Received 30 April 2026; revised 22 August 2026; accepted 1 September 2026. Date of publication 3 September 2026; date of current version 10 September 2026.

Digital Object Identifier 10.1109/OJITS.2026.3730477

A Two-Level Data-Driven Event Graph Approach for Performance Evaluation of a Driver Advisory System for Freight Trains

JULIA H. MICHELS[1], ZIYULONG WANG[1], JAN ANNE ANNEMA[2], JELLE VAN LUIPEN[3], AND ROB M. P. GOVERDE[1] (Member, IEEE)

[1]Department of Transport and Planning, Delft University of Technology, 2628 CN Delft, The Netherlands
[2]Department of Engineering Systems and Services, Delft University of Technology, 2628 BX Delft, The Netherlands
[3]Department of Innovation and Development, ProRail, 3511 EP Utrecht, The Netherlands

CORRESPONDING AUTHOR: Z. WANG (z.wang-19@tudelft.nl)

The work of Ziyulong Wang and Rob M. P. Goverde was supported by Europe's Rail Joint Undertaking through European Union's Horizon Europe Research and Innovation Program under Grant 101101973 (FP1—MOTIONAL).

ABSTRACT As rail networks face increasing pressure to improve punctuality and capacity under mixed-traffic conditions, digital support tools such as Connected Driver Advisory Systems (C-DAS) are increasingly deployed. Yet, empirical evidence of their operational effects in daily practice remains limited. This paper presents a data-driven two-level approach based on macroscopic and microscopic event graph representations, constructed from the same underlying corridor and observed train movements, to evaluate the performance of C-DAS. It fuses timetable event records, track section occupation and release data, and C-DAS usage logs to enable comparison between trains operated with and without C-DAS at timetable point and block section levels. Derived from the event graphs, differences in timetable deviation, running time deviation, and the occurrence of unplanned yellow signal aspects are assessed using non-parametric statistical tests. The approach is demonstrated through its application to RouteLint, a C-DAS deployed for freight operations in the Netherlands, using one year of data from a mixed-traffic mainline corridor comprising 23,573 freight train runs. The results show modest but consistent differences in operational performance between trains operated with and without RouteLint. Statistically significant differences in timetable deviation are identified at multiple key locations, with RouteLint-equipped trains showing approximately 30–70 s lower deviations at major stations and junctions. At the microscopic level, running time deviations exhibit substantial local variability, while lower deviations are observed on selected block sections for RouteLint-equipped trains. In addition, RouteLint-equipped trains show approximately 6% lower overall odds that an observed yellow signal aspect is unplanned, while the association varies across operating contexts.

INDEX TERMS Driver advisory system, rail freight, operational performance, statistical analysis.

I. INTRODUCTION

RAILWAY systems worldwide are undergoing a profound shift as digitalisation and data-driven operational support become increasingly important for sustaining performance in dense and heterogeneous networks [1]. Mixed traffic corridors illustrate these pressures most clearly, since passenger and freight trains have fundamentally different operating characteristics yet must share capacity-limited train paths in a tightly coordinated timetable [2]. The Netherlands provides a representative example, where several of the main corridors in the 2026 timetable are already expected to experience structural capacity shortages under growing transport demand [3]. Within such constrained conditions, freight trains are particularly susceptible to delays. Their longer acceleration and braking distances and lower running speeds [4], combined with limited train path flexibility due to tactical planning decisions made up to 18 months in advance, leave little room to recover from small timing deviations, which then propagate through the network and reduce effective capacity [5]. Although additional tracks and route upgrades remain essential at highly utilised bottlenecks,

The review of this article was arranged by Associate Editor Yihui Wang.

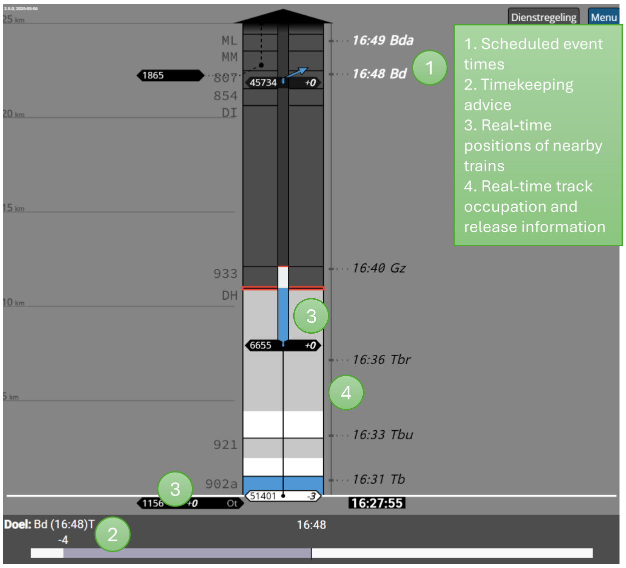

FIGURE 1. RouteLint dashboard showing scheduled event times, timekeeping advice, nearby train positions, and track occupation and release information.

these pressures have also intensified interest in Intelligent Transport Systems (ITS) that use operational data and real-time information to support reliable and efficient railway operations within the limits of the existing infrastructure [6].

Railway ITS span enabling functions such as precise train positioning and interoperable data architectures, as well as operational applications that connect traffic management information with train-level decision support [7], [8], [9]. As one prominent example within this broader ITS field, the Driver Advisory System (DAS) provides in-cab, drivable guidance that supports train drivers in regulating their driving in line with operational targets. The fundamental functions of a DAS are to present the target event times (for passages, arrivals or departures) at designated network locations, monitor the movement of the train to keep the advice valid, and, when supported, compute an energy-efficient trajectory that still satisfies these targets [10]. Existing implementations range from stand-alone DAS, where timetable and route data are preloaded before departure, to networked DAS that receives occasional updates from control centres, and Connected-DAS (C-DAS), which maintains real-time communication with traffic management and continuously adjusts its advice based on the evolving network state [11]. In the Netherlands, the railway infrastructure manager ProRail has developed RouteLint, a C-DAS that provides freight drivers with scheduled event times, up-to-date timekeeping advice, real-time positions of nearby trains, and information on upcoming track occupation and release. RouteLint is now widely deployed by European freight operators on cross-border corridors, and its real-time information on nearby train positions is also used operationally by the principal Dutch passenger operator. Figure 1 shows the RouteLint interface, including scheduled event times, timekeeping advice, nearby train positions, and information on track occupation and release.

The development of DAS has stimulated extensive research on both its theoretical potential and its behavioural effects. Existing studies can broadly be divided into two strands. The first focuses on optimising the algorithms that generate driving advice, aiming to produce energy-efficient and punctual train trajectories [11], [12], [13], [14], [15], [16]. The second investigates the behavioural and operational implications of DAS through either driving simulator studies or stated preference studies supported by experimental trials with human drivers [17], [18]. Although these contributions provide important insights, they generally concern prototype developments, controlled environments, or short-term experiments and therefore offer limited evidence on C-DAS performance under the heterogeneous conditions encountered in daily railway operations. Large-scale operational data are thus needed to examine repeated train runs over an extended period, capture variation across locations and traffic conditions, and assess performance at both the macroscopic level, where deviation is measured at timetable points, and the microscopic level, where track occupation and release are recorded at the block section level. However, no established approach currently integrates C-DAS usage records with timetable event and track occupation and release data to support such a two-level evaluation. As a consequence, despite promising theoretical and experimental findings, empirical evidence on the performance associated with C-DAS use in day-to-day freight operations remains limited. This gap motivates the central question of this paper: How can the operational performance of C-DAS-equipped trains be quantified using large-scale empirical data from real-world railway operations?

To address this question, this paper develops a structured and scalable approach to evaluate the operational performance associated with railway DAS use using real-world empirical data. We integrate one year of RouteLint usage logs with timetable event data and track occupation and release records from a mixed-traffic mainline corridor. These data are analysed at two complementary levels using event graph representations constructed from the same underlying corridor and observed train movements: the macroscopic level reflects timetable point performance, while the microscopic level describes driving deviations at the block section level. For each edge in both graphs, we estimate conflict-free average running times. This two-level approach enables statistical comparison across multiple indicators between DAS-equipped and non-equipped trains operating on the same corridor, including transport volume, timetable deviation, unplanned yellow signals, and various train interaction scenarios. It is adopted because timetable adherence and running time deviations are defined at different levels of granularity and require distinct aggregation, computation, and visualisation. Differences in timetable and running time deviation, treated as continuous variables, are evaluated using the Mann-Whitney U test, with random subsampling applied as a robustness check. We analyse unplanned yellow signals and their associated causes as categorical variables using Pearson's Chi-squared tests.

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

1) We develop a data-driven two-level approach that enables statistical comparison of train operation performance at timetable point and block section levels, based on macroscopic and microscopic event graph representations.
2) We apply the approach to compare operations of trains with and without C-DAS using one year of fused real-world operational data, providing an evidence-based quantification of C-DAS performance under mixed-traffic conditions.
3) We provide policy implications on the operational contexts in which C-DAS use is associated with measurable performance differences, and identify potential functionality enhancements for RouteLint as currently deployed in the Nethers.

The remainder of this paper is organised as follows. Section II reviews the relevant literature and technological developments and further elaborates on the research gap. Section III presents the approach developed for this paper. Section IV introduces the case study, and Section V presents the main findings, while Section VI discusses the limitations and policy implications. Finally, Section VII summarises the key insights and outlines directions for future research.

II. LITERATURE REVIEW

This section provides an overview of the existing literature and current practices related to DAS. It first discusses the state-of-the-art developments in DAS technology and research, with emphasis on system architectures and advisory algorithms, and then reviews current applications and implementations within European rail networks.

A. STATE-OF-THE-ART OF DAS

As noted earlier, DAS are commonly grouped into three categories: stand-alone, networked, and connected [10], [11]. Stand-alone DAS operates independently by loading the necessary timetable and route information onto the train at the start of the mission, i.e., before the first departure, after which the timetable data remain static and cannot be updated. Networked DAS introduces communication with one or more traffic control centres and can receive updates of the traffic plan. However, these updates are not continuous and therefore do not provide fully real-time adaptation. C-DAS represents the most advanced category, enabling real-time bi-directional communication with the traffic control centres in the operating area. This allows the system to receive updates to the real-time traffic plan dynamically and to generate revised advice that helps the driver avoid potential track occupation conflicts and remain aligned with the evolving traffic conditions.

Three main architectures for C-DAS driving functions can be distinguished in the literature [10], [11], [19]. In the DAS-onboard configuration, the onboard subsystem computes the train trajectory, derives the driving advice, and presents it directly to the driver via the driver machine interface. In the DAS-intermediate configuration, the train trajectory is computed by trackside subsystems and transmitted to the train, where the onboard unit derives the corresponding speed advice before displaying it to the driver. In the DAS-central configuration, all computation takes place trackside, and the onboard subsystem functions solely as a display device.

Independent of these categorical and architectural differences, all DAS rely on specific forms of advice to guide the driver. These include target speed advice, timekeeping advice such as predicted deviation from target event times at designated locations, and control action advice [10]. Some implementations may also include energy-efficiency advice aimed at reducing traction energy consumption in both undisturbed and disturbed traffic conditions.

C-DAS relies on real-time traffic management information. Recent work in the Europe's Rail FP1 MOTIONAL project has emphasised the interaction between traffic management systems and C-DAS [20], [21]. The traffic management system generates a real-time traffic plan describing the current ordering of trains, their routes, and their planned event times at timetable points [22], [23]. This plan can be used to compute a train path envelope for each train, identifying feasible time or speed windows within which conflict-free trajectories may be generated onboard [24], [25]. The exchange of such information between traffic management and C-DAS is supported by the Smart Communications for Efficient Rail Activities (SFERA) protocol developed by the International Union of Railways (UIC), which establishes standardised interfaces for interoperable data exchange [26].

Although DAS can draw on the extensive body of research on optimal and energy-efficient train trajectory optimisation (e.g., [27], [28], [29], [30], [31], [32]), this paper focuses specifically on developments and evaluations of DAS rather than on trajectory optimisation itself. Within the DAS-centred literature, a first line of research concentrates on improving advisory algorithms for either energy-efficient driving under nominal conditions or delay-recovery under disturbed conditions. For instance, a traction-distance-based algorithm for generating advisory information was designed and tested on a hardware-in-the-loop platform to compare driving with and without DAS [14]. A bi-level optimal control approach combined a minimum-time trajectory derived from Pontryagin's Maximum Principle with a coasting profile computed using a Lagrange multiplier technique to generate real-time advice to the driver [16]. The feasibility of onboard implementations was demonstrated by an Android-based prototype that applied dynamic programming to compute energy-efficient advice in practice [12]. In addition, a dynamic offline-online trajectory optimisation method was developed as a practical DAS solution that pre-computed benchmark trajectories and adapted them during the run to balance energy efficiency and timetable adherence [13]. For freight operations, an algorithmic architecture integrating buffer stairway prediction, freight train movement prediction, merging window detection, and merging optimisation was created to support conflict-aware driving in mixed-traffic conditions [11].

