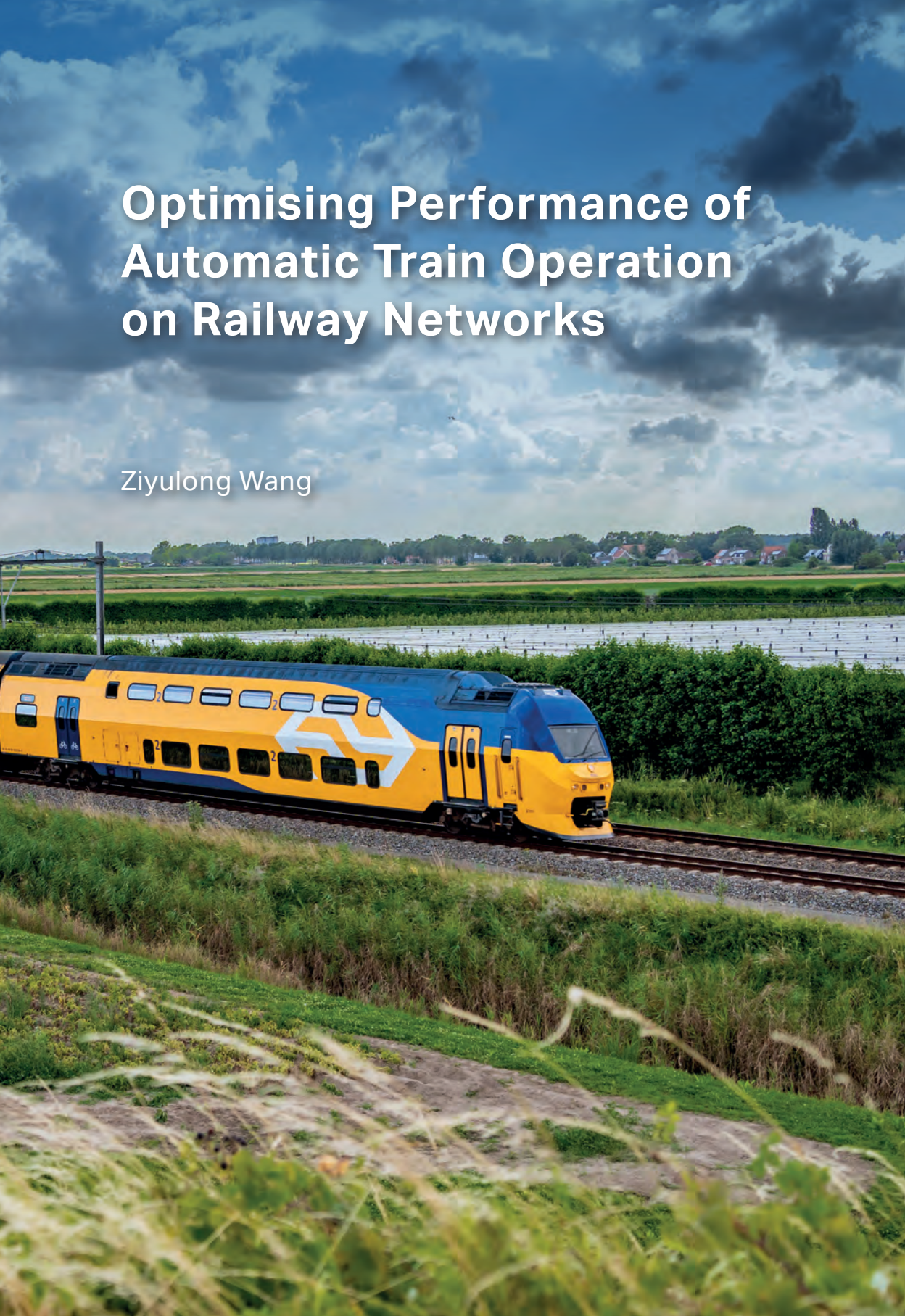
Optimising Performance of Automatic Train Operation on Railway Networks

