintaining the overall reliability of the statistical inference. Without such a correction, performing multiple tests increases the likelihood of obtaining false positives, as even random variation can appear statistically significant when many hypotheses are tested simultaneously. The adjusted significance threshold $p_{\mathbf{BH}}$ is calculated as:

$$p_{\mathbf{BH}} = \frac{i}{m} \cdot Q, \tag{3.14}$$

where i is the rank of the individual *p*-value, m the total number of tests, and Q the desired false discovery rate (typically 0.05). This procedure ensures that significant results reported in the analysis are statistically robust and not due to random fluctuations from repeated testing.

In summary, this chapter has outlined the methodological framework applied in this study, including the data preparation, event graph construction, and statistical analysis techniques. The following chapter presents the results of these analyses, with the operational differences observed between RTL users and non-users on both the macroscopic and microscopic levels.

4

Results

This chapter outlines the data preprocessing steps leading up to the model analysis and presents the results. Specifically, for this analysis, three types of data were used: the VOS, TROTS and RouteLint (RTL) datasets as described in Chapter 3. Appendix A provides an overview of all datasets used in this research.

4.1. Data cleaning and fusion

This subsection describes the datasets used in the analysis. Additional data descriptions, including value counts and distributions, can be found in Appendix A.

All columns and their data types of the datasets are described in Tables 4.1, 4.2, and 4.3. The first step in preprocessing was selecting the appropriate time period: 1 August 2024 to 21 July 2025. The period from 17 to 24 April 2025 was excluded due to a technical malfunction in the RTL application, which failed to provide usable driving advice during that time. In the VOS dataset, all records with missing pass_time values were removed, and similarly, rows with missing signal_passage values were deleted from the TROTS dataset. These missing events typically corresponded to planned but unexecuted train movements, caused by downstream disturbances, rolling stock failures, or significant delays.

In the datasets, individual trains t were identified based on individual train number and traffic date combinations. Based on the sequence of signals and timetable points, the direction of each train was decided, either heading west or east. Trains heading from Tilburg and heading towards Eindhoven, Tilburg Industrie, or Vught Aansluiting were classified as eastbound. Likewise, trains operating from Vught towards Eindhoven were also considered eastbound. Additionally, trains entering the corridor at Eindhoven and heading towards Vught or Tilburg were classified as westbound, as were trains travelling from Vught Aansluiting in the direction of Tilburg. This classification ensures a consistent directional interpretation across the datasets to avoid cycles in the graphs.

The datasets were cleaned according to the steps described in Chapter 3. Before filtering, the VOS dataset contained information on 603,184 trains and TROTS on 310,581 trains. After cleaning, the resulting datasets contained 246,304 records in VOS and 697,881 in TROTS, corresponding to the 23,573 individual freight trains included in the analysis.

Table 4.1: Data description of dataset VOS.

Column Name	Data Type	Description
traffic_date	datetime64[ns]	Date of the train operation
operator	object	Operating company
trainnumber	int64	Train number
tp	object	Timetable point
act	object	Activity at timetable point
track	object	Track number
driving_characteristic	object	Train characteristic
original_schedule	object	Original schedule
o_dev_m	float64	Difference in minutes between original plan and actual plan
o_dev_dif_m	float64	Deviation measurement difference in minutes at the timetable point compared to the original plan
o_dev_s	float64	Difference in seconds between original plan and actual plan
o_dev_dif_s	float64	Deviation measurement difference in seconds compared to the original plan at timetable point
plan_actual	object	Actual schedule
a_dev_m	float64	Time difference in minutes between actual plan and execution
a_dev_dif_m	float64	Deviation measurement difference in minutes at the last timetable point
a_dev_s	float64	Time difference in seconds between actual plan and execution
a_dev_dif_s	float64	Deviation measurement difference in seconds at the last timetable point
pass_time	datetime64[ns]	Executed time of train of activity at the timetable point
length	float64	Train length
weight	float64	Train weight
next_point	object	Next timetable point in the actual schedule
direction	object	Direction of train (eastbound/westbound)
category	object	Category of train (heavy/regular)

Table 4.2: Data description of dataset TROTS.

Column Name	Data Type	Description
traffic_date	datetime64[ns]	Date of the train operation
trainnumber	int64	Train number
operator	object	Operator of the train
tp	object	Timetable point
act	object	Activity at the signal
driving_characteristic	object	Train characteristic
signal_id	object	Signal ID
signal_aspect	object	Signal aspect at time of passing (green, yellow, or red)
signal_passage	datetime64[ns]	Time the signal was passed
signal_clear	datetime64[ns]	Time the signal was set to safe (green)
signal_release_delay	int64	Delay in seconds between time of passing yellow signal and the time of clearance of next signal
type_IA	object	Signal type
tp_dev	float64	Deviation in seconds between the actual plan and executed plan at corresponding timetable point
cause	object	Cause of a yellow signal aspect
block_occupancy	object	Block occupied by train
preceding_train	int64	Preceding train in front of current train
block_clear	datetime64[ns]	Block release time
direction	object	Direction of train (eastbound/westbound)
category	object	category of train (heavy/regular)

Table 4.3: Data description of dataset RTL.

Column Name	Data Type	Description
Timestamo	object	Login time in the RTL system
Train number	int32	Train number
Date	datetime64[ns]	Date
User	object	Operator (person logging in)
Drivingadvice use	bool	True if driving advice was used, otherwise False
Use GPS	bool	True if GPS was used, otherwise False
GPS match	bool	True if GPS signal matched within 100 meters, otherwise False
Appversion	object	Version of the app used
File Date	datetime64[ns]	Date of the record file

To identify whether a train driver used RTL, RTL user data was checked for matches based on a combination of date, train number, and freight operator. This required all three datasets to have consistent naming conventions for the operators. Therefore, operator names were standardised across the datasets. Appendix A.4 shows a full list of operators and their abbreviations.

A new Boolean column was added to the VOS and TROTS datasets to indicate whether a specific train was found in the RTL user data. This column serves as a boolean to indicate whether RTL was used by the driver for that train.

- True: if the train driver used RTL;
- False: if the train driver did not use RTL.

The correlation heatmap in Figure 4.1 shows the relationships between several continuous variables. As expected, there is a strong correlation (around 1) between the columns o_dev_dif_m and o_dev_dif_s, since both describe the delay magnitude in minutes and seconds. In addition, a correlation of 0.065 is observed between o_dev_s and o_dev_dif_s. The former represents the deviation between the original and the actual timetable, whereas the latter captures the change in this deviation between consecutive timetable points. This low correlation indicates that, despite substantial deviations

between the actual and the original timetables, the deviation between successive timetable points does not increase. A weak but positive correlation is observed between o_dev_s and a_dev_s. This indicates that, although deviations from the original timetable occur, they do not necessarily impact the delay of the adjusted timetable. Train weight and length show a correlation of 0.3, as longer trains are expected to have a higher overall mass. However, when compared with the deviations from the adjusted timetable (a_dev_s), both variables show little to no correlation. This suggests that heavier trains do not necessarily experience greater deviations from their scheduled running times.

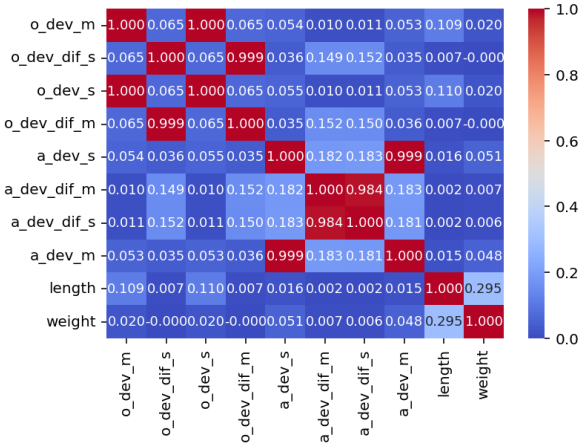


Figure 4.1: Heatmap of VOS continuous variables.

The descriptive statistics in Table 4.4 show notable directional differences. In particular, the mean of the deviation in seconds and minutes (a_dev_s and a_dev_m) is higher for eastbound traffic. Eastbound trains are also, on average, heavier than westbound trains. This finding is further illustrated in Figure 4.2, where the distribution of train weights reveals a clear asymmetry between east- and westbound operations. This distinction is relevant for the analysis, in which average running times are compared per edge, as described in Chapter 3. Identifying two groups of trains with distinct characteristics allows for more reliable and meaningful comparisons in the following sections.

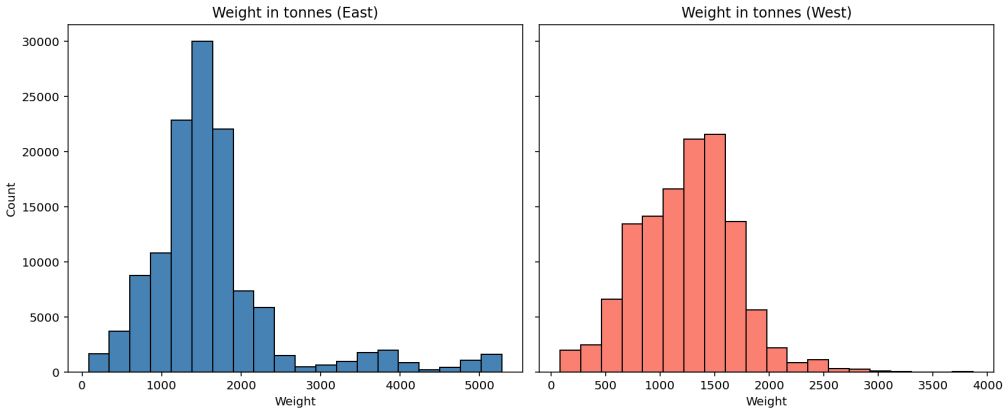


Figure 4.2: Histograms of weight in east and west direction

.

Table 4.4: Descriptive statistics of key variables per direction.

Variable	Direction	Count	Mean	Std	Min	24%	50%	75%	Max
a_dev_s	East	124,281	166.66	978.91	-6,600	-57	60	229	86,746
	West	121,997	98.43	622.43	-36,674	-68	49	203	15,387
a_dev_m	East	124,281	2.63	16.20	-110	0	1	3	1,445
	West	121,997	1.52	10.20	-611	-1	0	3	256
a_dev_dif_s	East	124,281	7.00	185.00	-6,734	-40	-5	24	14,857
	West	121,997	5.43	123.79	-9,411	-36	-2	31	9,371
a_dev_dif_m	East	124,281	0.12	3.06	-111	-1	0	0	246
	West	121,997	0.08	2.03	-156	-1	0	0	156
Lenght (m)	East	124,202	541.36	134.90	12	507	576	628	735
	West	121,945	542.03	137.87	12	508	578	629	736
Weight (t)	East	124,202	1,642.07	859.61	82	1,192	1,510	1,804	5,278
	West	121,945	1,243.60	446.11	83	904	1,269	1,537	3,869

The applied cleaning steps ensure sufficient data quality and coverage across the study period. The next sections use these datasets to evaluate RTL's operational effects on train performance.

4.2. Transport volume

Figure 4.3 shows how the total number of trains operating along the corridor has evolved over time. At the beginning of the 80-week maintenance period (November), the number of trains increased by almost 80 trains per day. During the week of 19 October 2024, however, no trains ran between Boxtel and Tilburg, and only services between Eindhoven and Vught were included in the dataset due to maintenance work. Towards the end of the year, train activity declined again around the holiday period and New Year, when only a limited number of passenger services and even fewer freight trains were operating. A further drop is observed in early June 2025, corresponding to the NS strikes on 6, 10, and 13 June, when no NS passenger trains ran in the southern province (i.e., the Brabantroute).

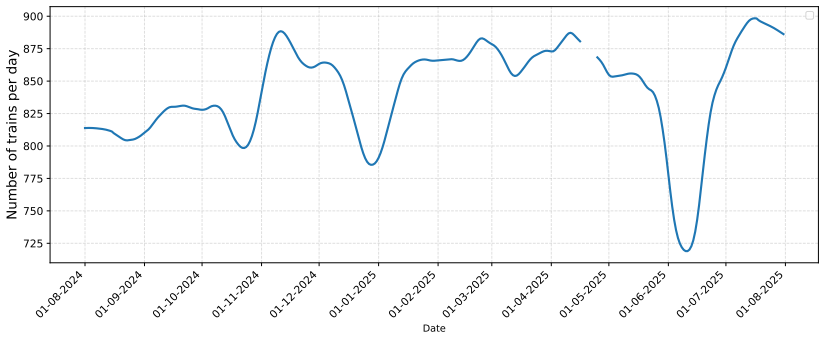
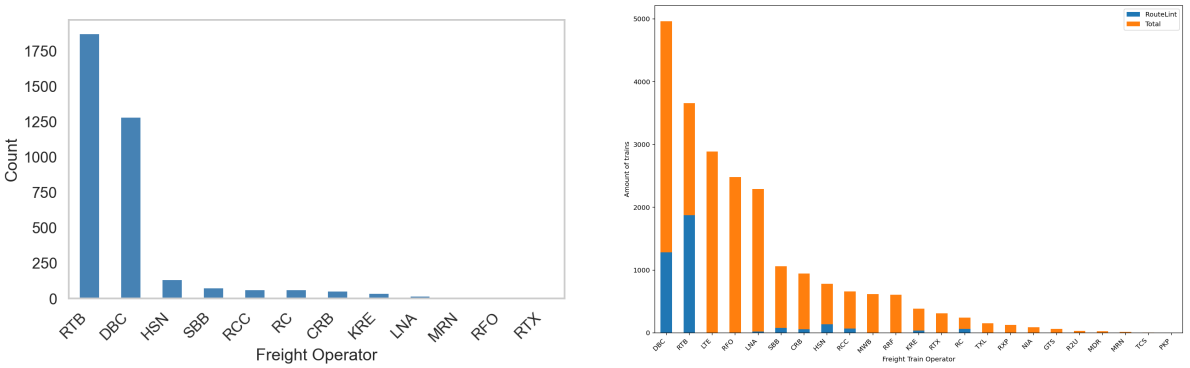


Figure 4.3: Total number of trains per day on the corridor Tilburg-Eindhoven.

Figure 4.4a presents the number of trains logged into RTL per freight operator. Out of a total of 23,528 trains running on the Brabantroute during the study period, 3,635 trains (15.45 %) made use of RTL. As shown in the figure, Rurtalbahn Cargo (RTB) and Deutsche Bahn Cargo (DBC) account for the highest usage share. Figure 4.4b also shows the relative share of RTL use per operator compared with the total number of trains operating on the corridor. This shows that some major freight operators, such as LTE and RFO, chose not to use RTL during the free trial period.

RTL was made freely available only from the start of the 80-week maintenance period. Before, DBC had already used an older version of the system (hereafter referred to as version 1.0), classified as a driver information system, as it displayed real-time timetable information but lacked features such as driving advice. As no usage data are available for version 1.0, it was excluded from the analysis. Between November 2024 and early May 2025, version 2.0 of RTL (the version equipped with DAS) was already in use by most freight operators and by a limited number of DB Cargo drivers. In early



(a) Number of RTL users per freight operating company.

(b) RTL usage in relation to the total number of trains on Brabantroute.

Figure 4.4: Overview of RTL usage per operator and in relation to total freight traffic.

May 2025, RTL 2.0 was officially released to all freight operators, including DB Cargo, and became the standard version in use for the remainder of the study period.

To examine the temporal development of RTL use, Figure 4.5 shows the evolution of the number of users throughout the year. Usage increased after the start of the 80-week period and rose sharply from May onwards. Since then, approximately 38% of all drivers operating on the Brabantroute have actively used RTL. Appendix A includes the figures of RTL usage per freight operator over time.

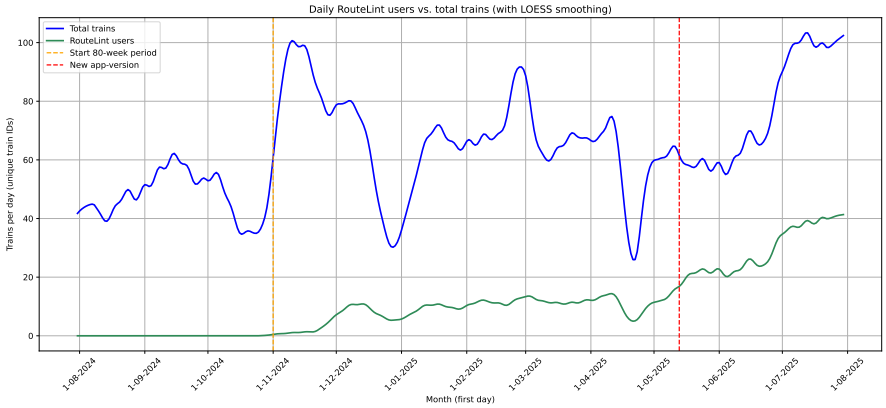


Figure 4.5: Growth in RTL version 2.0 usage over a year from August 2024 to August 2025.

In order to spatially see the transport volume, the aggregated event graphs were constructed, where each train is assigned a corridor direction, either eastbound or westbound, to prevent cycles within the graph.

Figures 4.6 and 4.7 show the macroscopic graphs, in which the edges represent the track sections between timetable points and include the total number of train movements along each track section. The figures display that the busiest part of the corridor lies between Tilburg and Eindhoven, with approximately 8,800 train movements per direction. Fewer trains operated between Vught–Eindhoven (945) and Vught Aansluiting–Tilburg (1,642), indicating that the sections from Tbr to Tba and from Btl to Tgra were the most intensively used.

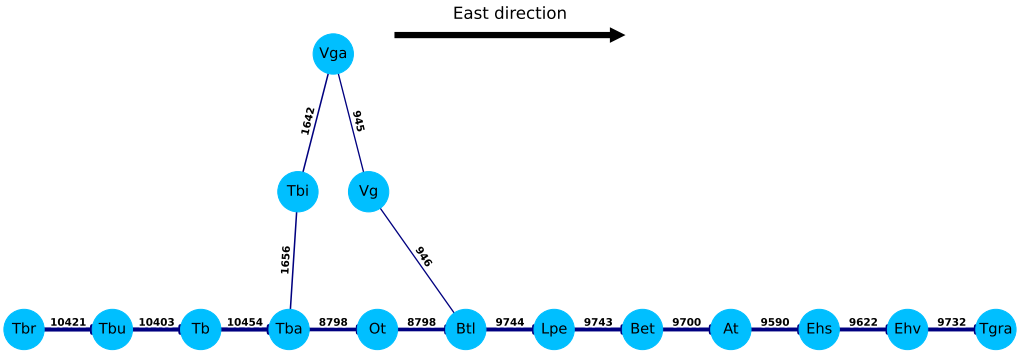


Figure 4.6: Travel volume in the east direction on the macroscopic level.

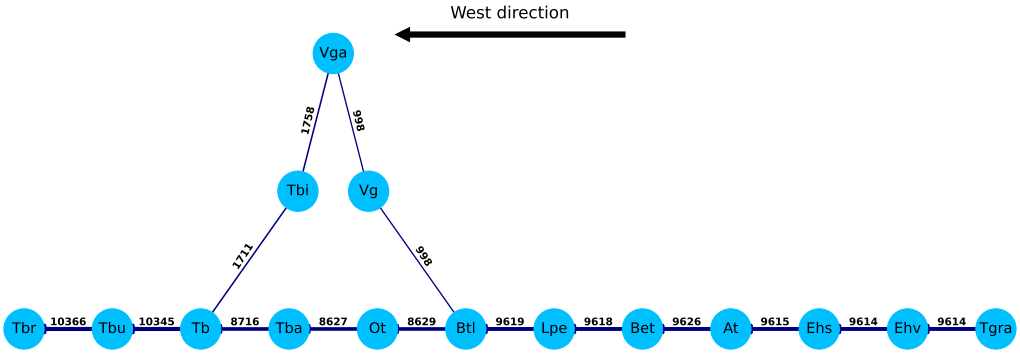


Figure 4.7: Travel volume in the west direction on the macroscopic level.

To obtain a more detailed view of where on the tracks freight trains most frequently operated, Figures 4.8 and 4.9 display the constructed microscopic graphs for eastbound and westbound directions. The eastbound network consists of 140 nodes and 209 edges, while the westbound network includes 140 nodes and 204 edges. Enlarged versions of both graphs are provided in Appendix C. In these graphs, the thickness of each line represents the frequency of train movements along that edge. In the eastbound direction, more than 80% of trains follow a similar route, except around Eindhoven, where numerous switches allow for alternative paths and greater heterogeneity occurs. In contrast, route variation in the westbound direction is less pronounced, except near Tilburg, where more alternative routes are used as three platforms (2–4) are reserved for westbound operations.

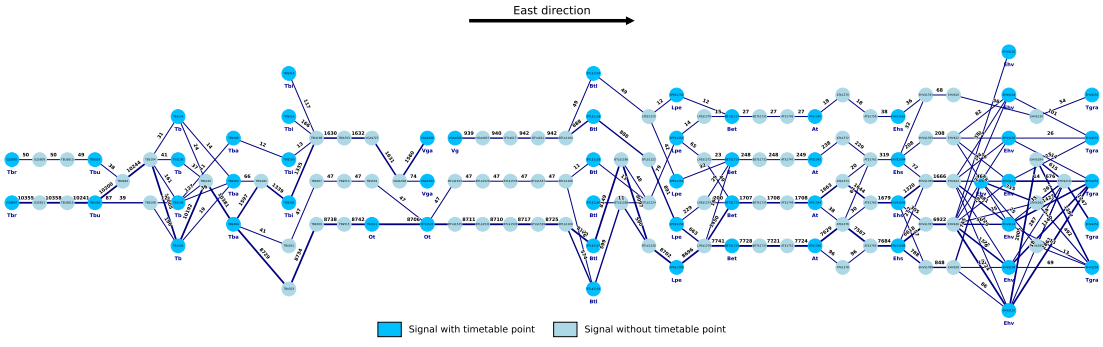


Figure 4.8: Transport volume in the east direction on the microscopic level.

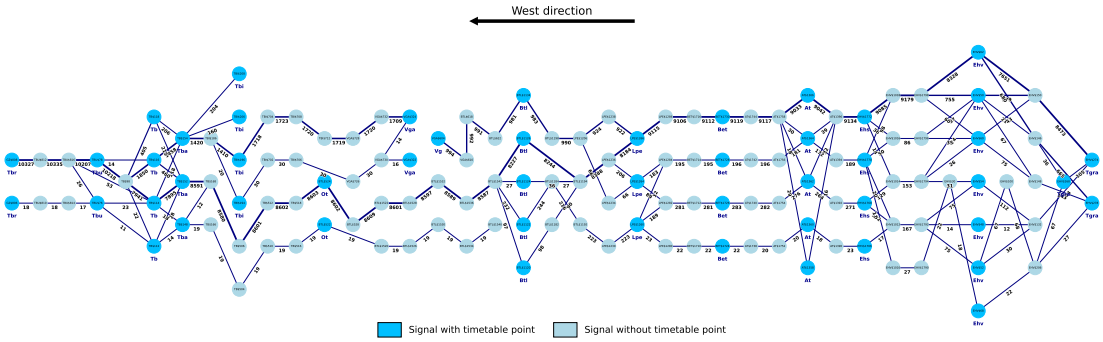


Figure 4.9: Transport volume in the west direction on the microscopic level.

Now that the number of users is discovered, and the busy corridors and tracks are identified, the next subsection will zoom in on the deviation per arc over time and in the graphs, to test whether there is a difference in deviation between RTL users and non-RTL users.

4.3. Deviation

Figure 4.10 shows the average absolute deviation on the Brabantroute between 1 August 2024 and 31 July 2025. At the start of the 80-week period, a relatively high average deviation is observed. From November onwards, the average deviation increases sharply to around 390 seconds per day, which coincides with a significant rise in transport volume during that period. During the holiday period in December, the average delay decreases due to the reduced freight activity. From January onwards, the average daily deviation lies around 280–300 seconds.

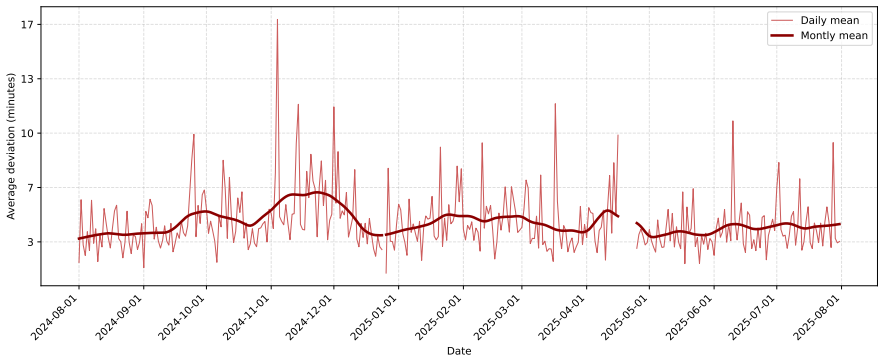


Figure 4.10: Average absolute deviation on Brabantroute over time.

To further examine the difference in deviation between RTL users and non-users, Figure 4.11 shows the daily average deviation for both groups, with a 7-day LOWESS smoothing applied to highlight underlying trends. Throughout most of the period, RTL users show lower average deviations compared to non-RTL users. However, around the end of November 2024, the average deviation of RTL users exceeds that of non-users. At the beginning of the 80 weeks, the use of RTL was limited only to a few trains per day (can be observed in Figure 4.5), which makes the average less reliable. On the 22nd of November 2024, only one train operated with RTL with a deviation of approximately 16 minutes. Furthermore, Figure 4.12 shows the same data smoothed with a 30-day LOWESS function to provide a more aggregated view of the trends.

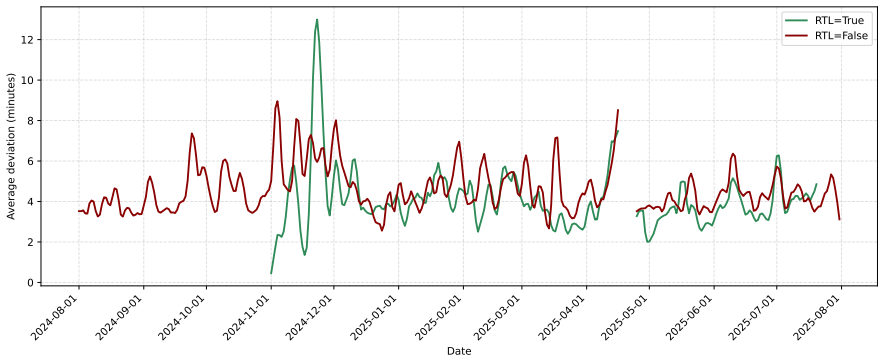


Figure 4.11: Weekly average absolute deviation from planned timetable split by RTL usage.

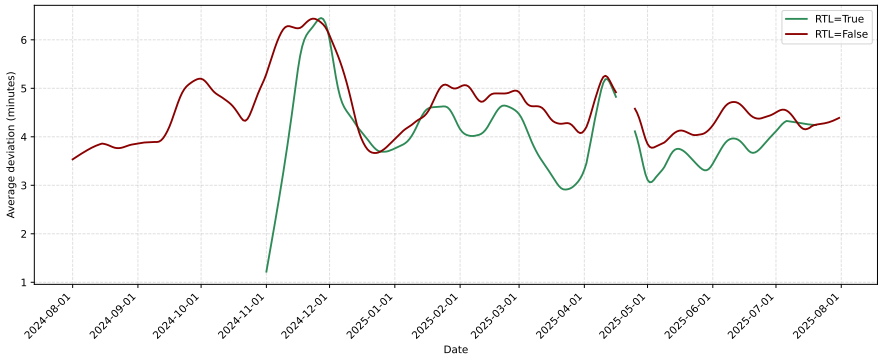


Figure 4.12: Monthly average absolute deviation from planned timetable split by RTL usage.

Both early and late arrivals can negatively affect the overall railway operation. Early arrivals may lead to operational conflicts when the route ahead is not yet cleared, while late arrivals can cause knock-on delays for following trains. Figure 4.13 illustrates how RTL trains differ in terms of early and late arrivals. In general, RTL-assisted trains experience fewer and smaller delays, although early arrivals still occur within this group. However, the average early deviations of RTL users typically remain below those of non-RTL users.

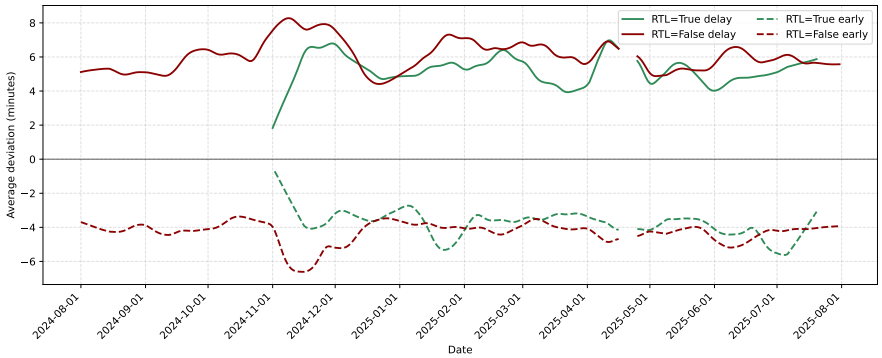


Figure 4.13: Monthly average late and early arrivals at timetable points split by RTL usage

To gain further insights into where these differences in delay and early arrivals occur, the following subsection analyses the deviations at both macro- and microscopic spatial levels.

4.3.1. Macroscopic level

At the macroscopic level, deviations are aggregated per timetable point. As described in Chapter 3, the VOS dataset contains both the planned and executed timestamps for each timetable point. The difference between these two values represents the deviation from the planned schedule. At the macroscopic level, all event times and corresponding deviations are aggregated per timetable point. Trains deviating by fewer than 180 seconds from the actual timetable are considered on time. Table 4.5 summarises how many trains per timetable point met the criteria of this punctuality threshold. Timetable points Tbi and Ehv show the lowest percentages of trains on time (52.88% and 53.56%, respectively), whereas Tbu and Tb exhibit the highest on-time rates (63.06% and 62.65%).

Table 4.5: Punctuality per timetable point (denoted as DRP in the table) based on arrival deviation (threshold = ±180 s).

d	n_{total}	$n_{\text{on\_time}}$	$n_{\text{deviation}}$	n_{delay}	n_{early}	%On time	%Deviation	%Delay	%Early
Tb	21012	13164	7847	5811	2036	62.65	37.35	27.66	9.69
Tbu	20788	13108	7679	5561	2118	63.06	36.94	26.75	10.19
Tbr	20787	12527	8259	6429	1830	60.26	39.73	30.93	8.80
Ehv	19525	10458	9065	6631	2434	53.56	46.43	33.96	12.47
Btl	19504	11271	8231	5692	2539	57.79	42.20	29.18	13.02
At	19376	10727	8647	5786	2861	55.36	44.63	29.86	14.77
Bet	19370	11218	8150	5453	2697	57.91	42.08	28.15	13.92
Lpe	19363	11621	7740	5216	2524	60.02	39.97	26.94	13.04
Tgra	19350	11182	8166	5450	2716	57.79	42.20	28.17	14.04
Ehs	19243	10418	8823	6132	2691	54.14	45.85	31.87	13.98
Tba	19171	11408	7758	5103	2655	59.51	40.47	26.62	13.85
Ot	17431	10616	6815	4546	2269	60.90	39.10	26.08	13.02
Vga	5342	3188	2152	1372	780	59.68	40.28	25.68	14.60
Tbi	4098	2167	1930	1015	915	52.88	47.10	24.77	22.33
Vg	1944	1077	866	576	290	55.40	44.55	29.63	14.92

While Table 4.5 provides an overview of punctuality per timetable point, Figure 4.14 offers a spatial perspective of these deviations along the network. The figure maps average deviations for eastbound and westbound flows, with and without RTL. The colour legend ranges from green (0–3 minutes, the punctuality threshold) to dark red (absolute deviation > 6 minutes), in steps of one minute. The graphs illustrate the directional differences in deviation between eastbound and westbound operations. In the eastbound direction, the largest deviations from the timetable occur around Eindhoven, whereas in the westbound direction, higher deviations are mainly observed near Tilburg (Tbi, Tbr). Overall, more nodes in the eastbound direction have deviations exceeding five minutes (4 when using RTL, 8 without the use of RTL), while in the westbound direction, only three nodes appear in red or dark red, only if RTL was not used.

Across almost all timetable points, the average absolute deviation is lower for RTL users. The largest improvement is found in the eastbound direction at Tbi, with a reduction of 143 seconds. This can be explained by the fact that, in this direction, most departing trains are through trains, and RTL is primarily used to facilitate entry into the network, resulting in smaller deviations per timetable point.