A second line of research examines how drivers perceive, interpret, and act upon DAS advice, using driving simulator experiments, stated preference studies, and human factors analyses. A recent systematic scoping review of rail human factors and automation [33] highlighted several gaps in this domain, including a predominant focus on drivers, small sample sizes, and limited attention to communication dynamics with traffic controllers. Early work showed that advisory information could support smoother driving, reducing unnecessary stops and associated energy consumption [34]. Because DAS reshapes traditionally unassisted driving tasks, researchers emphasised the need to understand how advisory information influences modern driving performance and to develop driver models that reflect contemporary operational demands [35]. Simulator experiments in the United Kingdom showed that novice drivers exhibited more erratic control behaviour, characterised by longer actuation periods and frequent switching between traction and braking [36]. More targeted ergonomics studies assessed specific forms of advice. Using eye tracking and mental workload measures, the study reported in [17] found that providing route context information or coasting advice did not negatively affect safety performance or attention allocation relative to baseline driving. Complementary evidence in [18] indicated that although DAS did not significantly alter job characteristics or occupational satisfaction, drivers with stronger task identity and professional commitment showed higher acceptance of and compliance with DAS guidance.

B. STATE-OF-PRACTICE OF DAS

Operational DAS has been deployed across several European railway networks over the past decade, developed by infrastructure managers, railway undertakings, or commercial system suppliers.

In many countries, infrastructure managers provide DAS, typically free of charge, that supply railway undertakings with real-time traffic information such as train positions, signal aspects, and updated event times. Under the European Union's vertically separated regulatory framework, infrastructure managers are prohibited from issuing prescriptive speed advice, as this would constitute operational control [37]. Therefore, speed-related advice remains the responsibility of the railway undertaking. Examples of such DAS provided by infrastructure managers include RouteLint in the Netherlands, infraDOAS in Austria, and the Infrabel DAS in Belgium.

Other networks employ supplier-based systems such as Siemens C-DAS or KeTech C-DAS in the United Kingdom, or GreenSpeed in Denmark, which integrate timetable, infrastructure, and real-time traffic data into onboard advisory modules. In Sweden, the Computer-Aided Train Operation (CATO) system, developed jointly by Transrail, Trafikverket, and LKAB, supports both connected and stand-alone operation, applicable to both freight and high-speed trains [38].

Several railway undertakings have also developed proprietary systems tailored to their operational needs [39]. For example, Nederlandse Spoorwegen (NS) uses TIMTIM to provide coasting-based advice for energy-efficient passenger operations [40], while Société Nationale des Chemins de fer Français (SNCF) deploys Opti-Conduite as a tablet application embedding the onboard C-DAS Energymiser to support energy minimisation and punctuality management [41].

C. SUMMARY AND IDENTIFIED RESEARCH GAPS

The development of DAS has progressed from conceptual advisory models to technically mature prototypes, and multiple full-scale deployments with diverse functionalities can already be observed across European networks. However, despite these advancements, existing research still relies predominantly on simulations, prototype evaluations, or controlled trials, offering only limited empirical insight into the realised performance of DAS under the heterogeneous and evolving conditions of daily railway operations. In particular, no established methodology exists for evaluating C-DAS using large-scale real-world data that capture both macroscopic timetable-level and microscopic block-section-level deviations, nor for systematically comparing DAS-equipped and non-equipped trains under mixed-traffic conditions. This creates a scientific gap in linking C-DAS usage records with timetable event and track occupation and release data to examine observed performance differences consistently across the two levels. In practice, ProRail, DB Cargo and other European freight railway undertakings need such evidence to assess deployed systems, identify where performance differences emerge and prioritise further development. As a result, the operational effects of C-DAS in day-to-day railway operations remain insufficiently evidenced in the existing literature. This paper aims to address these gaps.

III. METHODOLOGY

This section presents the datasets, data preprocessing and integration, the two-level event-graph approach, Key Performance Indicators (KPIs), and statistical comparison methods used to evaluate RouteLint. Figure 2 summarises the methodological workflow.

A. DESCRIPTION OF THE C-DAS UNDER STUDY AND AVAILABLE DATASETS

1) OVERVIEW OF ROUTELINT

RouteLint is developed and operated by the Dutch railway infrastructure manager ProRail and is currently provided to freight railway undertakings. It functions as a C-DAS that presents drivers with timekeeping advice, information about nearby trains, the upcoming route, and up-to-date event times at timetable points. For each upcoming event, RouteLint displays the expected deviation from the scheduled event time in minutes, with positive values indicating delay and negative values indicating earliness. This deviation is computed from an estimate of the current speed, derived from the travelled distance over elapsed time, combined with the remaining distance to the next timetable point. Since July 2024, the computation additionally incorporates applicable speed limits to improve the accuracy of the estimated arrival or passing time.

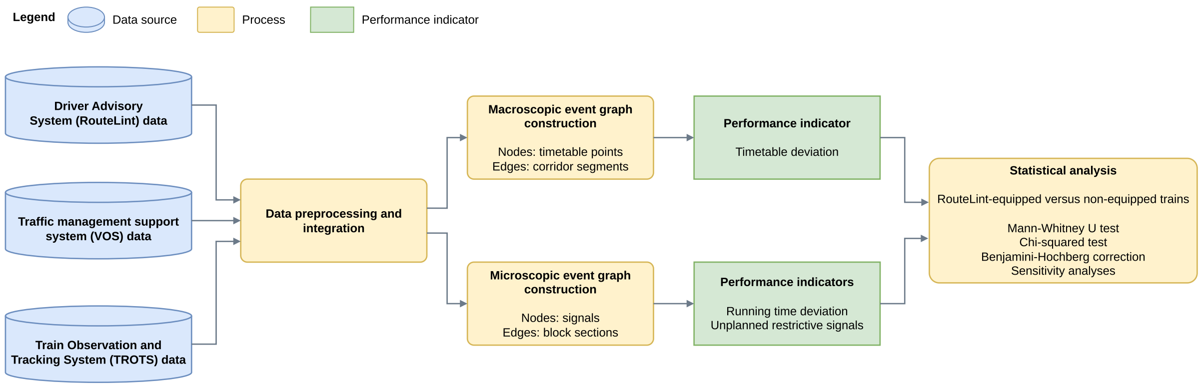

FIGURE 2. Overview of the approach for evaluating operational performance associated with RouteLint usage.

TABLE 1. Summary of variables in the RouteLint usage logs.

Variable	Type	Description
Session timestamp	datetime	Time at which the RouteLint login session started
Train number	int32	Identifier of the train linked to the login session
Date of operation	datetime	Operational date of the corresponding train run
RouteLint user identifier	string	Pseudonymised RouteLint user identifier including operator association
Driving advice used	bool	True if RouteLint driving advice was active during the session; False otherwise
GPS used	bool	True if GPS-based localisation was enabled; False otherwise
GPS-infrastructure match	bool	True if the GPS-derived position aligned with infrastructure data within 100 m; False otherwise
Application version	string	Version of the RouteLint application used during the session
Log file date	datetime	Date on which the corresponding log file was generated

RouteLint relies on several data sources to determine train position and route status. By default, train position is estimated using Global Positioning System (GPS) data, referenced against ProRail's GEO dataset (a geographic reference model containing the digital representation of the Dutch railway network). When GPS reception is unreliable, for example, due to shielding by steel freight locomotives or when trains pass through tunnels, RouteLint switches to trackside sensor data to infer train position. Information on the upcoming route is provided using a real-time infrastructure status feed (reporting signal aspects and switch positions), which, together with trackside track-clear detection, determines which track sections are currently occupied and which are cleared for the train.

As shown in Figure 1, the RouteLint interface displays the upcoming track using a colour scheme that distinguishes between the train's current position (blue), cleared sections on the open line with automatic signals (light grey), cleared sections in controlled interlocking areas (white), and uncleared sections where the train is not yet authorised to proceed (dark grey). The displayed schedule is based on the national system for railway capacity allocation and timetable management, and reflects any updates introduced by traffic controllers during operations.

Although RouteLint offers functionalities for timekeeping and situational awareness, several limitations are relevant for interpreting the results of this paper. In its current configuration, RouteLint does not explicitly incorporate the delay status of surrounding trains when producing timekeeping advice, provides time-based rather than speed-based guidance, and may occasionally exhibit minor inconsistencies when GEO-based positioning diverges from trackside sensor inputs. These characteristics are accounted for when processing the dataset and interpreting performance differences.

2) ROUTELINT USAGE LOGS

RouteLint usage is recorded in system logs that, for each login session, identify the train associated with the session and whether driving advice was activated during the run. The key variable relevant to this paper is whether a train operated with RouteLint guidance. To establish this, the RouteLint logs are linked to the timetable point records and the track section occupation and release records described below using train number, date of operation, and operator information. Because operator names are not fully consistent across the different datasets, naming conventions are harmonised prior to linkage. A Boolean indicator is then added to both records to enable subsequent comparisons, classifying each train as RouteLint-equipped (RouteLint usage recorded during the run) or non-equipped (no RouteLint usage recorded). Table 1 presents the variables available in the RouteLint usage logs used for this analysis.

3) PLANNED AND REALISED TIMETABLE POINT RECORDS

Macroscopic timetable information is obtained from the Dutch traffic management support system (Verkeersleiding Ondersteunend Systeem, VOS). For each train and each

TABLE 2. Summary of variables in the VOS timetable event dataset.

Variable	Type	Description
Date of operation	datetime	Operational date of the train run
Operator	string	Operating company associated with the train
Train number	int64	Identifier of the train
Timetable point	string	Identifier of the timetable point at which the event occurs
Event type	string	Arrival, departure, or passing event
Track	string	Identifier of the track at the timetable point
Driving characteristic	string	Train type used for filtering and classification
Planned event time	datetime	Scheduled event time in the published timetable
Adjusted event time	datetime	Rescheduled event time in the real-time traffic plan
Realised event time	datetime	Executed event time recorded during operations
Planned-adjusted deviation	float64	Difference between planned and adjusted event times (in minutes and seconds)
Adjusted-realised deviation	float64	Difference between adjusted and realised event times (in minutes and seconds)
Change in deviation	float64	Change in deviation relative to the previous timetable point (in minutes and seconds)
Train length	float64	Recorded train length [m]
Train weight	float64	Recorded train weight [t]
Next timetable point	string	Identifier of the timetable point that follows the current event in the real-time traffic plan
Direction	string	Direction of movement along the corridor
Train category	string	Category of train (heavy or regular)

TABLE 3. Summary of variables in the TROTS track section occupation and release dataset.

Variable	Type	Description
Date of operation	datetime	Operational date of the train run
Operator	string	Operating company associated with the train
Train number	int64	Identifier of the train
Timetable point	string	Identifier of the timetable point associated with the signal event
Event type	string	Recorded event at the signal (primarily passing; rare arrival/departure/shunting)
Driving characteristic	string	Train type used for filtering and classification
Signal identifier	string	Identifier of the signal at which the event occurs
Signal aspect	string	Aspect displayed at the moment of passage (green, yellow, red)
Signal passage time	datetime	Time at which the signal was passed
Signal clearance time	datetime	Time at which the next signal first displayed a permissive aspect
Signal release delay	int64	Delay between passage of a signal and clearance of the next signal [s]
Signal type	string	Signal classification according to infrastructure specifications (automatic, controlled)
Deviation at timetable point	float64	Deviation between adjusted plan and realised time at the associated timetable point [s]
Cause of yellow aspect	string	Assigned coded cause for a restrictive (yellow) aspect
Preceding train	int64	Identifier of the train that previously occupied the next track section
Direction	string	Direction of movement along the corridor
Train category	string	Category of train (heavy or regular)

timetable point, VOS records the planned event times from the published timetable, the adjusted event times in the real-time traffic plan, and the realised execution times. These records enable the construction of timetable point deviations and support the characterisation of running time behaviour between successive timetable points in later stages of the analysis. Table 2 summarises the variables used from the VOS dataset, including identifiers, event attributes, planned, adjusted, and realised event times, deviation indicators supplied by the system, and supplementary train characteristics used for filtering and classification. The train category is classified based on the bimodal distribution of train weight in the data, with trains labelled as heavy or regular.

4) TRACK SECTION OCCUPATION AND RELEASE RECORDS

Microscopic operational data are obtained from the train observation and tracking system (Trein Observatie en Tracking Systeem, TROTS), which monitors train movements at the level of signals and track sections. TROTS records passage times and corresponding signal aspects, as well as track section occupation and release events. These records enable detailed reconstruction of running times between signals, identification of restrictive aspects, and analysis of train interactions at the block section level, where each block section is represented by one or more track sections. Similar analyses based on train describer data, as now implemented in TROTS, have been used to identify and analyse route conflicts, delay propagation, and capacity bottlenecks [42]. The Dutch mainline railway network uses a two-block, three-aspect signalling system (NS'54), in which a signal displays red when the block ahead is occupied, yellow when the next signal is red, and green otherwise [43]. Approximately 5% of the signal aspect records in TROTS are missing. Let I denote the set of observed train runs and S_i the ordered set of signals passed by train i for which the clearance time of the next downstream signal is available. For each $s \in S_i$, $s+1$ denotes the next downstream signal in the direction of travel. To infer these missing aspects, the passage time of train i at signal s is compared with the time at which the downstream signal $s+1$ changes from red to a permissive aspect. Let $t_{i,s}^{\text{pass}}$ denote the time at which train i passes signal s, and let $t_{i,s+1}^{\text{clear}}$ denote the recorded time at which signal $s+1$ changes from red to a permissive aspect before train i passes that signal. For all $i \in I$ and $s \in S_i$, the signal release delay is defined as

$$d_{i,s} = \max(0,\ t_{i,s+1}^{\text{clear}} - t_{i,s}^{\text{pass}}). \tag{1}$$

A zero delay ($d_{i,s} = 0$) indicates that signal s displayed green at the moment of passage, whereas a strictly positive delay corresponds to a yellow aspect. Table 3 summarises the variables used from the TROTS dataset.