Ziyulong Wang



Optimising Performance of Automatic Train Operation on Railway Networks

Ziyulong WANG

Optimising Performance of Automatic Train Operation on Railway Networks

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus, Prof. dr. ir. H. Bijl
chair of the Board for Doctorates
to be defended publicly on
Wednesday 11 March 2026 at 10:00

by

Ziyulong WANG

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus	Chairperson
Prof. dr. R.M.P. Goverde	Delft University of Technology, promotor
Dr. ing. E. Quaglietta	Delft University of Technology, copromotor

Independent members:

Prof. dr. ir. B. van Arem	Delft University of Technology
Prof. dr. O. Cats	Delft University of Technology
Prof. dr. ir. B.H.K. De Schutter	Delft University of Technology
Prof. dr. ing. A. Oetting	Technical University of Darmstadt, Germany
Dr. M. Joborn	Linköping University, Sweden

The research leading to this dissertation has received funding from the Dutch railway infrastructure manager ProRail. The Netherlands Research School for Transport, Infrastructure and Logistics TRAIL is greatly acknowledged.

Cover image: Ziteng Li

TRAIL Thesis Series no. T2026/6, The Netherlands Research School TRAIL

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

ISBN: 978-90-5584-382-4

Printed in the Netherlands

Bridging the track between research and practice

Preface

A PhD is often described as a journey. In my case, quite literally, a train journey. While my research focuses on optimising the performance of Automatic Train Operation on railway networks, my PhD life turned out to be an exercise in optimising my own performance within an extraordinary real-life network as well. Plans were made carefully, yet rarely executed exactly as scheduled. Disturbances occurred, priorities shifted, and occasional rescheduling became unavoidable. Over time, I learned that progress is not about adhering to a single exact plan, but about continuing to move forward despite disturbances and delivering results within acceptable time bounds that can still be regarded as punctual. Much like a long football season, success was determined less by individual matches than by consistency, grit, proper teamwork, and the ability to recover when things did not go according to plan. This dissertation marks the end of that trajectory, and it is also the result of the support, trust, and guidance of many people, whom I would like to thank here.

I would like to begin by thanking the Netherlands for offering an academic and societal environment in which openness, critical thinking, and collaboration across cultures are genuinely valued. I arrived here as a 22-year-old student to begin my Master's studies, and over the years, this country allowed me to study and work at the intersection of theory and practice, engaging with research questions that matter beyond academia. The strong emphasis on practicality has influenced me most. As Johan Cruijff famously said, "Voetbal is simpel, maar simpel spelen is het moeilijkst". In research, this translates to the challenge of turning elegant concepts into solutions that remain robust and workable in complex operational settings. I am grateful to this country for shaping me into a better researcher and a better person.

Rob, as my promotor, you have had, and continue to have, the single most significant impact on my academic development and future career. There is a saying that every obstacle can be overcome through rigour, and your consistency, integrity, and dedication exemplify this principle. You combined high expectations with genuine trust and intellectual freedom. Your guidance went far beyond supervising research, extending to training me to become an independent researcher who can develop and defend ideas rigorously. I thank you for the time you invested, the opportunities you created, and the confidence you placed in me throughout my PhD.

Egidio, as my copromotor, thank you for your guidance and care throughout this journey. Your presence and support were always thoughtful, and we shared many valuable experiences and moments of joy across both the Netherlands and China through the BJTU–TU Delft collaborative BSc programme. I still remember an Italian proverb you mentioned early on, "La gatta frettolosa fa i gattini ciechi", which taught me the value of patience and careful judgement. Over time, I came to appreciate your measured advice and the care you showed, especially at moments when direction mattered more than speed. Thank you for supporting me along the way and for lightening this journey.