In the westbound direction, the largest improvement is observed at Vga (105 seconds), where trains diverge from ’s-Hertogenbosch towards Vught. The smallest improvements in the eastbound direction are found at Tbu (24 s), Tbr (23 s), and Ot (20 s). In the westbound direction, the smallest differences occur at Tgra (23 s) and Tba (26 s), both of which are merging points: at Tba, routes from Oisterwijk and Tilburg Industrie converge, while at Tgra, routes from Helmond and Geldrop merge.

Before the statistical testing, the deviation distributions were examined using both histograms and Q–Q plots (Figure 4.15), as discussed in Chapter 3. The histogram shows a peak around zero with a long right-hand tail, indicating that most trains operated close to schedule, while a smaller subset experienced substantial delays. Therefore, the Mann–Whitney U test was used to compare RTL vs. non-RTL per timetable point.

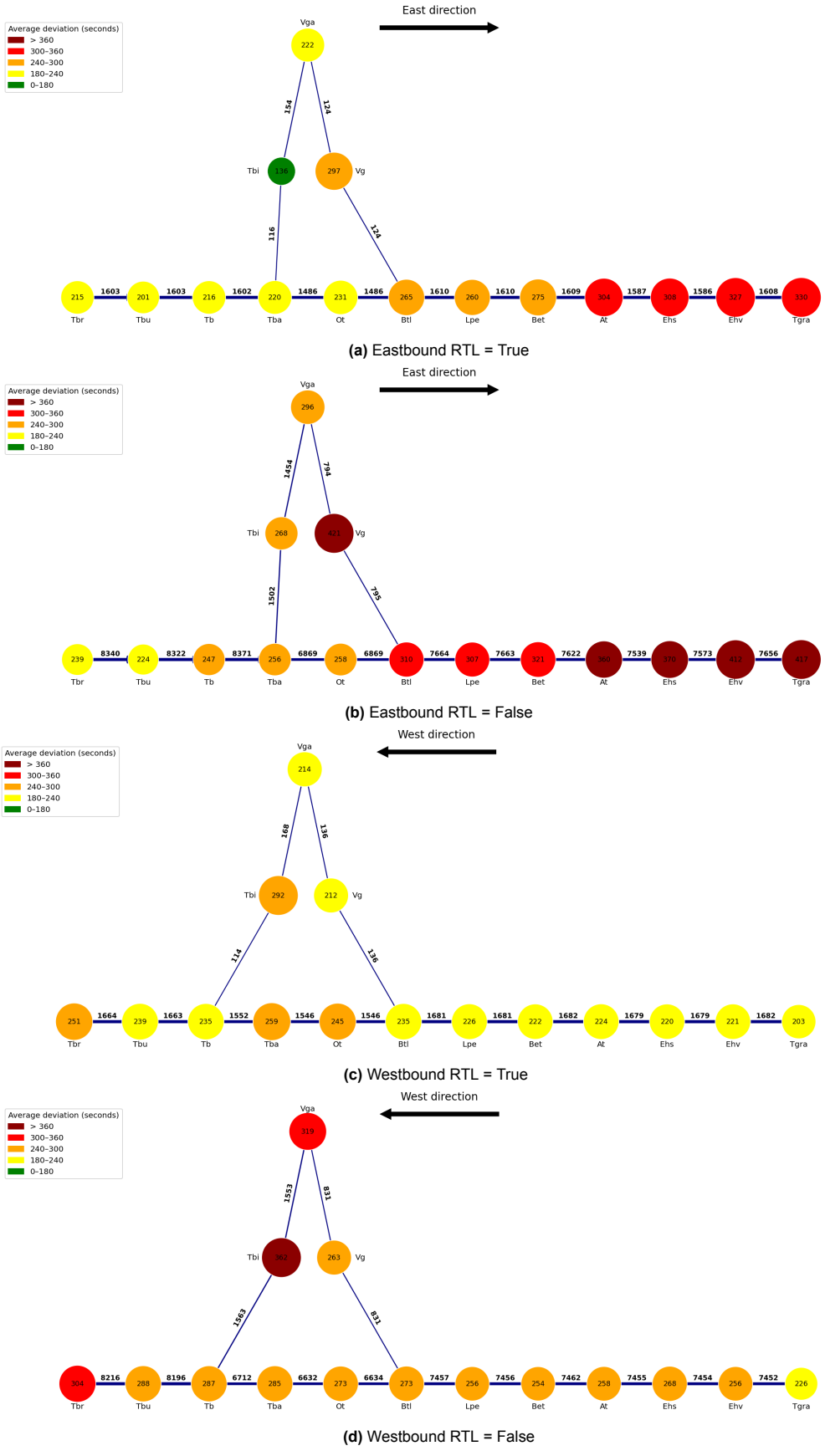
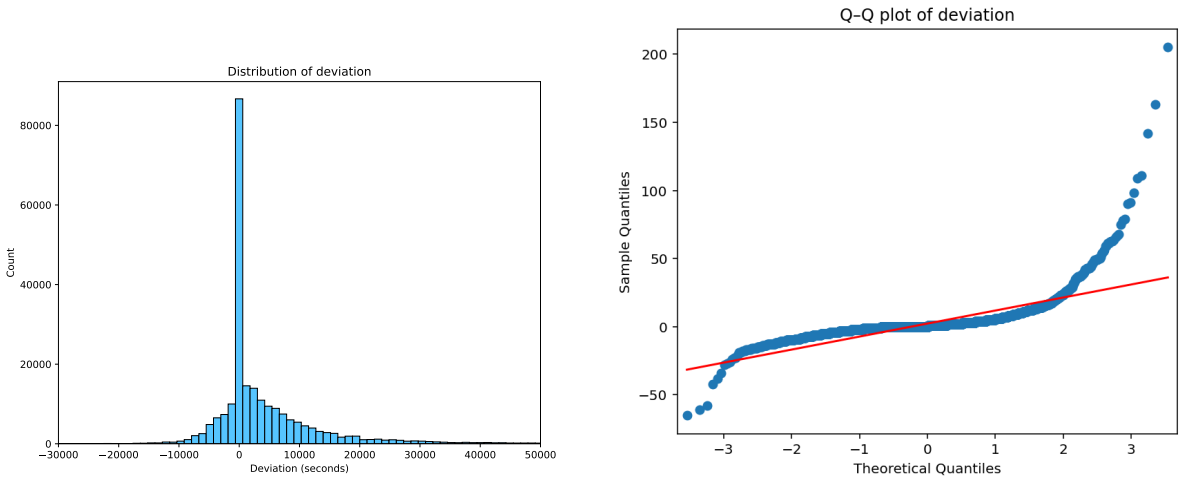


Figure 4.14: Average absolute deviations from the timetable per timetable point for eastbound and westbound directions, comparing trains with and without RTL.



(a) Histogram of the running time deviation in seconds.

(b) Q–Q plot of the running time deviation.

Figure 4.15: Distribution of running time deviations: (a) histogram and (b) Q–Q plot against the theoretical normal distribution.

Table 4.6 summarises the median of the deviations, U statistics, effect sizes (effect), and (Benjamini–Hochberg adjusted) p-values (p and p_{BH}). The largest significant difference in delay between drivers with and without RTL was found in the eastbound direction at Tbi (U = 74112.5, p < 0.05). The median delay was 66.5 seconds lower among RTL users, indicating that the typical deviation in this group was smaller. Other significant differences for eastbound traffic were observed at major passenger stations Ehv (-25s) and Tb (-10 s), and large junctions Tgra (-26.5s), Tba (-18s), and Vga (-31.5s), after correcting for false positives (p_{BH}<0.05). The smallest median differences can be found at Tbr (-2.5 s) and Tbu (-4.5s), the two timetable points before the large passenger station Tilburg; however, these small differences were not significant.

In the opposite direction, the largest significant difference can be observed at Ehs with a median difference of 33 seconds. Other large significant differences can be observed at stations Ehv (-28s), Tbr (-26s), and Tb (-23s). Other substantial significant differences were found at stations Ehv (–28 s), Tbr (–26 s), and Tb (–23 s). The smallest median differences were observed at Ot (–3 s) and Tba (–5 s). The p-values for these timetable points exceed 0.05, and thus the null hypothesis cannot be rejected. At station Tbi, the difference in medians was 15.5 seconds, indicating that RTL users typically showed slightly larger deviations than non-users. However, with a p-value of 0.5094, this difference is not statistically significant and is likely due to random variation.

In both directions, large and significant differences in the medians of deviations were found at the major stations Ehv and Tb, as well as at the timetable points located directly after these stations. For the eastbound direction, these include the diverging junctions Tba and Tgra, while in the westbound direction, they comprise the stations Ehs, Tbu, and Tbr. The westbound direction contained a higher number of timetable points (five more), with significant differences between RTL users and non-users. However, the largest median differences occurred in the eastbound direction, with the most pronounced difference of 66.5 seconds at Tbi. Thus, while the westbound direction exhibited more consistent differences in deviation between users and non-users, the eastbound direction showed larger median differences. Although the median differences at Vught were substantial in both directions (–20s eastbound and –33s westbound), the Mann–Whitney U test results were not significant, indicating that the rank distributions of the two groups largely overlapped. In both directions, effect sizes were small (effect ≤ 0.10), suggesting that the distributions still overlapped considerably, despite a consistent shift towards lower deviations for RTL users. Smaller intermediate stations, such as Ot, show no significant differences due to the homogeneous track layout with few switches or alternative routes.

While the macroscopic analysis provides an overview of where significant differences in deviations occur across timetable points, it does not capture how these differences relate to individual track segments. Therefore, the following subsection examines the data at the microscopic level to gain more detailed insight into the spatial characteristics of these deviations.

Table 4.6: Mann–Whitney U test results per timetable point and direction (east/west), comparing deviation between trains with (RTL=True) and without RTL(RTL=False).

D	n_T	n_F	Median$_T$ (s)	Median$_F$ (s)	Δ(T-F)	U	p	effect	p_{BH}
Direction: East									
Tbi	156	1259	71.5	138.0	-66.5	74112.5	$5.619{\cdot}10^{-9*}$	0.1330	0.00001*
Tgra	1608	6467	159.5	186.0	-26.5	4837096.5	$1.48{\cdot}10^{-5*}$	0.0482	0.0001*
Ehv	1598	6510	170.0	195.0	-25.0	4915951	0.0007*	-0.064 0.045	0.00035*
Tba	1601	7063	101.0	119.0	-18.0	5367087.0	0.0015*	0.0341	0.0056*
Vga	278	1485	92.5	124.0	-31.5	182507.0	0.0021*	0.0731	0.0064*
Tb	1602	7093	101.0	111.0	-10.0	5463861.0	0.0165*	0.0257	0.0412*
Bet	1609	6471	134.0	151.0	-17.0	5037025.0	0.0437*	0.0224	0.0936
Ot	1485	5988	113.0	120.0	-7.0	4303258.5	0.0550	0.0222	0.1030
Tbu	1602	7031	107.5	112.0	-4.5	5482614.0	0.0974	0.0178	0.1623
Ehs	1587	6391	158.0	172.0	-14.0	4948751.5	0.1358	0.0167	0.1874
Lpe	1609	6470	129.0	135.0	-6.0	5080749.5	0.1374	0.0165	0.1874
At	1612	6452	163.0	171.0	-8.0	5087677.0	0.1779	0.0150	0.2041
Btl	1619	6528	141.0	146.0	-5.0	5173516.5	0.1905	0.0145	0.2041
Vg	124	484	135.0	155.0	-20.0	27701.5	0.1864	0.0536	0.2041
Tbr	1602	7031	128.5	131.0	-2.5	5534672.5	0.2805	0.0116	0.2805
Direction: West									
Ehs	1679	6304	126.0	159.0	-33.0	4723762.0	$1.254{\cdot}10^{-11*}$	0.0758	$1.88{\cdot}10^{-10*}$
Ehv	1702	6376	123.0	151.0	-28.0	4924466.5	$4.924{\cdot}10^{-9*}$	0.0653	$3.32{\cdot}10^{-8*}$
Tbr	1664	7006	112.0	138.0	-26.0	5339560.0	$9.682{\cdot}10^{-8*}$	0.0573	$4.84{\cdot}10^{-7*}$
Tb	1671	7119	103.0	126.0	-23.0	5502900.0	$1.868{\cdot}10^{-6*}$	0.0508	$7.01{\cdot}10^{-6*}$
Tbu	1664	7007	107.0	124.0	-17.0	5427214.0	$1.154{\cdot}10^{-5*}$	0.0471	$3.46{\cdot}10^{-5*}$
Btl	1689	6342	134.0	149.0	-15.0	5019105.5	$6.99{\cdot}10^{-5*}$	0.0444	0.0002*
At	1689	6321	130.0	147.0	-17.0	5009308.5	$9.844{\cdot}10^{-5*}$	0.0435	0.0002*
Bet	1682	6310	122.0	139.0	-17.0	4987710.5	0.00015*	0.0424	0.0003*
Lpe	1681	6304	124.0	137.0	-13.0	5041965.0	0.0023*	0.0342	0.0034*
Tgra	1682	6304	102.0	111.0	-9.0	5045481.5	0.0023*	0.0341	0.0034*
Vga	304	1562	130.0	151.5	-21.5	216434.5	0.0146*	0.0565	0.0199*
Ot	1546	5849	141.0	144.0	-3.0	4399058.0	0.1016	0.0190	0.1246
Vg	136	465	130.0	163.0	-33.0	28756.5	0.1080	0.0656	0.1246
Tba	1552	5902	149.0	154.0	-5.0	4487535.0	0.2205	0.0142	0.2363
Tbi	184	1363	231.5	216.0	15.5	129149.0	0.5094	0.0168	0.5094

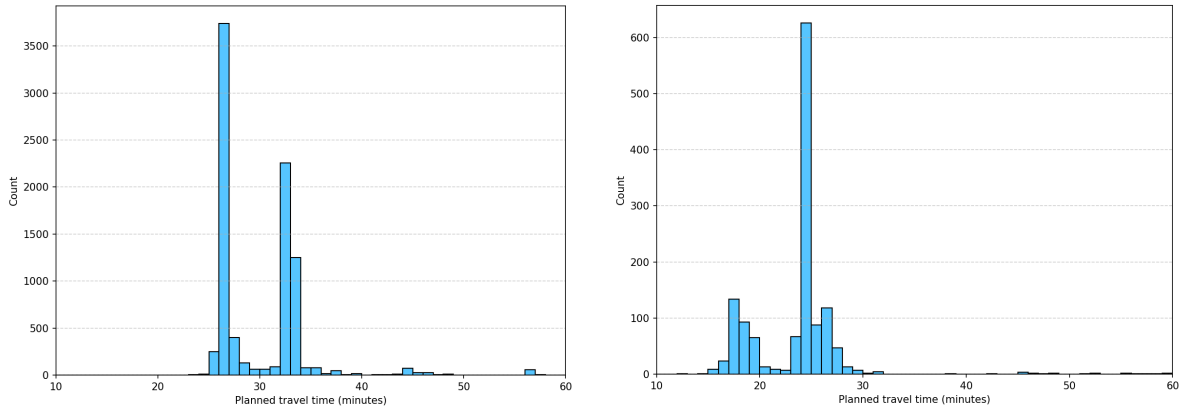
*Significant at $p < 0.05$.

4.3.2. Microscopic level

At the microscopic level, the analysis focuses on the travel times over the tracks between signals, represented as edges in the network. Before computing the average running times per edge, the distribution of the originally planned running times, derived from the original timetable of the VOS dataset prior to train operations, between timetables was examined in more detail to identify train categories. Figure 4.16 shows the bimodal distributions of two segments. From these distributions, two main categories of trains can be identified: heavy and regular. A train classified as heavy does not necessarily refer to its actual weight, but rather to the path reserved for heavy trains in the timetable. In the daily planning, two paths per hour are reserved for heavy trains to accommodate slower freight movements and prevent potential delays to following services. The most distinct heavy train paths can be found in the eastbound direction between Tilburg and Eindhoven, where planned running times exceed 30 minutes, mainly due to coal transport from the Port of Rotterdam to Germany. Similarly, westbound heavy trains are identified between Vught and Tilburg when the planned running time exceeds 22 minutes. All remaining trains are classified as regular.

Before calculating the average running time for each edge, the dataset was filtered to ensure representativeness. Edges used fewer than 15 times during the entire study period were excluded, as these typically correspond to alternative paths used only during large disruptions. For the remaining edges, only trains meeting the criteria described in the methodology, see Chapter 3. These constraints were defined to ensure that the computed average running time $\bar{r}_e$ reflects nominal operating conditions rather than disturbance-affected runs. In total, the average running time was computed for 105 eastbound and 104 westbound edges.

Figures 4.17–4.18 show the spatial pattern of average running times $\bar{r}_e$, while Figures 4.19–4.20 highlight the highest average absolute deviations, where the line thickness corresponds to the magnitude of the deviation. Two spatial patterns stand out. First, the longest running times occur in the corridor between Tbi - Vga for both directions: in the eastbound direction (TBI$703→VGA$727), the average running time is 449s for heavy trains and 439s for regular trains; in the westbound direc-



(a) East direction, Tilburg to Eindhoven

(b) West direction, Vught aansluiting to Tilburg

Figure 4.16: Distribution of the original planned running times between timetable points.

tion (VGA$728→TBI$721), the corresponding averages are 322s and 301s. Consistently, these long edges also show large average deviations 50s eastbound; 42s westbound), which is explained by the large inter-signal distances (≈10km), which increases the variability in running times. Second, around large stations Tilburg and Eindhoven, both the running times and the average deviations increase along their sections. The effect sizes, derived from the U statistic, are predominantly negligible to small (effect ≈ 0.03–0.13). Two eastbound edges exhibit stronger effects: AT$1346 → AT$1378 (effect ≈ 0.34) and TB$138 →TB$150 (effect ≈ 0.29), corresponding to small–medium magnitudes in this non-parametric context. These sections contain diverging/merging movements and possibly more frequent restrictive situations (unplanned yellow aspects), which further increase variability even under nominal operating conditions.

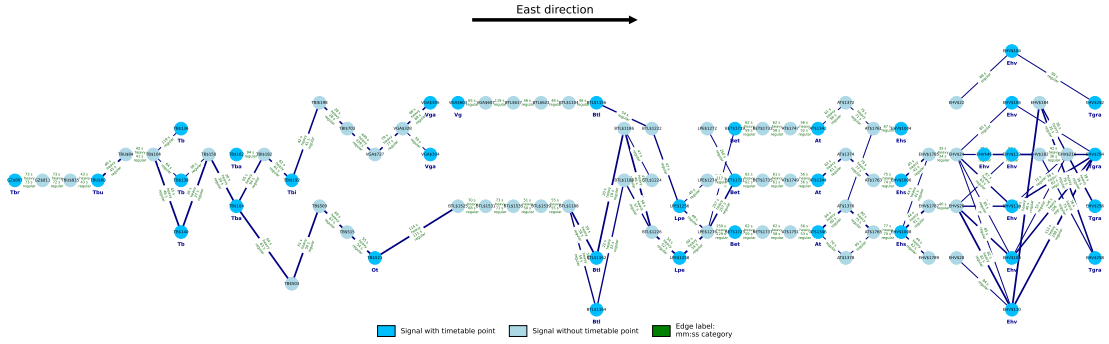


Figure 4.17: Average running time per edge per category in the east direction.

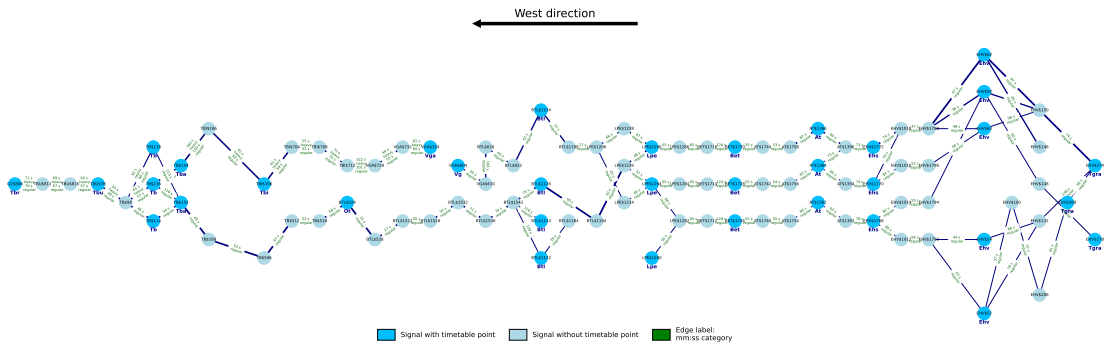


Figure 4.18: Average running time per edge per category in the west direction.

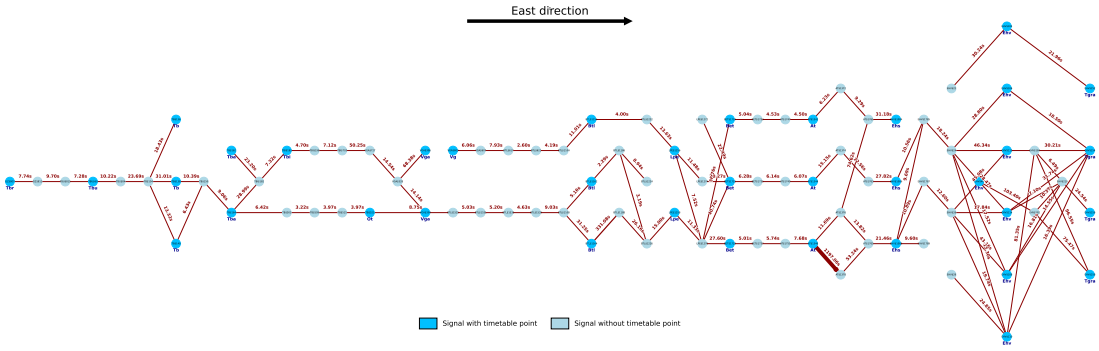


Figure 4.19: Average deviation per edge in the east direction.

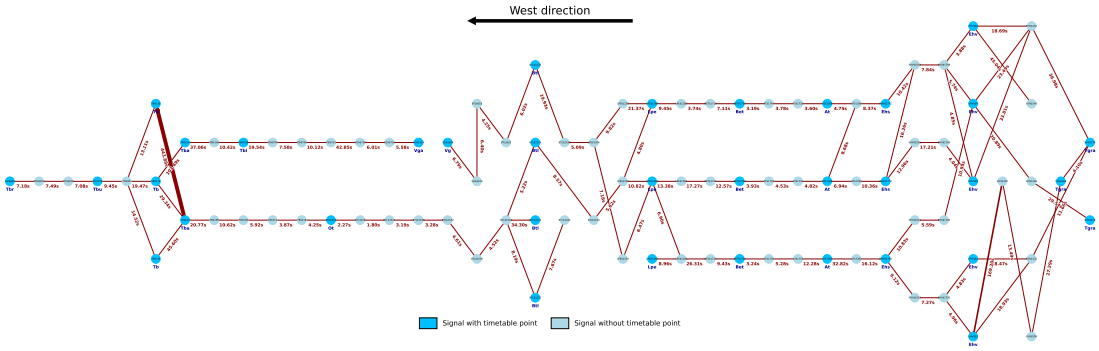


Figure 4.20: Average deviation per edge in the west direction.

To assess the association between RTL usage and punctuality, the absolute deviation distributions for RTL users and non-users were compared per edge using the Mann–Whitney U test. A small set of edges (14.3%) shows statistically significant differences, and after FDR correction, only 6.7% were significant (Table 4.7). A table with all edges can be found in the appendix D. The significant edges are highlighted in green in Figures 4.21 and 4.22. The largest difference in medians of -177 seconds can be found in the eastbound direction on edge AT$1346 → AT$1378, a track that is additionally used as a shunting area for trains to shunt or wait for passenger trains to pass. This difference corresponds to a medium effect size (effect = 0.335). The following paragraphs will explain the main findings by combining the results of average running times, average deviations, and significant differences on the microscopic level for both directions.

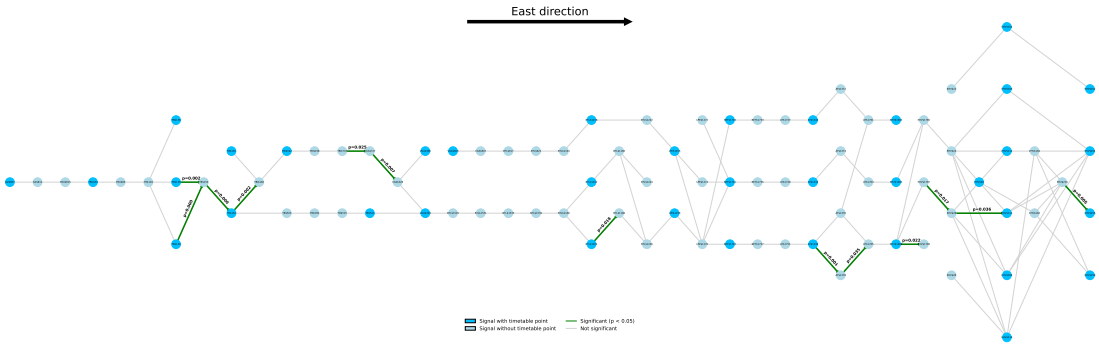


Figure 4.21: Edges on the section level showing statistically significant differences between RTL users and non-users (eastbound direction)

In the eastbound direction, the highest deviation around the timetable point Tb is observed on edge TB$104 → TB$138. This edge represents an alternative route for eastbound traffic, as the track is normally scheduled for westbound operations (platform 2). The higher delays can be explained by

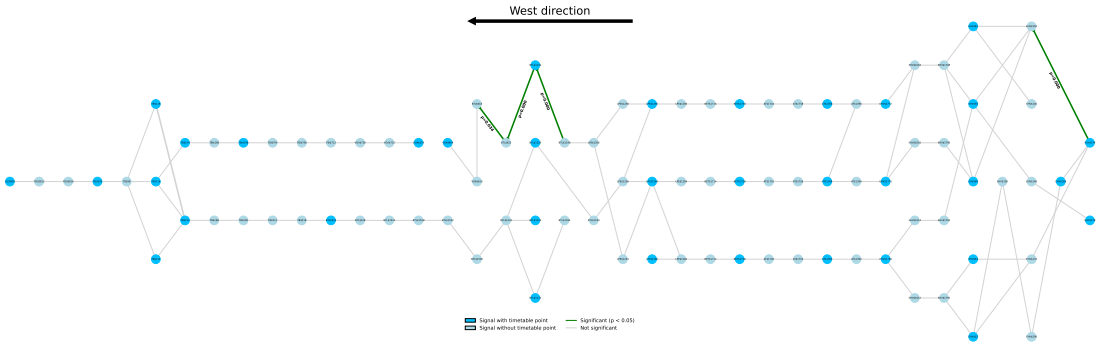


Figure 4.22: Edges on the section level showing statistically significant differences between RTL users and non-users (westbound direction)

the fact that blocks further along the line may still be reserved for opposing traffic, which have to be rerouted via platform 3. This edge also shows a significant difference: the median delay among RTL users is 3 seconds lower, while the average running time is approximately 30 seconds, indicating a relative improvement of around 10%.

The main route of the eastbound traffic through station Tb (TB$104 → TB$140 → TB$150 → TB$164), shows small significant differences of 0.52 and 0.10 seconds, with small effect values < 0.10, indicating a consistent yet marginal improvement.

The edge TB$164 → TBI$182 has a significant improvement of 2.07 seconds, with an average running time of 94 seconds (2.2%), and an average deviation of 28 seconds. This is a diverging movement from Tba towards Tbi that requires crossing.

From Ehs onwards, many edges show higher average deviations due to the high route heterogeneity around Eindhoven. Such as the edge EHV$110–EHV$184, part of a waiting/overtaking track with a crossing of three tracks, has an average deviation of 81 seconds, and edge EHV$184–EHV$258 with an average deviation of 96.55 seconds. In the eastbound significance map, Ehv shows several small but significant improvements (0–2 seconds) on short sections, consistent with micro-stabilisation rather than large time savings on edges EHV$1787 → EHV$26, EHV$26 → EHV$114, and towards Tgra EHV$210 → EHV$256.

The largest differences in medians can be found at a shunting area at At. AT$1346 → AT$1378 (and onward to AT$1765) functions as a shunting/holding siding. Traffic volume is low (96 trains/year) and deviations are high (197 seconds), reflecting enforced waiting or slow running to sequence with other trains. RTL usage is associated with significantly smaller deviations: −117 seconds (AT$1346–AT$1378) and −41 seconds (AT$1378–AT$1765). About 20% of trains on this segment use RTL, yet the effect remains detectable.

In the westbound, Ehv has way fewer different routes, and the most used westbound path is EHV$274 → EHV$150 (36 seconds) → EHV$62 (18s) and EHV$146 → EHV$62 (45 seconds), whose secondary function is waiting/overtaking track, and therefore faces higher average deviations. RTL users have statistically 5 seconds lower medians on EHV$274 → EHV$150. Edge EHV$268 → EHV$206 shows the highest average deviation at Ehv (169.2 seconds), reflecting an alternative route, with two crossings, that cause large deviations.

The other significant edges can be found around the diverging track section at Btl to Vg. BTL$1198→ BTL$1134 has a significant but negligible median difference (−0.01s) on $\bar{r}_e$ = 36 seconds, BTL$1134→BTL$662 has a median difference of −0.67 seconds on $\bar{r}_e$ = 62 seconds, and BTL$662→BTL$618 −0.005s on 63 seconds running time. Together, these results suggest that just downstream of the Boxtel diverging area, RTL is associated with slightly more stable driving, however, with small effect sizes.

On the westbound corridor, high average deviations occur around timetable point Tb, mainly due to the heterogeneity of routes and operational movements in this area. The largest deviation of 443 seconds occurs on edge TB$152→TB$118. This track section serves as a secondary waiting/overtaking track (platform 4) where freight trains frequently pause or slow down to allow passenger services to pass.

In short, the results indicate that a higher number of alternative routes and complex operational movements, such as crossovers, diverging, and merging, are associated with increased deviations.

This pattern is most evident in the eastbound direction around timetable point Ehv and in the westbound direction around Tb. Overall, the eastbound graph exhibits higher deviations and a greater number of significant edges than the westbound graph.

Table 4.7: Edges with significant differences between RTL=True and RTL=False.

Direction	s	$s+1$	n_{T}	n_{F}	Median$_T$	Median$_F$	Δ(T-F)	p	U	effect	p_{BH}
West	EHV$274	EHV$150	1472	6997	29.31	34.31	-5	0.00001*	4704332	0.057	0.00001*
West	BTL$1134	BTL$622	134	846	3.00	3.67	-0.67	0.00004*	44786	0.125	0.0021*
West	BTL$1198	BTL$1134	133	849	1.86	2.14	-0.28	0.0001*	45330	0.117	0.0036*
East	TB$150	TB$164	1596	8784	3.41	3.42	-0.01	0.00002*	6561782	0.040	0.0022*
East	TB$140	TB$150	1568	8623	3.74	4.26	-0.52	0.0001*	6357345	0.037	0.0040*
East	TB$164	TBI$182	111	1486	14.93	17.00	-2.07	0.0022*	69135	0.071	0.0436*
East	TB$138	TB$150	14	82	3.09	6.09	-3	0.0023*	301	0.290	0.0436*
East	AT$1346	AT$1378	15	69	4.47	121.47	-117	0.0011*	255	0.335	0.0353*
East	EHV$210	EHV$256	208	1039	3.89	4.89	-1	0.0055*	96007	0.072	0.087
East	VGA$727	VGA$328	154	1477	6.62	9.09	-2.47	0.0066*	99956	0.061	0.091
East	BTL$1164	BTL$1188	75	524	3.29	4.29	-1	0.018*	16732	0.085	0.197
East	EHV$1787	EHV$26	986	4874	5.18	5.82	-0.64	0.0166*	2299800	0.028	0.197
East	EHV$1008	EHV$1789	48	740	3.49	5.51	-2.02	0.022*	14687	0.072	0.212
East	TBI$703	VGA$727	154	1476	36.92	45.50	-8.58	0.025*	102736	0.049	0.216
East	AT$1378	AT$1765	16	68	12.31	54.19	-41.88	0.035*	384	0.199	0.267
East	EHV$26	EHV$114	54	511	4.50	5.24	-0.74	0.036*	11748	0.012	0.267

*Significant at $p < 0.05$.

4.3.3. Robustness check

To validate the robustness of the findings and test whether the two samples (RTL = True and RTL = False) are independent, the Mann–Whitney U test was repeated at both the macroscopic and microscopic levels using random subsampling. Approximately 30% of the data were randomly drawn for each test iteration, to confirm that the observed differences are not driven by a few extreme values or imbalanced sample sizes.