B. DATA PREPROCESSING AND INTEGRATION

The data processing begins by selecting all train movements over the Brabantroute corridor between Tilburg and Eindhoven during the period 1 August 2024 to 21 July 2025 in the Netherlands. This corridor contains approximately 280 signals and 413 block sections, providing a sufficiently rich network structure for constructing an event graph and analysing running time behaviour at a microscopic level. The week from 17 to 24 April 2025 is excluded because a technical malfunction in RouteLint prevented the system from providing valid driving advice during that period.

In the VOS timetable event dataset, records with missing realised event times were removed, as these correspond to unexecuted or cancelled movements. Similarly, in the TROTS track section dataset, records without successive signal passage times were discarded, as running times between signals cannot be reconstructed from incomplete passages. In addition, running times between two successive recorded signals were discarded if passage times at intermediate signals were missing. Missing signal aspects in the TROTS dataset were inferred using the signal release delay estimation introduced earlier, which derives aspects from the relative timing between signal passage and downstream signal clearance. Because Dutch passenger trains operated by NS use other advisory systems, this paper focuses on freight operations. Therefore, only trains labelled as freight in the driving characteristic field were retained. Variable formats were standardised across datasets, and unused or empty columns were removed.

The datasets were matched using the combination of train number and date of operation as a unique identifier. Duplicate train numbers occurring on the same day were removed, and only records with consistent identifiers across VOS, TROTS, and RouteLint were retained. The RouteLint usage indicator introduced earlier was then appended to the matched VOS and TROTS datasets so that each train carries a flag denoting whether RouteLint was used during the run. Finally, each train was assigned a direction (eastbound or westbound) based on the ordered sequence of timetable points and signals along the corridor. Before filtering, the VOS dataset contained 603,184 trains corresponding to 4,067,541 timetable event records, while TROTS contained 310,581 trains with 5,679,604 signal-level records. After applying the cleaning and integration steps described above, the final datasets consist of 246,304 VOS records and 697,881 TROTS records, corresponding to 23,573 freight train runs included in the analysis.

C. TWO-LEVEL APPROACH BASED ON EVENT GRAPH REPRESENTATIONS

Graph-based representations provide a structured means of combining spatial network structure with temporally varying transport observations [44]. To structure train movements and enable comparisons between RouteLint-equipped and non-equipped trains, we adopt a two-level approach based on separate macroscopic and microscopic event graph representations. For both levels, the graph topology is determined by the same underlying corridor, while the associated data represent the same observed train movements at different levels of granularity. An event graph is defined as a directed acyclic graph $G = (V, E)$, where nodes V correspond to operational events and edges E describe the progression between successive events along the corridor. The distinction between macroscopic and microscopic levels follows established practice in railway operations research [45], [46].

At the macroscopic level, train movements are described at timetable points, where operational interactions such as meeting, overtaking, or connections may occur. The macroscopic graph $G_{\text{macro}} = (V_{\text{macro}}, E_{\text{macro}})$ captures movements between consecutive timetable points, where nodes represent timetable points, while edges describe the corridor segments connecting them. Planned, adjusted, and realised event times from the VOS dataset are associated with timetable point nodes for each observed train, while running times between successive timetable points are carried by the edges. These values are aggregated across trains to analyse timetable adherence and the spatial evolution of deviations along the corridor.

At the microscopic level, train movements are described at the resolution of block sections between consecutive signals. The microscopic graph $G_{\text{micro}} = (V_{\text{micro}}, E_{\text{micro}})$ captures movements between consecutive signals, with nodes corresponding to signals and edges denoting block sections. Signal aspects, passage times, and inferred signal release delays are associated with signal nodes for each observed train, while running times over block sections are attributed to edges. Aggregating these values across trains allows the examination of running time variability, restrictive signal aspects, and train interactions at the block section level.

Furthermore, separate graphs are constructed for eastbound and westbound movements to ensure directional consistency. In contrast to modelling specific realised train movements, aggregation across trains at both levels preserves the operational structure of the corridor and provides a consistent basis for statistical comparison between RouteLint-equipped and non-equipped trains. Additionally, the use of separate macroscopic and microscopic graphs enables level-specific edge weight computation, supports independent evaluation of the corresponding indicators, and facilitates visualisation of performance measures at their respective levels of granularity. Both graphs are implemented using the Python NetworkX package [47].

D. KEY PERFORMANCE INDICATORS

The main KPIs of railway operations relate to timetable adherence and reliability [48], [49]. Extensive research has been devoted to their quantification at different operational levels, including timetable performance assessment [50], benchmarking of railway traffic simulators [51], and the evaluation of onboard positioning systems [52]. RouteLint is designed to support timetable adherence during daily operations and, through this, to contribute to a more reliable use of available network capacity. In line with this objective, the KPIs are selected to capture operational performance at two complementary levels. At the macroscopic level, timetable deviation is adopted as an established indicator. At the microscopic level, running time deviation is adapted to the block section level using a nominal reference running time, while unplanned restrictive signal aspects are identified from the recorded signal aspects. As RouteLint currently does not include explicit energy-efficiency functionalities, energy-related indicators are not considered.

1) TIMETABLE DEVIATION AT TIMETABLE POINTS

For each observed train run $i \in I$, let $V^i_{\text{macro}} \subseteq V_{\text{macro}}$ denote the set of timetable points visited by that train. For each $tp \in V^i_{\text{macro}}$, let $A_{i,tp}$ denote the set of recorded event types of train i at timetable point tp, comprising arrival, departure, or passing events where applicable. For each $a \in A_{i,tp}$, let $T^{\text{plan}}_{i,tp,a}$ denote the planned event time and $T^{\text{real}}_{i,tp,a}$ the realised event time recorded in the VOS dataset. The timetable deviation is defined for all $i \in I$, $tp \in V^i_{\text{macro}}$, and $a \in A_{i,tp}$ as

$$\Delta T_{i,tp,a} = T^{\text{real}}_{i,tp,a} - T^{\text{plan}}_{i,tp,a}. \tag{2}$$

Positive values indicate late running relative to the planned timetable, whereas negative values indicate early running. For the statistical analysis, timetable adherence is assessed using the absolute deviation $\left|\Delta T_{i,tp,a}\right|$.

2) RUNNING TIME DEVIATION AT BLOCK SECTIONS

At the microscopic level, operational performance is analysed by comparing realised running times on individual block sections with a reference running time representing nominal conflict-free operation. The reference is constructed from train runs that are punctual at the bounding timetable points and for which no restrictive signal aspect is identified over the corresponding signal sequence.

Let $r_{i,e}$ denote the realised running time of train run i over block section $e \in E_{\text{micro}}$. For each block section e, let $I^{\text{obs}}_e \subseteq I$ denote the set of observed train runs for which a running time $r_{i,e}$ is available, and let $I_e \subseteq I^{\text{obs}}_e$ denote the subset used to compute the reference running time under the nominal operating conditions specified below. The reference running time $\bar{r}_e$ is computed as the average realised running time of the train runs in I_e

$$\bar{r}_e = \frac{1}{|I_e|} \sum_{i \in I_e} r_{i,e}, \qquad \forall e \in E_{\text{micro}}. \tag{3}$$

Timetable points are indexed according to their order along the considered direction, such that $tp + 1$ denotes the next timetable point. To define I_e, the two consecutive timetable points surrounding block section e are denoted by tp and $tp + 1$, and $S(tp, tp + 1)$ denotes the ordered set of signals between them. Specifically, an ordered correspondence between timetable points in the macroscopic event graph and sequences of signals in the microscopic event graph is defined based on the same corridor infrastructure data. Nominal operation is characterised using a punctuality threshold τ_{punc} and a minimum required sample size n_{min}. The set I_e is constructed by selecting train runs $i \in I^{\text{obs}}_e$ that satisfy the following criteria

$$\left|T^{\text{real}}_{i,tp,\text{passing}} - T^{\text{plan}}_{i,tp,\text{passing}}\right| \leq \tau_{\text{punc}}, \tag{4}$$

$$\left|T^{\text{real}}_{i,tp+1,\text{passing}} - T^{\text{plan}}_{i,tp+1,\text{passing}}\right| \leq \tau_{\text{punc}}, \tag{5}$$

$$d_{i,s} = 0, \quad \forall s \in S(tp, tp + 1). \tag{6}$$

After constructing I_e, the minimum sample size requirement is enforced as

$$|I_e| \geq n_{\text{min}}. \tag{7}$$

Conditions (4)-(5) restrict the reference set to train runs that are punctual at the two timetable points bounding the microscopic segment. Stopping or dwelling events at the bounding timetable points are excluded by selecting passing events at both locations, as the analysis focuses on freight trains, which typically pass intermediate stations and only stop at yards. As a result, block sections adjacent to timetable points where trains stop are excluded from the reference set to avoid bias in the estimation of running times. Condition (6) excludes runs affected by restrictive signal aspects by requiring zero signal release delay, such that the observed running time reflects undisturbed train movement. Finally, condition (7) enforces a minimum sample size to ensure statistical robustness by excluding block sections whose reference set contains insufficient train runs.

Using the reference running time $\bar{r}_e$, the absolute running time deviation of train run i on block section e is defined as

$$\Delta r_{i,e} = \left|r_{i,e} - \bar{r}_e\right|, \qquad \forall e \in E_{\text{micro}}, \ \forall i \in I^{\text{obs}}_e. \tag{8}$$

The absolute formulation is adopted because both slower and faster traversal relative to nominal conditions represent deviations from nominal operation and are therefore equally relevant in the assessment.

3) UNPLANNED RESTRICTIVE SIGNAL ASPECTS

Restrictive signal aspects are classified as planned or unplanned using the recorded cause of the yellow aspect in the TROTS dataset. Planned restrictive aspects arise from intended operating conditions, such as speed restrictions near switches or when passing through stations, whereas unplanned restrictive aspects are due to operational conflicts, most notably with preceding trains. Unplanned restrictive signal aspects capture disturbances that are not part of the intended traffic plan and may lead to unnecessary braking, halting, and delay propagation. For each train run, the number of unplanned restrictive signal aspects encountered along the corridor is counted using the signal release delay defined in Eq. (1). Specifically, a restrictive signal aspect is identified when $d_{i,s} > 0$, which indicates that train i passes signal s before the downstream signal clears to a green aspect. Unplanned restrictive signal aspects are further analysed by their recorded causes and by the type of the preceding train (freight, intercity, regional, empty, or other services). This enables a more detailed assessment of the operational contexts in which train interactions differ between RouteLint-equipped and non-equipped trains.

E. STATISTICAL COMPARISON METHODS

The statistical analysis distinguishes between continuous KPIs related to timetable adherence and running times, and categorical indicators describing the occurrence and causes of unplanned restrictive signal aspects. Because individual driver identifiers are not available due to privacy constraints, the independence of observations for the continuous variables cannot be formally verified. Accordingly, a subsampling-based robustness check is applied to examine whether the statistical conclusions remain consistent. In addition, we apply a correction for multiple hypothesis testing when multiple statistical tests are performed simultaneously.

1) NON-PARAMETRIC TESTS FOR CONTINUOUS VARIABLES

Differences in continuous KPIs are evaluated using the Mann-Whitney U test, a non-parametric test for comparing two independent samples without assuming normality or homoscedasticity [53]. The hypotheses are formulated as:

- Null hypothesis (H_0): The distribution of absolute timetable deviations $|\Delta T_{i,tp,a}|$ at the macroscopic level or absolute running time deviations $\Delta r_{i,e}$ at the microscopic level does not differ between RouteLint-equipped and non-equipped trains.
- Alternative hypothesis (H_1): The distribution of absolute timetable deviations $|\Delta T_{i,tp,a}|$ or absolute running time deviations $\Delta r_{i,e}$ differs between RouteLint-equipped and non-equipped trains.

Let n_1 and n_2 denote the sample sizes of the RouteLint-equipped and non-equipped groups, respectively, and let R_1 and R_2 denote their corresponding sums of ranks. For each test, the Mann-Whitney U statistic is computed as

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1, \tag{9}$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2, \tag{10}$$

$$U = \min(U_1, U_2). \tag{11}$$

The smaller of U_1 and U_2 is used for significance assessment. All tests are evaluated at a significance level of $p < 0.05$.

In addition to statistical significance, effect sizes are reported to quantify the magnitude of observed differences. For the Mann-Whitney U test, the effect size η is computed from the standardised test statistic Z as

$$\eta = \frac{Z}{\sqrt{N}}, \tag{12}$$

where $N = n_1 + n_2$ is the total number of observations included in the test. Following [54], effect sizes of 0.10, 0.30, and 0.50 are interpreted as small, medium, and large effects, respectively.