I would like to thank the members of my PhD committee for accepting the invitation and for engaging with my dissertation. Bart van Arem, thank you for the discussions on automation

in road traffic versus railways, and for the many informal conversations over the years, often conducted in Dutch, which I greatly enjoyed. Oded Cats, you were one of my first stepping stones at TU Delft. You introduced me to the broader field of public transport research and helped me view research problems from a wider perspective. I am grateful for your guidance, openness, and continued collaboration, and I thank you for funding and supporting my participation in TRB 2024, allowing me to experience this remarkable conference before it changed. You have been a role model of intellectual depth, leadership, and hard work, while also showing how to combine a fulfilling professional career with a happy personal life. Bart De Schutter, thank you for contributing your expertise from the world of control theory. I greatly appreciate the time and effort you devoted to reviewing my work. Andreas Oetting, I still remember our interaction during the online RailBeijing presentation, where I presented work that later became Chapter 2 on ATO system architecture. Your insightful comments at the time left a lasting impression. Thank you for your engagement and interest in my work. Martin Joborn, it has been a great pleasure working together within the Europe's Rail project. Our discussions were always constructive, and I greatly admire your research on improving real-world railway system performance and on Connected Driver Advisory Systems. Your work consistently bridges theory and practice, which I value deeply.

I would like to thank the Dutch railway infrastructure manager, ProRail, for supporting this PhD and for enabling research that remains closely connected to real-world railway operations. Working with ProRail ensured that my research maintained a strong link to practice and real-world impact, and it also opened my eyes to the complexity of the operational railway system. I am grateful to Johanna Knijff, whose support in securing the funding for this project made this PhD possible. I sincerely appreciate her trust and commitment to enabling this research. I would also like to thank Maarten Bartholomeus for his dedicated collaboration and mentorship at ProRail. I am grateful for his expertise in signalling systems and ERTMS, which provided important technical grounding for this work. His decision to qualify as a train driver in order to combine ERTMS standardisation and research with first-hand operational experience is particularly noteworthy and reflects a strong commitment to bridging theory and practice. Among the many important people at ProRail, I would especially like to thank Jelle van Luipen. I am grateful to have met you at ProRail. You not only shared your railway knowledge, but also provided crucial support at a decisive moment by securing additional research funding, which allowed me to strengthen the connection between academia and practice. In doing so, you enabled me to work on research topics that we both genuinely enjoyed and considered meaningful. Beyond research, you taught me valuable life lessons and continue to do so, including how to navigate the complexities of professional and personal life with perspective and balance. I am thankful for having met you and many other inspiring people at ProRail, and for the numerous discussions that helped ground academic concepts in operational reality.

I thank Europe's Rail for providing a broader research platform and a collaborative environment that extended beyond national and disciplinary boundaries. Being involved in Europe's Rail projects allowed me to situate my work within a larger European vision for future railway systems and to learn from colleagues across institutions and countries.

A special thanks goes to my colleagues at the Digital Rail Traffic Lab and to my officemates in room 4.17. Thank you for the stimulating discussions, shared frustrations, moments of joy, and collective enthusiasm for railway operations research. This lab provided an environment in which ideas could be tested, challenged, and refined, and where support was always close by. I am deeply grateful for the sense of community and collegiality that made this journey both

productive and enjoyable. I especially cherish the many conversations over coffee and the feeling of coming into the office and feeling at home, with a warmth and familiarity that reminded me of my parents' living room.

I would also like to thank my colleagues at the Department of Transport & Planning for the open, welcoming, and international environment. To avoid the risk of omitting names, I refrain from listing individuals here. Instead, I would like to thank all of you who are reading this thesis. As a member of this department, I had the opportunity to interact with a broad range of expertise and to place my research within a wider academic and societal context.

My paranymphs, Renate and Renzo, have been particularly close friends throughout this PhD journey. Renate, you are always kind and thoughtful, and a reliable sounding board with a remarkable ability to listen. You and your husband are wonderful people who consistently show hospitality and kindness to those around you, and I am very grateful for the time we have spent together. Renzo, we shared many memorable and glorious moments, from our time in Washington D.C., where we enjoyed excellent food, including repeated visits to the same Chinese restaurant, and where you kindly helped me refill my water bottle countless times, to Tel Aviv, which I could not attend myself, and where you quite literally saved the day by presenting a paper you were not a co-author of and answering a challenging question with confidence and clarity. I am thankful for the friendship that you both brought into my life, and for the many restaurants, cuisines, and shared moments along the way.