At the macroscopic level, the subsampling results in Table 4.8 reproduce the patterns identified earlier in Table 4.6. The timetable points showing the strongest differences in the main analysis also display similar results after subsampling. In the westbound direction, Ehv originally showed a median difference of –28 seconds, which became –31 seconds after subsampling; Tbr changed from –26 to –19 seconds; Tb from –23 to –33 seconds; and Btl from –15 to –22.5 seconds. In the eastbound direction, Tbi (–66.5 to –71 seconds), Tgra (–26.5 to –54 seconds), Ehv (–25 to –51 seconds), and Tb (–10 to –25 seconds) again yield statistically significant results in the same direction of effect. In all these cases, the median deviation remains lower for RTL users, with effect sizes ranging between effect ≈ 0.05 and effect ≈ 0.17, comparable to the original estimates. A few timetable points, such as At (–17 to +2 seconds) and Btl (–15 to +6 seconds) in the eastbound direction, lost significance after subsampling, an expected outcome given the smaller sample sizes. Overall, the subsampled results confirm that the observed performance gap between RTL and non-RTL trains is consistent and not dependent on any particular subset of the data.

At the microscopic level, the results of the original test and the subsample are compared and presented in Table 4.9. Edges that were significant in the main test, such as EHV$274 → EHV$150 with median differences of -5 and -6 seconds, and BTL$1134 → BTL$622 with median differences of -0.67 and -1.33 seconds on the westbound, and TB$164→TBI$182, with median differences of -2.07 and -1 seconds, and TBI$703→VGA$727 eastbound, with median differences of -8.58 and -0.19 seconds. All edges retain the same direction of the median difference as in the main analysis (all negative). For some edges with very small deviations, such as BTL$1198 → BTL$1134 (-0.67 s) and TB$150 → TB$164 (-0.01 s), the difference rounds to zero after subsampling. No test results were obtained for AT$1346 → AT$1378 and AT$1378 → AT$1765, as the random subsample did not include any trains with RTL usage on these edges.

Taken together, the results demonstrate that the earlier findings are robust: most of the same timetable points and edges showing significant or near-significant differences in the full dataset remain so in the reduced samples, with consistent effect directions and comparable magnitudes. This strengthens confidence in the reliability of the Mann–Whitney U outcomes and confirms that the observed association between RTL usage and smaller deviations is statistically consistent rather than sample-specific.

Table 4.8: Original test and subsample results for eastbound and westbound directions on the macroscopic level.

	Original test results					Subsample test results				
d	n_{total}	Δ(T–F)	p	effect	p_{BH}	n_{total}	Δ(T–F)	p	effect	p_{BH}
east										
Tbi	1415	-66.5	$5.619{\cdot}10^{-9*}$	0.1330	0.00001*	410	-71.0	0.00071*	0.16720	0.00534*
Tgra	8075	-26.5	$1.48{\cdot}10^{-5*}$	0.0482	0.0001*	2468	-54.0	$9.05{\times}10^{-8*}$	0.07880	0.00136*
Ehv	8108	-25.0	0.0007*	0.045	0.00035*	2359	-51.0	0.00295*	0.06122	0.01473*
Tba	8664	-18.0	0.0015*	0.0341	0.0056*	2580	-7.0	0.17880	0.02647	0.36167
Vga	1763	-31.5	0.0021*	0.0731	0.0064*	504	-15.0	0.19289	0.05802	0.36167
Tb	8695	-10.0	0.0165*	0.0257	0.0412*	2650	-25.0	0.01181*	0.04891	0.04429*
Bet	8080	-17.0	0.0437*	0.0224	0.0936	2468	-11.5	0.79491	0.00523	0.82530
Ot	7473	-7.0	0.0550	0.0222	0.1030	2253	-10.0	0.45798	0.01564	0.67111
Tbu	8633	-4.5	0.0974	0.0178	0.1623	2604	-3.5	0.40329	0.01638	0.67111
Ehs	7978	-14.0	0.1358	0.0167	0.1874	2454	-2.0	0.71846	0.00728	0.82530
Lpe	8079	-6.0	0.1374	0.0165	0.1874	2413	-25.0	0.01602*	0.04903	0.04806*
At	8064	-8.0	0.1779	0.0150	0.2041	2458	2.0	0.77172	0.00585	0.82530
Btl	8147	-5.0	0.1905	0.0145	0.2041	2379	6.0	0.49215	0.01408	0.67111
Vg	608	-20.0	0.1864	0.0536	0.2041	165	-14.0	0.12988	0.11805	0.32469
Tbr	8633	-2.5	0.2805	0.0116	0.2805	2580	-7.0	0.82530	0.00435	0.82530
west										
Ehs	7983	-33.0	$1.254{\cdot}10^{-11*}$	0.0758	$1.88{\cdot}10^{-10*}$	2413	-18.0	0.00543*	0.05660	0.01536*
Ehv	8078	-28.0	$4.924{\cdot}10^{-9*}$	0.0653	$3.32{\cdot}10^{-8*}$	2432	-31.0	0.00018*	0.07594	0.00271*
Tbr	8670	-26.0	$9.682{\cdot}10^{-8*}$	0.0573	$4.84{\cdot}10^{-7*}$	2605	-19.0	0.00447	0.05569	0.01536
Tb	8790	-23.0	$1.868{\cdot}10^{-6*}$	0.0508	$7.01{\cdot}10^{-6*}$	2606	-33.0	0.00357*	0.05708	0.01536*
Tbu	8671	-17.0	$1.154{\cdot}10^{-5*}$	0.0471	$3.46{\cdot}10^{-5*}$	2569	-17.0	0.05750	0.03748	0.09583
Btl	8031	-15.0	$6.99{\cdot}10^{-5*}$	0.0444	0.0002*	2433	-22.5	0.00226*	0.06190	0.01536*
At	8010	-17.0	$9.844{\cdot}10^{-5*}$	0.0435	0.0002*	2320	-17.0	0.00614*	0.05689	0.01536*
Bet	7992	-17.0	0.00015*	0.0424	0.0003*	2397	-19.0	0.04338*	0.04126	0.08876
Lpe	7985	-13.0	0.0023*	0.0342	0.0034*	2413	-1.0	0.45166	0.01532	0.52115
Tgra	7986	-9.0	0.0023*	0.0341	0.0034*	2378	-10.5	0.13290	0.03082	0.16613
Vga	1866	-21.5	0.0146*	0.0565	0.0199*	551	-36.5	0.06849	0.07762	0.10274
Ot	7395	-3.0	0.1016	0.0190	0.1246	2248	-11.0	0.04734*	0.04183	0.08876
Vg	601	-33.0	0.1080	0.0656	0.1246	184	-46.5	0.11569	0.11609	0.15776
Tba	7454	-5.0	0.2205	0.0142	0.2363	2295	-1.5	0.77001	0.00610	0.82501
Tbi	1547	15.5	0.5094	0.0168	0.5094	453	3.0	0.88129	0.00704	0.88129

*Significant at $p < 0.05$.

Table 4.9: Microscopic-level Mann–Whitney U test results for original and subsample data.

			Original test results					Subsample test results				
d	s	$s+1$	n_{total}	Δ(T–F)	p	effect	p_{FDR}	n_{total}	Δ(T–F)	p	effect	p_{FDR}
West	EHV$274	EHV$150	8469	-5	0.00001*	0.057	0.00001*	2552	-6	$1.73 \cdot 10^{-5*}$	0.081	0.0013*
West	BTL$1134	BTL$622	980	-0.67	0.00004*	0.125	0.0021*	257	-1.33	0.0012*	0.18	0.043*
West	BTL$1198	BTL$1134	982	-0.28	0.0001*	0.117	0.0036*	274	0	0.049*	0.09	0.99
East	TB$150	TB$164	10380	-0.01	0.00002*	0.040	0.0022*	3033	0	0.0064*	0.045	0.166
East	TB$140	TB$150	10191	-0.52	0.0001*	0.037	0.0040*	3064	-0.16	0.0027*	0.05	0.167
East	TB$164	TBI$182	1597	-2.07	0.0022*	0.071	0.0436*	482	-1	0.16	0.043	0.99
East	TB$138	TB$150	96	-3	0.0023*	0.290	0.0436*	31	-5.08	0.27*	0.45	0.51
East	AT$1346	AT$1378	84	-117	0.0011*	0.335	0.0353*	-	-	-	-	-
East	EHV$210	EHV$256	1247	-1	0.0055*	0.072	0.087	378	-1.5	0.046*	0.086	0.894
East	VGA$727	VGA$328	1631	-2.47	0.0066*	0.061	0.091	477	-0.09	0.53	0.023	0.99
East	BTL$1164	BTL$1188	599	-1	0.018*	0.085	0.197	185	-2.63	0.02487*	-0.25	0.3880
East	EHV$1787	EHV$26	5860	-0.64	0.0166*	0.028	0.197	1737	0	0.42	0.004	0.99
East	EHV$1008	EHV$1789	788	-2.02	0.022*	0.072	0.212	184	-18.4	0.0046*	0.17	0.116
East	TBI$703	VGA$727	1630	-8.58	0.025*	0.049	0.216	464	-12.34	0.01376*	-0.19	0.2683
East	AT$1378	AT$1765	84	-41.88	0.035*	0.199	0.267	-	-	-	-	-

*Significant at $p < 0.05$.

4.4. Yellow signals

During train operation, drivers pass yellow signals. In several of these cases, yellow signals are planned, for example, due to a speed restriction at a switch or when passing through a station. However, in the event of train disruptions or disturbances, yellow signals appear due to short headways or failures in signalling systems.

Figures 4.23 and 4.24 present the spatial distribution of unplanned yellow signals along the corridor. For both directions, the highest concentration of unplanned yellow signals is observed just before the major stations of Eindhoven and Tilburg. In the eastbound direction, these are mainly located around the signals near Tilburg University, while in the westbound direction, they appear near the junction of Tilburg aansluiting. This pattern is likely explained by the fact that freight trains in these areas often have to wait for delayed passenger services, which make scheduled stops at these major stations. The signal TB$114, located at timetable point Tilburg, records the highest number of unexpected yellow aspects (1412). Multiple inbound routes (e.g., TB$152→TB$116 and TB$152→TB$118) merge towards TB$98, tightening headways and increasing the likelihood of yellow aspects and block waiting, consistent with the observed dispersion. TB$104 also shows a high rate of unexpected yellow aspects (626), consistent with local congestion and route swapping near Tilburg University and downstream towards TBU$060/TBU$084/TB$104.

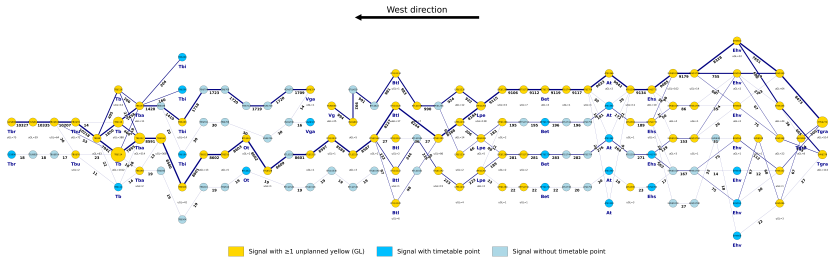


Figure 4.23: Spatial occurrence of unplanned yellow for the eastbound.

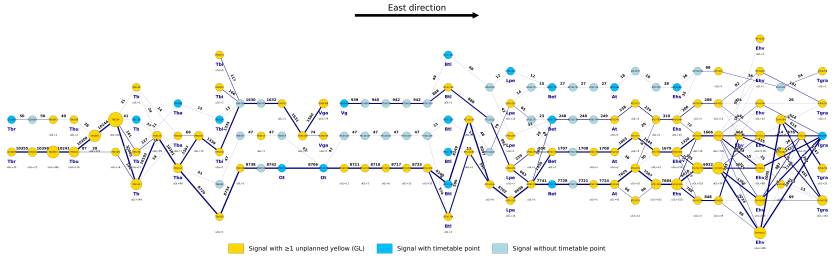


Figure 4.24: Spatial occurrence of unplanned yellow for the westbound.

To examine whether the occurrence of unplanned yellow signals differs significantly between RTL users and non-users, a chi-squared test was applied. The test result indicates a statistically significant difference ($p = 0.025$), suggesting that trains using RTL passed unplanned yellow signals less frequently than those without. From the 65910 yellow signals, 56343 had planned yellow aspects, and 9567 were unplanned (14.5%). However, the observed effect size is small: the probability of passing an unplanned yellow signal is 6% lower for RTL users, indicating that the practical difference between the two groups remains limited (Table 4.10).

These results indicate that the occurrence of unplanned yellow signals is closely related to traffic density and station approaches. To assess whether such interactions between trains contribute to these events, the next subsection analyses train interactions in more detail.

4.5. Train interaction

In this subsection, the focus lies on identifying differences in the underlying causes of yellow signals between RTL users and non-users. The TROTS dataset contains the causes for yellow aspects, which are described in Appendix A.5. These causes include, among others, technical or operational issues

Table 4.10: Results of Chi-squared test to identify a difference in unplanned yellow signal passage between RTL and non-RTL users.

	Planned Yellow	Unplanned Yellow
RTL = True	113,563	1,496
RTL = False	574,751	8,071

χ^2	5.03
p_{χ^2}	0.025
Odds ratio	0.94

such as a system error resulting in a route being released too late for an approaching train, incorrect manual adjustments by the traffic controller leading to trains being scheduled too closely together, or block occupation caused by a preceding delayed train.

With RTL, drivers should be able to anticipate more effectively, due to the greater visibility of the upcoming track, as the system allows them to look up to 25 kilometres in advance, helping them anticipate more effectively. To investigate a difference in driving behaviour, this subsection analyses both the underlying causes of yellow signals and the preceding traffic scenarios that may have influenced them. Using the TROTS dataset, preceding trains in front of the current freight trains were identified, and thereby inferred the type of train responsible for the block occupancy.

Table 4.12 divides the preceding trains into four categories: freight trains 1,431 occurrences as the preceding train), passenger trains (524,094), empty material movements (782), and others (175). The latter category includes all train types not typically operating on the Brabant corridor during nominal conditions, such as international services, night trains, and maintenance trains. An overview of the specific train types belonging to each category is provided in Table 4.11.

Table 4.11: Overview of train abbreviations, their types, main categories, and frequencies as preceding trains.

Abbreviation	Type of preceding train	Main category	Frequency preceding
GO	Freight train	Freight trains	127,236
IC	Intercity	Passenger trains	227,115
SPR	Sprinter	Passenger trains	286,532
LL	Empty locomotive	Empty trains	5,458
LLR	Empty locomotive (passenger)	Empty trains	103
LM	Empty rolling stock	Empty trains	24,771
ECD	Eurocity Direct	Others	2
EST	Eurostar	Others	10
HSN	High-speed national	Others	6,295
ICE	Intercity Express	Others	1,759
INT	International	Others	79
MP	Motorpost	Others	164
NJ	NightJet	Others	601
OME	Maintenance train	Others	240
OMD	Maintenance train	Others	1,817
STM	Steam train	Others	78
THA	Thalys	Others	77
UFM	Ultrasonic train	Others	215
UTS	Ultrasonic train	Others	275
VST	Ultrasonic train	Others	663
WTR	Work train	Others	8

For each unplanned yellow signal passage of a freight train, Table 4.12 reports the results of the statistical test, including the unplanned yellow signal rates for both groups, the difference (T–F) between those rates, the p_{chi}, and the corresponding odds ratio (OR) for each preceding train category. A statistically significant difference was only found when the preceding train was a freight train. In this case, the likelihood of passing an unplanned yellow signal was 16% lower when RTL was used compared to

not using RTL. Therefore, the null hypothesis can be rejected, indicating a difference between the two groups.

The category of preceding passenger trains includes both intercity and sprinter services, which have distinct operational characteristics. Sprinters typically stop at several intermediate stations. In this case study, they stop at six more stations than intercity services, while intercity trains only stop at Tb and Ehv. To account for these different train characteristics, a new test is performed, and the results are presented in Table 4.13. This additional test with more specific train categories shows that when a sprinter train precedes a freight train, the odds of passing an unplanned yellow signal are about 15% lower (OR=0.85) when RTL is used compared to when it is not, with ($p_{\text{BH}} = 0.007$), similar to the observed effect for preceding freight trains.

Table 4.12: Results of Chi-squared test to identify a difference in unplanned yellow signal passage per preceding train category between RTL and non-RTL users.

Preceding train	n_{T}	n_{F}	Rate T (%)	Rate F (%)	Δ (T-F)	OR [95% CI]	p_{chi}
Freight	281	1550	1.24%	1.48%	-0.24	0.84 [0.74, 0.95]	0.007*
Passenger	85187	439907	1.26%	1.39%	-0.07	0.96 [0.90, 1.02]	0.183
Empty	130	652	2.67%	2.56%	0.11	1.04 [0.83, 1.93]	0.693
Others	26	149	1.73%	1.37%	0.36	1.27 [0.86, 1.26]	0.324

*Significant at $p < 0.05$.

Table 4.13: Results of Chi-squared test to identify a difference in unplanned yellow signal passage per specific preceding train category between RTL and non-RTL users.

Category	n_{T}	n_{F}	Rate T (%)	Rate F (%)	Δ (T-F)	OR [95% CI]	p_{chi}	p_{BH}
SPR	466	2849	1.01%	1.19%	-0.18	0.85 [0.77, 0.93]	0.001*	0.007*
Freight	281	1550	1.24%	1.48%	-0.24	0.84 [0.74, 0.95]	0.007*	0.023*
IC	692	2862	1.58%	1.51%	0.07	1.05 [0.96, 1.14]	0.322	0.748
ICE	8	26	2.68%	1.78%	0.89	1.52 [0.68, 3.38]	0.428	0.748
LL	16	77	1.98%	1.66%	0.33	1.20 [0.70, 2.07]	0.606	0.849
OMD	6	35	2.62%	2.20%	0.42	1.19 [0.50, 2.87]	0.874	0.952
LM	114	575	2.81%	2.78%	0.03	1.01 [0.83, 1.24]	0.950	0.952

*Significant at $p < 0.05$.

Table 4.15 presents the different causes of unplanned yellow signals, comparing RTL users and non-RTL users across different train categories; the symbols representing the causes' operational movements are explained in 4.14.

Table 4.14: Overview of symbols of yellow aspect causes.

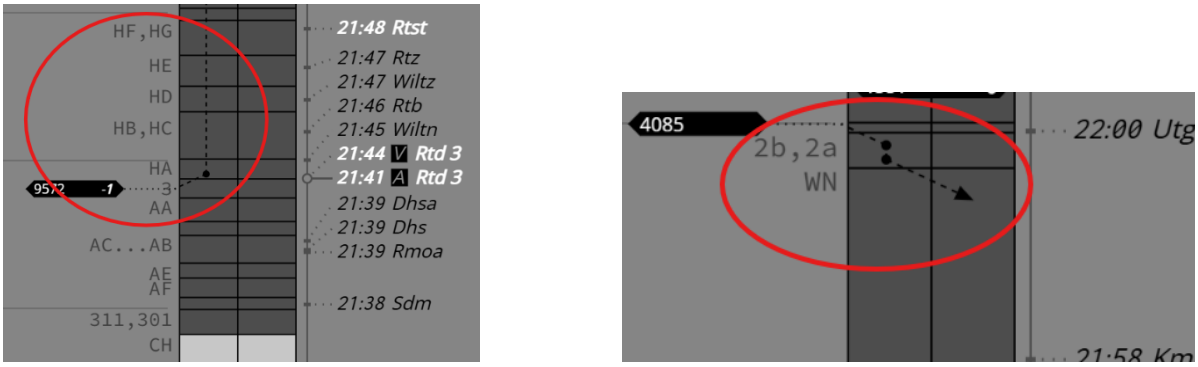
Cause symbol	**Explanation**
W	Order of trains is swapped.
>	Merging train movement in the same direction.
<	Diverging train movement in the same direction.
|	Following train movement in the same direction.
=	Hindrance from another train holding the next block.
X	Crossing train movement in the opposite direction.
!	Subsequent train movement in the opposite direction.
-	Crossing train movement in the same running direction.
$\wedge$	Executing train movement in the opposite direction.
TC l/L	Traffic Control scheduled l minutes for operation, where L minutes were necessary.
ARI	Route set too late; ARI (Automatic Route Setting) eventually sets the route.
PPR	Route set too late; the dispatcher ultimately sets the route manually

The analysis reveals that the only significant difference, after correcting for false discovery rates,

is found in the case of a preceding merging passenger train. This scenario caused 4.49 percentage more unplanned yellow aspects when using RTL, corresponding to an odds ratio of 1.31, and p_{BH} of 0.08. Causes such as TC 2/3 |, ARI, and PPR correspond to failures in route-setting systems or traffic control operations.

Table 4.16 highlights the train interactions that caused a yellow signal passage, including a distinction between intercity and sprinter trains. Although none of these interactions remain statistically significant after the FDR correction, it is notable that operational movements that include merging or crossing are usually associated with higher rates for the RTL users (Freight X 2.24, IC > 4.65, SPR W > 0.23, Empty > 0.92). On the other hand, the following scenario shows lower rates for RTL users (Empty locomotive -9.25, Freight -2.07, SPR -2.77).

In summary, this result reflects a complex pattern that includes one significant case, a preceding merging passenger train, alongside several non-significant interaction types. These patterns can be interpreted in light of how RTL visualises preceding trains. In RTL, drivers can see preceding trains on their dashboard; however, when these trains are still in the process of merging onto the main line, such as in the example shown in Figure 4.25, the driver cannot accurately assess how far ahead the merging train actually is. In such cases, ambiguity may arise: the RTL interface represents the merging train as a dotted line and displays its current delay or earliness based on the previous timetable, but it does not indicate whether that train's position or delay is likely to cause a headway conflict.



(a) Merging movement

(b) Crossing movement in the opposite direction

Figure 4.25: Merging and crossing movements displayed in RTL's dashboard.

This chapter has outlined the main empirical findings of this study based on the four KPIs: transport volume, deviation, yellow signals and train interaction. The results indicate that trains using RTL generally operate with lower deviation and fewer unplanned yellow signals compared to non-users. While the observed effects vary between directions and train categories, the general trend suggests improved anticipation and smoother traffic flow among RTL users.

In the following chapter, these outcomes are discussed in more depth, focusing on possible explanations, their policy implications, and how they relate to existing research on DAS.

Table 4.15: Significant test differences in cause within unplanned yellow signals per category and cause: (RTL users vs non-RTL users)

Category	Cause	n_{T}	n_{F}	Rate T (%)	Rate F (%)	Δ (T-F)	OR	p_{χ^2}	p_{bh}
Passenger	>	278	1218	26.28%	21.33%	+4.95	1.31	0.001*	0.008*
Freight	TC 2/3 |	7	10	2.49%	0.65%	+1.85	3.93	0.009*	0.051
Freight	X	11	26	3.91%	1.68%	+2.24	2.39	0.026*	0.079
Empty	ARI	11	24	8.46%	3.68%	+4.78	2.42	0.030*	0.178
Freight	W |	9	83	3.20%	5.35%	-2.15	0.58	0.170	0.341
Passenger	W >	38	258	3.59%	4.52%	-0.93	0.79	0.204	0.645
Passenger	W |	44	188	4.16%	3.29%	+0.87	1.27	0.183	0.645
Passenger	W X	7	68	0.66%	1.19%	-0.53	0.55	0.177	0.645
Passenger	|	307	1808	29.02%	31.66%	-2.64	0.88	0.096	0.645
Passenger	ARI split	24	91	2.27%	1.59%	+0.68	1.43	0.152	0.645
Freight	|	124	761	44.13%	46.19%	-2.07	0.92	0.566	0.849
Freight	>	29	175	10.32%	11.29%	-0.97	0.90	0.710	0.851
Other	|	6	26	23.08%	17.45%	+5.63	1.42	0.682	0.863
Other	>	7	46	26.92%	30.87%	-3.95	0.82	0.863	0.863
Freight	PPR	19	100	6.76%	6.45%	+0.31	1.05	0.950	0.950
Passenger	<	13	53	1.23%	0.93%	+0.30	1.33	0.457	0.999
Passenger	W TC X 2/3	10	45	0.95%	0.79%	+0.16	1.20	0.736	0.999
Passenger	PPR	67	324	6.33%	5.67%	+0.66	1.12	0.440	0.999
Passenger	=	7	47	0.66%	0.82%	-0.16	0.80	0.724	0.999
Passenger	-	21	98	1.98%	1.72%	+0.27	1.16	0.628	0.999
Passenger	TC 3/4 |	20	124	1.89%	2.17%	-0.28	0.87	0.641	0.999
Passenger	TC | 3/4	22	134	2.08%	2.35%	-0.27	0.88	0.674	0.999
Passenger	ARI	17	104	1.61%	1.82%	-0.21	0.88	0.721	0.999
Passenger	W ARI	8	40	0.76%	0.70%	+0.06	1.08	1.000	1.000
Passenger	W -	6	37	0.57%	0.65%	-0.08	0.87	0.926	1.000
Passenger	TC 3/4 >	11	65	1.04%	1.14%	-0.10	0.91	0.904	1.000
Empty	PPR	14	71	10.77%	10.89%	-0.12	0.99	1.000	1.000
Passenger	X	14	73	1.32%	1.28%	+0.05	1.04	1.000	1.000
Empty	X	8	30	6.15%	4.60%	+1.55	1.36	0.597	1.000
Empty	ARIsplit	7	37	5.38%	5.98%	-0.60	0.89	0.952	1.000
Empty	>	7	37	5.38%	5.67%	-0.29	0.95	1.000	1.000
Empty	|	18	94	13.85%	14.42%	-0.57	0.95	0.974	1.000
Passenger	$\wedge$	5	30	0.47%	0.53%	-0.05	0.90	1.000	1.000

*Significant at $p < 0.05$.

Table 4.16: Significant test differences in cause within unplanned yellow signals per category and cause: (RTL users vs non-RTL users), passenger trains split in intercity and sprinter services.

Category	Cause	n_{T}	n_{F}	Rate T (%)	Rate F (%)	Δ T-F	OR	p_{χ^2}	p_{bh}
Freight	X	11	26	3.91%	1.68%	+2.24	2.39	0.026*	0.079
IC	W >	24	179	4.05%	6.25%	-2.20	0.63	0.048*	0.265
IC	>	260	1124	43.92%	39.27%	+4.65	1.21	0.040*	0.265
Freight	W |	9	83	3.20%	5.35%	-2.15	0.58	0.170	0.341
LL	|	6	77	37.50%	46.75%	-9.25	0.68	0.689	0.689
IC	W X	6	44	1.01%	1.54%	-0.52	0.66	0.434	0.780
IC	W |	12	45	2.03%	1.57%	+0.45	1.30	0.540	0.780
IC	-	18	72	3.04%	2.52%	+0.52	1.22	0.557	0.780
IC	|	95	433	16.05%	15.13%	+0.92	1.07	0.615	0.780
SPR	<	11	38	2.36%	1.33%	+1.03	1.79	0.135	0.809
SPR	W |	32	143	6.87%	5.02%	+1.85	1.40	0.123	0.809
Freight	|	124	716	44.13%	46.19%	-2.07	0.92	0.566	0.849
Freight	>	29	175	10.32%	11.29%	-0.97	0.90	0.710	0.851
SPR	|	212	1375	45.49%	48.26%	-2.77	0.89	0.289	0.892
SPR	>	18	94	3.86%	3.30%	+0.56	1.18	0.627	0.892
IC	X	12	53	2.03%	1.85%	+0.18	1.10	0.905	0.905
SPR	W >	14	79	3.00%	2.77%	+0.23	1.09	0.897	1.000
LM	>	7	30	6.14%	5.22%	+0.92	1.19	0.863	1.000
LM	X	8	30	7.02%	5.22%	+1.80	1.37	0.586	1.000
SPR	=	5	30	1.07%	1.05%	+0.02	1.02	1.000	1.000
LM	|	12	58	10.53%	10.09%	+0.44	1.05	1.000	1.000

*Significant at $p < 0.05$.

5 Discussion

This chapter interprets the empirical findings presented in Chapter 4 and situates them within the broader context of Driver Advisory Systems (DAS) and railway operations. It addresses sub-question five by discussing the implications of the results for freight operators and infrastructure managers, while also outlining methodological limitations.

The analysis showed that RouteLint (RTL) users generally exhibited better operational performance, reflected in smaller deviations from the planned timetable and fewer unplanned yellow signal aspects. Although these effects are modest in magnitude, they are consistent and primarily visible at the macroscopic level. The reduction in deviations and restrictive signal occurrences among RTL users suggests that the system helps drivers to anticipate operational conditions more effectively, particularly around major stations with complex routing patterns. This finding aligns with previous simulation-based studies (Dong et al., 2018; Wang et al., 2019; Xiao et al., 2021), which demonstrated that DASs can reduce behavioural variability and improve operational stability.

At the macroscopic level, a greater number of significant differences were found westbound, while the largest median deviations occurred eastbound. The eastbound direction also exhibited higher overall deviation values, which likely increased the statistical power to detect group differences at both scales. In contrast, several westbound differences identified at the macroscopic level were not reflected microscopically.

This discrepancy can be explained by aggregation effects. Macroscopic results represent averages over multiple timetable points, which smooth out random variation and reveal consistent patterns more clearly. At the microscopic level, where data capture smaller spatial and temporal units (signal-to-signal edges), variability is higher, and small systematic differences become masked by noise. Given that the observed macroscopic deviations are typically within tens of seconds, detecting comparable differences per edge becomes statistically challenging.

Overall, the relationship between scale and sensitivity indicates that larger, network-level deviations are more likely to produce detectable microscopic effects, whereas smaller macro-level patterns are diluted by high local variance. This underscores how the level of aggregation influences the visibility of operational effects in statistical testing.

5.1. Limitations of the research

While the findings offer useful insights into the operational impact of DAS, several methodological and data-related limitations should be acknowledged.

The study relied on realisation data from the TROTS (event times at signals) and VOS datasets (event times at timetable points) to construct aggregated event graphs for both eastbound and westbound operations. These graphs provided an overview of where traffic density is concentrated and visualised how trains traverse the network. At the macroscopic level, deviation values could be directly obtained from the VOS dataset, whereas at the microscopic level, reference running times between signals were estimated using trains that ran on time and passed no yellow signals. To control for dif-

ferences in planned performance, trains were categorised by their scheduled running times as a proxy for weight class. Although this classification was verified to be broadly consistent, heavier categories showed longer reference times; some misclassifications may remain, introducing a limited bias. Such effects are expected to be minor and unlikely to affect the main conclusions.

Each observation in the dataset represents a unique train–run combination, but individual driver identities were unavailable. As a result, complete statistical independence between observations cannot be guaranteed, since the same driver may appear multiple times within either the user or non-user group. From July 2025 onwards, an additional interface option allowed users to indicate whether they were actively driving or merely viewing the system; 85% confirmed being the driver. However, this information was only available for the final month of the analysis period. Earlier entries may therefore include a small share of non-driving use (e.g., system testing or passive viewing), which could slightly underestimate the observed differences between groups.

Furthermore, the analysis could not account for individual differences in driving style or experience. More skilled or proactive drivers might operate more smoothly and adhere more closely to the timetable, irrespective of RTL use. If such drivers were also more likely to adopt the system, the estimated performance advantage of RTL could be overstated; conversely, if less experienced drivers relied more on the system, the effect might be understated. This unobserved heterogeneity should therefore be kept in mind when interpreting the results.

Multiple hypothesis tests were conducted to detect location- and interaction-specific effects. A False Discovery Rate (FDR) correction was applied to limit false positives, though this reduces statistical power and may prevent smaller but genuine effects from reaching significance. Hence, non-significant results should not be taken as evidence of no effect, but rather as an indication of uncertain or weak effects within the available data.

Finally, the analysis focused on performance metrics directly related to RTL's operational goals, timetable adherence, and signal aspects. Potential effects on energy efficiency, braking wear, or driver workload were not examined, although smoother driving patterns observed among users could plausibly influence them.

5.2. Policy implications

The findings from this data-driven analysis provide concrete evidence that DAS, like RTL, can improve train operations. Over a full year of real-world data (23,573 train runs), RTL usage was associated with less deviation from the schedule and smoother traffic flow. Based on these insights, several recommendations can be made for different stakeholders in the rail sector.

For Infrastructure Managers (IM), like ProRail, the first point is to encourage broad DAS adoption by Railway Undertaking companies. The analysis showed that trains using RTL ran more punctually and passed fewer restrictive signals than those without. For example, RTL users had significantly lower median deviations at most timetable points where punctuality is assessed (e.g., 66 seconds shorter at Tilburg Industrie (Tbi) in the eastbound direction, and around 20–30 seconds shorter at major stations like Tilburg and Eindhoven). They also faced approximately 6% fewer unexpected yellow signal stops overall, reflecting the benefits of assisted driving for improved traffic harmonisation. These gains, although modest per train, accumulate to improvements in network performance. By promoting or even mandating the use of DAS, ProRail could improve overall punctuality and result in effective line capacity without costly infrastructure investments. The 80-week Betuweroute maintenance period provides a clear example: even with an increase of around 80 trains per day on the Brabantroute, RTL adoption helped to maintain punctual train operation.