2) ROBUSTNESS CHECKS USING SUBSAMPLING

The Mann-Whitney U test assumes independence between observations. To assess whether the test results are sensitive to potential dependence among observations, a robustness check based on random subsampling is applied. For each test, a random subset containing 30% of the filtered observations used in the test is drawn without replacement, and the test is re-evaluated on this reduced sample. Subsampling has been used as a diagnostic tool for examining the sensitivity of statistical results under possible dependence [55]. If the direction and statistical significance of the results remain unchanged, the corresponding conclusions are considered robust. This subsampling procedure is used as a diagnostic safeguard and does not replace the original hypothesis testing or attempt to correct dependence among observations.

3) CHI-SQUARED TESTS FOR CATEGORICAL VARIABLES

Categorical KPIs are analysed using Pearson's Chi-squared test of independence, which, unlike the Mann-Whitney U test, does not rely on distributional assumptions for continuous data and evaluates associations between categorical variables based on aggregated frequencies [56]. Hence, the null and alternative hypotheses are formulated as:

- Null hypothesis (H_0): The classification of an observed yellow signal aspect as planned or unplanned is independent of RouteLint usage.
- Alternative hypothesis (H_1): The classification of an observed yellow signal aspect as planned or unplanned is associated with RouteLint usage.

Because the analysis is based on a 2×2 contingency table of discrete frequencies, Yates' continuity correction is applied to reduce the approximation error arising from the use of the continuous Chi-squared distribution. The Yates-corrected Chi-squared statistic is therefore computed as

$$\chi^2 = \sum_{r=1}^{2} \sum_{c=1}^{2} \frac{(|O_{rc} - E_{rc}| - 0.5)^2}{E_{rc}}, \tag{13}$$

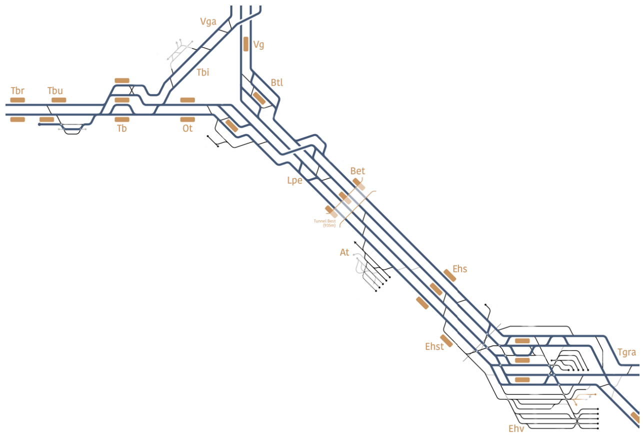

FIGURE 3. Infrastructure layout of the Eindhoven (Ehv)–Tilburg (Tb) corridor.

where O_{rc} and E_{rc} denote the observed and expected frequencies, respectively, in cell (r, c) of the contingency table. The index r distinguishes RouteLint-equipped and non-equipped trains, while c distinguishes planned and unplanned yellow signal aspects. A statistically significant result ($p < 0.05$) indicates an association between RouteLint usage and whether an observed yellow signal aspect is planned or unplanned. Beyond the aggregate comparison, we repeat the test after stratifying the observations by the recorded cause of the unplanned restrictive signal aspect and by the type of preceding train, to examine how the association with RouteLint usage varies across strata.

4) CORRECTION FOR MULTIPLE HYPOTHESIS TESTING

When multiple statistical tests are conducted simultaneously, the probability of obtaining false positives increases. To control for this effect, we use a False Discovery Rate (FDR) correction based on the Benjamini-Hochberg (BH) procedure [57]. The correction is applied to both the Mann-Whitney U tests and the Chi-squared tests whenever multiple hypotheses are evaluated in parallel. The BH procedure controls the expected proportion of falsely rejected null hypotheses among all rejected null hypotheses. Let $p_{(1)} \leq \cdots \leq p_{(m)}$ denote the m raw p-values arranged in ascending order. The BH-adjusted p-value corresponding to the ith ordered raw p-value is computed as

$$p_{\mathrm{BH},(i)} = \min\left\{1, \min_{i \leq j \leq m}\left(\frac{m}{j} p_{(j)}\right)\right\}, \qquad (14)$$

where i denotes the rank of the raw p-value for which the adjustment is calculated, j denotes the rank considered in the adjustment, and m is the total number of tests. A result is considered statistically significant when $p_{\mathrm{BH},(i)} < Q$, where the desired FDR level Q is set to 0.05.

IV. CASE STUDY

The case study focuses on the Brabantroute corridor between Eindhoven and Tilburg, a key freight artery within the Dutch mainline railway network and a representative mixed-traffic corridor linking major ports in the Netherlands and Germany. A total of 68 km of infrastructure is analysed, comprising 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). Within this scope, 15 timetable points are considered: the major stations Tilburg (Tb) and Eindhoven (Ehv); the minor passenger stations Tbr, Tilburg Universiteit (Tbu), Oisterwijk (Ot), Btl, Vg, Best (Bet), and Eindhoven Strijp-S (Ehs); the siding areas Tilburg Industrie (Tbi), Acht (At), and Liempde (Lpe); and the critical junctions Tba, Vga, and Tgra. The corridor includes inflows and outflows to freight marshalling yards and is shared with dense passenger traffic. Figure 3 presents the infrastructure layout of the Ehv–Tb corridor.

Following the start of an 80-week maintenance programme on the Betuweroute, a dedicated freight corridor, in November 2024, the Brabantroute served as a principal alternative route and experienced a substantial increase in freight traffic, with an additional 5 to 75 freight trains per day. To support conflict-free operations under these intensified conditions, RouteLint was offered free of charge to freight operators. This context provides a suitable setting for evaluating the performance of RouteLint as a C-DAS under capacity-constrained and operationally heterogeneous conditions, while ensuring adequate samples for comparisons between trains with and without RouteLint.

The temporal scope and dataset composition are defined in Sections III and III-B. The analysis covers approximately one year of operations, from 1 August 2024 to 21 July 2025, and includes 23,573 freight train runs identified in the cleaned and fused datasets. For microscopic reference running time estimation, nominal operations are identified using a punctuality threshold of $\tau_{\mathrm{punc}} = 180$ s, consistent with Dutch policy practice, together with a minimum sample size of $n_{\min} = 15$ observations per edge. Sensitivity tests with alternative cut-offs (5, 10, 15, 20, and 25 observations) indicate that average running time estimates stabilise beyond this setting while preserving sufficient network coverage.

V. RESULTS

In this section, we present the empirical results based on approximately one year of fused operational data from a Dutch mixed-traffic corridor, organised according to the defined KPIs to examine differences associated with RouteLint usage. Prior to presenting the deviation results, the empirical distributions of the deviation measures were assessed to verify the suitability of non-parametric statistical tests. The observed distributions exhibited pronounced right skewness and heavy tails, indicating clear departures from normality and supporting the subsequent use of non-parametric methods.

A. OVERVIEW OF RAIL TRAFFIC AND C-DAS ADOPTION

Figure 4 illustrates the temporal evolution of the total number of trains on the case study corridor over the analysed period. A pronounced increase in the number of trains is observed following the closure of the Betuweroute, indicating the diversion of freight trains to the Brabantroute. Temporary

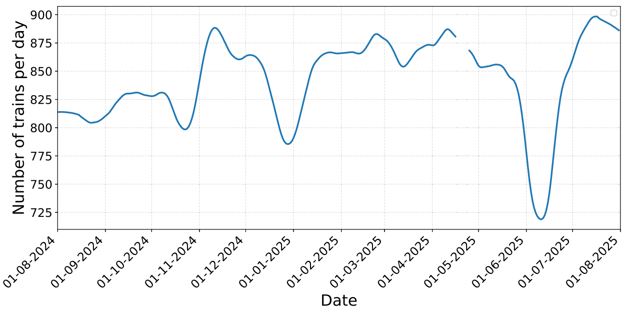

FIGURE 4. Temporal evolution of the total number of trains on the case study corridor per day.

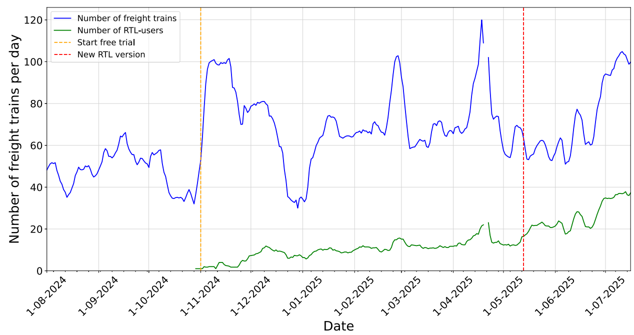

FIGURE 5. Temporal evolution of RouteLint usage in relation to total freight traffic on the Brabantroute over the study period. Vertical dashed lines indicate the start of the Betuweroute maintenance programme and the full rollout of RouteLint version 2.0 with timekeeping advice.

reductions in traffic occur during corridor-specific maintenance works in October 2024, around the end-of-year holiday period, and in early June 2025, coinciding with national passenger rail strikes on 6, 10, and 13 June. The gap in the time series corresponds to a technical malfunction in RouteLint between 17 and 24 April 2025, during which valid usage data were unavailable.

The temporal evolution of RouteLint adoption in relation to overall freight traffic is shown in Figure 5. Across the full study period, 3,635 out of 23,573 freight train runs (15.45%) were operated with RouteLint activated. Before November 2024, a limited number of drivers used an earlier version of the system that provided timetable information without timekeeping advice. This version does not qualify as a C-DAS and does not generate usage logs comparable to those of RouteLint version 2.0. Such runs are therefore excluded from the RouteLint-equipped group. The observation window is defined independently of system deployment, and effective C-DAS adoption emerges gradually from November 2024 onwards as an empirical outcome rather than as a predefined selection criterion. Nevertheless, RouteLint usage was not randomly assigned across train runs but reflected operational uptake. Consequently, the RouteLint-equipped and non-equipped groups may differ in characteristics that also affect operational performance. Following the start of the free trial period and the progressive rollout of RouteLint version 2.0, system usage increased steadily over time. From May 2025 onwards, RouteLint was actively used on approximately 38% of freight train runs on the corridor.

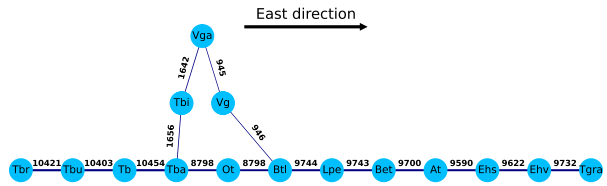

FIGURE 6. Macroscopic event graph of eastbound freight movements on the case study corridor.

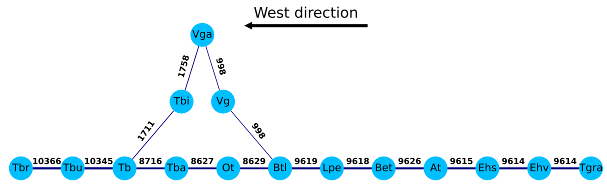

FIGURE 7. Macroscopic event graph of westbound freight movements on the case study corridor.

To complement the temporal analysis, the spatial distribution of freight traffic along the corridor is investigated using event graph representations with aggregated train movements on edges. At the macroscopic level, Figures 6 and 7 illustrate the distribution of train movements between successive timetable points, with edge widths proportional to the number of observed train movements. Freight traffic is predominantly concentrated on the central segment between the major stations Tb and Ehv, which carries approximately 8,800 train movements per direction over the study period. By contrast, the connecting sections between Vg and Ehv, and between Vga and Tb, exhibit substantially lower traffic volumes, reflecting the lesser use of branch lines compared with the main corridor.

At the microscopic level, Figures 8 and 9 present the distribution of freight traffic at the block section level for the eastbound and westbound directions. Nodes are labelled by signal identifiers, with colours distinguishing whether a node corresponds to a timetable point, while edges represent block sections between consecutive signals. As before, edge widths are proportional to the number of observed train movements over the study period. In the eastbound direction, freight trains largely follow a common routing along most of the corridor, whereas more diverse routing patterns emerge in the vicinity of Ehv due to the presence of multiple switches and platform options. In contrast, westbound trains show less route diversity overall, except near Tb, where several platforms are reserved for westbound operations and alternative paths are more frequently used.

B. TIMETABLE DEVIATION AT TIMETABLE POINTS

Figures 10-13 compare the average absolute timetable deviation at each timetable point, separated by direction (eastbound and westbound) and by RouteLint usage (without and with). The colour legend ranges from green (0-3 min, corresponding to deviations within the punctuality threshold) to dark red (absolute deviations exceeding 6 min).

Clear directional patterns emerge. In the eastbound direction, the largest average deviations are observed at Ehv,

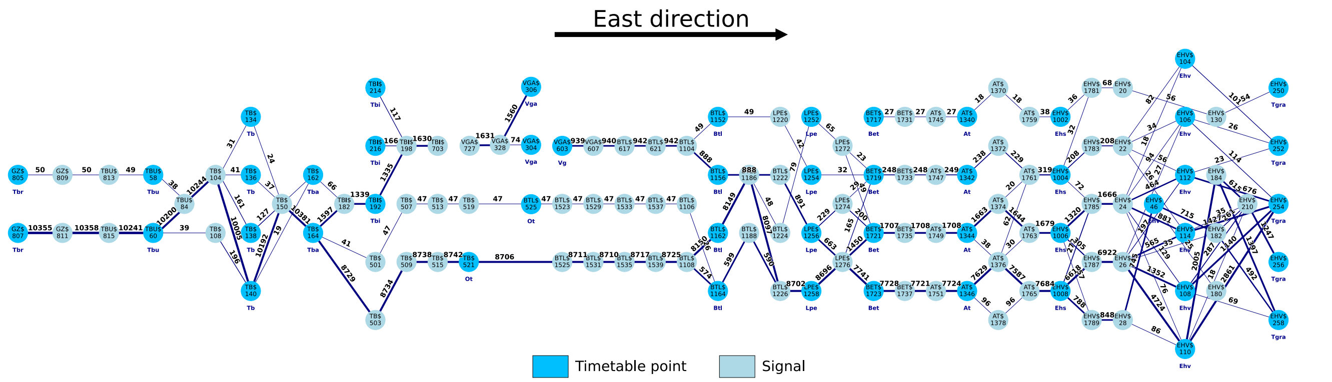

FIGURE 8. Microscopic event graph of eastbound freight movements on the case study corridor.