I would like to thank my family, whose support extended far beyond work and anchored me in the world beyond academia. I can proudly say that my parents and my aunts are the best in the world. Your unconditional love, encouragement, and trust have been my constant foundation. Without you, I am nothing. Even when my research felt abstract or distant, your support was always present and tangible. Even when I was physically far away, I never felt that our love became strained or distant. Even when things were difficult or uncertain, you gave me the freedom to choose my own path and the confidence to keep going. I am deeply grateful for your belief in me throughout this journey.

I would also like to thank my partner, Xinyan Zhao, for your patience, understanding, and support. Our journey was not without difficulties, but through reflection and growth, we learned how to move forward together. Your quiet strength, kindness, and steadfast presence gave me the confidence to keep going. Your support made the challenges lighter and the successes more meaningful.

Finally, I would like to thank my friends. Your presence reminded me that there is life beyond deadlines and revisions. In particular, my friends from basketball and football, with whom I shared countless moments on the field. Fighting side by side, sweating together, and simply enjoying the game brought us closer and gave me energy when it was most needed. The conversations, distractions, and shared laughter mattered more than you may realise.

To borrow the language of a train announcement, this is the final stop of my PhD journey. Please change here for future journeys. Please mind the gap between research and practice. Thank you to everyone who travelled with me along the way.

Ziyulong Wang
Rijswijk, January 2026

Contents

Chapter 1

Introduction

1.1 Context and background

Railway transport serves as a fundamental pillar of national and international transport networks, providing a high-capacity, sustainable, and efficient mode of mobility. It plays an essential role in boosting economic growth, reducing congestion, and shaping long-term mobility planning. As urbanisation spurs and the demand for well-coordinated, high-performance railway operations grows, rail networks worldwide are under increasing pressure to modernise and improve operational efficiency (Wang, Z. et al., 2023). In response to these pressing issues, many regions have embraced digitalisation and automation as key enablers for optimising railway performance, improving service reliability, and alleviating environmental impact (ERA, 2024; Federal Railroad Administration, 2022; Lu et al., 2019).

A crucial aspect of this transformation is the advancement of the Traffic Management System (TMS) and Automatic Train Operation (ATO), two technological domains that support railway automation and real-time operational efficiency. Although a universally accepted definition of TMS in railways remains elusive, TMS is a central decision-making entity for managing real-time railway traffic, aiming to mitigate the propagation of service disturbances and disruptions while ensuring network-wide fluidity (Cacchiani et al., 2014; Corman & Meng, 2015; Tschirner et al., 2014). ATO, on the other hand, assists or automates train control functions, including speed control and stopping, to enhance service reliability, energy efficiency, and punctuality (IEC, 2014). Despite its successful implementation in metro systems, extending ATO to mixed-traffic mainline railways presents far greater complexities. Unlike metro networks, which operate in controlled environments with homogeneous traffic, mainline railways must accommodate a diverse operational landscape, including the coexistence of passenger and freight services, diverse infrastructure configurations, and the involvement of multiple stakeholders with competing priorities (Wang, Z. et al., 2022).

Further complicating this challenge is the fragmented operational condition, where legacy signalling systems differ across national networks, and rolling stock characteristics vary significantly. In Europe, efforts are underway to define a standardised interface between TMS and ATO to establish a harmonised and interoperable framework (ERA, 2018; European Commission, 2022). However, TMS is typically not fully automated and often relies on human operators for decision-making. Similarly, while successful trials are conducted in mainline operations, large-scale deployment of ATO at the corridor or network level remains limited. At the core of these challenges is the absence of a coordinated, effective, and adaptive approach that seamlessly

integrates TMS with ATO in such complex operational environments. To address this, regional and international organisations have increasingly prioritised research into structured frameworks that align train operation with network-wide coordination. A key initiative in this domain is the Europe's Rail Joint Undertaking, established by the European Union to advance railway digitalisation and automation while improving system resilience and optimising network performance (European Commission, 2022).