Additionally, IM could adopt the analytical method described in this study as part of its ongoing performance monitoring. By systematically analysing large datasets of signal and timetable information, it becomes possible to identify where and how DAS implementations deliver operational benefits. Unlike simulation-based analysis, this data-driven approach provides concrete empirical evidence to guide strategic decisions. Regular evaluation of KPIs, such as deviations, signal aspects, and train interactions, would allow ProRail to measure the impact of innovations or policy changes. When systems such as RTL are deployed on new routes, adopted by additional railway undertakings, or enhanced with new features, comparable before-and-after analyses can be used to verify their effectiveness.

For DAS developers, the key policy implication is to integrate conflict prediction into the advice

generated by the system. The study's analysis highlights specific scenarios where RTL's functionality could be enhanced. One clear opportunity is to improve the advice for complex junction scenarios. Currently, RTL shows trains on merging or crossing routes as dotted lines, but without any estimate of their relative timing or distance. This makes it difficult for drivers to assess potential conflicts ahead. The findings of this study show that when a passenger train merged ahead of a freight, RTL-equipped drivers had a 30% higher chance of passing a yellow signal and braking, indicating they didn't get enough information to adjust in time. To address this, developers should implement predictive conflict indicators, for example, if a freight train is expected to catch up to a slower train at a junction, RTL should display this upcoming conflict to help drivers anticipate these interactions more accurately. This would reduce the uncertainty and unnecessary braking we observed with merging scenarios.

Beyond its operational benefits, driver involvement is essential to further improve DAS systems. Developers should actively involve drivers in the feedback process, as they are the end-users who can identify issues in the advice or information at an early stage. By reporting these observations, drivers can help developers understand where improvements are needed and what kind of enhancements would be most valuable. Some freight drivers in the trial have already shared informal feedback, noting that RTL helped them maintain steady speeds but that greater clarity at merging and crossing points is still needed. Establishing a formal feedback loop would allow such insights to be collected systematically. For instance, a simple in-app reporting function could enable drivers to flag unclear or incorrect advice immediately after a trip, helping developers to detect and resolve issues more quickly, on the condition that this will not disturb the safety by any means. This continuous exchange between users and developers will ensure that RTL evolves in line with real operational needs.

Railway Undertakings should invest in DAS such as RTL, which have proven to improve punctuality and reduce restrictive signal encounters. Trains equipped with RTL ran closer to schedule and experienced fewer unexpected yellow signals, without any negative effects on workload or safety (Verstappen et al., 2022). Because DAS is a software-based solution, it offers a cost-effective way to enhance operational performance. By helping drivers anticipate better and adhere to the timetable, DAS reduces delays, unnecessary braking, and fuel or energy consumption. These savings, combined with improved service reliability, translate directly into lower operational costs and better on-time performance for both freight and passenger services. Speed-based advice might be even more effective if the DAS were developed by a railway undertaking or collaborated with an RU, as operational conditions and driving practices could then be more accurately accounted for.

For train drivers, the introduction of RTL offers a valuable support tool to improve situational awareness. RTL results in more informed driving: instead of reacting last-minute to restrictive signals or catching up to a slower train ahead, drivers can manage their speed proactively. By embracing the DAS and actively using its advice, drivers can maintain smoother operations that keep them on time without unnecessary braking. It is important to note that RTL is designed as an assistive tool, not an automation. Drivers remain in control, but the system extends their visibility by providing insight into the track up to 25 km ahead. This is especially useful in busy areas around major stations such as Tilburg and Eindhoven, as it provides better insights, especially in following and diverging situations.

6

Conclusions

This chapter summarises the main findings of the research and discusses how they collectively address the overarching aim of evaluating the operational benefits of a railway DAS using empirical data. Each subsection revisits one of the research questions introduced in Chapter 1, outlining the corresponding methodological approach, key results, and implications. Together, these discussions provide an integrated understanding of how RouteLint (RTL), a Dutch DAS, performs in practice, how its effects can be measured quantitatively, and what these outcomes imply for future development and implementation of DAS in freight operations.

6.1. Answers to research questions

RQ1: What is the state-of-the-art and the state-of-practice of Driver Advisory Systems?
Current research identifies DAS as a key tool for optimising train performance, with studies highlighting its potential to improve energy efficiency and punctuality. For example, DAS algorithms use strategies like maximal coasting, reduced-cruise speed, or combined coasting-cruise profiles to minimise delays and energy use. Recent work has even explored AI-driven DAS frameworks (e.g., LLM-based advisory logic), though these advanced systems remain largely experimental and unvalidated in real-world service. In practice, several DAS have been deployed in Europe. The Dutch RTL system is a Connected - DAS (C-DAS) providing timekeeping advice and route information to the drivers. RTL is currently used by freight train drivers. Likewise, NS (the Dutch passenger railway undertaker) has developed TIMTIM, an onboard C-DAS that integrates GPS-based positioning and Related Train Information to offer coasting advice when on time. Other countries also have DAS initiatives: for instance, Austria's infraDOAS and Belgium's Infrabel DAS are free services from infrastructure managers, whereas vendors like Siemens and KeTech sell DAS tools to operators. Overall, the trend is toward hybrid architectures combining central traffic data processing with onboard interfaces. Many operational DAS, however, currently offer only timekeeping advice (not direct speed commands) and do not fully account for delays of other trains. In summary, the state-of-the-art literature has matured to include optimisation and even AI approaches, but mainly in simulations, whereas the state-of-practice shows emerging real-world systems (like RTL and TIMTIM) aimed at modest gains in punctuality and traffic flow.

RQ2: Which methods and KPIs can be used to quantify the effect of Driver Advisory Systems on freight train operations using empirical data?
This study employed a data-driven framework using four empirical KPIs that capture key aspects of freight train operation. The KPIs were defined as:

- Transport Volume
- Deviation from timetable
- Unplanned yellow signals
- Train interaction

Operational train event data at both timetable points and signals were mapped onto aggregated event graphs at two scales: a macroscopic graph (timetable points) and a microscopic graph (signals). This allowed each KPI to be computed network-wide under RTL-use versus non-use. To compare the groups, continuous KPIs (e.g., deviations, running times) were analysed with non-parametric Mann–Whitney U tests, accompanied by a random-subsampling robustness check using 30% of the dataset to mitigate potential dependence between repeated driver observations. Categorical outcomes (e.g., whether an unplanned yellow was encountered and the cause of yellow signals) were tested using chi-squared statistics. For both types of tests, Benjamini–Hochberg False Discovery Rate (FDR) correction was applied to control for multiple comparisons. In the microscopic graph, the Critical Path Method (CPM) was applied to identify timing dependencies and bottlenecks in the signal sequence. Together, these methods provide a data-driven way to link train performance metrics to DAS usage, enabling quantification of deviation, unplanned yellow signal passage, and train interactions.

RQ3: How do freight train operations with and without a Driver Advisory System differ?
The analysis focuses on the Brabantroute corridor between Eindhoven and Tilburg, a key freight artery within the Dutch railway network. In total, 68 km of track were analysed. The empirical results show that freight trains using the DAS operated more consistently in line with their planned schedules and encountered fewer restrictive signals than non-users. For example, at Tilburg Industrie, the median deviation was approximately 66 seconds shorter for DAS users. Notably, significant differences were mainly observed at major stations Eindhoven en Tilburg and at the first timetable point following these stations, typically ranging between 10 and 30 seconds. Correspondingly, the proportion of trains encountering unplanned yellow signals was lower under DAS operation, representing an overall reduction of around 6% in unexpected stops. Statistical tests confirmed that these differences were significant. However, the improvements were modest in magnitude, typically on the order of several tens of seconds and a few percentage points in conflict reduction. Moreover, the impact was context-dependent. In situations involving merging or crossing movements, where delayed passenger trains interact with freight services, RTL users did not always perform better and occasionally performed slightly worse. This likely reflects the current system's limitations, as RTL's advisory algorithm does not yet account for conflicting routes within such scenarios. In following scenarios a small reduction in unplanned yellow signal passages was particularly observed when preceding sprinter services were involved. These findings indicate that RTL's benefits are context-dependent and most pronounced under relatively stable traffic conditions. Under nominal conditions, a small reduction in unplanned yellow signal passages was particularly observed when preceding sprinter services were involved. Overall, however, the use of the DAS was associated with small but measurable improvements in punctuality, consistent with the system's ability to support drivers in anticipating and adapting to upcoming operational situations.

RQ4: How do the performance impacts of a Driver Advisory System vary across scales of operation and over space and time?
The empirical analysis revealed that RTL had a measurable but context-dependent positive impact on freight train operations.

With the introduction of RTL in November, the average deviation between RTL users and non-users initially varied considerably, with RTL use at times even corresponding to higher average deviations. However, as system adoption increased over time, the average absolute deviation among RTL users consistently remained below that of non-users, indicating a gradual improvement in operational consistency as drivers became more familiar with the system.

At the macroscopic level, significantly smaller medians were shown with RTL-usage after FDR correction at large passenger stations, Ehv (-25 seconds eastbound and -33 seconds westbound) and Tb (-10/-23 seconds), and at junctions with diverging scenarios Tgra (-26.5 seconds eastbound, -9 westbound) and Tba (-18 seconds eastbound) and Tbi (-66.5 seconds eastbound). In both directions, effect sizes were small ($effect \leq 0.10$), suggesting that the distributions still overlapped considerably, despite a consistent shift towards lower deviations for RTL users.

At the microscopic level, where individual signal passages were examined, significant though smaller effects were found on several edges, mostly near diverging and merging sections. For instance, eastbound trains between AT$1346 $\rightarrow$ AT$1378 showed an improvement of approximately 117 seconds, while other local effects averaged 1–5 seconds reductions in running time deviations. These outcomes suggest that RTL use supports smoother operation on conflict-prone track sections, though its benefits are partly masked by local variability.

Together, these results demonstrate that RTL provides small but consistent operational gains, especially under stable traffic conditions and around major junctions. While improvements are modest, typically on the order of tens of seconds per train and a few percentage points in conflict reduction, they accumulate across the network, evidencing that even limited advisory support can enhance flow stability and timetable adherence in mixed-traffic corridors.

RQ5: What policy implications can be drawn from the data-driven analysis
The findings point towards several policy recommendations. First, infrastructure managers (IMs) should actively promote the wider adoption of DAS among railway undertakings. The results indicate that trains equipped with RTL operated more punctually and showed smaller deviations from the planned timetable. Encouraging or even mandating system use, for example, by offering RTL free of charge to operators, could therefore enhance overall network performance without the need for costly infrastructure investments. During the test period, even under conditions of increased traffic density, the deviations observed with RTL were substantially smaller than those without the system. Second, IMs should institutionalise data-driven monitoring of DAS performance. The KPI framework applied in this study could be integrated into routine operational analysis: by continuously evaluating timetable and signal data, infrastructure managers can identify where and under what conditions DAS implementation yields the greatest benefits, thereby supporting evidence-based investment decisions. For system developers, the results highlight several improvement opportunities. In particular, predictive conflict alerts and enhanced situational feedback mechanisms should be incorporated to address the limitations observed in complex merging or crossing scenarios. By supporting drivers in anticipating operational conditions and adhering more closely to the timetable, DAS can help reduce delays, unnecessary braking, and energy consumption. These operational efficiencies, combined with more reliable service performance, translate into lower costs and improved punctuality for both freight and passenger services.

Lastly, the main research question is answered: *How can the operational performance of freight trains be evaluated using empirical data from Driver Advisory Systems such as RouteLint?*

This study shows that the operational performance of freight trains can be evaluated empirically by combining large-scale operational datasets with well-defined performance indicators and robust statistical methods. By integrating realised timetable data, signal-level observations, and RTL records, the analysis quantified key aspects of train operations, punctuality, signal behaviour, and train interactions, across both macroscopic and microscopic levels. Applying non-parametric statistical tests allowed meaningful comparison between train runs with and without RTL, revealing modest but consistent improvements in timetable adherence and fewer unplanned yellow signals for users. These results demonstrate that empirical data, especially can be effectively used to measure the real-world impact of Driver Advisory Systems, providing a scalable framework for evaluating their contribution to more reliable and efficient railway operations.

In summary, this study makes three main contributions. First, it introduces a comprehensive methodological framework for evaluating train operational performance between drivers using and not using a DAS. The framework combines multiple analytical methods and key performance indicators to capture different dimensions of railway operations. Second, the framework is applied to large-scale, real-world data from the Dutch rail network to quantify the operational impact of the RTL system on freight train performance along the Brabantroute corridor. This application provides one of the first empirical assessments of DAS effectiveness based on observed operational data rather than simulations. Finally, the study translates these empirical findings into actionable insights and policy recommendations for infrastructure managers, system developers, railway undertakings, and train drivers. Together, these contributions demonstrate how DAS can be leveraged to enhance network capacity, operational efficiency, and service reliability within existing infrastructure constraints.

6.2. Future research

Future research could build on this study in several complementary ways, addressing both methodological and functional aspects of DAS development.

The current approach combined macro- and micro-level levels to identify differences in driving behaviour between DAS users and non-users. While significant differences in median deviations, up to 66 seconds, were detected at the macroscopic level, such differences were harder to observe at the signal

level, where the spatial and temporal scales are much smaller. Nevertheless, the microscopic analysis was useful in showing where along the network deviations and yellow signal encounters occurred. A mesoscopic, train-path-based approach could offer a middle ground: it would preserve the detailed view of individual train movements needed to capture operational scenarios, while also aggregating data in a way that better highlights differences in driving behaviour than purely microscopic analysis allows. Such a method could therefore provide more interpretable insights into where and how DAS influences real-world operations. Moreover, a mesoscopic level of analysis could also make it possible to represent both eastbound and westbound directions within a single network graph. This would facilitate the identification of opposing traffic flows and interactions between trains running in opposite directions, which are difficult to capture at the microscopic level. Such an approach could therefore provide additional insights into where and how DAS influences real-world operations.

One promising direction for future research is to better understand when and how drivers use DAS in real operations. At present, the circumstances that prompt drivers to log into or consult RTL are not systematically recorded. It is likely that some drivers only activate the system in specific situations, such as when approaching a yellow signal, facing delays, or operating in congested areas, to gain additional situational awareness. If so, DAS usage may be biased toward more challenging runs, meaning that the observed performance gains could underestimate the system's full potential. To examine this, future studies should monitor DAS activation alongside operational data. Combining RTL logs with complementary data sources such as GPS traces, event logs, or in-cab, eye-tracking observations would make it possible to link DAS use directly to driving context and workload. Verstappen et al. (2022) has already shown the value of such human-factors analysis: using a simulator and eye-tracking, they found that route-context advice improved safety and did not increase workload. Applying similar techniques in live operations could reveal where and when drivers consult RTL. for example, near junctions or restrictive signals, and how their behaviour changes with the tool in use.

Further research could focus on improving DAS logic for complex operational situations. Using historical operational data, researchers could identify particular situations (merging junctions or crossings) where braking or delays occur, and then simulate whether alternative advisory strategies could have improved the outcome. For instance,Wang et al. (2019) proposes a connected DAS framework to help freight trains merge smoothly into a busy passenger line by predicting traffic gaps and advising an optimal approach speed. Incorporating similar logic into RTL could allow it to proactively assist in challenging scenarios like merging or running on tightly scheduled corridors.

So far, RTL has primarily aimed to keep trains on schedule, yet energy efficiency represents another major opportunity for DAS. Recent research shows that punctuality and energy savings can be achieved simultaneously by optimising train speed profiles. For instance, Z. Li et al. (2018) demonstrated a dynamic trajectory optimisation method that recovered delays while reducing energy use, effectively balancing schedule adherence and eco-driving objectives. Other DAS implementations have achieved similar results, optimising for energy-efficient operation without compromising punctuality (Dong et al., 2018; Xiao et al., 2021). Future versions of RTL could integrate coasting guidance or gentle braking strategies to conserve energy, particularly when trains are running early or have timetable margins, thus contributing to broader sustainability goals, provided that safety is not compromised and operational responsibility remains with the railway undertaking. Expanding the evaluation framework to include additional performance metrics such as fuel or electricity consumption, component wear, and emissions would provide valuable input for further algorithm development.

Finally, the limits of RTL's effectiveness could be explored through advanced simulation and scenario-based testing. In the automotive field, Y. Huang et al. (2024) applied Bayesian optimisation to identify worst-case situations for autonomous vehicle controllers, enabling targeted improvements to safety performance. A similar approach could be adopted for DAS in rail operations. By systematically varying simulation conditions, such as extreme schedule perturbations, unexpected route changes, or dense traffic near bottlenecks, and applying optimisation methods, researchers could pinpoint cases where the current RTL algorithm fails to provide effective guidance or yields little benefit. This would reveal the boundaries of the system's advice logic and highlight areas for refinement. Combined with field data, this could support the development of adaptive algorithms that detect when operations deviate from normal conditions and automatically switch to a specialised advisory mode tailored to those scenarios.

References

Abbott, M. (2017). Vertical integration, separation in the rail industry: a survey of empirical studies on efficiency. *European Journal of Transport and Infrastructure Research*, *17*, 207–224.

Abramović, B., Zitricky, V., & Biškup, V. (2016). Organisation of railway freight transport: case study CIM/SMGS between Slovakia and Ukraine. *European Transport Research Review*, *8*.

Albrecht, T. (2014). Energy-efficient railway operation. In I. A. Hansen & J. Pachl (Eds.), *Railway timetabling & operations* (pp. 91–114). Eurailpress.

B. Pascariu and J. V. Flensburg and P. Pellegrini and C. M. L. Azevedo. (2025). Formulation and solution framework for real-time railway traffic management with demand prediction. *IET Intelligent Transport Systems*, *19*.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300.

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

Buck, P. (2005). *The Betuweroute*. EUROPEAN RAILWAY REVIEW.

Bulková, Z., Gašparík, J., & Zitrický, V. (2024). The Management of Railway Operations during the Planned Interruption of Railway Infrastructure. *Infrastructures*, *9*(7), 119.

Cacchiani, V., Huisman, D., Kidd, M., Kroon, L., Toth, P., Veelenturf, L., & Wagenaar, J. (2014). An Overview of Recovery Models and Algorithms for Real-time Railway Rescheduling 1. *Transportation Research Part B: Methodological*, *63*, 15–37.

Cheng, Y. (1996). Optimal train traffic rescheduling simulation by a knowledge-based system combined with critical path method. *Railway Technical Research Institute, 4*(6), 399–413.

Clark, S. (2012). A history of railway signalling. *Proceedings of the IET Professional Development Course on Railway Signalling and Control Systems*, 6–25. https://doi.org/10.1049/ic.2012.0040

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd). Lawrence Erlbaum Associates.

Cunillera, A., Jonker, H. H., Scheepmaker, G. M., Bogers, W. H. T. J., & Goverde, R. M. P. (2023). Coasting advice based on the analytical solutions of the train motion model. *Journal of Rail Transport Planning & Management*, *28*.

Dong, H., Zhu, H., & Gao, S. (2018). An Approach for Energy-Efficient and Punctual Train Operation via Driver Advisory System. *IEEE Intelligent Transportation Systems Magazine*, *10*(3), 57–67.

European Commission. (2011). White Paper on Transport: Roadmap to a Single European Transport Area – Towards a competitive and resource efficient transport system [COM(2011) 144 final]. https://transport.ec.europa.eu/transport-themes/strategies/2011-white-paper-transport_en

Forum, I. T. (2019). *Efficiency in railway operations and infrastructure management* (tech. rep.) (International Transport Forum Discussion Paper No. 2019/01). OECD Publishing. Paris. https://www.itf-oecd.org/sites/default/files/docs/efficiency-railway-operations-infrastructure_1.pdf

George, N. (2022). Eco-Driving Strategy optimization for freight trains. *Journal of Mathematical Techniques and Computational Mathematics*, *1*(3).

Ghaviha, N., Bohlin, M., Holmberg, C., Dahlquist, E., Skoglund, R., & Jonasson, D. (2017). A driver advisory system with dynamic losses for passenger electric multiple units. *Transportation Research Part C: Emerging Technologies*, *85*, 11–130.

González-Gil, A., Palacin, R., & Batty, P. (2015). Optimal energy management of urban rail systems: Key performance indicators. *Energy Conversion and Management*, *90*, 282–291.

Goverde, R. M. P., Bešinović, N., Binder, A., Cacchiani, V., Quaglietta, E., Roberti, R., & Toth, P. (2016). A three-level framework for performance-based railway timetabling. *Transportation Research Part C: Emerging Technologies*, *67*, 62–83.

Goverde, R. M. P., Corman, F., & 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*, 78–94.

Goverde, R. M. P., & Hansen, I. A. (2013). Performance Indicators for Railway Timetables. *2013 IEEE International Conference on Intelligent Rail Transportation Proceedings*, 301–306.

Goya, J., Miguel, G. D., Arrizabalaga, S., Zamora-Cadenas, L., Adin, I., & Mendizabal, J. (2018). Methodology and key performance indicators (KPIs) for railway on-board positioning systems. *IEEE Transactions on Intelligent Transportation Systems*, *19*, 4035–4042.

Gueudar-Delahaye, C., Weidner, T., Vis, T., Dislaire, S., Smailji, A., Sabbaghian, M., Wachter, M., Yee, R., & Sierhuis, P. (2022). Slides – DAS and SFERA edition 2 workshop [Retrieved from internal workshop material: `slides-das_and_sfera_edition_2_workshop-10102022.pdf`]. *Interne SNCF Réseau / UIC*.

Hanusz, Z., Tarasinska, J., & Zielinski, W. (2016). Shapiro–Wilk Test with Known Mean. *REVSTAT – Statistical Journal*, *14*(1), 89–100.

Huang, P., Lessan, J., Wen, C., Peng, Q., Fu, L., Li, L., & Xu, X. (2020). A Bayesian network model to predict the effects of interruptions on train operations. *Transportation Research Part C: Emerging Technologies*, *114*, 338–358.

Huang, Y., Sun, J., & Tian, Y. (2024). A Bayesian Optimization Method for Finding the Worst-Case Scenarios of Autonomous Vehicles. *IEEE Transactions on Intelligent Transportation Systems*, *26*(1), 529–543.

Kecman, P., & Goverde, R. M. (2015a). Online data-driven adaptive prediction of train event times. *IEEE Transactions on Intelligent Transportation Systems*, *16*, 465–474.

Kecman, P., & Goverde, R. M. (2015b). Predictive modelling of running and dwell times in railway traffic. *Public Transport*, *7*, 295–319.

Kennisinstituut voor Mobiliteitsbeleid (KiM). (2020). *Inzicht in de kosten van het spoorgoederenvervoer* (KiM-20-A20). Den Haag, Nederland. https://www.kimnet.nl/publicaties/rapporten/2020/12/01/inzicht-in-de-kosten-van-het-spoorgoederenvervoer

Kennisinstituut voor Mobiliteitsbeleid (KiM). (2021). *Maatschappelijke kosten van verstoringen op het spoor* (tech. rep.) (KiM-publicatie, KiM-21-A23). Ministerie van Infrastructuur en Waterstaat. Den Haag, Nederland. https://www.kimnet.nl/publicaties/rapporten/2021/11/26/maatschappelijke-kosten-van-verstoringen-op-het-spoor

Knorr-Bremse. (n.d.). *Further customer benefits* (tech. rep.). https://rail.knorr-bremse.com/media/2000-products/product-broschures/p_1329_leader-2.pdf

Kristoffersson, I., & Palmqvist, C.-W. Improving commuter train punctuality using lead indicators. In: In *24th EURO Working Group on Transportation Meeting, Aveiro, Portugal*. 2021. www.sciencedirect.comwww.elsevier.com/locate/procedia2352-1465

Kyriacou, V., Englezou, Y., Panayiotou, C., & Timotheou, S. (2023). Bayesian Traffic State Estimation Using Extended Floating Car Data. *IEEE Transactions on Intelligent Transportation Systems*, *24*, 1518–1532.

Large, D., Golightly, D., & Taylor, E. (2017). Train-driving simulator studies: Can novice drivers deliver the goods? *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, *231*, 1186–1194.

Li, Z. C., Wen, C., Hu, R., Xu, C., Huang, P., & Jiang, X. (2021). Near-term train delay prediction in the Dutch railways network. *International Journal of Rail Transportation*, *9*, 520–539.

Li, Z., Chen, L., Roberts, C., & Zhao, N. (2018). Dynamic Trajectory Optimization Design for Railway Driver Advisory System. *IEEE Intelligent Transportation Systems Magazine*, *10*(1), 121–132.

Luo, Y. C., Xun, J., Wang, W., Zhang, R. Z., & Zhao, Z. C. (2025). A Driver Advisory System Based on Large Language Model for High-speed Train. https://doi.org/10.48550/arXiv.2501.07837

McHugh, M. L. (2013). The chi-square test of independence. *Biochemia Medica*, *23*(2), 143–149.

McKnight, P. E., & Najab, J. (2010). Mann–Whitney U Test. In I. B. Weiner & W. E. Craighead (Eds.), *The corsini encyclopedia of psychology*. John Wiley & Sons. https://doi.org/10.1002/9780470479216.corpsy0524

Meijer, S. A., Mayer, I. S., van Luipen, J., & Weitenberg, N. (2012). Gaming Rail Cargo Management: Exploring and Validating Alternative Modes of Organization. *Simulation & Gaming*, *43*(1), 85–101.

Meng, L., Abid, M. M., Jiang, X., Khattak, A., & Khan, M. B. (2019). Increasing Robustness by Reallocating the Margins in the Timetable. *Journal of Advanced Transportation*. https://doi.org/10.1155/2019/1382394

Nicholson, G. L., Kirkwood, D., Roberts, C., & Schmid, F. (2015). Benchmarking and evaluation of railway operations performance. *Journal of Rail Transport Planning & Management*, 274–293.

Ogawa, T., Yokouchi, T., Takeuchi, Y., & Saito, T. (2024). Verifying the Energy-saving Effect of a Driver Advisory System using Speed Estimation for Freight Trains. *IEEJ Journal of Industry Applications*, *13*, 754–760.

Pachl, J. (2020). *Railway Signalling Principles*. Technische Universität Braunschweig. https://doi.org/10.13140/RG.2.2.14777.60004/1

Panou, K., Tzieropoulos, P., & Emery, D. (2013). Railway driver advice systems: Evaluation of methods, tools and systems. *Journal of Rail Transport Planning & Management*, *3*, 150–162.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine Series 5*, *50*(302), 157–175.

Pineda-Jaramillo, J., Bigi, F., Bosi, T., Viti, F., & D'ariano, A. (2023). Short-Term Arrival Delay Time Prediction in Freight Rail Operations Using Data-Driven Models. *IEEE Access*, *11*, 46966–46978.

ProRail. (2024). *Werkzaamheden aan het spoor in Duitsland* (tech. rep.). Ministerie van Infrastructuur en waterstaat.

Quaglietta, E., Pellegrini, P., & Goverde, R. M. P. (2016). The ON-TIME real-time railway traffic management framework: A proof-of-concept using a scalable standardised data communication architecture. *Transportation Research Part C: Emerging Technologies*, *63*, 23–50.

Radtke, A. (2014). Infrastructure Modelling. In I. A. Hansen & J. Pachl (Eds.), *Railway timetabling & operations* (2nd revised and extended edition, pp. 47–62). DVV Media Group GmbH, Eurailpress.

Razali, N., & Wah, Y. B. (2011). Power comparisons of shapiro–wilk, kolmogorov–smirnov, lilliefors and anderson–darling tests. *Journal of Statistical Modeling and Analytics*, *2*(1), 21–33.

Salden, M., Tiekstra, N., Bogers, W., Vis, T., & Jonker, H. (2024). *Visie c-das* (tech. rep.). Nederlandse Spoorwegen.

Sandblad, B., Andersson, A. W., & Tschirner, S. (2015). Information Systems for Cooperation in Operational Train Traffic Control. *Procedia Manufacturing*, *3*, 2882–2888.

Scheepmaker, G. M., & Goverde, R. M. P. (2021). Multi-objective railway timetabling including energy-efficient train trajectory optimization. *European Journal of Transport and Infrastructure Research*, *21*, 1–42.

Scheepmaker, G. M., Goverde, R. M. P., & Kroon, L. G. (2017). Review of energy-efficient train control and timetabling. *European Journal of Operational Research*, *257*, 355–376.

Scheepmaker, G. M., Willeboordse, H. Y., Hoogenraad, J. H., Luijt, R. S., & Goverde, R. M. P. (2020). Comparing train driving strategies on multiple key performance indicators. *Journal of Rail Transport Planning & Management*, *13*.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3-4), 591–611.

Sogin, S. L., Lai, Y.-C. R., Dick, C. T., & Barkan, C. P. L. (2013). Comparison of Capacity of Single- and Double-Track Rail Lines Using Simulation Analyses. *Transportation Research Record*, *2374*(1), 111–118.

Spanninger, T., Trivella, A., Büchel, B., & Corman, F. (2022). A review of train delay prediction approaches. *Journal of Rail Transport Planning & Management*, *22*, 100312.

Stenström, C., Parida, A., & Galar, D. (2012). Performance Indicators of Railway Infrastructure. *International Journal of Railway Technology*, *1*, 1–18.

Surname, I., Surname, I., & Surname, I. (2000a). The title of the article. *The Title of the Journal*, *1*(2), 123–456.

Surname, I., Surname, I., & Surname, I. (2000b). *The title of the book* (8th ed.). Publisher.

Surname, W. < I., Surname, I., & Surname, I. (2000). *Title of the website*. Retrieved December 24, 2020, from https://example.com

Tiong, K. Y., & Palmqvist, C. W. (2023). Quantitative methods for train delay propagation research. *Transportation Research Procedia*, *72*, 80–86.

van Infrastructuur en Waterstaat, M. (2020). *3.5 Openbaar Vervoer en Spoor – Artikel 16* (tech. rep.) (Beleidsregel inzake punctualiteitsprestaties op het spoor). Rijksoverheid. https://wetten.overheid.nl/

Verstappen, V., Pikaar, E., & Zon, R. (2022). Assessing the impact of driver advisory systems on train driver workload, attention allocation and safety performance. *Applied Ergonomics*, *100*, 103645.

Vo, K.-P. (1988). Dag—a program that draws directed graphs. *Software: Practice and Experience*, *18*. https://doi.org/10.1002/spe.4380181104

Wang, P., & Goverde, R. M. P. (2016). Multiple-phase train trajectory optimization with signalling and operational constraints. *Transportation Research Part C: Emerging Technologies*, *69*, 255–275.

Wang, P., Goverde, R. M. P., & van Luipen, J. (2019). A connected driver advisory system framework for merging freight trains. *Transportation Research Part C: Emerging Technologies*, *105*, 203–221.

Wang, P. (2017). Train Trajectory Optimization Methods for Energy-Efficient Railway Operations. *Dissertation (TU Delft), Delft University of Technology*. https://doi.org/10.4233/uuid:ce04a07d-89fc-470a-9d1a-b6fae9182dae

Wang, P., & Goverde, R. M. P. (2017). Development of A Train Driver Advisory System: ETO. *2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, 140–145. https://doi.org/10.1109/MTITS.2017.8005654

Weeda, V., & van Onna, J. (2019). *Geen boze treinreizigers meer?* (Tech. rep.). ProRail and NS.

Wen, C., Huang, P., Li, Z., Lessan, J., Fu, L., Jiang, C., & Xu, X. (2019). Train dispatching management with data-driven approaches: A comprehensive review and appraisal. *IEEE Access*, *7*, 114547–114571.

Xiao, Z., Wang, Q., Sun, P., Zhao, Z., Rao, Y., & Feng, X. (2021). Real-Time Energy-Efficient Driver Advisory System for High-Speed Trains. *IEEE Transactions on Transportation Electrification*, *7*, 3163–3172.

Zhang, Y., Li, R., Guo, T., Li, Z., Wang, Y., & Chen, F. (2019). A conditional Bayesian delay propagation model for large□scale railway traffic networks. *Proceedings of the Australasian Transport Research Forum (ATRF)*, 1–12.

Zhu, H., Sun, X., Chen, L., Gao, S., & Dong, H. (2016). *Analysis and design of Driver Advisory System (DAS) for energy-efficient train operation with real-time information*. IEEE.