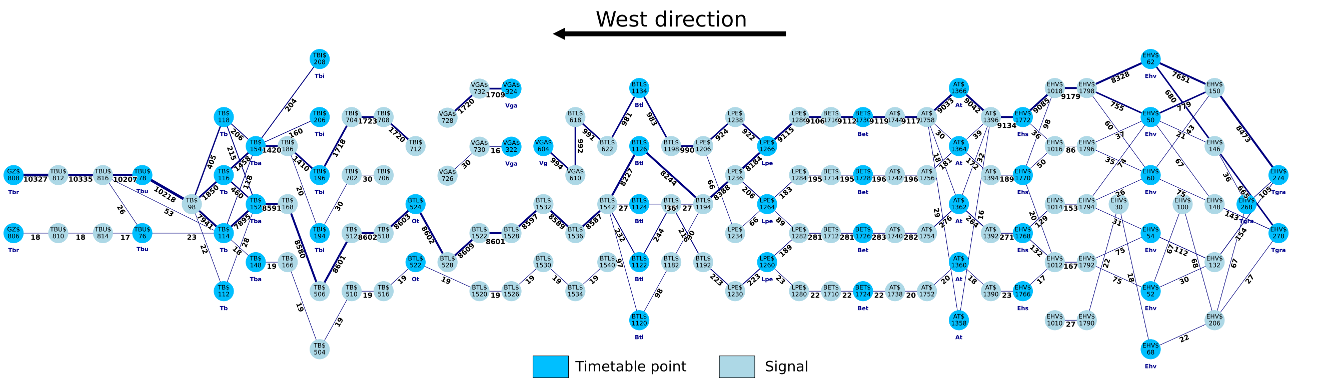

FIGURE 9. Microscopic event graph of westbound freight movements on the case study corridor.

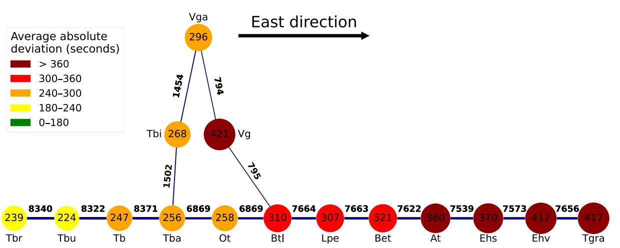

FIGURE 10. Average absolute timetable deviation per timetable point for eastbound trains without RouteLint.

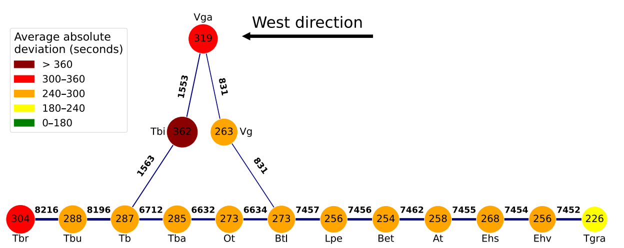

FIGURE 12. Average absolute timetable deviation per timetable point for westbound trains without RouteLint.

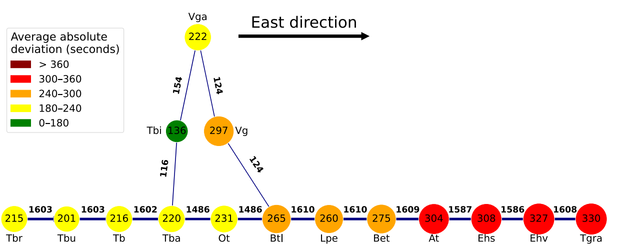

FIGURE 11. Average absolute timetable deviation per timetable point for eastbound trains with RouteLint.

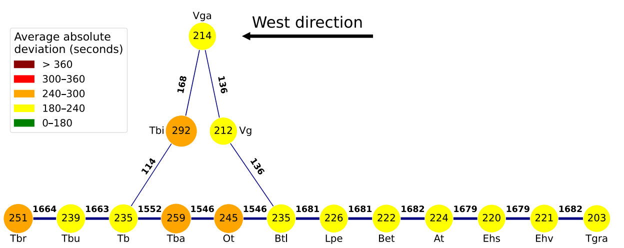

FIGURE 13. Average absolute timetable deviation per timetable point for westbound trains with RouteLint.

while in the westbound direction, the highest deviations occur around Tbi. In both cases, average deviations exceed 6 min. These timetable points correspond to major stations and adjacent junctions, where traffic density is high and operational interactions are most pronounced.

Across all timetable points, lower average deviations are consistently observed for trains operated with RouteLint compared with non-equipped trains. The size of this difference, however, varies substantially across locations. The largest difference is observed at Tbi in the eastbound direction, where the average deviation is 132 s lower for RouteLint-equipped trains, while in the westbound direction the largest difference occurs at Vga (105 s lower). In contrast, at timetable points associated with smaller stations

TABLE 4. Mann-Whitney U test results (with vs without RouteLint), reported only for timetable points significant after BH correction.

tp	$\text{Median}_{\text{with}}$ [s]	$\text{Median}_{\text{without}}$ [s]	ΔMedian [s]	η	p_{BH}
East					
Tbi	71.5	138.0	-66.5	0.1330	0.00001
Tgra	159.5	186.0	-26.5	0.0482	0.0001
Ehv	170.0	195.0	-25.0	0.0450	0.00035
Tba	101.0	119.0	-18.0	0.0341	0.0056
Vga	92.5	124.0	-31.5	0.0731	0.0064
Tb	101.0	111.0	-10.0	0.0257	0.0412
West					
Ehs	126.0	159.0	-33.0	0.0758	$1.88 \cdot 10^{-10}$
Ehv	123.0	151.0	-28.0	0.0653	$3.32 \cdot 10^{-8}$
Tbr	112.0	138.0	-26.0	0.0573	$4.84 \cdot 10^{-7}$
Tb	103.0	126.0	-23.0	0.0508	$7.01 \cdot 10^{-6}$
Tbu	107.0	124.0	-17.0	0.0471	$3.46 \cdot 10^{-5}$
Btl	134.0	149.0	-15.0	0.0444	0.0002
At	130.0	147.0	-17.0	0.0435	0.0002
Bet	122.0	139.0	-17.0	0.0424	0.0003
Lpe	124.0	137.0	-13.0	0.0342	0.0034
Tgra	102.0	111.0	-9.0	0.0341	0.0034
Vga	130.0	151.5	-21.5	0.0565	0.0199

and simpler infrastructure layouts, differences between the two groups are more limited. In the eastbound direction, the smallest differences are observed at Tbr (24 s) and Tbu (23 s), whereas in the westbound direction smaller differences occur at Tgra (23 s) and Tba (26 s), both of which are merging locations.

To assess whether the differences observed in average absolute deviations are reflected in differences in the deviation distributions, macroscopic-level comparisons were conducted between trains operated with and without RouteLint, separately by direction. Table 4 reports only those timetable points for which differences remain statistically significant after BH correction, together with the corresponding median deviations, standardised effect size η, and adjusted p-values.

In both directions, significant median deviation reductions ($p_{\text{BH}} < 0.05$) are observed at the major stations Ehv (-25 s eastbound and -28 s westbound) and Tb (-10 s eastbound and -23 s westbound), as well as at timetable points closely associated with these stations. In the eastbound direction, these include the diverging junctions Tba (-18 s) and Tgra (-26.5 s), while in the westbound direction, significant differences are observed at Ehs (-33 s), Tbu (-17 s), and Tbr (-26 s). The westbound direction contains a larger number of timetable points with statistically significant differences. However, the largest median reduction is observed in the eastbound direction, with a pronounced difference of 66.5 s at Tbi. This asymmetry indicates that westbound improvements are more evenly distributed across timetable points, whereas eastbound improvements are more spatially concentrated. A contributing factor is the difference in train characteristics: eastbound freight trains are, on average, heavier (mean 1,642.07 t) and exhibit higher variability in timetable deviation (mean 2.63 min, standard deviation 16.20 min) than westbound trains (mean weight 1,243.60 t, mean deviation 1.52 min, standard deviation 10.20 min). Higher variability reduces statistical power at individual locations, making distributional shifts harder to detect consistently across timetable points, even when local improvements are substantial.

Across all reported timetable points, effect sizes remain small, indicating substantial overlap between the deviation distributions. Nevertheless, the consistent negative median shifts indicate a corridor-wide tendency towards lower median deviations for RouteLint-equipped trains. At timetable points without statistically significant differences, infrastructure simplicity and limited routing options may contribute to the smaller observed differences between the two groups. This is particularly evident at smaller passenger stations such as Ot. An exception is Vg, where median differences are substantial in both directions (-20 s eastbound and -33 s westbound) but not statistically significant, indicating strong overlap between the rank distributions of the two groups.

C. RUNNING TIME DEVIATION AT BLOCK SECTIONS

Figures 14-15 present the average absolute running time deviations per block section at the microscopic level for the eastbound and westbound directions, respectively, with edge thickness proportional to deviation size. Large deviations are primarily concentrated around major operational nodes, in particular stations Tb and Ehv. The largest eastbound average absolute running time deviation is observed around station Tb (1,644.35 s), whereas the largest westbound value occurs at the siding area Lpe (1,799.23 s). Both locations are characterised by extensive routing flexibility and frequent train interactions.

To quantify the association between RouteLint usage and microscopic running time deviations, absolute deviation distributions were compared between RouteLint-equipped and non-equipped trains on each block section using the Mann-Whitney U test. Across all analysed edges, 14.3% reveal statistically significant differences, while 7.6% remain significant after BH correction, indicating that statistically detectable differences are confined to a subset of block sections rather than being present along the entire corridor. Within this subset, statistically significant differences are observed on a broader set of block sections in the eastbound direction, including several locations with pronounced median reductions and moderate effect sizes. On the contrary, the westbound direction exhibits fewer significant block sections, with differences that are more localised and generally smaller in magnitude, indicating weaker distributional shifts. Table 5 reports all block sections with statistically significant differences, together with the median difference in absolute running time deviation, the corresponding standardised effect size η, and the BH-adjusted p-value, with an asterisk indicating significance after BH correction. Figures 16 and 17 mark these edges in green along the corridor for the eastbound and westbound directions, respectively.

In the eastbound direction, the largest differences are observed at the siding area At. The block section between AT$1346 and AT$1378 shows a median reduction in running

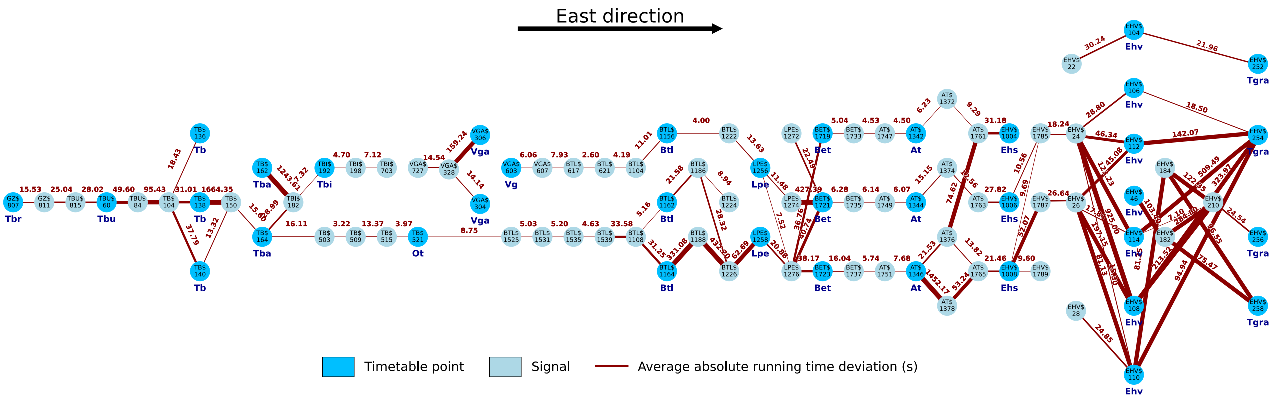

FIGURE 14. Average absolute running time deviation per block section in the eastbound direction.