At present, railway traffic management relies on deterministic timetable (re)planning, where the TMS (re)schedules trains based on discrete target event times at scheduled timing points, i.e., departures, arrivals or passings at predefined discrete network locations within stations and junctions. This approach assumes strict adherence to planned time-distance train paths, allowing only limited flexibility for operational deviations. At the same time, most real-world traffic management remains rule-based and human-supervised, which limits its adaptability to real-time uncertainties to an even greater extent. While ATO automates train driving functions and mitigates human driving variations, it is inclined to favour energy-efficient driving strategies, including coasting or optimised cruising. However, such detailed ATO-driven train trajectory optimisation and execution may not align with traffic management models, which often assume fixed train speeds. These variations in train behaviour can be further influenced by real-time operational conditions, deepening the mismatch between scheduled plans and actual execution. This disconnect between timetable (re)scheduling and real-time train operation presents a fundamental challenge in achieving seamless interaction between TMS and ATO.

To bridge this misalignment, Train Path Envelopes (TPEs) are proposed to define a set of timing points along the train route, each associated with a time target or time window, allowing trains to adhere to the schedule while improving energy efficiency. These envelopes are computed based on the (updated) timetable provided by the TMS and forwarded to individual trains as train trajectory computation constraints for a designated corridor. In addition to scheduled timing points, TPEs may incorporate additional control timing points to account for tight headways or diverse driving strategies aimed at maintaining conflict-free operations. Rather than enforcing target passage times, TPEs provide a range of feasible trajectories with controlled flexibility, serving as a critical interface between TMS and ATO. They allow automated trains to dynamically adjust their speed while remaining compliant with the timetable as long as they operate within the defined envelope.

Nevertheless, the literature on TPEs remains scarce, with no established methodology for their computation. Existing studies have explored the use of TPEs for train trajectory optimisation (Wang, P. & Goverde, 2016, 2017, 2019), yet they do not explicitly propose a formalised method for determining them. In particular, the computation and optimisation of TPEs for successively scheduled trains within a railway corridor, their interdependencies, their adaptability to updated timetables and real-time train status feedback, as well as their role in synchronising traffic management with train operation, are not systematically addressed. Moreover, current standardisation endeavours and specifications, including those within the European Rail Traffic Management System (ERTMS) (ERA * UNISIG * EEIG ERTMS USERS GROUP, 2023b,c) and UIC Smart Communications for Efficient Rail Activities (SFERA) (UIC, 2020), have yet to specify a structured approach for TPE implementation. Instead, the responsibility for defining TPEs has been deliberately left open for railway infrastructure managers, who oversee capacity planning and traffic regulation, to develop at the trackside. Without clear guidelines for computing and integrating TPEs within the system architecture, their practical implementation remains unclear, posing challenges for realising harmonious TMS-ATO interaction. Furthermore,

key aspects such as the allocation of timing points, their responsiveness to evolving operational conditions, and the effective utilisation of timetable allowances (such as buffer times and running time supplements) lack investigation and answers. As a result, TPEs continue to be an underdefined yet critical component in advancing railway digitalisation and automation. To address this critical gap, this research proposes architectures and methodologies to harmonise the interface between TMS and ATO while analysing their potential impact on operational performance.

1.2 Research objective and research questions

Based on the identified research gap, the railway research community and industry require a structured approach to effectively aligning railway traffic management and ATO. TPEs have emerged as a promising interface to facilitate this alignment, enabling coordinated decision-making and seamless interaction between timetable (re)scheduling and speed-based train operation. This thesis aims to develop architectures, methodologies, and analytical approaches to improve the interaction between traffic management and ATO, ensuring consistency between (re-)planned schedules and real-time train trajectory generation. A central focus is the formulation of TPEs as constraints that guide train trajectories within a feasible time-distance range, allowing flexibility in train operation while maintaining adherence to the timetable. Additionally, this research investigates how event time flexibility can enhance network-wide railway traffic management by incorporating train rescheduling strategies and improving adaptability to real-time variations.

By addressing these aspects, this work contributes to both train control and traffic management by providing a framework that supports conflict-free operations, improves energy efficiency, and strengthens the robustness of railway scheduling and control. This framework bridges the gap between TMS and ATO while advancing both theoretical insights and practical railway applications. Methodologically, it develops methods to align ATO with traffic management, including the assessment of different ATO implementations, TPE computation and optimisation, and timetable rescheduling with flexibility. Practically, it examines the feasibility of these methods in achieving effective alignment while also proposing TPE generation as a functional module to support this goal.