Quantifying the Operational Benefits of a Railway Driver Advisory System

Julia Hannah Michels
MSc Transport, Infrastructure and Logistics
Delft University of Technology, The Netherlands

ABSTRACT As rail networks face increasing pressure to improve punctuality and capacity under mixed-traffic conditions, digital support tools such as Driver Advisory Systems (DAS) are becoming essential for more efficient operations. Yet, despite extensive simulation-based research, little is known about how these systems perform in real-world freight operations. This study addresses that gap by providing a data-driven evaluation of RouteLint, a Connected DAS implemented on the Dutch Brabantroute. Using one year of operational data, the analysis combines timetable and signal-level information to compare train performance with and without RouteLint. Non-parametric statistical tests are applied across macroscopic and microscopic scales to assess differences in timetable adherence, signal aspects, and train interactions. The results show measurable but modest improvements in punctuality and a reduction in unplanned yellow signals for RouteLint users, achieved without additional infrastructure investments. RTL users show modest but consistent gains in punctuality: median deviations are significantly smaller at major nodes, Eindhoven (-33 seconds), Tilburg (-23 seconds), and at key junctions (Tgra -26.5 seconds), Tba (-18 seconds), and Tbi (-66.5 seconds), with small effect sizes. Unplanned yellow signals are also lower for RTL users (overall 6% reduction). The findings show that DAS, like RTL, can play a key role in improving reliability and efficiency in railway operations.

INDEX TERMS Driver Advisory System, freight rail, RouteLint, Operational performance, punctuality, signalling, quantitative analysis

1. Introduction

The rail network of the Netherlands is one of the busiest and most complex railway systems in the world (Verstappen, Pikaar, and Zon, 2022). With increasing traffic demand and a largely saturated infrastructure, the system faces growing challenges in maintaining reliability and operational efficiency (P. Wang and R. M. P. Goverde, 2016). Rail freight transport is expected to grow by 50% in 2030 and 100% in 2050 compared to 2015, and for passenger transport, a growth of 30% is forecasted for 2030. Several lines in the Dutch railway network are expected to face potential capacity shortages on a standard day in the 2026 timetable year, particularly in the Randstad, the southern corridors, and the routes towards Belgium and Germany (European Commission, 2011).

These capacity pressures are particularly evident on mixed-traffic lines, where dense passenger services leave limited time slots for freight trains. Although freight trains operate according to pre-designed timetables, their services are often less punctual than passenger services (P. Wang, R. M. P. Goverde, and Luipen, 2019). Compared with passenger trains, freight services differ substantially in power, mass, braking performance, and acceleration behaviour (Sogin et al., 2013). Their greater mass leads to longer acceleration and braking curves, higher energy consumption during speed changes (George, 2022), and lower operational speeds on some tracks to reduce rail wear. Consequently, freight train operations should be as conflict-free as possible to minimise delay and its propagation.

One promising approach to improve operational efficiency without costly infrastructure investments is the use of Driver Advisory Systems (DAS). A DAS is an onboard decision-support tool that provides real-time guidance to train drivers, typically in the form of optimal speed profiles or schedule adherence advice (P. Wang, R. M. P. Goverde, and Luipen, 2019). By helping drivers anticipate upcoming signals and conflicts, DAS can reduce unnecessary braking and stopping, thereby improving punctuality and energy efficiency. In the Netherlands, the Infrastructure Manager (IM) ProRail has deployed a DAS known as RouteLint (RTL), which provides train drivers with up-to-date timekeeping advice and information on the status of upcoming routes and nearby trains. RTL's guidance is generated centrally based on real-time network conditions and communicated to locomotives,

aiming to help drivers stay on schedule and avoid red or yellow signal aspects.

Prior studies have demonstrated the theoretical potential of DAS to enhance timetable adherence, reduce energy usage, and even lower driver workload (Panou, Tzieropoulos, and Emery, 2013; Dong, Zhu, and Gao, 2018; Verstappen, Pikaar, and Zon, 2022). However, most of these findings are based on simulation experiments, prototype trials, or driving simulator tests rather than large-scale operational data. In practice, there is still limited empirical evidence of how much systems like RTL improve performance in daily operation, especially for freight trains on busy, mixed-traffic networks. This gap between findings from controlled or simulation-based studies and actual operational outcomes motivates a data-driven evaluation of DAS performance in real-world operation.

Building on this, the following main research question is formulated: *How can the operational performance of freight trains be evaluated using empirical data from Driver Advisory Systems such as RouteLint?* The main question is explored through five sub-questions: What is the state-of-the-art and the state-of-practice of Driver Advisory Systems? Which methods and KPIs can be used to quantify the effect of Driver Advisory Systems on freight train operations using empirical data? How do freight train operations with and without a Driver Advisory System differ? How do the performance impacts of a Driver Advisory System vary across scales of operation and over space and time? What policy implications can be drawn from the data-driven analysis?.

We examine the operational performance differences using four key performance indicators (KPIs): transport volume, timetable deviation, unplanned yellow signals, and train interaction. To capture spatial patterns in these indicators, aggregated event graphs were constructed to map train events at both the macroscopic level (timetable points) and microscopic level (signals). For each track edge, average running times were estimated under conflict-free conditions, defined as on-time passages, through movements, and green signal aspects, and absolute deviations from these reference values were computed per train. Differences between RTL users and non-users with respect to timetable deviation were tested per edge using the Mann–Whitney U test, with random subsampling applied as a robustness check. Unplanned yellow signal passages and their associated causes were analysed using Pearson's Chi-squared tests. Real-world operational data covering one year were used to provide a data-driven assessment of DAS effects under mixed-traffic conditions.

In summary, the main contributions of this paper are as follows:

1) We present a comprehensive methodological framework to evaluate train operational performance between two groups, drivers with and without DAS, using multiple analytical methods and key performance indicators.
2) We apply this framework to large-scale real-world data from the Dutch network to quantify the operational effects of the RTL system on freight train performance along the Brabantroute corridor.
3) Based on the empirical findings, we provide practical insights and policy recommendations for infrastructure managers, system developers, railway undertakings, and train drivers on how DAS can be used to enhance network capacity, efficiency, and service reliability.

The remainder of this paper is structured as follows: Section 2 provides a review of the relevant literature and further elaborates on the research gap. Section 3 describes the methodological framework developed for this study. Section 4 introduces the case study and presents the main results. Finally, Section 5 summarises the key findings, discusses the policy implications, and outlines directions for future research.

2. Literature review

This section reviews the existing literature and current practices related to DAS. DAS can be grouped into three types: Standalone-DAS (S-DAS), Networked-DAS (N-DAS), and Connected-DAS (C-DAS) (P. Wang, R. M. P. Goverde, and Luipen, 2019; Panou, Tzieropoulos, and Emery, 2013).

- S-DAS operates independently by pre-loading all necessary data onto the train before or at the start of the journey. It functions without real-time communication with the traffic control centres and does not adapt to changes in the operational plan during the train run.
- N-DAS, on the other hand, is capable of communicating with one or more traffic control centres, allowing it to receive updates regarding schedules or routing information. While more complex than S-DAS, the updates in N-DAS are typically periodic rather than continuous, and thus not truly real-time.
- C-DAS represents the most advanced form, featuring real-time bi-directional communication with traffic control centres in the areas where the train is operating. C-DAS can dynamically receive updated timetable targets and generate a corresponding updated speed profile to avoid conflicts, providing advice to the driver to adhere to the updated schedule.

We describe C-DAS as a system that distributes intelligence between the traffic control centre and the train, offering a practical solution for managing real-time traffic and schedule disturbances. According to Panou, Tzieropoulos, and Emery (2013), three main architectural clusters of DAS can be distinguished. In the DAS-Onboard configuration, the onboard unit computes the optimal speed profile, derives the driving advice, and presents it directly to the driver. In the DAS-Distributed configuration, computation occurs at trackside units (TSUs), and only the resulting advice is transmitted to the train, where it is displayed to the driver.

In the DAS-Central configuration, all processing occurs at the TSUs, leaving no onboard computation.

A critical aspect of C-DAS design is showing accurate advice to the driver. Panou, Tzieropoulos, and Emery (2013) also describes the different forms of advice:

- Suggested speed: This is the most straightforward form of advice, providing optimal speed instructions for the driver to follow. It is particularly useful when specific schedule targets need to be met.
- Timekeeping: This advice shows the deviation between the actual time at a specific point on the track and the scheduled time on the timetable, helping the driver manage delays.
- Control action advice: Including suggestions for coasting, tractive power, or brake pressure to achieve the most efficient operation.
- Energy savings: Besides speed or time suggestions, DAS can also advise the driver with the objective of minimising energy usage.

Several authors have shown via simulation and prototypes that DAS can improve both energy efficiency and timetable adherence under various conditions. For example, Dong, Zhu, and Gao (2018) demonstrated with a hardware-in-the-loop experiment that a DAS can optimize a train's speed profile to reduce energy use in nominal conditions and then switch to delay-recovery driving in disturbed conditions, thereby improving punctuality. Xiao et al. (2021) found in a driver-in-the-loop simulator test that even a single high-speed train equipped with DAS improved its on-time performance and used less energy under minor delays, compared to driving without DAS. Similarly, P. Wang, R. M. P. Goverde, and Luipen (2019) showed that under normal operations, a real-time DAS could help maintain an optimal speed trajectory that minimizes running time deviations and prevents unnecessary braking. These studies collectively indicate that DAS guidance can smooth out operations: avoiding excessive stops or speed fluctuations, which in turn saves energy and reduces knock-on delays.

Recent studies further explore the integration of artificial intelligence into DAS. For instance, Luo et al. (2025) propose an intelligent DAS framework leveraging large language models to generate context-specific advice for high-speed trains. Although such AI-driven systems are promising, capable of learning from operational data to prevent incidents, they remain largely experimental and unvalidated in live settings. Other researchers have refined traditional algorithms: Ghaviha et al. (2017) developed a smartphone-based S-DAS that models dynamic traction efficiency more accurately to improve energy estimates, while Li et al. (2018) introduced a hybrid online–offline optimisation that precomputes optimal trajectories and adapts them in real time as conditions evolve. We observe that the literature demonstrates a steady progression from basic advisory concepts toward sophisticated systems incorporating optimisation and AI. However, most of these advances have been validated only in simulations or small-scale trials, offering valuable insight but limited empirical evidence from daily railway operations.

Among the DAS implementations in Europe, RTL is the system developed by the Dutch IM, ProRail, and serves as the primary case study in this research. It is a real-world example of a C-DAS, in which all advice is centrally generated and transmitted to the train. The onboard functionality is limited to displaying the prepared advice to the driver (Panou, Tzieropoulos, and Emery, 2013). Nowadays, RTL is widely used by several European freight companies that deliver goods between the Netherlands and its neighbouring countries. It provides timekeeping advice and gives additional information on the upcoming route, scheduled event times at critical timetable points, and the real-time positions of nearby trains. Figure 1 shows the dashboard of the RTL app and its available advice and information.

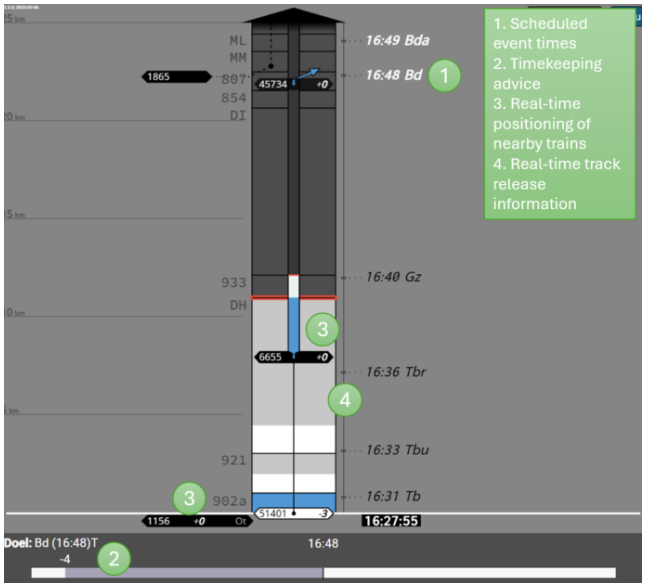


FIGURE 1: RTL dashboard showing scheduled times, timekeeping advice, nearby trains, and track release status.

While theoretical studies and small-scale demonstrations have extensively explored DAS functionality, empirical evaluations of its real-world impact remain scarce, particularly for freight operations. Most prior work relies on simulation models, controlled field experiments, or driving simulators with human participants rather than on large-scale operational data. For instance, Panou, Tzieropoulos, and Emery (2013) and Quaglietta, Pellegrini, and R. M. P. Goverde (2016) analysed potential improvements and timing strategies, yet without direct observation of long-term operations. Experimental prototypes, such as those presented by Pengling Wang and Rob M. P. Goverde (2017) and Large, Golightly, and Taylor (2017), demonstrate potential under controlled conditions, and Verstappen, Pikaar, and

Zon (2022) evaluated DAS effects on driver workload using a high-fidelity simulator. Similarly, Dong, Zhu, and Gao (2018) and Xiao et al. (2021) validated DAS concepts through hardware- and driver-in-the-loop experiments, confirming operational and energy benefits under idealised circumstances. These contributions have been crucial for understanding DAS mechanisms and refining algorithms and interfaces. However, their findings reflect best-case scenarios rather than typical railway conditions, where stochastic delays, diverse driver behaviour, and complex traffic interactions prevail.

In practice, DAS performance can be influenced by numerous operational factors: drivers may not fully adhere to the advice, network dynamics introduce uncertainty, and integration with existing systems may be imperfect. To our knowledge, few large-scale, data-driven studies have examined DAS effectiveness under everyday operational conditions, and almost none focus on freight operations.

3. Methodology

This section first outlines the problem description and assumptions 3.1 datasets used in the analysis and describes the procedures for data cleaning and merging (Section 3.2). Subsequently, the key performance indicators (KPIs) are introduced in Section 3.3, followed by an explanation of the method used to construct the event graphs (Section 3.4). Finally, the analytical approaches employed to evaluate train performance are presented in Sections 3.5 and 3.6.

3.1. Problem description and assumptions

We aim to quantify the operational effect of a C-DAS on freight trains. Using one year of observational data, we compare trains that used Routelint (RTL) to trains that did not, across multiple performance indicators and spatial scales, without altering infrastructure or timetables. Because the study relies on operational (non-experimental) data, inferences are associational: we estimate distributional and rate differences between groups under realistic conditions.

To ensure the validity of the analysis, several methodological and operational assumptions are made. First, we assume that an active login on RTL during a train run indicates that the logged-in person is the active driver and that, on average, the advice provided by the system is followed.

The construction of the reference set for "conflict-free" passages, defined as on-time, through movements under green (GR) signal aspects, is assumed to produce representative baseline running times for each edge. A minimum threshold of fifteen valid train passages per edge ensures statistical stability of the reference estimates.

Because individual train runs may not be statistically independent, drivers may appear multiple times, operators may follow consistent driving styles, and trains can influence each other through network effects, we treat observations as independent for the purpose of test statistics.

3.2. Preprocessing and variable selection

Data cleaning involved removing records with missing timestamps and harmonising operator names across datasets. To ensure statistical robustness, only edges with at least 15 valid train runs should be included. Missing signal aspects were reconstructed using the Signal Release Delay indicator (Equation 1). We define the Signal Release Delay as the time between passing s and the subsequent clearing of $s{+}1$:

$$\text{Signal Release Delay} \;=\; \max\!\big(0,\; m^{\text{clear}}_{t,s+1} - m^{\text{pass}}_{t,s}\big) \qquad (1)$$

Where:

- $m^{\text{pass}}_{t,s}$ is the timestamp when train t passes signal s,
- $m^{\text{clear}}_{t,s+1}$ is when signal $s{+}1$ first shows a green of yellow aspect,
- $\max(0, \cdot)$ ensures that negative values are set to zero, which correspond to s being green upon approach.

Classification rule:

- Signal Release Delay $= 0$ Green (GR) at s;
- Signal Release Delay > 0 Yellow (GL) at s.

Finally, a directional attribute was assigned to each train based on its sequence of event times and signal order along the corridor, either west or eastbound.

3.3. Key Performance indicators

Four Key Performance Indicators (KPIs) were defined:

1) Transport volume
2) Timetable deviation at timetable points
3) Occurrence of unplanned yellow aspects
4) Train interaction (preceding train influence)

To answer the second sub-question of how freight train performance differs between operations with and without a DAS, a network representation of the corridor is provided through an aggregated event graph.

In the macroscopic graph $G_{macro}(V, E)$, the nodes represent the timetable points and the edges are the corridors in between. The nodes contain the information per train of the original planned event times, the real-time planned event times, the executed event times, and the deviation from the planned and original timetable. In contrast, in the microscopic graph $G_{micro}(V, E)$, the nodes represent signals, both automatic and controlled, and contain information about the signal aspect, passage time, clearance time, release delay, and the cause of the signal aspect for each train. The edges represent the track sections between the signals and include the running times for each train on those tracks.

3.4. Aggregated Event Graphs

Both graphs were implemented using the NetworkXPython package, which allows for the visualisation of large directed acyclic graphs. The nodes and edges were defined based on the infrastructure topology, using the ordered sequence of

timetable points or signals to establish directional connectivity.

Train-specific events were then mapped onto these edges/nodes, and the aggregated values form the basis for the statistical analysis presented in the next section. From the aggregated event graphs, the four KPIs are derived to evaluate the operational performance of freight trains with and without a DAS.

3.5. Mann-Whitney U test

The Mann-Whitney U test tests whether the distribution of deviations for RTL users systematically differs from that of non-users. The test assumes that the two groups are independent. Although this assumption cannot be formally verified, a robustness check based on random subsampling is conducted after each test to assess the stability of the results. In this procedure, 30% of the data are randomly sampled to confirm that the conclusions remain consistent across subsamples. This step, therefore, serves as a robustness validation of the statistical findings.

The test is applied per edge in the event graph, with the following hypotheses:

- Null hypothesis (H_0): There is no significant difference in driving performance between drivers with and without RTL.
- Alternative hypothesis (H_1): Drivers using RTL show significantly lower deviation from average running time.

The Mann–Whitney U statistic is calculated as:

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \tag{2}$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 \tag{3}$$

$$U = \min(U_1, U_2) \tag{4}$$

where n_1 and n_2 represent the sample sizes of the two groups, and R_1 and R_2 denote the sums of ranks for each group. The smaller of the two U values is used to determine the level of significance.

The tests are evaluated at a significance level of $p < 0.05$, meaning that differences with a lower p-value are regarded as statistically significant. In addition to statistical significance, effect sizes were computed to assess the magnitude of observed differences. For non-parametric comparisons performed using the Mann–Whitney U test, the effect size was derived from the standardized test statistic (Z) according to Cohen's definition:

$$\text{effect} = \frac{Z}{\sqrt{N}} \tag{5}$$

where N is the total number of observations included in the test. Following the conventions proposed by Cohen (1988), effect sizes of effect $= 0.10$, effect $= 0.30$, and effect $= 0.50$ were interpreted as small, medium, and large effects, respectively.

3.5.1. Macroscopic level

For the macroscopic evaluation, the deviation between the planned and executed event times is calculated at each timetable point for all trains. This comparison provides an overview of the punctuality performance across the corridor.

3.5.2. Microscopic level

On the microscopic level, the signals have no planned event times, only passing times. Using the Critical Path Method (CPM), as described by Cheng (1996), the sequence of events will be modelled to identify the critical path and quantify deviations from the ideal flow. CPM is selected for its ability to reveal timing dependencies and operational bottlenecks in the considered railway corridors. This allows for a comparative analysis of driver behaviour based on RTL usage. To examine the deviations on the track-based level, the average running time $\bar{r}_e$ is determined for $e \in E$ in the event graph.

To ensure a reliable reference running time, the average running time per category is calculated using only trains operating under nominal conditions. The VOS and TROTS datasets are combined to identify trains that fulfil these conditions. Some nodes are linked to timetable points based on geographical location, and only trains meeting the following criteria are included in the reference group:

- The train must have passed both consecutive timetable points with a deviation of no more than 180 seconds (3 minutes) from the scheduled time (Infrastructuur en Waterstaat, 2020). This condition corresponds to equations (7) and (8).
- Only trains that did not encounter yellow or red aspects at any of the signals are included, ensuring that the measured running times reflect conflict-free operating conditions unaffected by restrictive signalling (equation 9).
- The activity type a at both the current and next timetable point must be a through passage (DR), as expressed in equation (10).
- Each edge must contain at least 15 trains satisfying these criteria to ensure statistical reliability (equation 11).

The threshold of 15 trains per edge was selected after testing alternative cut-offs (10, 15, and 20), which showed that average running time estimates stabilised beyond 15 observations while preserving sufficient network coverage.

The average running time for each edge is computed as:

$$\bar{r}_e = \frac{\sum_{t \in T} z_{t,e} \cdot r_{t,e}}{\sum_{t \in T} z_{t,e}} \qquad \forall e \in E \tag{6}$$

Under the following conditions:

$$|p^{\text{plan}}_{t,d} - p^{\text{real}}_{t,d}| \leq 180 \qquad \forall t \in T, d \in D \tag{7}$$

$$|p^{\text{plan}}_{t,d+1} - p^{\text{real}}_{t,d+1}| \leq 180 \qquad \forall t \in T, d \in D \tag{8}$$

$$asp_{t,s} = GR \qquad \forall t \in T, \forall s \in S_t(d, d{+}1) \tag{9}$$

$$a_{t,d}, a_{t,d+1} = DR \qquad \forall t \in T, d \in D \tag{10}$$

$$\sum_{t \in T} z_{t,e} \geq 15 \qquad \forall e \in E \tag{11}$$

Where:

- E – set of all edges between consecutive signals $(s, s{+}1)$, $e \in E$
- T – set of trains considered in the analysis, $t \in T$
- S – set of signals along the corridor, $s \in S$
- D – set of timetable points along the corridor, $d \in D$
- $d(s)$ – timetable points located at the position of signal s
- $S(d, d{+}1)$ – set of all signals located between two consecutive timetable points d and $d{+}1$
- $S_t(d, d{+}1) \subset S(d, d{+}1)$ – subset of signals that were actually passed by train t between timetable points d and $d{+}1$
- $r_{t,e}$ – running time of train t over edge e [s]
- $a_{t,d}$ – activity of event of train t at timetable point d
- $p^{plan}_{t,d}$ – planned passing time of train t at timetable point d [s]
- $p^{real}_{t,d}$ – real (observed) passing time of train t at timetable point d [s]
- $asp_{t,s}$ – aspect shown by signal s for train t
- $z_{t,e}$ – binary variable, 1 if train t is included in the running time calculation for edge e, 0 otherwise

With the average running times per edge, the average absolute deviation from the average running time is computed as:

$$\bar{\Delta r}_e = \frac{1}{|T_e|} \sum_{t \in T_e} |r_{t,e} - \bar{r}_e| \qquad \forall e \in E \tag{12}$$

where r is the observed running time for a train over a given edge $e \in E$, and $\bar{r}_e$ is the average running time for that edge, as computed from a filtered reference group. This yields a distribution of absolute running time deviations per train and per edge. The absolute formulation is adopted because both early and late arrivals are considered deviations from the planned schedule and equally relevant in assessing punctuality. Once all deviations are determined, the aggregated event graphs are used to identify locations (edges) with the highest deviations. For each edge, the significance of performance differences between RTL users and non-users is then evaluated using the Mann–Whitney U test.

3.6. Chi-squared test

To determine whether RTL usage influences the frequency of unplanned yellow signals, a Pearson's Chi-squared test of independence is applied. The Chi-squared test is suitable for this analysis because it evaluates the association between two categorical variables, in this case, RTL usage (user vs. non-user) and signal aspect (unplanned vs. planned yellow and green), without assuming any specific data distribution (McHugh, 2013; Pearson, 1900). It therefore provides an appropriate method for testing whether the proportion of unplanned yellow signals differs significantly between the two groups.

The test compares the observed frequencies O_i in each category to the frequencies E_i expected under the null hypothesis of independence.

The null and alternative hypotheses are formulated as:

- H_0: There is no association between RTL usage and the occurrence of unplanned yellow signals.
- H_1: RTL usage is associated with a different frequency of unplanned yellow signals.

The Chi-squared statistic χ^2 is calculated as:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}, \tag{13}$$

where k is the number of categories. Large deviations between observed and expected frequencies result in a higher χ^2 value, indicating a more substantial likelihood that the variables are not independent.

A significant result ($p < 0.05$) suggests that the distribution of yellow signal types differs between RTL users and non-users, indicating that RTL usage may influence how often trains encounter unplanned yellow signals. Additionally, the analysis evaluates whether different preceding-train scenarios are associated with variations in the causes of yellow signal encounters and whether these effects differ between RTL users and non-users.

To account for multiple hypothesis testing, the False Discovery Rate (FDR) correction is applied using the Benjamini–Hochberg (BH) procedure (Benjamini and Hochberg, 1995). This correction is applied to both the Mann–Whitney U and Chi-squared test results whenever more than five independent tests are performed simultaneously. The BH method controls the expected proportion of falsely rejected null hypotheses (false positives) among all significant results, thus maintaining the overall reliability of the statistical inference. Without such a correction, performing multiple tests increases the likelihood of obtaining false positives, as even random variation can appear statistically significant when many hypotheses are tested simultaneously. The adjusted significance threshold $p_{\mathbf{BH}}$ is calculated as:

$$p_{\mathbf{BH}} = \frac{i}{m} \cdot Q, \tag{14}$$

where i is the rank of the individual p-value, m the total number of tests, and Q the desired false discovery rate (typically 0.05). This procedure ensures that significant results reported in the analysis are statistically robust and not due to random fluctuations from repeated testing.

4. Results

This section presents the empirical results comparing freight train operations with and without RTL on the Brabantroute. Using one year of matched timetable, signal, and RouteLint (RTL) log data, we evaluate performance differences across macroscopic (timetable-point) and microscopic (signal-to-signal) scales.

Section 4.1 outlines the case study and data coverage, followed by Section 4.2 describing traffic volumes and RTL usage. Section 4.3 analyses timetable deviations using Mann-Whitney U tests, with robustness checks through random subsampling. Section 4.4 examines unplanned yellow signal passages via χ^2 tests, and lastly, Section 4.5 explores train interaction effects.

4.1. Case study

The analysis focuses on the Brabantroute corridor between Eindhoven and Tilburg, a key freight artery within the Dutch railway network. In total, 68 km of track were analysed: 45 km between Tilburg Reeshof (Tbr) and Tongelre aansluiting (Tgra), 16 km between Vught aansluiting (Vga) and Tilburg aansluiting (Tba), and 7 km between Vught (Vg) and Boxtel (Btl). Along this corridor, 15 timetable points are located, including two major stations Tilburg (Tb) and Eindhoven (Ehv), seven minor passenger stations Tbr, Tilburg Universiteit (Tbu), Oisterwijk (Ot), Btl, Vg, Best (Bet), Eindhoven Strijp-S (Ehs), siding areas Tilburg Industrie (Tbi), Acht (At) and Liempde (Lpe), and some critical junctions Tba, Vga, and Tgra. The study covers one year of operational data, from August 2024 to July 2025, encompassing the 80-week Betuweroute diversion period, during which freight traffic intensity on this corridor increased due to maintenance works elsewhere.

Three primary data sources were used: (1) the VOS dataset, containing timetable and passing-point records; (2) the TROTS dataset, providing signal-level observations; and (3) RTL user logs, indicating whether a driver used the system. Records from these datasets were matched based on train number, date, and operator. In total, 23,573 individual freight trains were successfully identified across the datasets and included in the analysis.

4.2. Transport volume

Figure 2 illustrates the evolution of train activity along the corridor. Following the start of the Betuweroute diversion, the number of daily freight trains increased by nearly 80, reflecting the corridor's intensified use.

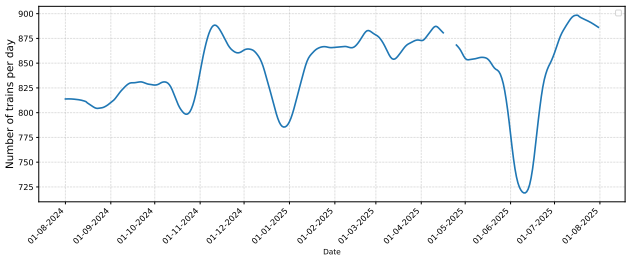


FIGURE 2: Spatial presentation of the total number of trains on the corridor.

Figures 3 and 4 show that out of the 23,528 freight trains, 3,635 (15.45%) used RTL, primarily among RTB and DBC operators. Usage increased after the start of the 80-week period and rose sharply from May onwards. Since then, approximately 38% of all drivers operating on the Brabantroute have actively used RTL.

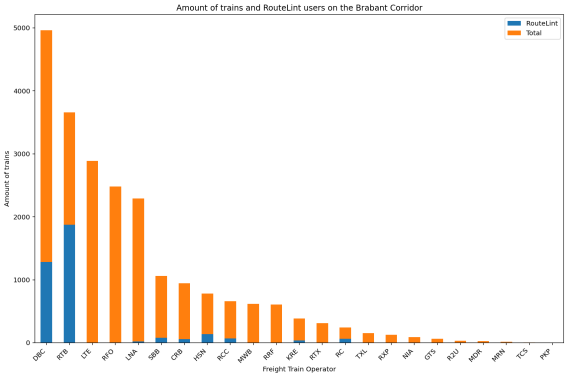


FIGURE 3: Overview of RTL usage per operator and in relation to total freight traffic.

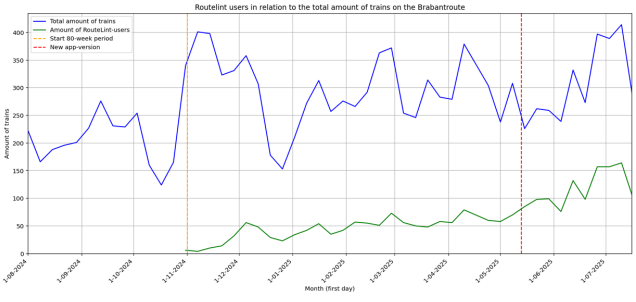


FIGURE 4: Growth in RTL usage over a year in relation to total freight trains.

4.3. Deviation

Figure 5 shows how RTL-users have, on average, lower absolute deviation than non-users, with exceptions during November, due to small RTL samples (e.g., 22 November 2024).

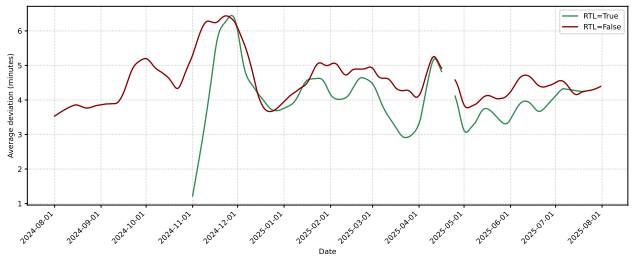


FIGURE 5: Monthly average absolute deviation from planned timetable split by RTL usage.

4.3.1. Macroscopic level

The differences in average deviations per direction and split by RTL usage are presented in Figures 6-9. In the eastbound direction, the largest deviations from the timetable occur around Ehv, whereas in the westbound direction, higher deviations are mainly observed near Tb. Across most points, the average absolute deviation is lower with RTL, with the largest average reduction at Tbi eastbound (-143 seconds), and Vga in westbound (-105 seconds).