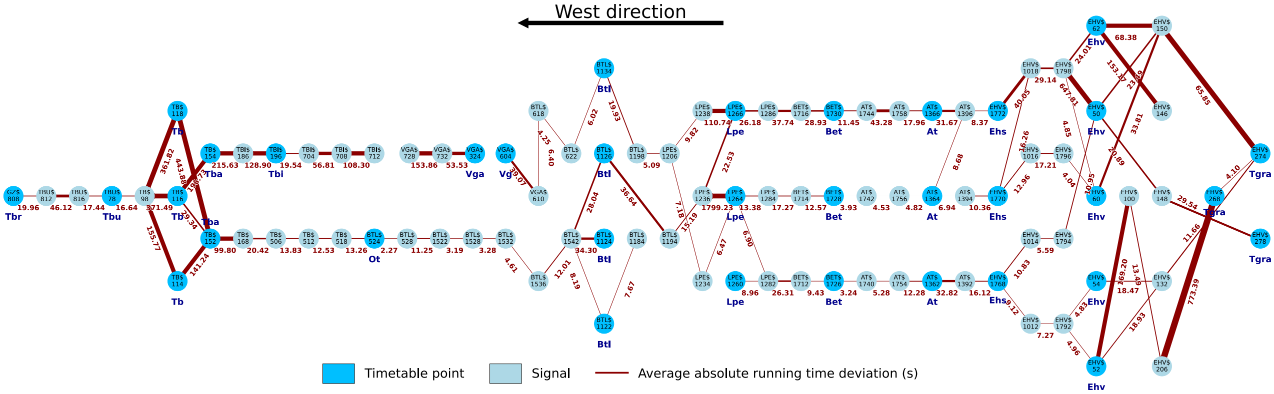

FIGURE 15. Average absolute running time deviation per block section in the westbound direction.

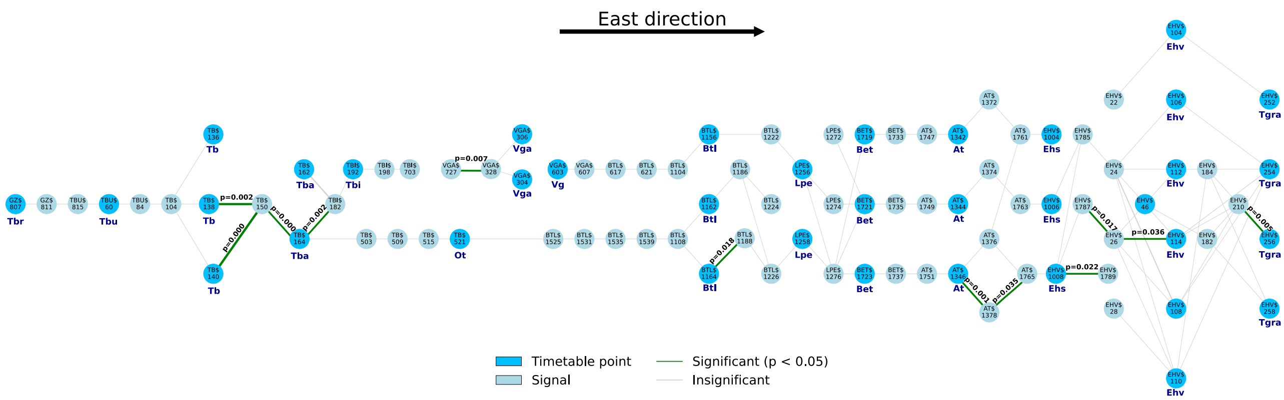

FIGURE 16. Edges with statistically significant differences in absolute running time deviation in the eastbound direction.

time deviation of 117 s for RouteLint-equipped trains, corresponding to a medium effect size of 0.335. The subsequent block section between AT$1378 and AT$1765 also exhibits a large median reduction of 41.88 s, with a smaller effect size of 0.199. Both sections are characterised by large absolute deviations and frequent use for operational waiting and train ordering, where RouteLint usage is associated with reduced dispersion in running time deviations relative to non-equipped trains.

Several statistically significant block sections are identified around station Tb in the eastbound direction. Along the main eastbound route through station Tb (TB$140-TB$150-TB$164), consecutive block sections exhibit statistically significant differences. However, median reductions remain below 1 s and effect sizes are consistently below 0.10, indicating small but recurring shifts in the distributions of running time deviations rather than substantive improvements at the block section level. Within the Tb area, the largest local effect occurs on the block section between TB$138 and TB$150, which corresponds to an alternative eastbound routing using infrastructure typically assigned to westbound movements. Here, the median reduction of 3 s is substantial relative to its short average running time of approximately 30 s and is reflected in an effect size of 0.29. Further downstream, the block section between TB$164 and TBI$182 shows a statistically significant median reduction of 2.07 s

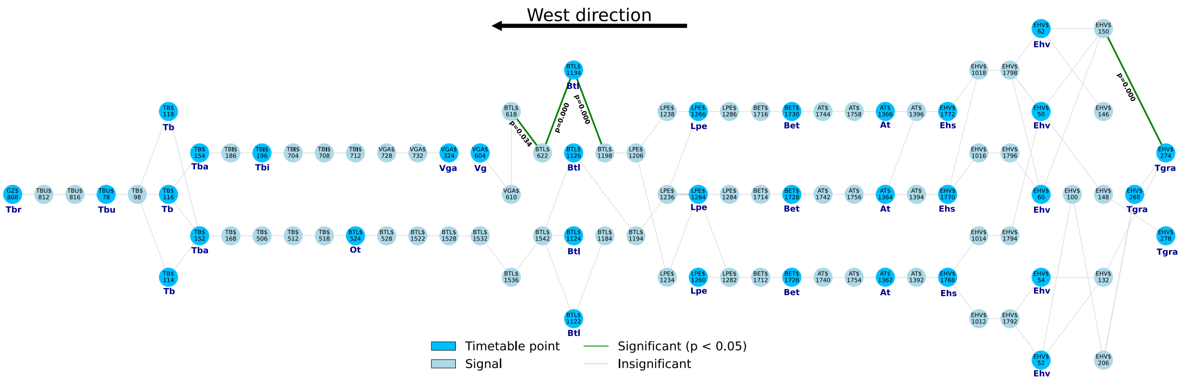

FIGURE 17. Edges with statistically significant differences in absolute running time deviation in the westbound direction.

TABLE 5. Block sections with statistically significant differences in absolute running time deviation between RouteLint-equipped and non-equipped trains, with an asterisk indicating significance after BH correction.

Direction	s	$s{+}1$	ΔMedian [s]	η	p_{BH}
West	EHV$274	EHV$150	-5.00	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	-3.00	0.290	0.0436*
East	AT$1346	AT$1378	-117.00	0.335	0.0353*
East	EHV$210	EHV$256	-1.00	0.072	0.087
East	VGA$727	VGA$328	-2.47	0.061	0.091
East	BTL$1164	BTL$1188	-1.00	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	AT$1378	AT$1765	-41.88	0.199	0.267
East	EHV$26	EHV$114	-0.74	0.012	0.267

on an average running time of 94 s, with an effect size of 0.071, corresponding to a diverging movement from Tba towards Tbi involving crossing movements. From station Ehs onwards, several eastbound block sections also indicate statistically significant differences, consistent with increased route heterogeneity arising from the presence of waiting or overtaking tracks and multiple crossings.

In the westbound direction, the most pronounced effect is observed in the Ehv area on block section EHV$274-EHV$150, where RouteLint-equipped trains exhibit a median reduction in running time deviation of 5 s, with a small effect size of 0.057. This block section is frequently used for operational waiting or overtaking within the westbound routing and is associated with higher running time deviations among non-equipped trains. Additional statistically significant block sections are identified around the Btl diverging area. On BTL$1198-BTL$1134 and BTL$1134-BTL$622, median deviation reductions of 0.28 s and 0.67 s are observed on average running times of 36 s and 62 s, respectively. These block sections correspond to diverging movements, where alternative routing options and interactions with other trains introduce additional running time variability. Although statistically significant, the associated effect sizes remain small (0.117 and 0.125), indicating limited separation between the deviation distributions of RouteLint-equipped and non-equipped trains on these short block sections.

D. ROBUSTNESS CHECK

To examine whether the Mann-Whitney U test outcomes are sensitive to potential dependence among observations, the subsampling procedure described in Section III-E was applied. For each comparison, the test was repeated on random subsamples containing 30% of the original observations drawn without replacement.

At both the macroscopic and microscopic levels, most timetable points and block sections that exhibit significant or near-significant differences in the full dataset retain the same direction of effect under subsampling, with comparable magnitudes. In particular, median running time deviations on block sections remain lower for RouteLint-equipped trains in the subsampled tests, and the negative median differences observed in the full dataset are preserved. At the microscopic level, a small number of block sections with very limited median differences in the full sample (e.g., BTL$1198-BTL$1134 and TB$150-TB$164) yield median differences that round to zero after subsampling. For timetable points At and Btl in the westbound direction, as well as block sections AT$1346-AT$1378 and AT$1378-AT$1765, statistical significance is lost, or no test result is obtained due to the limited number of RouteLint-equipped trains in the random subsample.

E. UNPLANNED YELLOW SIGNALS

Unplanned yellow signals occur mainly on approaches to major stations, where multiple routes converge, and headways tighten, most notably around Tb and Ehv. In the eastbound direction, higher occurrences are observed near signals around Tbu, while in the westbound direction, they are more pronounced near Tba.

To examine whether the occurrence of unplanned yellow signals differs between RouteLint-equipped and non-equipped trains, a chi-squared test was applied to all observed yellow signal aspects. Out of 65,910 yellow signals, 56,343 correspond to planned yellow aspects, while 9,567 are unplanned (14.5%). The test indicates a statistically

TABLE 6. **Observed frequencies and chi-squared test results for unplanned yellow signal aspects by RouteLint usage.**

	Planned yellow signals	Unplanned yellow signals
With RouteLint	9,293	1,496
Without RouteLint	47,050	8,071
χ^2 statistic		4.32
p-value		0.038
Odds ratio (with vs. without)		0.94

significant association between RouteLint usage and the classification of an observed yellow signal aspect as planned or unplanned ($\chi^2 = 4.32$, $p = 0.038$). The corresponding odds ratio is 0.94, equivalent to approximately 6% lower odds that an observed yellow signal aspect is unplanned for RouteLint-equipped trains than for non-equipped trains. Although statistically significant, the aggregate association is modest. Table 6 summarises the observed frequencies and test statistics.

Further delving into preceding traffic scenarios and the underlying causes of unplanned yellow signal aspects, the TROTS dataset is used to identify the train immediately preceding each freight train at the moment an unplanned yellow aspect occurs. This allows the preceding train type associated with prolonged track occupancy to be inferred. RouteLint provides drivers with enhanced visibility of upcoming track occupation through a forward view of up to 25 km, which may support earlier adaptation of driving behaviour before restrictive signals are encountered.

Table 7 presents the results of chi-squared tests for unplanned yellow signal aspects stratified by preceding train type, excluding types with insufficient sample sizes. For each type, the table reports the number of unplanned yellow signal aspect observations for RouteLint-equipped and non-equipped trains, the observed rate of unplanned yellow signal aspects among all corresponding signal aspect observations (%), the rate difference between the two groups (RouteLint-equipped minus non-equipped), and the corresponding odds ratio, together with the BH-adjusted p-value. After BH correction, statistically significant differences are observed when the preceding train is a freight train or a Sprinter train. In both cases, RouteLint-equipped trains exhibit lower observed rates of unplanned yellow signal aspects. Consistently, the corresponding odds ratios of 0.84 and 0.85 indicate lower odds of encountering an unplanned yellow signal aspect for RouteLint-equipped trains compared with non-equipped trains under the same preceding train type.

For different causes of unplanned yellow signal aspects, the TROTS dataset distinguishes a range of operational situations, including train reordering, merging, diverging, following, and crossing movements, as well as route-setting-related causes. After analysing these causes separately and applying the BH correction, only one statistically significant difference remains. Specifically, when an unplanned yellow signal is associated with a preceding merging passenger train, RouteLint-equipped trains exhibit a higher occurrence rate than non-equipped trains. This case corresponds to an increase of 4.49 percentage points in the unplanned yellow signal rate for RouteLint-equipped trains, with an odds ratio of 1.31 and a BH-adjusted p-value of 0.008. One possible explanation for this pattern concerns how RouteLint visualises preceding trains during merging movements. Merging trains are displayed as dotted trajectories, for which the headway impact on the following train cannot yet be determined. Although the current delay or earliness of the merging train is shown, RouteLint does not indicate whether the merge will result in a restrictive signal for the following train. In such situations, the available information may suggest a clear route, while subsequent route setting and train interactions may still lead to an unplanned yellow signal. For all other causes, no statistically significant differences remain after BH correction. In particular, train-following scenarios tend to be associated with lower unplanned yellow signal rates for RouteLint-equipped trains, but these differences are not statistically conclusive and are not interpreted further.

F. TEMPORAL SENSITIVITY ANALYSIS

The gradual uptake of RouteLint may contribute to temporal and operational confounding over the observation period. We therefore repeated the statistical comparisons for June 2025, when both RouteLint-equipped and non-equipped trains were sufficiently represented. The three national passenger rail strike dates in June were excluded, leaving 652 RouteLint-equipped and 1,125 non-equipped freight train runs, corresponding to a RouteLint share of 36.7%. At the macroscopic level, RouteLint-equipped trains show lower median absolute timetable deviations at all 15 eastbound timetable points, with 10 differences remaining statistically significant after BH correction. In the westbound direction, lower median deviations are observed at 10 of the 15 timetable points, although none remain statistically significant after correction. At the microscopic level, six block sections remain statistically significant after BH correction, comprising four eastbound and two westbound sections around Vga and Btl. For these six block sections, RouteLint-equipped trains show lower median absolute running time deviations, with differences ranging from 1.0 to 4.0 s. For unplanned yellow signal aspects, the June comparison is not statistically significant ($\chi^2 = 1.49$, $p = 0.223$, odds ratio $= 1.11$), and no significant differences by preceding train type or assigned cause remain after BH correction. Thus, the main association with lower timetable deviations persists in the June sensitivity analysis, while the microscopic and unplanned yellow signal results show more variation across the two observation periods.