To achieve this objective, the main research question is thus formulated as:

How can ATO be aligned effectively with the traffic plan to achieve conflict-free, energy-efficient, and flexible railway operations?

To answer this main question, the thesis is structured around the following key sub-research questions:

1. What are the functional and operational requirements for ATO to interact seamlessly with TMS? (Chapter 2)
2. How can ATO account for train driving variations while ensuring feasibility, energy efficiency, and punctuality? (Chapter 3)
3. How sensitive is ATO to real-time variations in train statuses and traffic plan? (Chapter 4)
4. How can event time flexibility enhance the alignment between network-wide rescheduling and ATO-enabled train operations? (Chapter 5)

1.3 Thesis contributions

This thesis advances research on TPEs as a key interface between railway traffic management and ATO, contributing to both the scientific community and society as described in the following subsections. It addresses the methodological, theoretical, and practical aspects of TPE computation, optimisation, and integration into automated and real-time railway operations.

1.3.1 Scientific contributions

The main scientific contributions of this thesis are structured into five key areas:

1. *Assessment of Automatic Train Operation architectural configurations*.
   Multiple ATO driving function architectures are proposed, analogous to the state-of-the-art Connected Driver Advisory System (C-DAS) in the literature. A systematic evaluation of these configurations identifies trade-offs in the distribution of intelligence between trackside and onboard subsystems. The findings contribute to the ongoing standardisation of ERTMS/ATO and provide insights into the impact of governance structures, signalling systems, and communication technologies on ATO implementation (Chapter 2).
2. *Formulation and optimisation of Train Path Envelopes*.
   A framework and models for computing TPEs are developed to provide trajectory generation constraints for ATO. The framework considers multiple punctual train driving strategies and computes the corresponding blocking times based on those to identify potential conflicts. Conflict resolution is performed in two stages: the first applies a linear programming model to derive conflict-free blocking time constraints, and the second proposes a critical-block strategy to optimise the placement of timing points along designated railway corridors. The resulting timing points and associated time windows form the TPE, which enables ATO to operate punctually, conflict-free, and energy-efficiently. This methodological development also provides managerial insights for railway stakeholders by presenting a TPE generation approach which can be positioned either at an active trackside ATO subsystem or at the TMS, and serves to establish a well-defined interface between traffic management and train operation (Chapter 3).
3. *Sensitivity analysis of Train Path Envelopes*.
   A sensitivity analysis using the elementary effects method is proposed to examine how TPEs respond to timetable updates and real-time variations in train operation, particularly the impact on timing point locations and their associated time windows. Timetable updates are represented as headway adjustments, while real-time variations in train operation are modelled through speed, time, and distance changes derived from ATO status reports. The findings provide insights into the robustness of TPE as train trajectory generation constraints and their role in aligning traffic management with train automation (Chapter 4).
4. *Incorporation of network-level event time flexibility in train timetable rescheduling*.
   A Mixed Integer Linear Programming (MILP) problem formulation is developed to incorporate event time flexibility as the network-level counterpart to TPEs in real-time railway traffic management. This model introduces time windows for departure, arrival, and passing events at critical timetable points, replacing conventional target event times to enhance adaptability. By structuring event time flexibility within a mathematical optimisation model with rescheduling strategies, the approach absorbs minor deviations in train movements without triggering immediate conflict resolution, thereby reducing computa-

tional complexity in real-time timetable rescheduling and strengthening alignment between network-wide rescheduling and ATO-enabled as well as human-based train operations (Chapter 5).

5. *Development and implementation of a Train Path Envelope generator*.
   A prototype generator is developed and implemented as a software tool to support the computation of TPEs. The tool aligns with the standardisation efforts outlined in Chapter 2 and incorporates Train Path Slot computation, timing point placement, and time window determination from Chapters 3-4. Reaching Technology Readiness Level (TRL) 4, the prototype is being integrated with the FRISO railway simulation tool to facilitate interaction between TMS and ATO. This software provides a foundation for future research in effectively aligning traffic management with train operation.