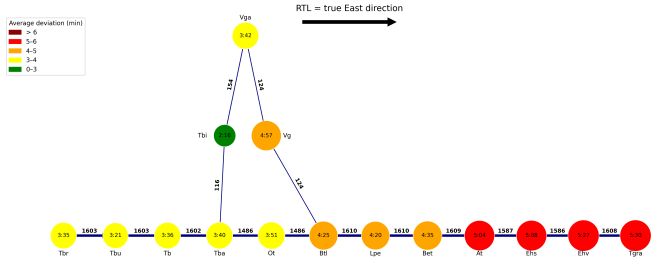


FIGURE 6: Average Deviation east direction with RTL

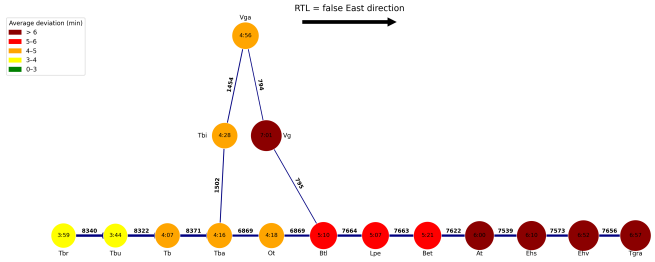


FIGURE 7: Average Deviation east direction without RTL

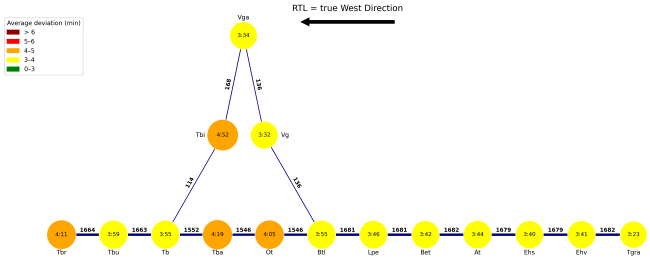


FIGURE 8: Average Deviation west direction with RTL

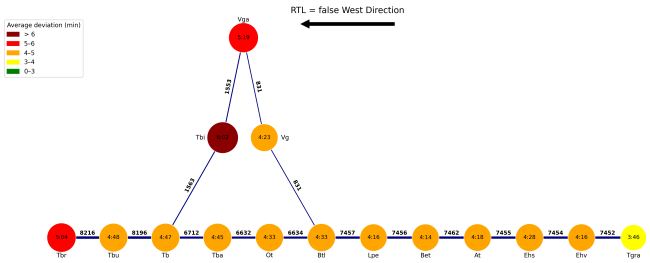


FIGURE 9: Average Deviation west direction without RTL

TABLE 1: Mann–Whitney U test results per timetable point and direction (east/west), comparing deviation between RTL=True and RTL=False trains.

d	Median$_T$	Median$_F$	Δ(T–F)	r_{effect}	p_{BH}
East					
Tbi	71.5	138	-66.5	0.1330	0.00001
Tgra	159.5	186	-26.5	0.0482	0.0001
Ehv	170	195	-25	0.045	0.00035
Tba	101	119	-18.0	0.0341	0.0056
Vga	92.5	124	-31.5	0.0731	0.0064
Tb	101	111	-10.0	0.0257	0.0412
West					
Ehs	126	159	-33.0	0.0758	$1.88{\cdot}10^{-10}$
Ehv	123	151.0	-28.0	0.0653	$3.32{\cdot}10^{-8}$
Tbr	112	138	-26.0	0.0573	$4.84{\cdot}10^{-7}$
Tb	103	126	-23.0	0.0508	$7.01{\cdot}10^{-6}$
Tbu	107	124	-17.0	0.0471	$3.46{\cdot}10^{-5}$
Btl	134	149	-15.0	0.0444	0.0002
At	130	147	-17.0	0.0435	0.0002
Bet	122	139	-17.0	0.0424	0.0003
Lpe	124	137	-13.0	0.0342	0.0034
Tgra	102	111	-9.0	0.0341	0.0034
Vga	130	151.5	-21.5	0.0565	0.0199

The Mann–Whitney U results (Table 1) show significantly smaller medians with RTL-usage after FDR correction at large passenger stations, Ehv (-25 seconds eastbound and -33 seconds westbound) and Tb (-10/-23 seconds), and at junctions with diverging scenarios Tgra (-26.5 seconds

eastbound, -9 westbound) and Tba (-18 seconds eastbound) and Tbi (-66.5 seconds eastbound). In both directions, effect sizes were small ($effect \leq 0.10$), suggesting that the distributions still overlapped considerably, despite a consistent shift towards lower deviations for RTL users.

4.3.2. Microscopic performance

Figures 10–11 highlight the highest average absolute deviations, where the line thickness corresponds to the magnitude of the deviation. The longest average running times occur in the corridor between Tbi - Vga for both directions: in the eastbound direction (TBI$703→VGA$727), the average running time is 449s for heavy trains and 439s for regular trains; in the westbound direction (VGA$728→TBI$721), the corresponding averages are 322s and 301s. Consistently, these long edges also show large average deviations 50s eastbound; $\tilde{4}$2s westbound), which is explained by the large inter-signal distances (≈10km), which increases the variability in running times.

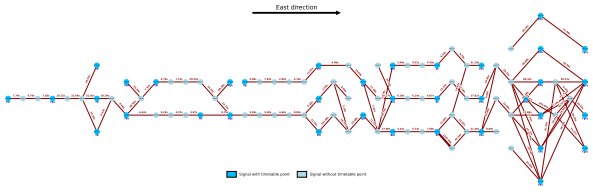


FIGURE 10: Average deviation per edge in the east direction.

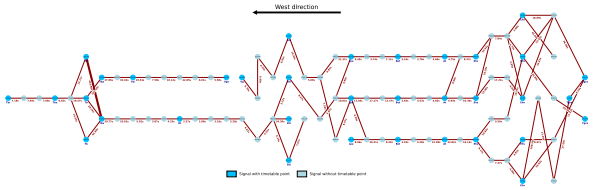


FIGURE 11: Average deviation per edge in the west direction.

To assess the association between RTL usage and punctuality, the absolute deviation distributions for RTL users and non-users were compared per edge using the Mann–Whitney U test. A small set of edges (14.3%) shows statistically significant differences, and after FDR correction, only 6.7% were significant (Table 2). The significant edges are shown on the corridor in figures 12 and 13.

In the eastbound direction, the highest deviation around the timetable point Tb is observed on edge TB$104 → TB$138. This edge represents an alternative route for eastbound traffic, as the track is normally scheduled for westbound operations (platform 2). The higher delays can be explained by the fact that blocks further along the line may still be reserved for opposing traffic, which have to be rerouted via platform 3. This edge also shows a significant difference: the median delay among RTL users is 3 seconds lower, while the average running time is approximately 30 seconds, indicating a relative improvement of around 10%.

The main route of the eastbound traffic through station Tb (TB$104 → TB$140 → TB$150 → TB$164), shows small significant differences of 0.52 and 0.10 seconds, with small effect values < 0.10, indicating a consistent yet marginal improvement.

The edge TB$164 → TBI$182 has a significant improvement of 2.07 seconds, with an average running time of 94 seconds (2.2%), and an average deviation of 28 seconds. This is a diverging movement from Tba towards Tbi that requires crossing.

From Ehs onwards, many edges show higher average deviations due to the high route heterogeneity around Eindhoven. Such as the edge EHV$110–EHV$184, part of a waiting/overtaking track with a crossing of three tracks, has an average deviation of 81 seconds, and edge EHV$184–EHV$258 with an average deviation of 96.55 seconds. In the eastbound significance map, Ehv shows several small but significant improvements (0–2 seconds) on short sections, consistent with micro-stabilisation rather than large time savings on edges EHV$1787 → EHV$26, EHV$26 → EHV$114, and towards Tgra EHV$210 → EHV$256.

The largest differences in medians can be found at a shunting area at At. AT$1346 → AT$1378 (and onward to AT$1765) functions as a shunting/holding siding. Traffic volume is low (96 trains/year) and deviations are high (197 seconds), reflecting enforced waiting or slow running to sequence with other trains. RTL usage is associated with significantly smaller deviations: -117 seconds (AT$1346-AT$1378) and -41 seconds (AT$1378-AT$1765). About 20% of trains on this segment use RTL, yet the effect remains detectable.

In the westbound, Ehv has way fewer different routes, and the most used westbound path is EHV$274 → EHV$150 (36 seconds) → EHV$62 (18s) and EHV$146 → EHV$62 (45 seconds), whose secondary function is waiting/overtaking track, and therefore faces higher average deviations. RTL users have statistically 5 seconds lower medians on EHV$274 → EHV$150. Edge EHV$268 → EHV$206 shows the highest average deviation at Ehv (169.2 seconds), reflecting an alternative route, with two crossings, that cause large deviations.

The other significant edges can be found around the diverging track section at Btl to Vg. BTL$1198→BTL$1134 has a significant but negligible median difference (-0.01s) on $\bar{r}_e$ = 36 seconds, BTL$1134→BTL$662 has a median difference of -0.67 seconds on $\bar{r}_e$ = 62 seconds, and BTL$662→BTL$618 -0.005s on 63 seconds running time. Together, these results suggest that just downstream of the Boxtel diverging area, RTL is associated with slightly more stable driving, however, with small effect sizes.

On the westbound corridor, high average deviations occur around timetable point Tb, mainly due to the heterogeneity of routes and operational movements in this area. The largest deviation of 443 seconds occurs on edge TB$152→TB$118. This track section serves as a secondary waiting/overtaking track (platform 4) where freight trains frequently pause or slow down to allow passenger services to pass.

TABLE 2: Edges with significant differences between RTL=True and RTL=False.

Direction	s	$s+1$	Δ(T-F)	effect	p_{BH}
West	EHV$274	EHV$150	-5	0.057	0.00001*
West	BTL$1134	BTL$622	-0.67	0.125	0.0021*
West	BTL$1198	BTL$1134	-0.28	0.117	0.0036*
East	TB$150	TB$164	-0.01	0.040	0.0022*
East	TB$140	TB$150	-0.52	0.037	0.0040*
East	TB$164	TBI$182	-2.07	0.071	0.0436*
East	TB$138	TB$150	-3	0.290	0.0436*
East	AT$1346	AT$1378	-117	0.335	0.0353*
East	EHV$210	EHV$256	-1	0.072	0.087
East	VGA$727	VGA$328	-2.47	0.061	0.091
East	BTL$1164	BTL$1188	-1	0.085	0.197
East	EHV$1787	EHV$26	-0.64	0.028	0.197
East	EHV$1008	EHV$1789	-2.02	0.072	0.212
East	TBI$703	VGA$727	-8.58	0.049	0.216
East	AT$1378	AT$1765	-41.88	0.199	0.267
East	EHV$26	EHV$114	-0.74	0.012	0.267

*Significant at $p < 0.05$.

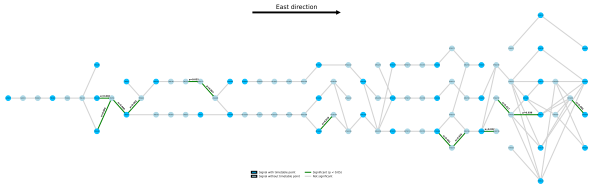


FIGURE 12: Result statistical analysis RTL use, east direction.

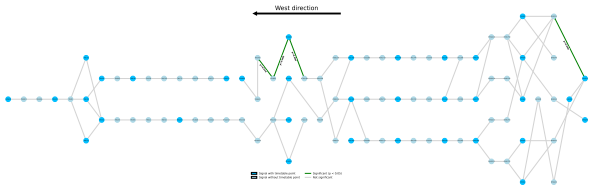


FIGURE 13: Result statistical analysis RTL use, West direction.

4.3.3. Robustness

To validate the robustness of the findings and test whether the two samples (RTL = True and RTL = False) are independent, the Mann–Whitney U test was repeated at both the macroscopic and microscopic levels using random subsampling. Approximately 30% of the data were randomly drawn for each test iteration, to confirm that the observed differences are not driven by a few extreme values or imbalanced sample sizes. The results recover the main patterns; marginal median shifts with very small absolute differences may lose significance under reduced n, as expected.

4.4. Yellow signals

To examine whether the occurrence of unplanned yellow signals differs significantly between RTL users and non-users, a chi-squared test was applied. The test result indicates a statistically significant difference ($p = 0.025$), suggesting that trains using RTL passed unplanned yellow signals less frequently than those without. From the 65910 yellow signals, 56343 had planned yellow aspects, and 9567 were unplanned (14.5%). However, the observed effect size is small: the probability of passing an unplanned yellow signal is 6% lower for RTL users, indicating that the practical difference between the two groups remains limited (Table 3).

TABLE 3: Chi-squared test unplanned yellow vs. RTL-use

	Planned Yellow	**Unplanned Yellow**
RTL = False	574,751	8,071
RTL = True	113,563	1,496
χ^2	5.03	
p_{χ^2}	0.025	
Odds ratio	0.94	

4.5. Train interaction

A significant difference was found in how often a train passed an unplanned yellow signal when a freight train and a Sprinter service was the preceding train occupying the upcoming block. As shown in Table 4, the likelihood of passing an unplanned yellow signal were 15% and 16% lower when RTL was used compared to not using RTL.

TABLE 4: Comparison of unplanned yellow signals per category between RTL users and non-users

Preceding train	n_T	n_F	Δ (pp)	OR	p_{chi}	p_{BH}
SPR	466	2849	-0.18	0.85	0.001*	0.007*
Freight	281	1550	-0.24	0.84	0.007*	0.023*
IC	692	2862	0.07	1.05	0.322	0.748
ICE	8	26	0.89	1.52	0.428	0.748
LL	16	77	0.33	1.20	0.606	0.849
OMD	6	35	0.42	1.19	0.874	0.952
LM	114	575	0.03	1.01	0.950	0.952

*Significant at $p < 0.05$.

Table 6 presents the different causes of unplanned yellow signals, comparing RTL users and non-RTL users across different train categories; the symbols representing the causes' operational movements are explained in Table 5.

The analysis reveals that the only significant difference, after correcting for false discovery rates, is found in the case of a preceding merging passenger train. This scenario caused 4.49 percentage more unplanned yellow aspects when using RTL, corresponding to an odds ratio of 1.31, and p_{BH} of 0.08.

TABLE 5: Overview of symbols of yellow aspect causes.

Cause symbol	Explanation
W	Order of trains is swapped.
>	Merging train movement in the same direction.
<	Diverging train movement in the same direction.
|	Following train movement in the same direction.
=	Hindrance from another train holding the next block.
X	Crossing train movement in the opposite direction.
TC l/L	Traffic Control scheduled l minutes for operation, where L minutes were necessary.
ARI	Route set too late; ARI (Automatic Route Setting) eventually sets the route.
PPR	Route set too late; the dispatcher ultimately sets the route manually

5. Conclusions and discussion

Using one year of real-world freight operations on the Brabantroute (23,573 runs), the Driver Advisory System RouteLint (RTL) was associated with smaller deviations from the planned timetable and fewer unplanned yellow signal aspects. Effects are modest per train but consistent, and clearest at the macroscopic scale, indicating that RTL helps drivers anticipate conditions around complex nodes (e.g., Eindhoven, Tilburg). This aligns with simulation evidence that DAS reduces behavioural variability and stabilises operations (Xiao et al., 2021; Dong, Zhu, and Gao, 2018; P. Wang, R. M. P. Goverde, and Luipen, 2019). Aggregation across timetable points reveals systematic improvements that are harder to detect at signal-to-signal resolution due to higher local variance. Directional differences also emerge: eastbound traffic shows larger absolute deviations (boosting power to detect differences), whereas westbound flows display more consistent macro-effects.

TABLE 6: Significant test differences in cause within unplanned yellow signals per category and cause: (RTL users vs non-RTL users)

Category	Cause	Δ (T-F)	OR	p_{χ^2}	p_{bh}
Passenger	>	+4.95	1.31	0.001*	0.008*
Freight	TC 2/3 |	+1.85	3.93	0.009*	0.051
Freight	X	+2.24	2.39	0.026*	0.079
Empty	ARI	+4.78	2.42	0.030*	0.178
Freight	W |	-2.15	0.58	0.170	0.341
Passenger	W >	-0.93	0.79	0.204	0.645
Passenger	W |	+0.87	1.27	0.183	0.645
Passenger	W X	-0.53	0.55	0.177	0.645
Passenger	|	-2.64	0.88	0.096	0.645

*Significant at $p < 0.05$.

5.1. Policy implications

For Infrastructure Managers (IM), like ProRail, should encourage broad DAS adoption by Railway Undertaking companies. Our analysis showed that trains using RTL ran more punctually and passed fewer restrictive signals than those without. For example, RTL users had significantly lower median deviations at most timetable points where punctuality is assessed (e.g., 66 seconds shorter at Tilburg Industrie (Tbi) in the eastbound direction, and around 20–30 seconds shorter at major stations like Tilburg and Eindhoven). They also faced approximately 6% fewer unexpected yellow signal stops overall, reflecting the benefits of assisted driving for improved traffic harmonisation. These gains, although modest per train, accumulate to improvements in network performance. By promoting or even mandating the use of DAS, ProRail could improve overall punctuality and result in effective line capacity without costly infrastructure investments.

For DAS developers, the key policy implication is to integrate conflict prediction into the advice generated by the system. The study's analysis highlights specific scenarios where RTL's functionality could be enhanced. One clear opportunity is to improve the advice for complex junction scenarios. Currently, RTL shows trains on merging or crossing routes as dotted lines, but without any estimate of their relative timing or distance. This makes it difficult for drivers to assess potential conflicts ahead. The findings of this study show that when a passenger train merged ahead of a freight, RTL-equipped drivers had a 30% higher chance of passing a yellow signal and braking, indicating they didn't get enough information to adjust in time. To address this, developers should implement predictive conflict indicators, for example, if a freight train is expected to catch up to a slower train at a junction, RTL should display this upcoming conflict to help drivers anticipate these interactions more accurately. This would reduce the uncertainty and unnecessary braking we observed with merging scenarios.

Railway undertakings should invest in DAS as a cost-effective lever for punctuality and smoother operations, with expected spillovers to energy/fuel savings and wear reduction. Evidence shows no safety/workload penalty (Verstappen, Pikaar, and Zon, 2022).

For train drivers, the introduction of RTL offers a valuable support tool to improve situational awareness. RTL results in more informed driving: instead of reacting last-minute to restrictive signals or catching up to a slower train ahead, drivers can manage their speed proactively. By embracing the DAS and actively using its advice, drivers can maintain smoother operations that keep them on time without unnecessary braking. It is important to note that RTL is designed

as an assistive tool, not an automation. Drivers remain in control, but the system extends their visibility by providing insight into the track up to 25 km ahead. This is especially useful in busy areas around major stations such as Tilburg and Eindhoven, as it provides better insights, especially in following and diverging situations.

5.2. Limitations of the research

Observations are train runs, not drivers; repeated drivers may reduce independence. Early logs may include some non-driving sessions, likely biasing results toward conservative estimates of RTL benefit. Driver skill/experience is unobserved and could confound effects. Multiple testing was controlled via FDR, which lowers power; hence, non-significance implies uncertainty rather than absence of effect.

5.3. Conclusion

In this study, we evaluated the operational impact of the Dutch Driver Advisory System RouteLint (RTL) on freight trains along the Brabantroute using one year of empirical data. We integrated timetable and signal-level datasets to compare the performance of trains with and without RTL through a data-driven framework combining macroscopic and microscopic event graphs. Using non-parametric statistical tests, we quantified differences in punctuality, signal behaviour, and train interactions. Our analysis shows that RTL use leads to modest but consistent improvements in timetable adherence and a reduction in unplanned yellow signal aspects, particularly around major junctions and under stable traffic conditions. These gains demonstrate that digital advisory tools can enhance operational consistency without requiring additional infrastructure. However, we also found that RTL's benefits diminish in complex merging or crossing scenarios, where the current system cannot yet anticipate short-term conflicts. By highlighting these limitations, we identify opportunities to improve conflict prediction and adaptive guidance in future DAS algorithms. Methodologically, we contribute a scalable framework for evaluating DAS performance using large-scale operational data and robust statistical testing. Empirically, we provide one of the first real-world assessments of a connected DAS in freight operations. Overall, our findings show that when widely adopted and further refined, DAS like RTL can play a key role in improving reliability and efficiency in railway operations.

References

Benjamini, Yoav and Yosef Hochberg (1995). "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 57.1, pp. 289–300.

Cheng, Yu (1996). "Optimal train traffic rescheduling simulation by a knowledge-based system combined with critical path method". In: *Railway Technical Research Institute,* 4.6, pp. 399–413.

Cohen, Jacob (1988). *Statistical Power Analysis for the Behavioral Sciences*. 2nd. Hillsdale, NJ: Lawrence Erlbaum Associates.

Dong, Hairong, Hainan Zhu, and Shigen Gao (2018). "An Approach for Energy-Efficient and Punctual Train Operation via Driver Advisory System". In: *IEEE Intelligent Transportation Systems Magazine* 10.3, pp. 57–67.

European Commission (2011). *White Paper on Transport: Roadmap to a Single European Transport Area – Towards a competitive and resource efficient transport system*. COM(2011) 144 final. Luxembourg. URL: https://transport.ec.europa.eu/transport-themes/strategies/2011-white-paper-transport_en.

George, Nevin (2022). "Eco-Driving Strategy optimization for freight trains". In: *Journal of Mathematical Techniques and Computational Mathematics* 1.3.

Ghaviha, Nima et al. (2017). "A driver advisory system with dynamic losses for passenger electric multiple units". In: *Transportation Research Part C: Emerging Technologies* 85, pp. 11–130.

Infrastructuur en Waterstaat, Ministerie van (2020). *3.5 Openbaar Vervoer en Spoor – Artikel 16*. Tech. rep. Beleidsregel inzake punctualiteitsprestaties op het spoor. Rijksoverheid. URL: https://wetten.overheid.nl/.

Large, D.R., D. Golightly, and E. Taylor (2017). "Train-driving simulator studies: Can novice drivers deliver the goods?" In: *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit* 231 (10), pp. 1186–1194.

Li, Zhu et al. (2018). "Dynamic Trajectory Optimization Design for Railway Driver Advisory System". In: *IEEE Intelligent Transportation Systems Magazine* 10.1, pp. 121–132.

Luo, Y. C. et al. (2025). *A Driver Advisory System Based on Large Language Model for High-speed Train*. DOI: 10.48550/arXiv.2501.07837. arXiv: 2501.07837 [cs.AI].

McHugh, Mary L. (2013). "The Chi-square test of independence". In: *Biochemia Medica* 23.2, pp. 143–149.

Panou, K., P. Tzieropoulos, and D. Emery (2013). "Railway driver advice systems: Evaluation of methods, tools and systems". In: *Journal of Rail Transport Planning & Management* 3 (4), pp. 150–162.

Pearson, Karl (1900). "On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling". In: *Philosophical Magazine Series 5* 50.302, pp. 157–175.

Quaglietta, E., P. Pellegrini, and R. M. P. Goverde (2016). "The ON-TIME real-time railway traffic management framework: A proof-of-concept using a scalable standardised data communication architecture". In: *Transportation Research Part C: Emerging Technologies* 63, pp. 23–50.

Sogin, Samuel L. et al. (2013). “Comparison of Capacity of Single- and Double-Track Rail Lines Using Simulation Analyses”. In: *Transportation Research Record* 2374.1, pp. 111–118.

Verstappen, V., E. Pikaar, and R. Zon (2022). “Assessing the impact of driver advisory systems on train driver workload, attention allocation and safety performance”. In: *Applied Ergonomics* 100, p. 103645.

Wang, P. and R. M. P. Goverde (2016). “Multiple-phase train trajectory optimization with signalling and operational constraints”. In: *Transportation Research Part C: Emerging Technologies* 69, pp. 255–275.

Wang, P., R. M. P. Goverde, and J. van Luipen (2019). “A connected driver advisory system framework for merging freight trains”. In: *Transportation Research Part C: Emerging Technologies* 105, pp. 203–221.

Wang, Pengling and Rob M. P. Goverde (2017). “Development of A Train Driver Advisory System: ETO”. In: *2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*. IEEE, pp. 140–145. ISBN: 9781509064847. DOI: 10.1109/MTITS.2017.8005654.

Xiao, Zhuang et al. (2021). “Real-Time Energy-Efficient Driver Advisory System for High-Speed Trains”. In: *IEEE Transactions on Transportation Electrification* 7 (4), pp. 3163–3172.

A

Additional data description and cleaning

Table A.1: Data description of datasets VOS timetable points

Column Name	Data Type	Description
traffic_date	datetime64[ns]	Date of the train operation
operator	object	Operating company
trainnumber	int64	Train number
tp	object	timetable point
act	object	Act at timing point
track	object	Track number
driving_characteristic	object	Train characteristic (freight)
original_schedule	object	Original schedule
o_dev_m	float64	Difference in minutes between original plan and actual plan
o_dev_dif_s	float64	Delay measurement difference in seconds compared to the original plan at timetable point in seconds
o_dev_s	float64	Difference in seconds between original plan and actual plan
o_dev_dif_m	float64	Delay measurement difference in minutes at the last control point compared to the original plan
plan_actual	object	Actual schedule
a_dev_s	float64	Time difference in seconds between actual plan and execution
a_dev_dif_s	float64	Deviation measurement difference in seconds at the last control point
a_dev_m	float64	Time difference in minutes between actual plan and execution
a_dev_dif_m	float64	Deviation measurement difference in minutes at the last control point
pass_time	datetime64[ns]	Actual time the train passed the control point
length	float64	Train length
weight	float64	Train weight
next_point	object	Next timetable point in the actual schedule
direction	object	Direction of train (East / West)
category	object	Category of train (heavy/regular)

Table A.2: Data description of datasets TROTS Brabantroute

Column Name	Data Type	Description
traffic_date	datetime64[ns]	Date of the train operation
trainnumber	int64	Train number
operator	object	Operator of the train
tp	object	Timetable point
act	object	Action at the signal
driving_characteristic	object	Train characteristic (freight)
signal_id	object	Signal ID
signal_aspect	object	Signal aspect at time of passing (green, yellow, or red)
signal_passage	datetime64[ns]	Time the signal was passed
signal_clear	datetime64[ns]	Time the signal was set to safe (green)
signal_release_delay	int64	Seconds delay in setting signal to safe aspect
type_IA	object	Signal type
tp_dev	float64	Delay measurement in seconds at the last control point before signal passage
cause	object	Cause of a yellow signal aspect
block_occupancy	object	Block occupied by train
preceding_train	int64	Preceding train in front of current train
block_clear	datetime64[ns]	Block release time
direction	object	Direction of train (East / West)
category	object	Category of train (heavy/regular)

Table A.3: RouteLint user data

Column Name	Data Type	Description
Tijdstip	object	Login time in the RouteLint system
Treinnummer	int32	Train number
Datum	datetime64[ns]	Date
Gebruiker	object	Operator (person logging in)
GebruiktRijadvies	bool	True if driving advice was used, otherwise False
GebruiktGps	bool	True if GPS was used, otherwise False
GpsGematcht	bool	True if GPS signal matched within 100 meters, otherwise False
AppVersie	object	Version of the app used
bestandsdatum	datetime64[ns]	Date of the record file

	count	mean	std	min	25%	50%	75%	max
o_dev_m	238695	74.65	181.24	-2255.00	-1.00	6.00	105.00	1789
o_dev_dif_s	238695	27.95	756.22	-24062.00	-38.00	-3.00	28.00	75500
o_dev_s	238695	4490.81	10885.67	-135344.00	-60.00	394.00	6315.00	107393
o_dev_dif_m	238695	0.47	12.59	-401.00	-1.00	0.00	1.00	1258
a_dev_s	246278	132.86	822.59	-36674.00	-56.00	55.00	216.00	86746
a_dev_dif_m	246278	0.10	2.61	-156.00	-1.00	0.00	0.00	246
a_dev_dif_s	246278	6.22	157.68	-9411.00	-38.00	-3.00	27.00	14857
a_dev_m	246278	2.08	13.57	-611.00	0.00	0.00	3.00	1445
length	246147	541.69	136.38	12.00	513.00	577.00	628.00	736
weight	246147	1444.66	714.94	82.00	1052.00	1391.00	1662.00	5278
time_loss	693703	23.02	709.12	0.00	0.00	0.00	0.00	146805
tp_dev	697819	116.86	793.93	-36685.00	-73.00	42.00	205.00	86729

Table A.4: Overview of Train Operating Companies and RouteLint Abbreviations

Abbreviation	Company Name	RouteLint Abbreviation
CRB	CrossRail Benelux	cro
DBC	DB Cargo	db
GTS	General Transport Service	gen
HSN	HSL Netherlands	hsl
KRE	KombiRail Europe	kom
LNA	Lineas NV	lin
LTE	LTE bv	lte
MDR	MDrivers bv	mdr
MRN	Metrans Rail Netherlands bv	met
MWB	Medway Belgium	med
NIA	Niederrheinische Verkehrsbetreibe AG	nie
R2U	Rail2U BV	rai
RC	Rhein Cargo	rhe
RCC	Rail Cargo Carrier	rai
RFO	Rail Force One	rai
RTB	RTB Cargo	rtb
RTX	Railtrax NV	rai
RXP	Railexpress bv	rai
SBB	SBB Cargo International	sbb

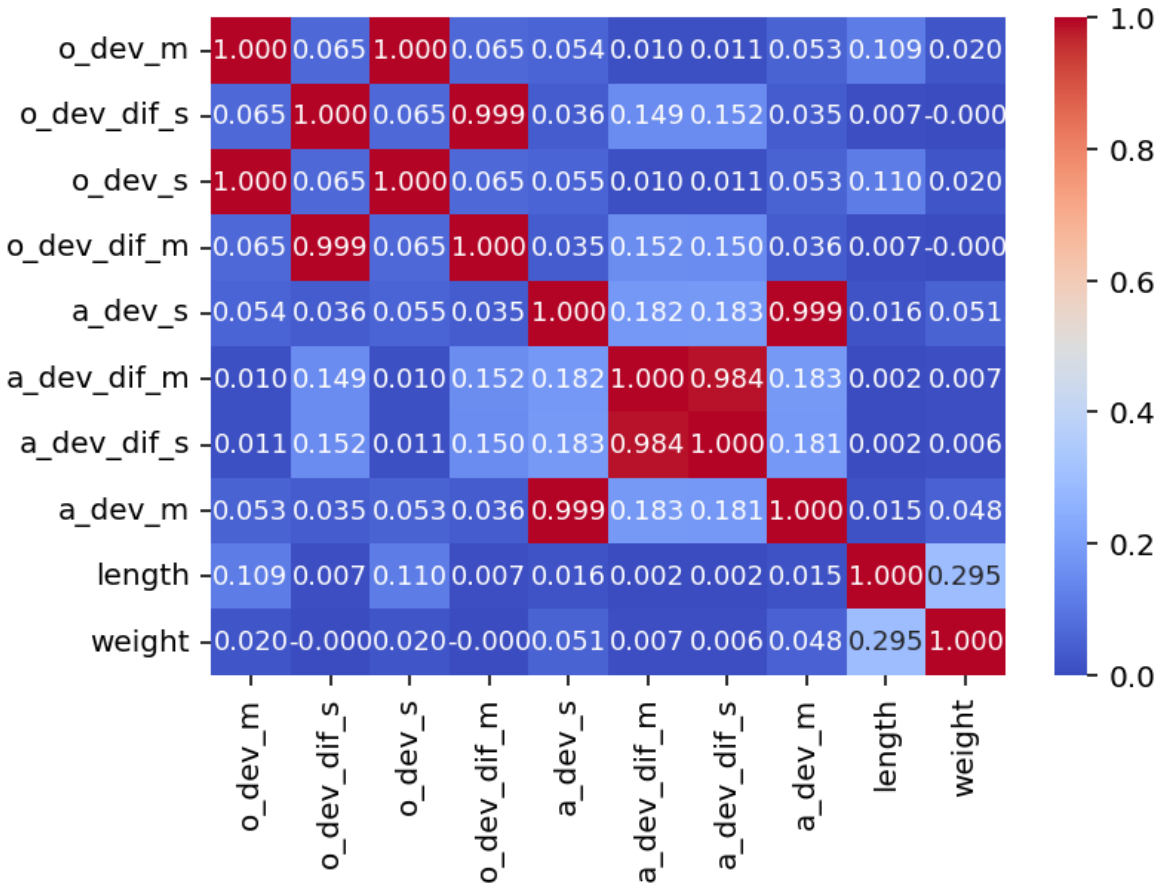


Figure A.1: Heatmap variables VOS dataset

Table A.5: Explanation of the causes provided for yellow signals

Cause	Explanation
trainV =	Hindrance from another train holding the next block; order of trains is swapped.
train =	Hindrance from another train holding the next block.
train !	Subsequent train movement in the opposite direction.
train -	Crossing train movement in the same running direction.
train ^	Executing train movement in the opposite direction.
trainV |	Subsequent train movement in the same direction; order of trains swapped.
train |	Subsequent train movement in the same direction.
train <	Diverging train movement in the same direction.
trainV <	Diverging train movement in the same direction; order of trains swapped.
train >	Merging train movement in the same direction.
trainV >	Merging train movement in the same direction; order of trains swapped.
rijweg_PPR	Route set too late; the dispatcher ultimately sets the route manually in the ProcessPlan Routes module.
rijweg_ARI	Route set too late; ARI (Automatic Route Setting) eventually sets the route automatically.
rijweg_ARIopgedeeld	ARI route set in split segments.
train_drgVL | ¾	Hindrance from another train occupying the next block in a successive movement. At least one train was scheduled/modified by Traffic Control, and the planned times do not meet the generic conflict norms from the Network Statement.
trainV x	Crossing train movement in the opposite direction; order of trains swapped.
train_drgVL ¾	Hindrance from another train occupying the next block in a merging movement. At least one train was scheduled/modified by Traffic Control and the planned times do not meet the generic conflict norms from the Network Statement.
trainV ∧	Executing train movement in the opposite direction; order of trains swapped.
trainV_tevroeg |	Hindrance from another train occupying the next block in a successive direction. The causing train was on time and the hindered train was early; order of trains swapped.
train_drglVL | ⅔	Hindrance from another train occupying the next block in a following movement. At least one train was scheduled/modified by Traffic Control and the planned times do not meet the generic conflict norms from the Network Statement.