As an additional descriptive check, we compared the observed characteristics of RouteLint-equipped and non-equipped trains. RouteLint shares are similar across train categories, with 35% for heavy trains and 37% for regular trains, and are 37% in both directions. RouteLint usage varies across hours of the day, from 24.0% at 02:00 and 19:00 to 63.8% at 05:00, compared with an overall share of 36.7%. The realised path analysis shows RouteLint shares of approximately 36% to 40% on many commonly traversed

TABLE 7. Unplanned yellow signal aspects by preceding train type, with an asterisk indicating significance after BH correction.

Preceding train type	n_{with}	n_{without}	$\text{Rate}_{\text{with}}$ (%)	$\text{Rate}_{\text{without}}$ (%)	ΔRate (percentage points)	Odds ratio	p_{BH}
Sprinter	466	2849	1.01	1.19	-0.18	0.85	0.007*
Freight	281	1550	1.24	1.48	-0.24	0.84	0.023*
Intercity	692	2862	1.58	1.51	0.07	1.05	0.748
Intercity Express	8	26	2.68	1.78	0.89	1.52	0.748
Empty locomotive	16	77	1.98	1.66	0.33	1.20	0.849
Maintenance train	6	35	2.62	2.20	0.42	1.19	0.952
Empty rolling stock	114	575	2.81	2.78	0.03	1.01	0.952

block sections, whereas some infrequently used alternative branches show larger deviations from this share. The clearest difference occurs across operators, where RouteLint uptake varies substantially, with some operators showing little or no use, while others exceed 80%. These descriptive results hence support the empirical comparison across several observed characteristics, while the noted differences in operator uptake and some local routing patterns should still be considered when interpreting the observed performance differences.

VI. DISCUSSION

The results show lower deviations from the planned timetable among RouteLint-equipped trains. For unplanned yellow signals, the aggregate association is modest, while the differences that remain statistically significant after BH correction are concentrated in two preceding train types and one operational cause. Although the magnitude of the timetable deviation differences varies across locations, these differences are consistently observed at the macroscopic level and are therefore relevant from a deployment perspective. In dense mixed-traffic corridors, performance degradation typically arises from the accumulation of small running time deviations and traffic interactions, particularly around major stations and junctions with complex routing. In such situations, the information provided by RouteLint may support drivers in adjusting their behaviour earlier in response to downstream operational conditions.

Directional patterns provide further insight into aggregation effects. More statistically significant differences are observed westbound at the macroscopic level, whereas the largest median deviations occur eastbound, and several macroscopic differences are not reflected at the microscopic level. Macroscopic timetable deviation captures the cumulative outcome of multiple interactions across successive block sections, thereby decreasing random variation and allowing systematic differences to become visible. In contrast, microscopic running time deviations are measured at the level of individual track sections, where substantial local variability can mask small systematic effects. Running time deviation is measured relative to a reference representing nominal conflict-free operation on the same block section and therefore mainly reflects differences in train running behaviour. However, accumulated running times also affect subsequent train interactions, route setting and signal aspects, which may in turn influence running times on later block sections. The observed microscopic differences therefore cannot be attributed to RouteLint alone.

While a single graph could represent train movements and include both timetable point and signal events, the use of separate macroscopic and microscopic event graph representations follows from the distinct definitions of the considered KPIs and facilitates interpretation at two different levels. Specifically, timetable deviation is defined at timetable points based on differences between realised and planned event times, whereas running time deviation is defined on block sections relative to empirically constructed nominal conflict-free running times. These differences require separate graph representations and level-specific edge weight computations, and cannot be consistently represented and visualised within a single integrated graph. Thus, this separation allows both timetable-point-level delays and block-section-level running time variability to be analysed across the corridor, providing a more comprehensive assessment of the operational performance associated with RouteLint usage.

For railway infrastructure managers, the results provide empirical evidence relevant to decisions concerning wider DAS adoption by railway undertakings and the more consistent use of shared, cooperative information for traffic harmonisation. In the case study presented in this paper, the closure of the Betuweroute led to a substantial increase in freight traffic on the Brabantroute. During this period, the free trial and uptake of RouteLint provide evidence on the operational performance associated with advisory system use under increased demand. Infrastructure managers may also consider the approach introduced in this study as a tool for ongoing performance monitoring and system evaluation across different corridors and operating conditions. By fusing routinely collected operational data, including timetable data, track occupation and release records, and DAS usage information, it becomes possible to assess where and under which operating conditions advisory system use is associated with measurable performance differences, and to track performance changes following software updates or policy changes using empirical evidence rather than simulation alone.

With respect to railway undertakings, the findings provide evidence relevant to DAS investment and deployment decisions, particularly by identifying operating contexts in which measurable performance differences are observed. In these contexts, lower running time deviations and fewer unplanned yellow signals may support smoother freight train operation and reduce the need for unnecessary braking, with potential benefits for energy use and the reliability of mixed-traffic services. If a DAS is developed and deployed by a railway undertaking, regulatory constraints permit the provision of direct driving instructions, which may allow advisory systems to be more closely aligned with operator-specific driving practices and operational objectives.

For DAS developers, the results indicate that incorporating conflict monitoring and prediction from the traffic management system into advisory generation is a relevant direction for further development. Presenting raw situational information alone may be insufficient in operational contexts where merging or diverging movements can lead to restrictive signalling. Providing drivers with clearer indications of potential conflicts ahead would support earlier and more informed driving adjustments. Such functionality could be supported by concepts such as train path envelopes to align traffic management decisions with onboard train trajectory generation at a finer operational level [24], [25].

Nevertheless, we acknowledge several limitations of this study. First, each observation represents a unique train run combination, while 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 RouteLint-equipped or non-equipped group. A robustness check based on subsampling was therefore applied to assess the sensitivity of the results to potential dependence. However, this procedure does not remove dependence, and the statistical inference may still be affected. Second, the analysis could not account for individual differences in driving style or experience. More skilled or proactive drivers may operate more smoothly and adhere more closely to the timetable regardless of RouteLint usage, which may introduce unobserved heterogeneity into the comparison. Third, RouteLint adoption increased gradually during the study period and was not based on random assignment. The observed differences may therefore reflect both RouteLint usage and differences in group composition or operating conditions, such as operator composition and variations in realised routing. The sensitivity analysis for June 2025 shows that the main association with lower timetable deviations persists, while the microscopic and unplanned yellow signal results vary. The results are hence interpreted as performance differences associated with RouteLint usage rather than as effects attributable solely to RouteLint. Finally, from July 2025 onwards, an additional interface option allowed users to indicate whether they were actively driving or merely viewing the system, with approximately 85% confirming active driving after data processing. This information was only available for the final month of the analysis period. Earlier records may therefore include a small share of non-driving usage, such as system testing or passive viewing, which could lead to a slight underestimation of the observed differences between groups.

VII. CONCLUSION

This paper evaluated the operational performance associated with the use of a Connected Driver Advisory System (C-DAS), RouteLint, using approximately one year of real-world freight operations data from a dense mixed-traffic corridor in the Netherlands, comprising 23,573 train runs. By fusing timetable event records, track section occupation and release data, and DAS usage logs, a data-driven two-level approach based on macroscopic and microscopic event graph representations was developed to enable comparison between trains operated with and without C-DAS. Constructed from the same underlying corridor and observed train movements, the approach uses edge weights derived from aggregated nominal conflict-free running times for performance evaluation.

Using non-parametric statistical tests, the results show modest but consistent differences in operational performance associated with RouteLint usage across timetable adherence, running time deviation, and unplanned yellow signal aspects. At the macroscopic level, statistically significant differences in median timetable deviation are observed at multiple key locations, with RouteLint-equipped trains showing approximately 30–70 s lower deviations at major stations and junctions. At the microscopic level, running time deviations exhibit substantial local variability, while lower median deviations are observed for RouteLint-equipped trains on selected block sections, with the largest observed difference of 117 s in a siding area. Furthermore, the odds that an observed yellow signal aspect is unplanned are approximately 6% lower for RouteLint-equipped trains. Statistically significant differences are concentrated in specific operating contexts, with lower odds when following freight trains or regional passenger trains, but higher odds in cases involving a preceding passenger train in a merging movement. The results show that the observed performance differences depend strongly on the level of aggregation. While macroscopic indicators capture the cumulative moderation of deviations along the corridor, microscopic indicators are more sensitive to local routing variability, which can mask small systematic differences. The proposed approach provides a data-driven and evidence-based basis for quantifying and interpreting empirical C-DAS deployment results, and supports discussion among infrastructure managers, railway undertakings, and system developers.

Future research may build on these findings in several complementary directions. First, integrating richer conflict-related information from the traffic management system into C-DAS may further enhance driver support in operational contexts such as merging and diverging situations. Second, combining large-scale empirical analysis with targeted human-factors studies would enable deeper investigation of how advisory information is interpreted by drivers and how behavioural heterogeneity influences observed operational effects. Finally, the potential of C-DAS may be further explored through optimisation and simulation studies, which can be used to refine the advisory generation algorithm and decision support under controlled operating scenarios.

ACKNOWLEDGMENT

The authors would like to thank Tasos Bithas from Delft University of Technology for verifying the signal locations and their number, and Vincent Weeda and Maya Tielman from ProRail for providing the data used in this article.

REFERENCES

[1] EU-RAIL. (2022). *Europe's Rail Master Plan*. [Online]. Available: https://rail-research.europa.eu/wp-content/uploads/2022/03/EURAIL_Master-Plan.pdf

[2] Z. Wang, J. Aoun, C. Szymula, and N. Bešinović, "Promising solutions for railway operations to cope with future challenges—Tackling COVID and beyond," *J. Rail Transp. Planning Manage.*, vol. 28, Dec. 2023, Art. no. 100405.

[3] ProRail. (2025). *Timetable Redesign: Capacity Model 2027 Netherlands*. Jan. 24, 2025. [Online]. Available: https://www.prorail.nl/siteassets/homepage/toekomst/europa/documenten/ttr-capacity-model-timetable-2027_final.pdf

[4] S. L. Sogin, Y.-C. Lai, C. T. Dick, and C. P. L. Barkan, "Comparison of capacity of single- and double-track rail lines," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2374, no. 1, pp. 111–118, Jan. 2013.

[5] C.-W. Palmqvist, A. Lind, and V. Ahlqvist, "How and why freight trains deviate from the timetable: Evidence from Sweden," *IEEE Open J. Intell. Transp. Syst.*, vol. 3, pp. 210–221, 2022.

[6] M. Nold and F. Corman, "How will the railway look like in 2050? A survey of experts on technologies, challenges and opportunities for the railway system," *IEEE Open J. Intell. Transp. Syst.*, vol. 5, pp. 85–102, 2024.

[7] F. Rocco Di Torrepadula, S. Di Martino, N. Mazzocca, and P. Sannino, "A reference architecture for data-driven intelligent public transportation systems," *IEEE Open J. Intell. Transp. Syst.*, vol. 5, pp. 469–482, 2024.

[8] R. Frolow, L. Zhang, and V. Schwieger, "Precise train positioning with unscented Kalman filter and low-cost sensors," *IEEE Open J. Intell. Transp. Syst.*, vol. 7, pp. 304–312, 2026.

[9] J. T. Borecka, D. R. Sánchez, and F. Corman, "A simulation-based greedy search algorithm for real-time mitigation of power peaks in railway networks," *IEEE Open J. Intell. Transp. Syst.*, vol. 7, pp. 822–837, 2026.

[10] K. Panou, P. Tzieropoulos, and D. Emery, "Railway driver advice systems: Evaluation of methods, tools and systems," *J. Rail Transp. Planning Manage.*, vol. 3, no. 4, pp. 150–162, Nov. 2013.

[11] P. Wang, R. M. P. Goverde, and J. van Luipen, "A connected driver advisory system framework for merging freight trains," *Transp. Res. C, Emerg. Technol.*, vol. 105, pp. 203–221, Aug. 2019.

[12] N. Ghaviha, M. Bohlin, C. Holmberg, E. Dahlquist, R. Skoglund, and D. Jonasson, "A driver advisory system with dynamic losses for passenger electric multiple units," *Transp. Res. C, Emerg. Technol.*, vol. 85, pp. 111–130, Dec. 2017.

[13] Z. Li, L. Chen, C. Roberts, and N. Zhao, "Dynamic trajectory optimization design for railway driver advisory system," *IEEE Intell. Transp. Syst. Mag.*, vol. 10, no. 1, pp. 121–132, Spring. 2018.

[14] H. Dong, H. Zhu, and S. Gao, "An approach for energy-efficient and punctual train operation via driver advisory system," *IEEE Intell. Transp. Syst. Mag.*, vol. 10, no. 3, pp. 57–67, Fall. 2018.