1.3.2 Societal contributions

This thesis contributes to the railway industry and society by addressing key challenges in integrating railway traffic management and train operation to enhance efficiency, sustainability, automation, and service quality. These contributions provide infrastructure managers, railway undertakings, policymakers, and passengers with societal benefits, including:

- Minimising unnecessary stops and delays to improve punctuality, reduce arrival and departure time uncertainty, and strengthen service reliability, network capacity utilisation, and passenger satisfaction.
- Incorporating event time flexibility into traffic rescheduling, reducing rigid (re)planning dependency and computational complexity, and making real-time railway operations more adaptive and efficient.
- Optimising train trajectories through energy-efficient driving strategies, lowering operational energy consumption, and contributing to sustainable railway transport and global sustainability efforts.
- Advancing automation to lower operational costs for railway undertakings, enhancing ATO execution, limiting human influences, and enabling workforce evolution through driver reskilling and expanded traffic management roles.
- Providing a scalable framework for better coordination between railway traffic management and train operation, supporting real-world deployment and standardisation efforts for railway automation and digitalisation.

1.4 Outline of the thesis

This thesis is organised into six chapters, with Chapters 2 to 5 each addressing a key aspect of TPEs and their role in integrating railway traffic management and train operation. Figure 1.1 provides a visual representation of the chapter interactions, while the following paragraphs outline their key contributions.

Chapter 2 examines the functional and operational role of ATO driving functions within existing system specifications. It proposes different ATO system architectures, assessing their implications for integrating ATO with traffic management. This chapter establishes the foundation for positioning TPEs as an interface between ATO and TMS.

Chapter 3 develops a methodological framework for defining and optimising TPEs as tra-

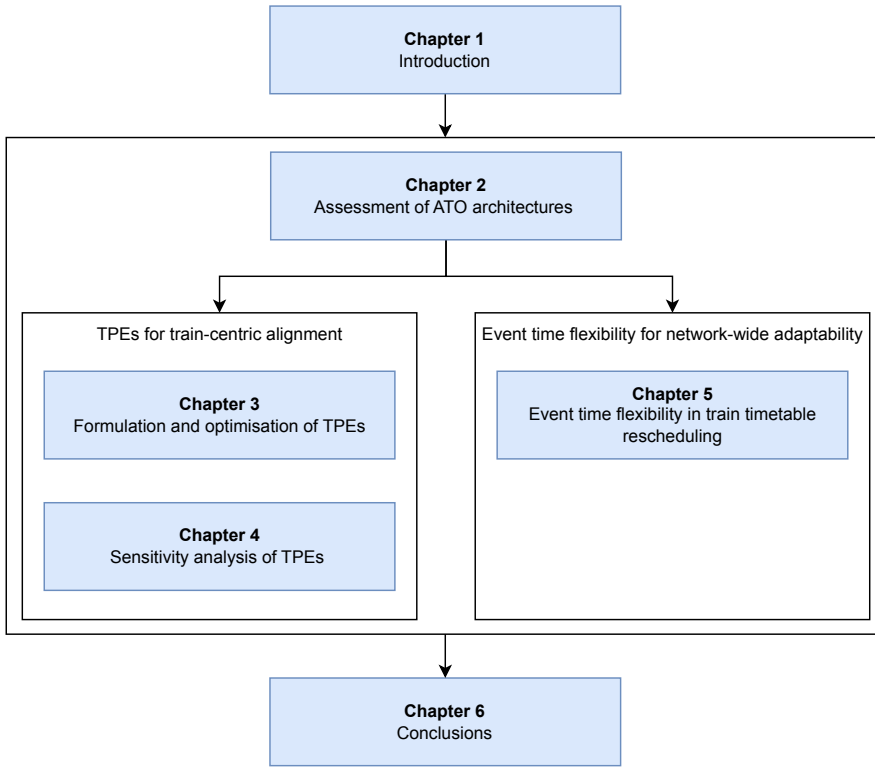


Figure 1.1: Overview of thesis structure to align traffic management and train operation

jectory generation constraints to ensure conflict-free and energy-efficient ATO execution. A two-stage optimisation-based approach is introduced to determine the placement of timing points and their associated time windows, providing controlled train driving flexibility by incorporating control timing points within the TPE while maintaining adherence to the timetable.