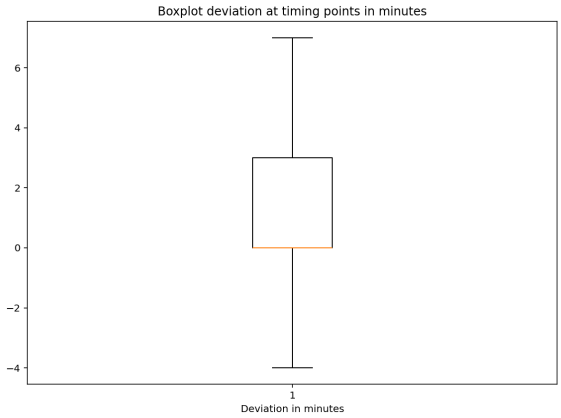
B

Data analysis

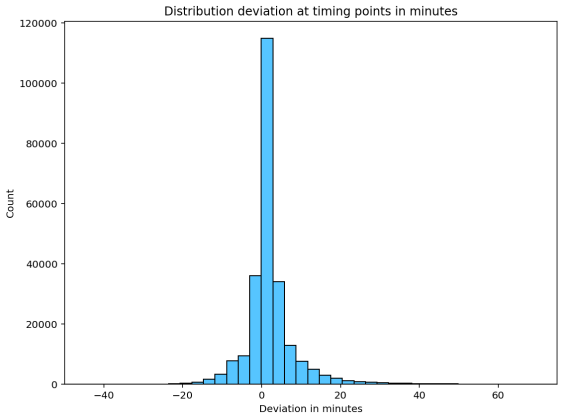
Table B.1: Mann–Whitney U-test with random-seed comparing deviation between RTL and non-RTL trains per edge on macroscopic level.

Direction	DRP	n_{true}	n_{false}	Median_T	Median_F	Δ (T–F)	U	p	r_{effect}	p_{BH}
Oost	Tgra	503	1,965	133.0	187.0	-54.0	438,370.0	9.05×10^{-8}	0.07880	0.00136
West	Ehv	493	1,939	122.0	153.0	-31.0	425,827.5	0.00018	0.07594	0.00271
Oost	Tbi	51	359	53.0	124.0	-71.0	6,473.5	0.00071	0.16720	0.00534
West	Btl	521	1,912	121.0	143.5	-22.5	454,672.5	0.00226	0.06190	0.01536
Oost	Ehv	437	1,922	151.0	202.0	-51.0	381,743.5	0.00295	0.06122	0.01473
West	Tb	463	2,143	94.0	127.0	-33.0	453,325.0	0.00357	0.05708	0.01536
West	Tbr	481	2,124	118.0	137.0	-19.0	468,480.5	0.00447	0.05569	0.01536
West	Ehs	513	1,900	137.0	155.0	-18.0	448,418.5	0.00543	0.05660	0.01536
West	At	521	1,799	126.0	143.0	-17.0	431,746.5	0.00614	0.05689	0.01536
Oost	Tb	478	2,172	90.0	115.0	-25.0	480,978.0	0.01181	0.04891	0.04429
Oost	Lpe	489	1,924	119.0	144.0	-25.0	437,283.5	0.01602	0.04903	0.04806
West	Bet	520	1,877	129.0	148.0	-19.0	459,808.5	0.04338	0.04126	0.08876
West	Ot	465	1,783	142.0	153.0	-11.0	389,825.0	0.04734	0.04183	0.08876
West	Tbu	482	2,087	109.0	126.0	-17.0	475,086.0	0.05750	0.03748	0.09583
West	Vga	88	463	112.5	149.0	-36.5	17,877.5	0.06849	0.07762	0.10274
West	Vg	42	142	135.0	181.5	-46.5	2,504.5	0.11569	0.11609	0.15776
Oost	Vg	43	122	110.0	124.0	-14.0	2,214.5	0.12988	0.11805	0.32469
West	Tgra	494	1,884	104.5	115.0	-10.5	444,935.0	0.13290	0.03082	0.16613
Oost	Tba	463	2,117	113.0	120.0	-7.0	470,564.0	0.17880	0.02647	0.36167
Oost	Vga	81	423	107.0	122.0	-15.0	15,567.5	0.19289	0.05802	0.36167
Oost	Tbu	472	2,132	105.5	109.0	-3.5	490,799.0	0.40329	0.01638	0.67111
West	Lpe	527	1,886	131.0	132.0	-1.0	486,318.0	0.45166	0.01532	0.52115
Oost	Ot	462	1,791	115.0	125.0	-10.0	404,468.0	0.45798	0.01564	0.67111
Oost	Btl	475	1,904	144.0	138.0	6.0	461,400.0	0.49215	0.01408	0.67111
Oost	Ehs	470	1,984	169.0	171.0	-2.0	461,260.0	0.71846	0.00728	0.82530
West	Tba	494	1,801	153.5	155.0	-1.5	441,032.0	0.77001	0.00610	0.82501
Oost	At	487	1,971	175.0	173.0	2.0	475,869.0	0.77172	0.00585	0.82530
Oost	Bet	508	1,960	136.5	148.0	-11.5	494,119.0	0.79491	0.00523	0.82530
Oost	Tbr	490	2,090	127.0	134.0	-7.0	508,773.5	0.82530	0.00435	0.82530
West	Tbi	57	396	234.0	231.0	3.0	11,147.5	0.88129	0.00704	0.88129

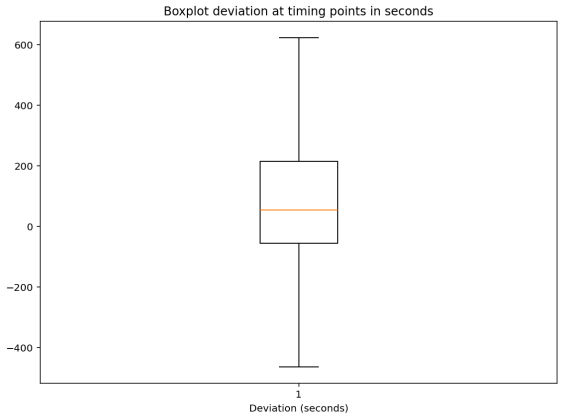
*Significant at $p < 0.05$.



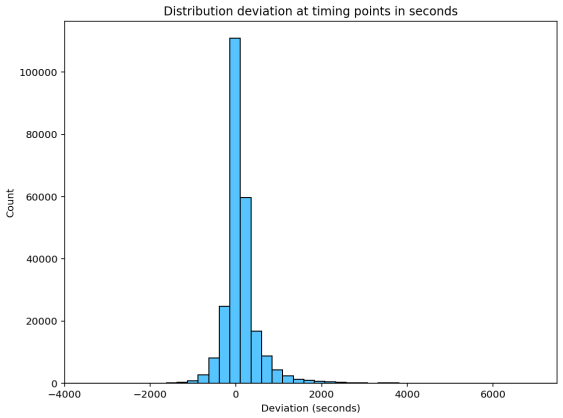
(a) Boxplot deviation from up-to-date timtable in seconds.



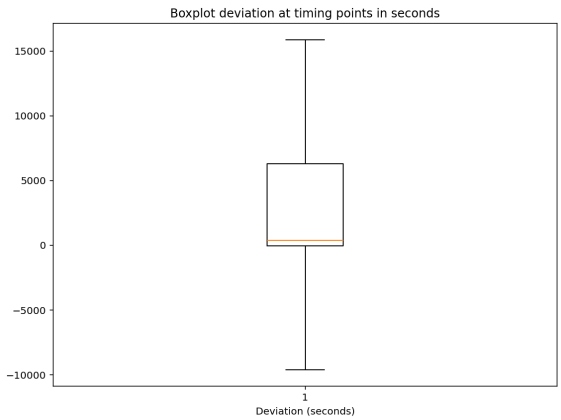
(b) Histogram deviation from up-to-date timtable in minutes.



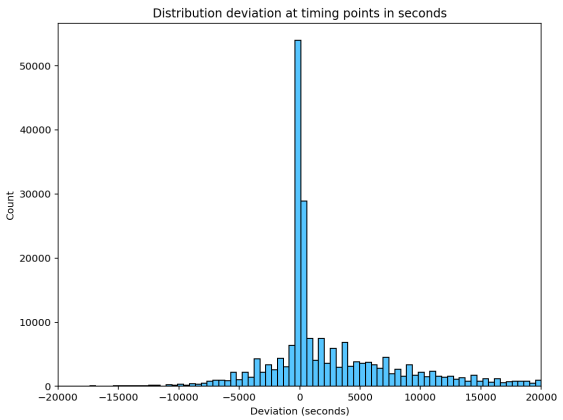
(c) Boxplot deviation from up-to-date timtable in seconds.



(d) Histogram deviation from up-to-date timetable in seconds.

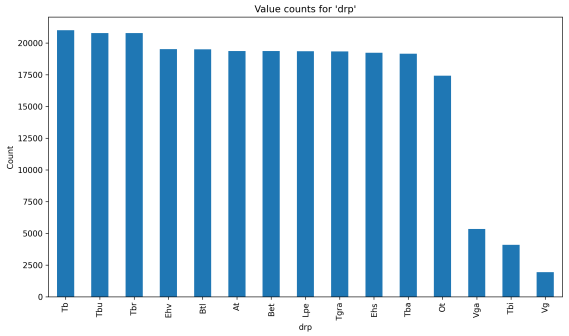


(e) Boxplot deviation from original timetable in seconds.

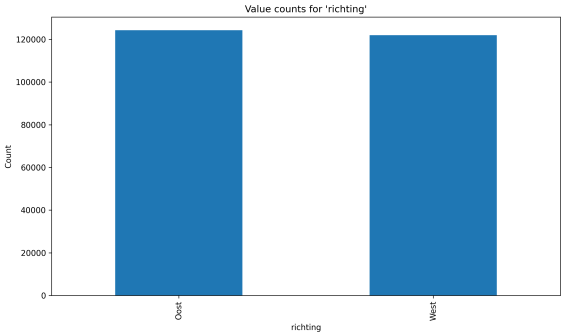


(f) Histogram devation from original timetable in seconds.

Figure B.1: Boxplots and histograms.

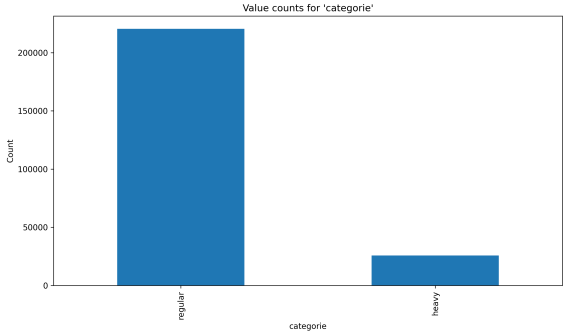


(a) Timetalbe point

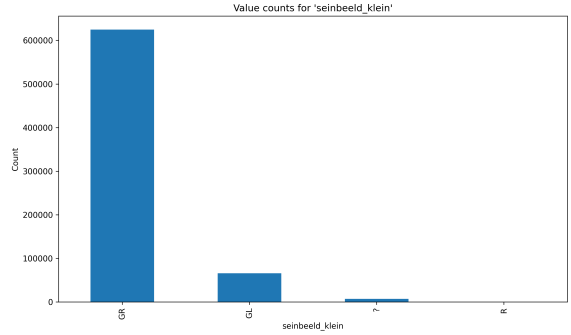


(b) Direction

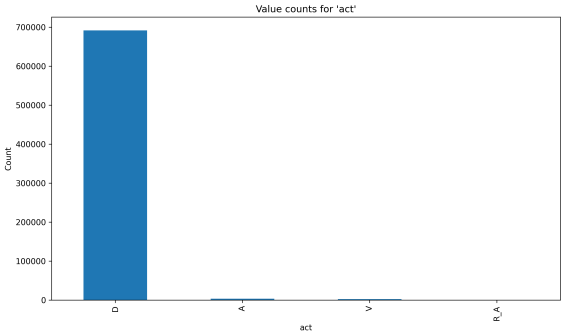
Figure B.2: Value counts of VOS variables



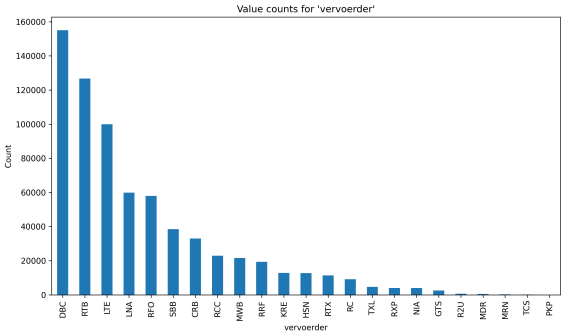
(a) Category



(b) Signal aspects

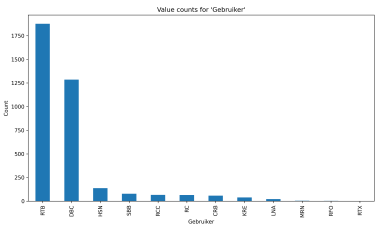


(c) Act at signals

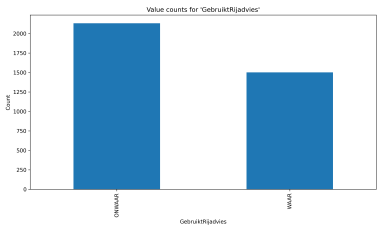


(d) Freight operators

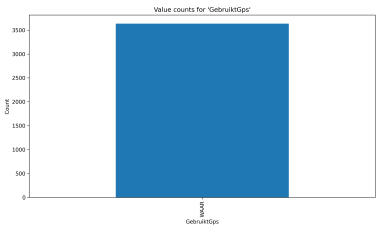
Figure B.3: Value counts of TROTS variables



(a) Freight operator log (RouteLint)

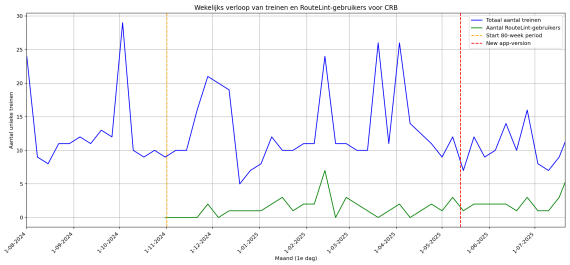


(b) Driving advice used

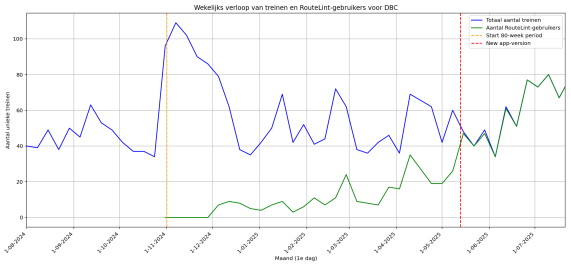


(c) GPS used

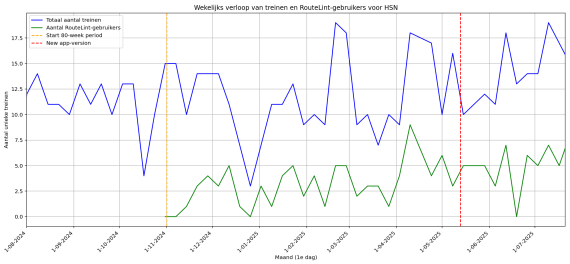
Figure B.4: Value counts of RouteLint data variables.



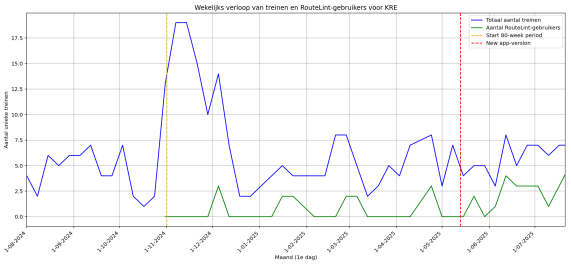
(a) CRB



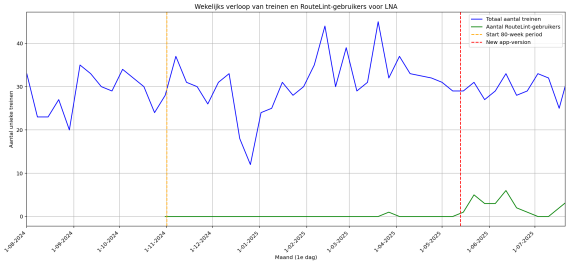
(b) DBC



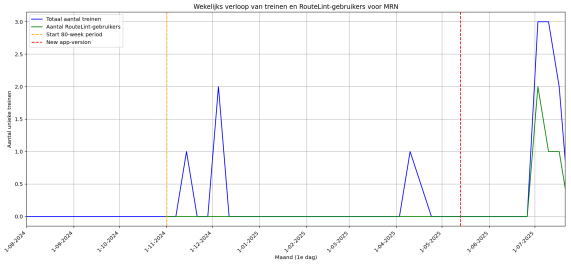
(c) HSN



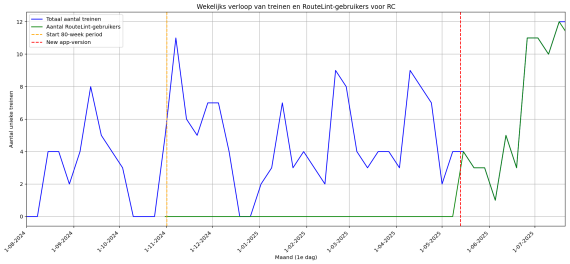
(d) KRE



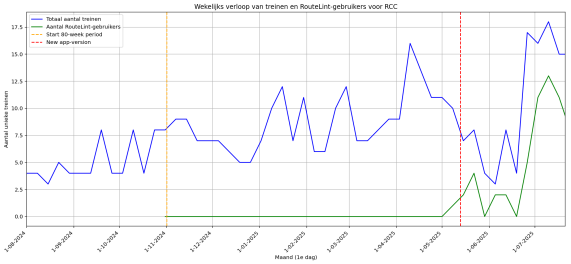
(e) LNA



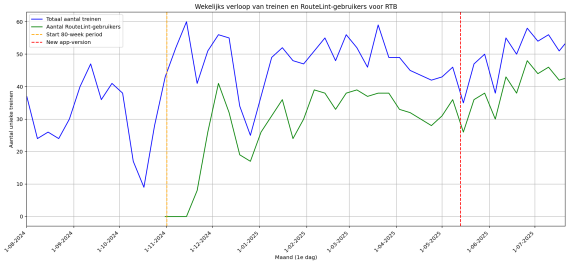
(f) MRN



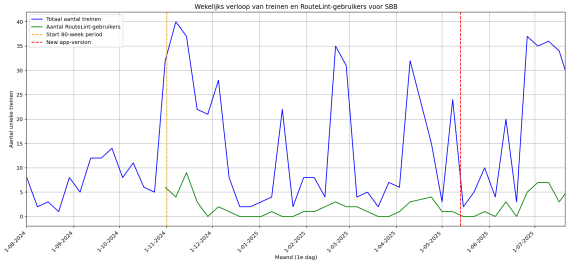
(g) GTS



(h) RCC

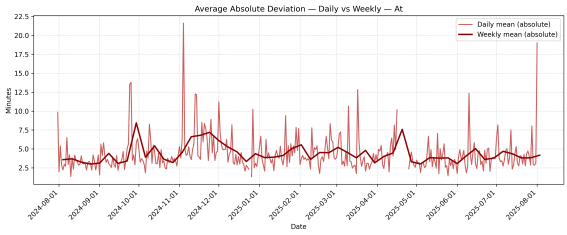


(i) RTB

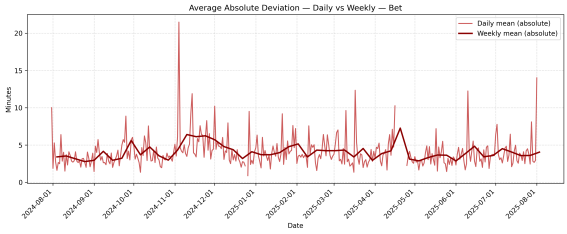


(j) SBB

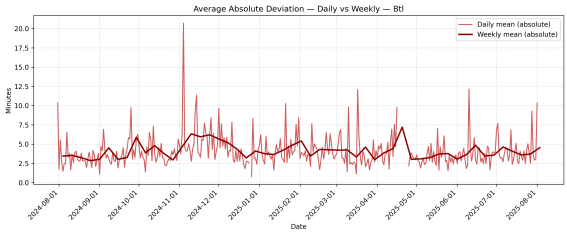
Figure B.5: Weekly RouteLint usage relative to the total number of trains per freight operator (Operators with no use are excluded).



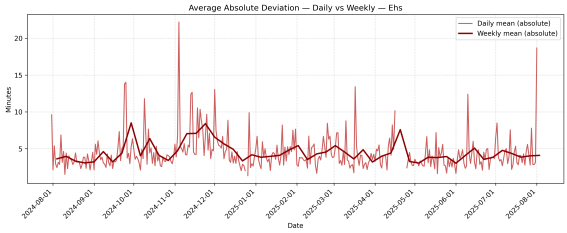
(a) CRB



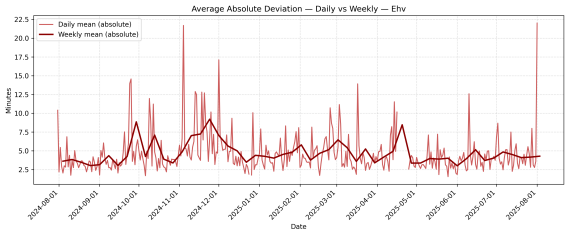
(b) DBC



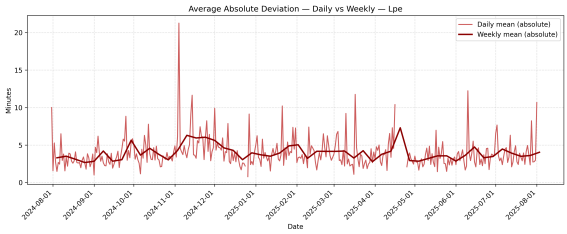
(c) HSN



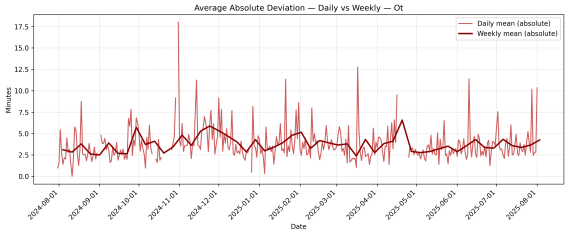
(d) KRE



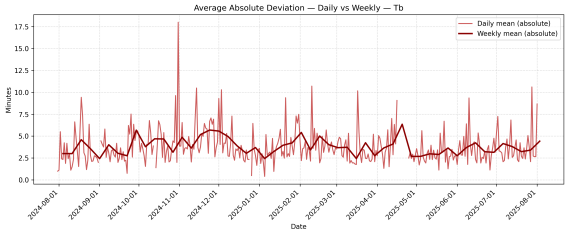
(e) LNA



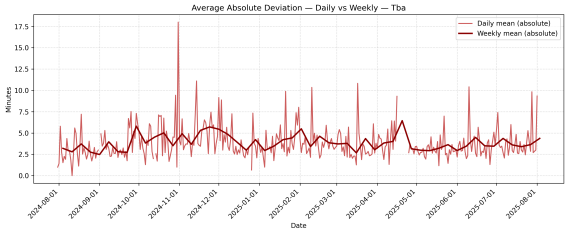
(f) Lpe



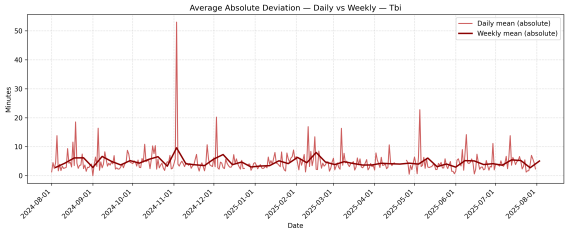
(g) Ot



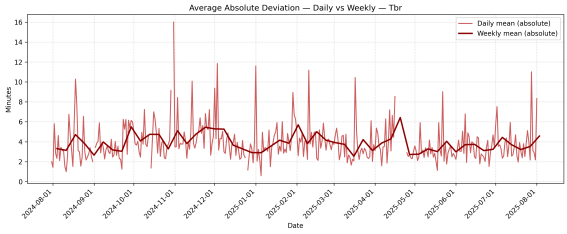
(h) Tb



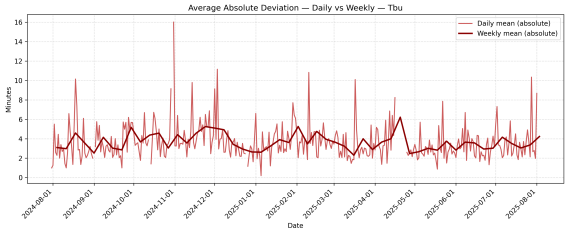
(i) Tba



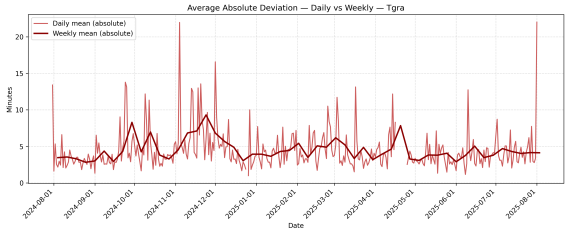
(j) Tbi



(k) Tbr



(l) Tbu



(m) Tgra

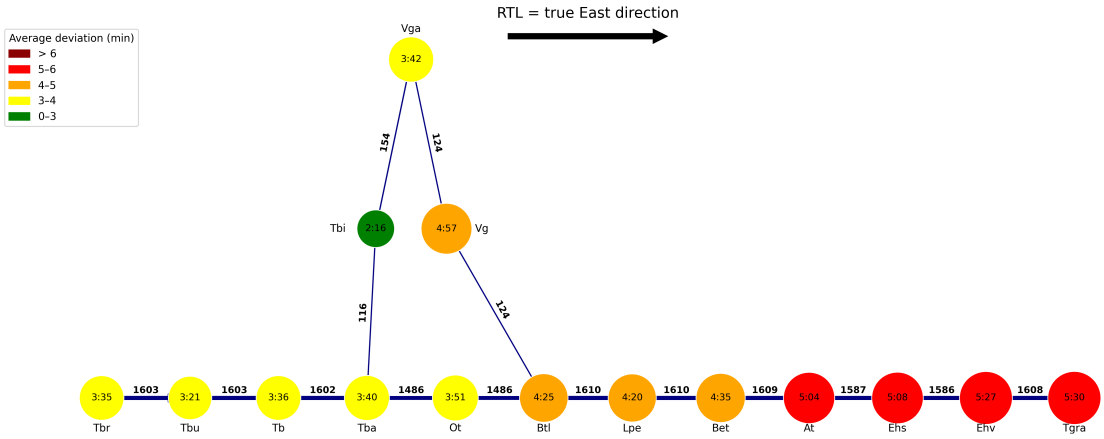


Figure B.7: RouteLint Users average deviation per timetable point in east direction

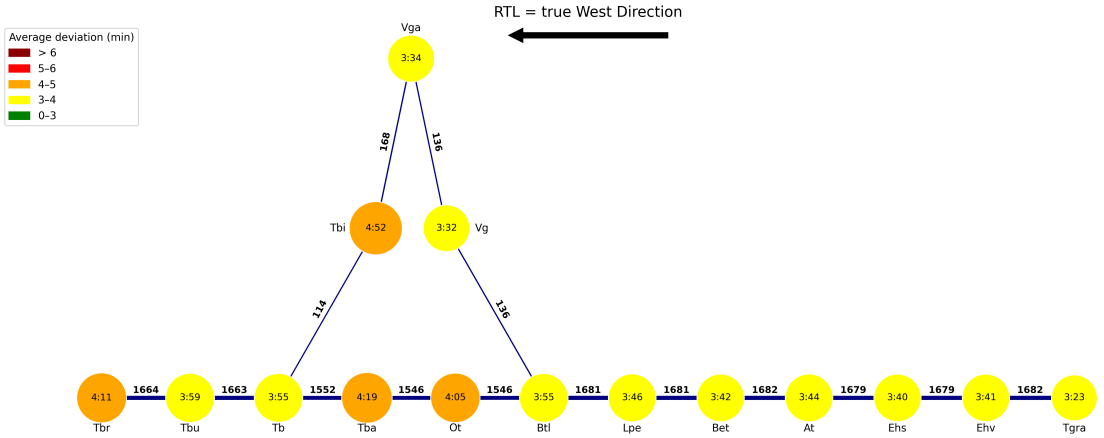


Figure B.8: RouteLint Users average deviation per timetable point in west direction

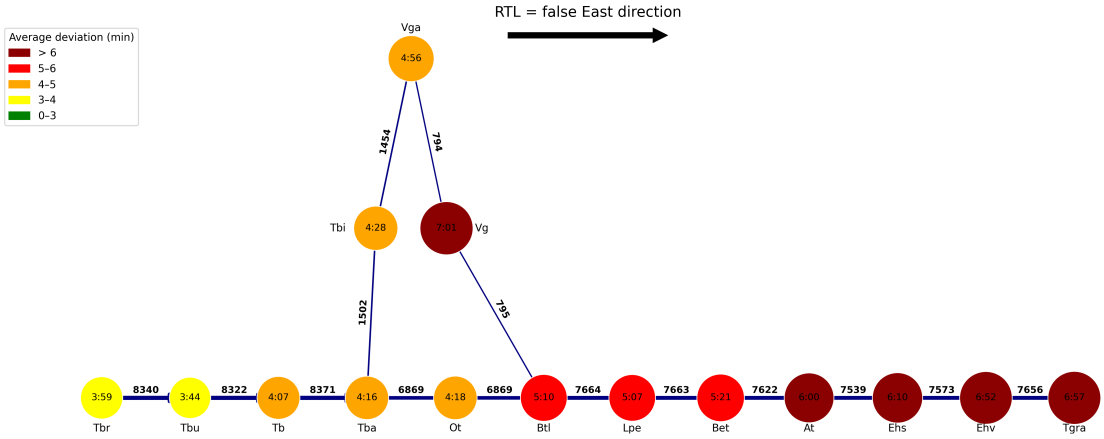


Figure B.9: Non RouteLint Users average deviation per timetable point in east direction

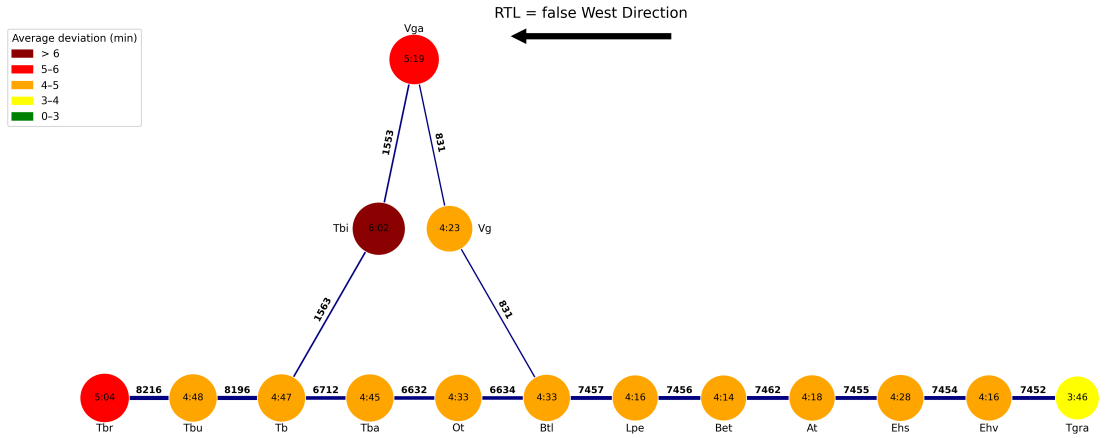
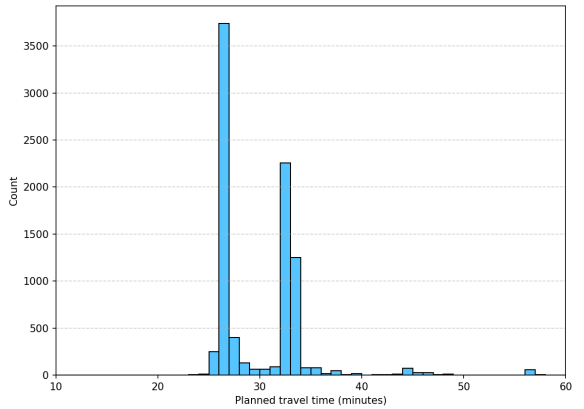


Figure B.10: Non RouteLint Users average deviation per timetable point in west direction

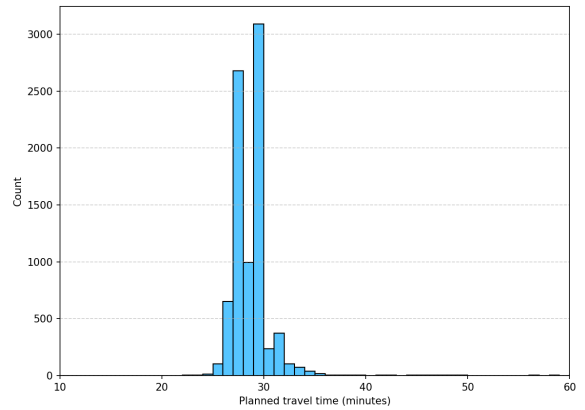
Table B.2: Mann–Whitney U-test with random-seed comparing deviation between RTL and non-RTL trains per edge on microscopic level.