[15] Z. Tian, N. Zhao, S. Hillmansen, C. Roberts, T. Dowens, and C. Kerr, "SmartDrive: Traction energy optimization and applications in rail systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 7, pp. 2764–2773, Jul. 2019.

[16] Z. Xiao, Q. Wang, P. Sun, Z. Zhao, Y. Rao, and X. Feng, "Real-time energy-efficient driver advisory system for high-speed trains," *IEEE Trans. Transport. Electrific.*, vol. 7, no. 4, pp. 3163–3172, Dec. 2021.

[17] V. J. Verstappen, E. N. Pikaar, and R. G. Zon, "Assessing the impact of driver advisory systems on train driver workload, attention allocation and safety performance," *Appl. Ergonom.*, vol. 100, Apr. 2022, Art. no. 103645.

[18] G. N. Schnücker, J. Salge, and L. Onnasch, "Exploring the influence of job characteristics on the adoption of driver advisory systems for energy-efficient driving: Insights from a longitudinal field study in rail operations," *Appl. Ergonom.*, vol. 129, Nov. 2025, Art. no. 104597.

[19] Z. Wang, E. Quaglietta, M. G. P. Bartholomeus, and R. M. P. Goverde, "Assessment of architectures for automatic train operation driving functions," *J. Rail Transp. Planning Manage.*, vol. 24, Dec. 2022, Art. no. 100352.

[20] MOTIONAL, *Requirements for the Deployment of TMS Linked With ATO/C-DAS*, Deliverable D15.1, FP1-MOTIONAL, Europe's Rail Joint Undertaking (EU-Rail), Brussels, Belgium, 2023.

[21] MOTIONAL, *TMS and ATO/C-DAS Timetable Test & Simulation Environment*, Deliverable D15.2, FP1-MOTIONAL, Europe's Rail Joint Undertaking (EU-Rail), Brussels, Belgium, 2024.

[22] S. Tschirner, B. Sandblad, and A. W. Andersson, "Solutions to the problem of inconsistent plans in railway traffic operation," *J. Rail Transp. Planning Manage.*, vol. 4, no. 4, pp. 87–97, Dec. 2014.

[23] E. Quaglietta et al., "The ON-TIME real-time railway traffic management framework: A proof-of-concept using a scalable standardised data communication architecture," *Transp. Res. C, Emerg. Technol.*, vol. 63, pp. 23–50, Feb. 2016.

[24] Z. Wang, E. Quaglietta, M. G. P. Bartholomeus, A. Cunillera, and R. M. P. Goverde, "Optimising timing points for effective automatic train operation," *Comput. Ind. Eng.*, vol. 206, Aug. 2025, Art. no. 111237.

[25] Z. Wang, E. Quaglietta, M. Bartholomeus, and R. M. P. Goverde, "Sensitivity of train path envelopes for automatic train operation," *IEEE Trans. Intell. Transp. Syst.*, vol. 26, no. 11, pp. 18921–18933, Nov. 2025.

[26] UIC, Data Exchange With Driver Advisory Systems (DAS) Following the Smart Communications For Efficient Rail Activities (SFERA) Protocol, Standard IRS-90940, 2020.

[27] R. Idrovo, P. Zabalegui, U. Diez de Ulzurrun, P. Ciaurriz, S. Arana, and J. Mendizabal, "Toward automatic train operation for GoA3/4: Novel architecture and applicability criteria of speed profile optimization strategies," *IEEE Trans. Intell. Transp. Syst.*, vol. 26, no. 11, pp. 18358–18384, Nov. 2025.

[28] G. M. Scheepmaker, R. M. P. Goverde, and L. G. Kroon, "Review of energy-efficient train control and timetabling," *Eur. J. Oper. Res.*, vol. 257, no. 2, pp. 355–376, Mar. 2017.

[29] A. Albrecht, P. Howlett, P. Pudney, X. Vu, and P. Zhou, "The key principles of optimal train control—Part 1: Formulation of the model, strategies of optimal type, evolutionary lines, location of optimal switching points," *Transp. Res. B, Methodol.*, vol. 94, pp. 482–508, Dec. 2016.

[30] A. Albrecht, P. Howlett, P. Pudney, X. Vu, and P. Zhou, "The key principles of optimal train control—Part 2: Existence of an optimal strategy, the local energy minimization principle, uniqueness, computational techniques," *Transp. Res. B, Methodol.*, vol. 94, pp. 509–538, Dec. 2016.

[31] R. M. P. Goverde, G. M. Scheepmaker, and P. Wang, "Pseudospectral optimal train control," *Eur. J. Oper. Res.*, vol. 292, no. 1, pp. 353–375, Jul. 2021.

[32] S. Su, Z. Tian, and R. M. P. Goverde, *Energy-Efficient Train Operation: A System Approach for Railway Networks*. Cham, Switzerland: Springer, 2023.

[33] S. A. Kusumastuti, T. H. J. Kolkman, J. C. Lo, and S. Borsci, "Charting the landscape of rail human factors and automation: A systematic scoping review," *Transp. Res. Interdiscipl. Perspect.*, vol. 30, Mar. 2025, Art. no. 101350.

[34] T. Albrecht and J. van Luipen, "What role can a driver information system play in railway conflicts?," *IFAC Proc. Volumes*, vol. 39, no. 12, pp. 251–256, Jan. 2006.

[35] A. Naweed, "Investigations into the skills of modern and traditional train driving," *Appl. Ergonom.*, vol. 45, no. 3, pp. 462–470, May 2014.

[36] D. R. Large, D. Golightly, and E. Taylor, "Train-driving simulator studies: Can novice drivers deliver the goods?," *Proc. Inst. Mech. Eng., F, J. Rail Rapid Transit*, vol. 231, no. 10, pp. 1186–1194, Nov. 2017.

[37] C. A. Nash, A. S. J. Smith, D. van de Velde, F. Mizutani, and S. Uranishi, "Structural reforms in the railways: Incentive misalignment and cost implications," *Res. Transp. Econ.*, vol. 48, pp. 16–23, Dec. 2014.

[38] L. Yang, T. Lidén, and P. Leander, "Achieving energy-efficiency and on-time performance with driver advisory systems," in *Proc. IEEE Int. Conf. Intell. Rail Transp.*, Aug. 2013, pp. 13–18.

[39] R. S. Luijt, M. P. F. van den Berge, H. Y. Willeboordse, and J. H. Hoogenraad, "5 years of Dutch eco-driving: Managing behavioural change," *Transp. Res. A, Policy Pract.*, vol. 98, pp. 46–63, Apr. 2017.

[40] A. Cunillera, H. H. Jonker, G. M. Scheepmaker, W. H. T. J. Bogers, and R. M. P. Goverde, "Coasting advice based on the analytical solutions of the train motion model," *J. Rail Transp. Planning Manage.*, vol. 28, Dec. 2023, Art. no. 100412.

[41] P. Howlett, P. Pudney, and A. Albrecht, "An optimal train journey with bounds on energy consumption during specified intermediate time intervals," *J. Rail Transp. Planning Manage.*, vol. 27, Sep. 2023, Art. no. 100391.

[42] R. M. P. Goverde and L. Meng, "Advanced monitoring and management information of railway operations," *J. Rail Transp. Planning Manage.*, vol. 1, no. 2, pp. 69–79, Dec. 2011.

[43] R. M. P. Goverde, F. Corman, and A. D'Ariano, "Railway line capacity consumption of different railway signalling systems under scheduled and disturbed conditions," *J. Rail Transp. Planning Manage.*, vol. 3, no. 3, pp. 78–94, Aug. 2013.

[44] D. V. Cuong, V. M. Ngo, P. Cappellari, and M. Roantree, "Analyzing shared bike usage through graph-based spatio-temporal modeling," *IEEE Open J. Intell. Transp. Syst.*, vol. 5, pp. 115–131, 2024.

[45] T. Schlechte, R. Borndörfer, B. Erol, T. Graffagnino, and E. Swarat, "Micro–macro transformation of railway networks," *J. Rail Transp. Planning Manage.*, vol. 1, no. 1, pp. 38–48, Nov. 2011.

[46] N. Bešinović, R. M. P. Goverde, E. Quaglietta, and R. Roberti, "An integrated micro–macro approach to robust railway timetabling," *Transp. Res. B, Methodol.*, vol. 87, pp. 14–32, May 2016.

[47] A. A. Hagberg, D. A. Schult, and P. J. Swart, "Exploring network structure, dynamics, and function using NetworkX," in *Proc. 7th Python Sci. Conf.*, Pasadena, CA, USA, 2008, pp. 11–15.

[48] C. W. Palmqvist and I. Kristoffersson, "A methodology for monitoring rail punctuality improvements," *IEEE Open J. Intell. Transp. Syst.*, vol. 3, pp. 388–396, 2022.

[49] P. Lapamonpinyo, S. Derrible, and F. Corman, "Real-time passenger train delay prediction using machine learning: A case study with amtrak passenger train routes," *IEEE Open J. Intell. Transp. Syst.*, vol. 3, pp. 539–550, 2022.

[50] R. M. P. Goverde et al., "A three-level framework for performance-based railway timetabling," *Transp. Res. C, Emerg. Technol.*, vol. 67, pp. 62–83, Jun. 2016.

[51] G. L. Nicholson, D. Kirkwood, C. Roberts, and F. Schmid, "Benchmarking and evaluation of railway operations performance," *J. Rail Transp. Planning Manage.*, vol. 5, no. 4, pp. 274–293, Dec. 2015.

[52] J. Goya, G. De Miguel, S. Arrizabalaga, L. Zamora-Cadenas, I. Adin, and J. Mendizabal, "Methodology and key performance indicators (KPIs) for railway on-board positioning systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 12, pp. 4035–4042, Dec. 2018.

[53] Y. Dodge, "Mann–Whitney test," in *The Concise Encyclopedia of Statistics*. Cham, Switzerland: Springer, 2008, pp. 327–329.

[54] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ, USA: Lawrence Erlbaum Associates, 1988.

[55] M. Sherman and E. Carlstein, "Replicate histograms," *J. Amer. Stat. Assoc.*, vol. 91, no. 434, pp. 566–576, Jun. 1996.

[56] Y. Dodge, "Chi–square test of independence," in *The Concise Encyclopedia of Statistics*. Cham, Switzerland: Springer, 2008, pp. 79–82.

[57] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *J. Roy. Stat. Soc. B, Stat. Methodol.*, vol. 57, no. 1, pp. 289–300, Jan. 1995.

JULIA H. MICHELS received the M.Sc. degree in transport infrastructure and logistics from Delft University of Technology, The Netherlands, in 2025. She conducted her graduation research in collaboration with Dutch railway infrastructure manager ProRail. Her research interests include transportation, train operations, optimization, and data-driven analysis.

ZIYULONG WANG received the M.Sc. degree in civil engineering and the Ph.D. degree in railway transport from TU Delft in 2020 and 2026, respectively. He is currently a Post-Doctoral Researcher with TU Delft, where his research focuses on improving the alignment between traffic management and train operation on mainline railway networks, in close collaboration with Dutch railway infrastructure manager ProRail. He has contributed to several railway research and industrial projects, including MOTIONAL, R2DATO, RAILS, and RouteLint. He is an Associate Member of the Institution of Railway Signal Engineering (IRSE) and a member of the International Association of Railway Operations Research (IAROR). His research interests include public transport, railway planning and operations, data analytics, and applied network science.

JAN ANNE ANNEMA received the M.Sc. degree in chemistry from Utrecht University, The Netherlands, in 1987, and the Ph.D. degree in transport policy evaluation from TU Delft, in 2008. He is currently an Associate Professor of transport policy with the Faculty of Technology, Policy and Management, TU Delft. He worked for various governmental institutions, mainly on ex post and ex ante transport policy evaluation, before joining TU Delft, in 2008. He is also the Scientific Director of TRAIL, the Research School for Transport, Infrastructure and Logistics, a Ph.D. network organization of six Dutch universities.

JELLE VAN LUIPEN received the M.Sc. degree in aerospace engineering from TU Delft, in 1996, and the Master of Business degree in rail systems from TU Delft. He developed the first version of RouteLint in close collaboration with train drivers and traffic controllers, and conducted both simulator-based and field testing to support its validation. He is currently a Program Manager with the Department of Innovation and Development, ProRail, where he focuses on digitalization in railway operations.

ROB M. P. GOVERDE (Member, IEEE) received the M.Sc. degree in mathematics from Utrecht University, The Netherlands, in 1993, and the Engineering Doctorate (Eng.D.) degree in mathematical modeling and decision support and the Ph.D. degree in railway transport from TU Delft, in 1996 and 2005, respectively. He is currently a Professor of railway traffic management and operations and the Director of the Digital Rail Traffic Laboratory, TU Delft. He has cooperated in many international research projects, including the EU projects ON-TIME, MOVINGRAIL, RAILS, PERFORMINGRAIL, SORTEDMOBILITY, and MOTIONAL. He has co-authored over 100 journal articles. His research interests include railway timetabling and traffic management, automatic train operation, and railway signaling. He is a fellow of the Institution of Railway Signal Engineering (IRSE). He is the President of the International Association of Railway Operations Research (IAROR). He served as an Associate Editor for IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS. He is an Associate Editor and the former Editor-in-Chief of *Journal of Rail Transport Planning and Management*.