Richting	s	$s+1$	p-value	U	n_{true}	n_{false}	$Median_T$	$median_F$	r_{bc}	p_{FDR}
West	BTL$1134	BTL$622	0.00011	3022.5	38	253	2.33	3.67	-0.37	0.0069
West	BTL$1198	BTL$1134	0.00019	3912.0	42	281	1.86	2.86	-0.34	0.0069
West	EHV$274	EHV$150	0.00086	424117.0	460	2034	29.31	33.31	-0.09	0.0213
East	EHV$210	EHV$256	0.00637	8187.5	63	324	2.89	4.89	-0.20	0.2484
East	TB$164	TBI$182	0.00545	5261.5	32	450	14.58	17.58	-0.27	0.2484
West	BTL$622	BTL$618	0.01277	3491.0	38	237	2.00	3.47	-0.22	0.2362
East	TBI$703	VGA$727	0.01376	8866.0	53	411	32.58	45.42	-0.19	0.2683
East	BTL$1164	BTL$1188	0.02487	1391.5	23	162	3.29	5.92	-0.25	0.3880
East	BTL$1108	BTL$1164	0.04514	872.0	16	147	5.00	8.36	-0.26	0.5399
East	TB$150	TB$164	0.04845	589540.5	466	2658	3.42	3.42	-0.05	0.5399
East	EHV$112	EHV$254	0.05587	1696.0	26	162	5.88	8.13	-0.19	0.5447
West	EHV$50	EHV$1798	0.06723	3960.5	44	210	4.37	4.37	-0.14	0.9951
West	EHV$146	EHV$62	0.08442	3404.0	43	183	4.28	5.28	-0.13	0.9999
East	BTL$1539	BTL$1108	0.12212	466862.0	440	2199	2.65	2.65	-0.03	0.8931
East	EHV$1008	EHV$1787	0.13140	279889.5	352	1653	5.47	5.53	-0.04	0.8931

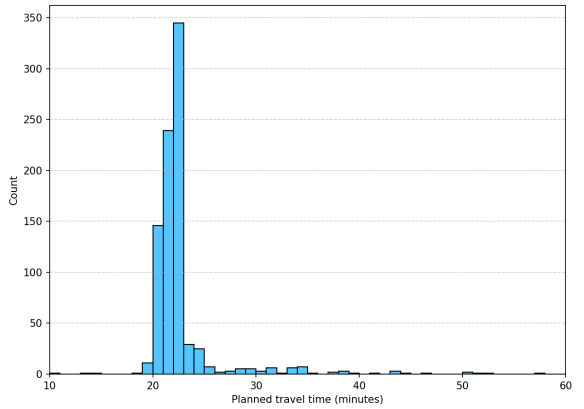
*Significant at $p < 0.05$.



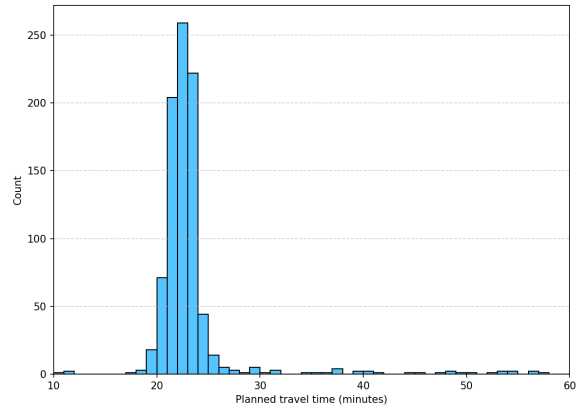
(a) East direction, Tilburg to Eindhoven



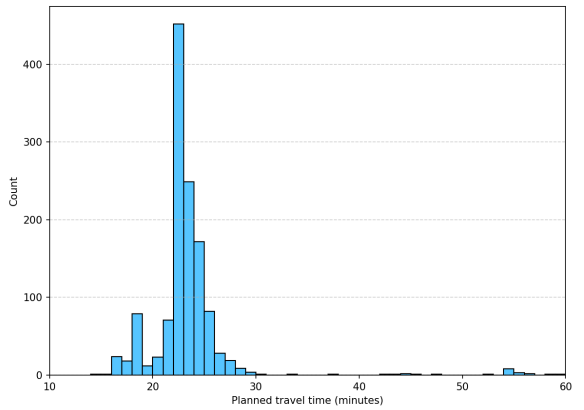
(b) West direction, Eindhoven to Tilburg



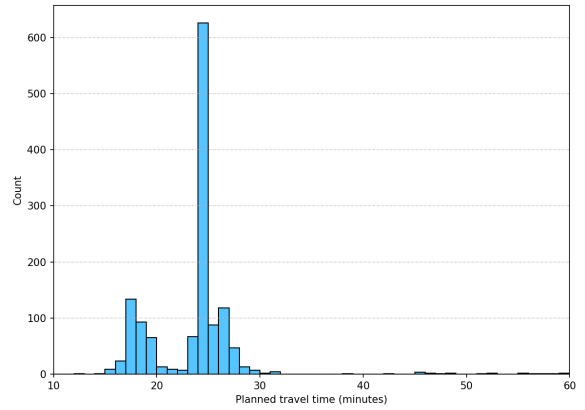
(c) East direction, Vught to Eindhoven



(d) West direction, Eindhoven to Vught



(e) East direction, Tilburg to Vught aansluiting

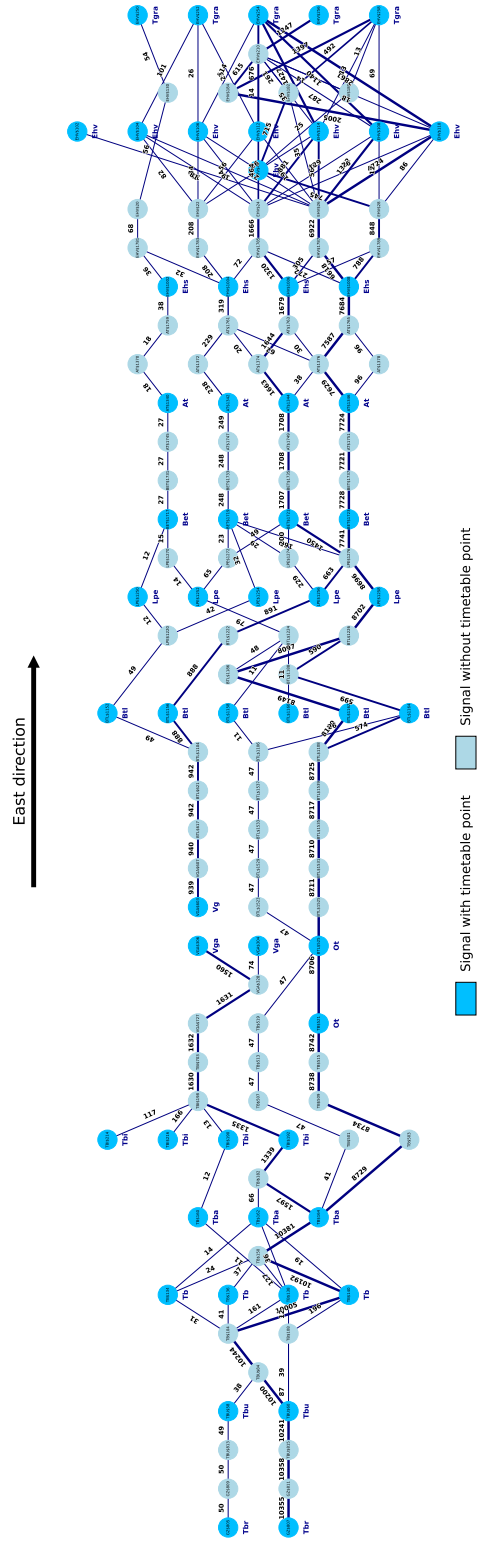


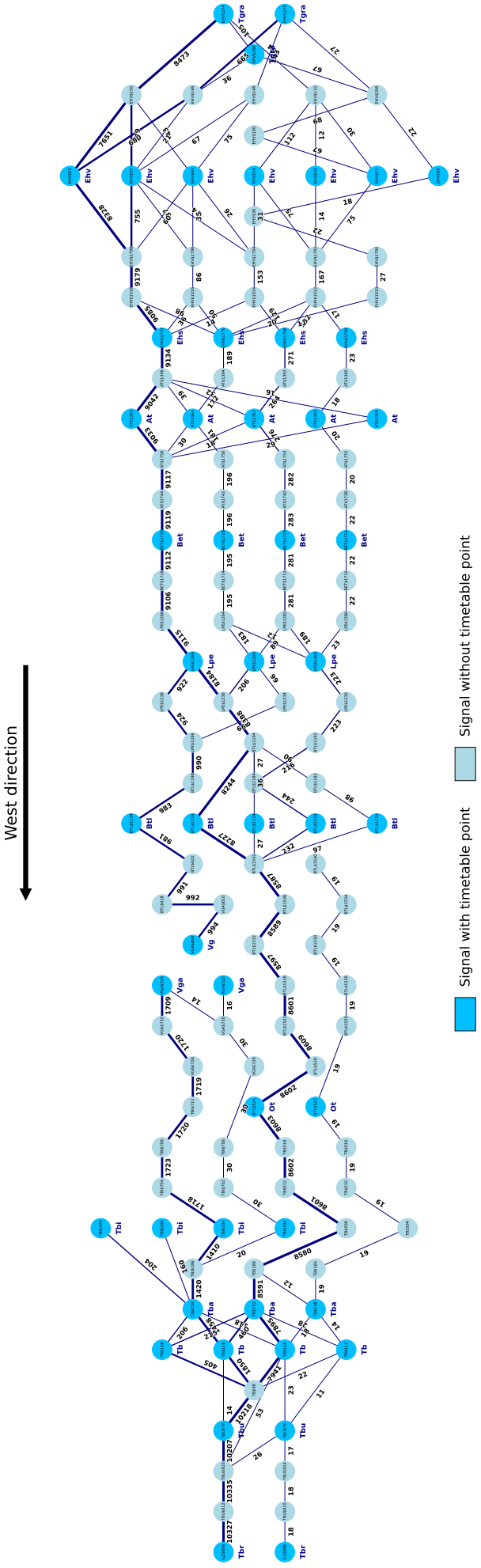
(f) West direction, Vught aansluiting to Tilburg

Figure B.11: Distribution of the original planned running times between timetable points.

C

Event graphs




West direction
Signal with timetable point
Signal without timetable point

D

Extensive results microscopic-level

Table D.1: Edge-level summary for average running time and average absolute deviation.

richting	u	v	n_{tot}	n_{avg}	n_T	n_F	$avg\_\bar{r}\_heavy$	$avg\_\bar{r}\_regular$	$median_T$	$median_F$	p
West	EHV$274	EHV$150	8470	156	1472	6997	NaN	73.6860	103	108	0
East	TB$150	TB$164	10380	4725	1596	8784	30.4240	30.4080	29	30	0
West	BTL$1134	BTL$622	981	366	134	846	NaN	62.3330	60	60	0
East	TB$140	TB$150	10191	4594	1568	8623	26.4060	26.7420	25	26	0
West	BTL$1198	BTL$1134	983	427	133	849	NaN	36.8640	36	36	0
East	AT$1346	AT$1378	96	17	15	69	NaN	69.5290	71	191	0.0010
East	TB$164	TBI$182	1597	728	111	1486	94.9330	94.2370	82	89	0.0020
East	TB$138	TB$150	125	35	14	82	NaN	30.0860	29.5	35	0.0020
East	EHV$210	EHV$256	1247	351	208	1039	39.1320	36.1050	39	39	0.0050
East	VGA$727	VGA$328	1631	254	154	1477	79.3290	77.0910	73	71	0.0070
East	BTL$1164	BTL$1188	599	127	75	524	22.2940	21.5820	21	24	0.0180
East	EHV$1787	EHV$26	6921	56	986	4874	NaN	39.8210	44	44	0.0170
East	EHV$1008	EHV$1789	788	201	48	740	44.2900	42.4880	43	42	0.0220
East	TBI$703	VGA$727	1630	375	154	1476	449.1580	439.4200	420	418	0.0250
East	AT$1378	AT$1765	96	16	16	68	NaN	94.8120	100.5	149	0.0350
East	EHV$26	EHV$114	565	137	54	511	89.8280	86.2410	86	88	0.0360
East	BTL$1539	BTL$1108	8722	4257	1480	7242	55.1320	54.6510	53	53	0.0640
East	BTL$1108	BTL$1164	574	112	75	499	57.7140	58.6430	64	77	0.0590
East	TB$104	TB$138	161	30	14	111	NaN	64.3000	69.5	80	0.0630
East	VGA$328	VGA$306	1558	240	150	1408	67.8730	68.2920	68	76	0.0830
East	EHV$114	EHV$254	1421	376	173	1248	86	81.8300	81	85	0.1130
East	BTL$621	BTL$1104	942	364	124	818	NaN	48.1320	47	47	0.1260
East	EHV$1008	EHV$1787	6614	523	1171	5443	43.9780	40.5320	46	45	0.1520
East	AT$1761	EHV$1004	319	16	24	228	NaN	66.7500	70.5	70	0.1680
East	TBU$60	TBU$84	10195	3203	1572	8623	45.9560	45.0410	46	46	0.1760
West	BTL$622	BTL$618	991	371	135	855	NaN	63.4720	62	62	0.0340
East	EHV$106	EHV$254	114	38	28	86	106	113.1430	115	110	0.2110
East	GZ$811	TBU$815	10356	3939	1593	8763	72.6470	72.2720	69	70	0.2180
East	LPE$1276	BET$1721	1450	785	273	1177	256.5250	256.6750	245	229	0.2510
East	AT$1342	AT$1372	238	78	37	201	60.9060	56.2390	60	57	0.9790
West	EHV$146	EHV$62	679	303	115	564	NaN	84.7190	84	84	0.5500
West	EHV$150	EHV$50	779	158	136	643	NaN	59	59	58	0.4890
West	EHV$150	EHV$60	43	13	10	33	NaN	97.6920	102.5	99	0.4030
West	EHV$150	EHV$62	7645	3815	1325	6319	NaN	87.3250	86	87	0.4460
West	EHV$1768	AT$1392	271	71	58	213	NaN	78.3520	74.5	76	0.8400
West	EHV$1770	AT$1394	189	39	31	158	NaN	74.1790	75	77	0.6170
West	EHV$1772	AT$1396	9134	4943	1592	7541	NaN	72.7420	71	73	0.9660
West	EHV$1792	EHV$1012	167	49	36	131	NaN	54.1020	53.5	54	0.9990
West	EHV$132	EHV$54	112	19	19	93	NaN	85.9470	87	83	0.5110
West	EHV$148	EHV$50	67	19	16	51	NaN	59.6840	58.5	57	0.1430
West	EHV$1014	EHV$1768	129	36	28	101	NaN	82.4720	79	82	0.4600
West	EHV$1018	EHV$1770	98	20	13	85	NaN	78.7500	75	82	0.4070
West	BET$1726	BET$1712	281	48	61	220	NaN	128.3120	125	126.5	0.7470
West	BET$1728	BET$1714	195	50	31	164	NaN	124.1200	125	123	0.8060

Continued on ne

Table D.1 (continued).

richting	u	v	n_{tot}	n_{avg}	n_T	n_F	$avg\_\bar{r}\_heavy$	$avg\_\bar{r}\_regular$	$median_T$	$median_F$	p
West	BET$1730	BET$1716	9109	4966	1586	7522	NaN	125.5070	123	124	0.7250
West	BTL$1122	BTL$1542	232	49	46	186	NaN	138.8370	137	139.5	0.2820
West	BTL$1126	BTL$1542	8224	4469	1472	6752	NaN	74.9620	74	73	0.0840
West	BTL$1184	BTL$1122	244	54	48	196	NaN	38.9810	38	39	0.2090
West	BTL$1194	BTL$1126	8241	4517	1472	6769	NaN	60.0870	60	60	0.9550
West	BTL$1522	BTL$528	8608	4574	1543	7065	NaN	26.5030	26	26	0.9950
West	BTL$1528	BTL$1522	8601	4568	1543	7058	NaN	49.5010	48	48	1
West	BTL$1532	BTL$1528	8597	4567	1543	7054	NaN	52.1330	51	51	0.9970
West	BTL$1536	BTL$1532	8589	4569	1542	7047	NaN	70.5760	69	70	0.9980
West	BTL$1542	BTL$1536	8587	4568	1541	7046	NaN	69.6270	69	69	1
West	BTL$524	TB$518	8602	4669	1543	7059	NaN	58.2740	56	57	1
West	BTL$528	BTL$524	8602	4575	1543	7059	NaN	31.4010	30	31	1
West	BTL$618	VGA$610	992	363	135	856	NaN	106.3280	104	104	0.2620
West	EHV$100	EHV$52	67	18	16	51	NaN	37.2220	36	37	0.2430
West	EHV$1012	EHV$1768	132	44	31	101	NaN	86.2950	82	84	0.8770
West	EHV$1018	EHV$1772	9082	4925	1582	7499	NaN	73.7430	72	74	0.6690
West	EHV$1794	EHV$1014	153	43	35	118	NaN	55.1400	53	53	0.8810
West	EHV$274	EHV$268	105	22	24	81	NaN	36	36.5	36	0.5810
West	EHV$1798	EHV$1018	9176	4946	1593	7582	NaN	46.7980	47	48	0.6860
West	TB$118	TB$98	404	94	37	367	92.3000	90.2300	91	90	0.5490
West	TB$152	TB$114	7886	2609	1444	6441	NaN	54.2580	56	56	1
West	TB$152	TB$116	460	95	64	396	NaN	115.5890	113	115	0.8360
West	TB$152	TB$118	215	11	25	190	NaN	113.0910	226	194.5	0.3190
West	TB$154	TB$116	1456	182	96	1360	66.7030	61.1590	100.5	77	0.8970
West	TB$168	TB$152	8582	1176	1538	7044	NaN	49.4820	56	56	0.9250
West	TB$506	TB$168	8579	1144	1536	7043	NaN	53.4780	53	53	1
West	TB$512	TB$506	8600	4550	1541	7059	NaN	67.4950	65	65	0.9990
West	TB$518	TB$512	8601	4659	1543	7058	NaN	51.6740	50	50	1
West	TB$116	TB$98	1841	547	157	1684	64.6100	57.4430	59	58	0.1800
West	TB$98	TBU$78	10217	5087	1644	8573	54.9160	48.8780	46	47	0.8030
West	TBI$196	TBI$186	1408	391	102	1306	61.4790	58.1510	55	53	0.4870
West	TBI$704	TBI$196	1718	307	165	1553	61.1490	58.5920	64	53	1
West	TBI$708	TBI$704	1722	426	165	1557	54.6270	52.0720	47	47	0.0750
West	TBI$712	TBI$708	1719	427	165	1554	73.2130	70.1680	62	63	0.7490
West	TBU$78	TBU$816	10204	4783	1645	8559	59.9780	57.3170	55	56	0.6040
West	TBU$812	GZ$808	10325	4605	1660	8665	70.7540	69.5200	68	68	0.9950
West	TBU$816	TBU$812	10329	4778	1662	8667	68.6720	67.0580	65	66	0.9930
West	VGA$324	VGA$732	1709	221	165	1544	62.3710	60.1850	57	57	0.0940
West	VGA$610	VGA$604	993	329	135	857	NaN	50.0550	49	48	0.2950
West	TBI$186	TB$154	1418	171	103	1315	162.2310	139.8790	145	134	0.9660
West	EHV$1796	EHV$1016	86	16	16	70	NaN	51.6880	51.5	53	0.9730
West	TB$114	TB$98	7929	3259	1442	6475	NaN	45.8720	43	43	0.9990
West	LPE$1284	LPE$1264	183	51	28	155	NaN	54.7450	54	54	0.8700
West	EHV$206	EHV$100	68	14	16	52	NaN	56.9290	60	60.5	0.5120
West	EHV$268	EHV$206	67	15	12	55	NaN	80.9330	93	81	0.9770
West	EHV$274	EHV$132	154	36	29	125	NaN	138.5	136	136	0.8110
West	BET$1716	LPE$1286	9103	4765	1584	7518	NaN	65.3640	64	64	0.8580
West	EHV$278	EHV$148	143	17	26	117	NaN	73.7650	68	67	0.3110
West	EHV$50	EHV$1794	74	23	14	60	NaN	61.4350	55.5	51.5	0.1900
West	EHV$50	EHV$1798	748	207	139	609	NaN	36.3670	33	33	0.7200
West	EHV$52	EHV$1792	75	25	16	59	NaN	54.7600	53.5	53	0.7570
West	EHV$54	EHV$1792	75	18	15	60	NaN	44.0560	47	43	0.7820
West	LPE$1286	LPE$1266	9113	4722	1585	7527	NaN	54.5860	53	54	0.9960
West	EHV$62	EHV$1798	8326	4664	1440	6885	NaN	47.2370	46	47	0.8610
West	LPE$1234	LPE$1206	66	18	11	55	NaN	44.7220	45	43	0.2660
West	LPE$1236	BTL$1194	8387	4505	1504	6883	NaN	50.1830	49	50	1
West	LPE$1238	LPE$1206	924	407	124	799	NaN	48.2140	43	44	0.4860
West	LPE$1264	LPE$1234	66	18	11	55	NaN	45.7220	46	45	0.3680
West	LPE$1264	LPE$1236	202	59	41	161	NaN	51.6950	54	51	0.3130
West	LPE$1266	LPE$1236	8181	4438	1462	6719	NaN	48.0550	47	47	1
West	LPE$1266	LPE$1238	921	380	124	796	NaN	46.5210	45	45	0.0810
West	LPE$1282	LPE$1260	189	48	35	154	NaN	56.0210	55	55	0.4950
West	LPE$1282	LPE$1264	89	32	26	63	NaN	67.5	66	67	0.6610
West	LPE$1206	BTL$1198	990	430	135	854	NaN	16.5330	16	16	0.1820
West	BET$1714	LPE$1284	195	48	31	164	NaN	64.9170	65	65	0.9640
West	AT$1362	AT$1754	276	82	60	216	NaN	68.1220	62	62	0.7940
West	AT$1758	AT$1744	9112	5031	1585	7526	NaN	53.3640	52	53	1

Continued on ne

Table D.1 (continued).

richting	u	v	n_{tot}	n_{avg}	n_T	n_F	$avg\_\bar{r}\_heavy$	$avg\_\bar{r}\_regular$	$median_T$	$median_F$	p
East	BTL$1188	BTL$1226	587	136	75	512	58.5	56.8310	51	54	0.3000
East	BTL$1222	LPE$1256	891	351	122	769	NaN	46.7210	46	45	0.9810
East	BTL$1226	LPE$1258	8697	3727	1472	7225	53.9160	51.6870	52	51	0.9890
East	BTL$1525	BTL$1531	8711	4408	1479	7232	69.8340	68.1090	66	66	1
East	BTL$1531	BTL$1535	8710	4407	1479	7231	72.8950	70.7240	69	69	1
East	BTL$1535	BTL$1539	8717	4348	1479	7238	51.3090	49.6500	50	49	0.9660
East	BTL$617	BTL$621	942	365	124	818	NaN	45.9260	45	45	0.5080
East	EHV$1006	EHV$1785	1320	203	222	1098	45.3640	41.7740	48	45	0.8790
East	EHV$1008	EHV$1785	272	78	39	233	43.7890	41.0510	42	42	0.4940
East	EHV$108	EHV$210	286	71	12	232	NaN	51.9300	65	59.5	0.7670
East	EHV$108	EHV$254	1138	385	65	1073	105.7940	106.2620	107	107	0.5410
East	EHV$110	EHV$184	2005	311	510	1495	39.7500	39.6240	40	40	0.7340
East	EHV$110	EHV$254	2859	1090	631	2228	111.5210	108.2490	108	108	0.9950
East	EHV$112	EHV$254	675	211	105	570	106.0120	103.7560	101	102	0.4980
East	EHV$1785	EHV$24	1666	227	279	1387	34.8360	33.7670	43	40	0.9970
East	EHV$182	EHV$210	261	61	17	244	42.5	36.1960	34	35	0.7110
East	EHV$182	EHV$258	492	70	24	468	76	75.4670	110.5	98	0.8080
East	BTL$1186	BTL$1226	8094	3745	1396	6698	47.3740	45.5420	45	45	0.9870
East	EHV$184	EHV$210	614	177	165	449	36.4380	34.8630	33	33	0.6810
East	BTL$1162	BTL$1186	8148	4200	1404	6744	19.5330	19.2320	19	19	0.8920
East	BTL$1108	BTL$1162	8150	4219	1403	6747	58.8500	57.9190	56	56	0.6490
East	AT$1344	AT$1374	1663	775	303	1360	62.4710	61.4920	60	58	0.8940
East	AT$1346	AT$1376	7628	3226	1242	6386	63.5710	59.6980	58	58	0.9990
East	AT$1372	AT$1761	229	76	37	192	74.8710	71.3110	74	71	0.6500
East	AT$1374	AT$1763	1644	687	298	1346	78.8660	77.2570	76	74	0.9370
East	AT$1376	AT$1765	7587	3062	1243	6344	79.7820	74.9520	74	73	0.9990
East	AT$1747	AT$1342	249	81	38	211	56.2350	51.8940	55	52	0.9930
East	AT$1749	AT$1344	1708	809	309	1399	56.1900	55.7390	54	53	0.9910
East	AT$1751	AT$1346	7724	3520	1259	6465	56.3850	53.0940	52	52	1
East	AT$1763	EHV$1006	1679	220	301	1378	75.2220	73.6800	73	72	0.6150
East	AT$1765	EHV$1008	7684	822	1262	6422	76.5810	69.6750	72	70	0.7090
East	BET$1719	BET$1733	248	87	38	210	62.2780	58.8430	60	57	0.8610
East	BET$1721	BET$1735	1707	830	309	1398	61.5520	61.5610	59	57	0.9120
East	BET$1723	BET$1737	7727	3607	1259	6468	61.9350	58.9700	56	57	1
East	BET$1733	AT$1747	248	86	38	210	61.9720	58.2800	60.5	57	0.8310
East	BET$1735	AT$1749	1708	818	309	1399	61.2880	60.9430	59	58	0.9990
East	BET$1737	AT$1751	7721	3571	1259	6462	61.6800	58.2570	57	57	0.9990
East	BTL$1104	BTL$1156	888	364	121	767	NaN	45.9560	44	44	0.7020
East	BTL$1156	BTL$1222	888	360	121	767	NaN	53.9170	52	52	0.8410
East	EHV$184	EHV$258	1397	157	345	1052	73.1540	76.2210	117	117	0.4950
East	EHV$24	EHV$106	94	12	18	47	NaN	85.1670	89.5	92	0.4940
East	EHV$24	EHV$108	128	21	14	77	NaN	85.1900	89.5	85	0.8970
East	TBI$192	TBI$198	1335	542	114	1221	41.8290	41.3960	38	39	0.6160
East	TBI$198	TBI$703	1630	397	154	1476	28.0780	27.8220	27	27	0.8770
East	TBU$815	TBU$60	10238	2441	1580	8658	43.3410	42.7040	40	41	0.5860
East	TBU$84	TB$104	10234	3851	1580	8654	41.5580	41.4290	40	40	0.6420
East	VGA$603	VGA$607	939	370	123	816	NaN	64.6300	64	64	0.7370
East	VGA$607	BTL$617	940	372	124	816	NaN	118.5540	117	117	0.6130
West	VGA$728	TBI$712	1711	302	165	1546	321.9470	301.2750	271	274	0.7470
West	AT$1364	AT$1756	181	51	29	152	NaN	63.0780	62	61.5	0.6060
West	AT$1366	AT$1758	9031	4979	1570	7460	NaN	62.1540	61	61	0.9980
West	AT$1392	AT$1362	264	73	57	207	NaN	63.2470	60	61	0.8550
West	AT$1394	AT$1364	172	39	24	148	NaN	59.9490	60.5	60	0.4490
West	AT$1396	AT$1366	9040	4953	1571	7468	NaN	60.0780	59	59	0.8650
West	AT$1740	BET$1726	283	86	61	222	NaN	55.3370	54	54	0.3260
West	AT$1742	BET$1728	196	57	31	165	NaN	54.9300	55	54	0.8420
West	AT$1744	BET$1730	9118	4996	1586	7531	NaN	54.7600	54	54	0.8810
West	AT$1754	AT$1740	282	90	60	222	NaN	55.1670	52.5	53	0.5520
West	AT$1756	AT$1742	196	59	31	165	NaN	53.7800	52	52	0.7320
East	TBI$182	TBI$192	1339	733	114	1225	61.8720	61.3570	58	60	0.5680
East	TB$521	BTL$1525	8706	4405	1477	7229	114.0880	111.0670	108	108	1
East	TB$515	TB$521	8742	4641	1481	7261	55.3470	54.2630	53	53	1
East	TB$509	TB$515	8737	4636	1481	7256	68.5440	67.0980	66	66	1
East	EHV$24	EHV$110	76	12	25	35	NaN	85	82	86	0.5920
East	EHV$24	EHV$112	464	94	79	385	91.3330	94.6070	101	98	0.5710
East	EHV$24	EHV$114	881	200	118	763	66.6090	62.5520	66	64	0.3320
East	EHV$26	EHV$108	1349	44	44	1125	NaN	60.1590	86	84	0.8470

Continued on ne

Table D.1 (continued).

richting	u	v	n_{tot}	n_{avg}	n_T	n_F	$avg\_\bar{r}\_heavy$	$avg\_\bar{r}\_regular$	$median_T$	$median_F$	p
East	EHV$26	EHV$110	4721	1121	1098	3623	82.3590	81.3370	83	83	0.7950
East	EHV$26	EHV$112	197	30	25	172	105.5	93.0500	112	108.5	0.9240
East	EHV$28	EHV$110	86	11	11	60	NaN	84.0910	89	85	0.7510
East	EHV$46	EHV$182	715	87	33	682	84.2140	75.2190	89	80	0.8950
West	BET$1712	LPE$1282	281	48	61	220	NaN	64.9580	73	68	0.9740
East	GZ$807	GZ$811	10355	3985	1595	8760	72.9040	72.6010	69	70	0.9930
East	LPE$1256	LPE$1276	663	324	93	570	NaN	48.9040	46	46	0.7630
East	LPE$1258	LPE$1276	8695	4051	1472	7223	60.0010	53.5510	50	49	1
East	LPE$1274	BET$1721	199	59	27	172	NaN	216.6270	205	209	0.9020
East	LPE$1276	BET$1719	165	64	32	133	287.1520	247.1290	256.5	244	0.7100
East	LPE$1276	BET$1723	7740	3705	1259	6481	249.9160	228.8400	216	217	1
East	TB$104	TB$140	10001	4116	1550	8451	44.8640	46.3910	41	42	0.6030
East	TB$164	TB$503	8728	4634	1481	7247	64.1930	63.1220	61	61	0.8720
East	TB$503	TB$509	8734	4638	1480	7254	51.4200	50.4250	49	49	0.9940
East	LPE$1256	LPE$1274	229	58	29	200	NaN	49.4480	44	46	0.7090
West	VGA$732	VGA$728	1716	222	165	1551	58.7290	55.0860	52	53	0.3740
East	AT$1376	AT$1761	62	18	3	58	NaN	100	113	94.5	NaN
East	BTL$1186	BTL$1224	48	19	5	32	NaN	47.1050	47	46	NaN
East	EHV$104	EHV$252	101	16	8	81	NaN	99	101.5	99	NaN
East	EHV$114	EHV$210	35	10	1	29	NaN	72	96	66	NaN
East	EHV$22	EHV$104	82	13	8	68	NaN	88.3850	90.5	86	NaN
East	LPE$1272	BET$1721	49	19	8	34	NaN	250.1050	236.5	241.5	NaN
East	TB$104	TB$136	41	10	3	21	NaN	118.8000	129	117	NaN
East	TB$162	TBI$182	65	13	3	46	NaN	94.0770	82	99.5	NaN
East	VGA$328	VGA$304	74	15	4	57	NaN	59.4670	60.5	59	NaN
West	AT$1396	AT$1364	39	14	9	30	NaN	72.5710	73	72.5	NaN
West	BTL$1124	BTL$1542	27	12	4	23	NaN	116	80.5	138	NaN
West	EHV$1016	EHV$1770	50	13	8	42	NaN	81.8460	73.5	79	NaN
West	EHV$132	EHV$52	30	11	9	21	NaN	79.4550	78	79	NaN
West	EHV$60	EHV$1796	35	13	6	29	NaN	45.5380	46.5	45	NaN
West	EHV$60	EHV$1798	60	34	7	53	NaN	47.5	48	47	NaN