Chapter 4 investigates the sensitivity of TPEs to timetable updates and real-time operational variations. Through a sensitivity analysis using elementary effects, it quantifies how TPEs respond to dynamic railway conditions, providing insights into their robustness and adaptability.

Chapter 5 extends the concept of TPEs to real-time railway traffic management by introducing a Train Rescheduling with Flexibility model. In this model, event time flexibility serves as the network counterpart to TPEs, represented as time windows at key timetable points. This approach provides a TMS rescheduling algorithm with tangible and dynamic bounds for train paths, enabling it to absorb disturbances while ensuring conflict-free operations. The initial optimisation model was developed as part of a Master's thesis under the author's supervision and has since been revised, with a modular codebase developed as part of the refined methodology presented in this chapter.

Finally, Chapter 6 synthesises the key findings of the thesis, discusses its theoretical and practical contributions, and identifies avenues for future research.

Chapter 2

Assessment of architectures for automatic train operation driving functions

A successful integration of TPEs into railway traffic management and train automation requires a clear understanding of how ATO driving functions are defined and allocated within existing system architectures. This chapter investigates the functional and operational role of ATO in mainline railways, particularly within the European Railway Traffic Management System (ERTMS), and evaluates different system configurations based on state-of-the-art architectures developed for Connected Driver Advisory Systems (C-DAS). Three ATO driving function architectures are proposed, each based on a different allocation of TPE generation, trajectory computation, and brake or traction control across onboard and trackside intelligent components. A Strengths, Weaknesses, Opportunities, and Threats (SWOT) analysis is used to assess the trade-offs between these alternatives in relation to railway governance models and existing signalling and communication equipment. The findings provide a structured basis for positioning TPEs as an interface between ATO and TMS, supporting their alignment within future railway automation and digitalisation efforts.

Apart from minor changes, this chapter has been published as:

Wang, Z., Quaglietta, E., Bartholomeus, M.G.P., & Goverde, R.M.P. (2022). Assessment of architectures for Automatic Train Operation driving functions. *Journal of Rail Transport Planning & Management*, 24, 100352.

2.1 Introduction

Railway transport demand is increasing worldwide. For instance, the Dutch railway system expects a 30% to 40% growth in passenger demand in the coming 20 years in the Netherlands (Government of the Netherlands, 2019). Given the limited space and railway infrastructure, Automatic Train Operation (ATO) is considered one of the ways to utilise the current infrastructure more efficiently by automatically generating real-time train control commands (i.e., accelerating, cruising, coasting, and decelerating) according to a train trajectory that conforms to a conflict-free Real-Time Traffic Plan (RTTP) determined by the Traffic Management System (TMS) and therefore reducing train running variations. This will shorten the headways between trains and consequently lead to a more efficient timetable (Yin et al., 2017).

ATO is not a freshly minted topic in the railway domain and has been implemented in urban railways for more than 50 years. However, it is much more complicated to be realised in mainline railways in terms of heterogeneous traffic, various stop distances, complex track layouts, network size, open environment, multiple stakeholders, and multi-operator involvement.

Currently, Driver Advisory System (DAS) has already been more widely studied and analysed for mainline railways, and several best practices can be observed throughout the world (ON-TIME, 2013a; Luijt et al., 2017). It essentially supports drivers with speed advice based on the timetable and the current train delay. The Standalone Driver Advisory System (S-DAS) relies on a fixed timetable and is thus not accurate under disturbances. The next innovation has been to connect the DAS to the TMS so that timetable adjustments can be communicated to the DAS. This so-called Connected Driver Advisory System (C-DAS) can hence provide speed advice based on the actual timetable (Wang, P. et al., 2019). In this way, drivers can anticipate their driving behaviour and achieve a higher level of punctuality and lower energy consumption (Panou et al., 2013). ON-TIME (2013a); Panou et al. (2013) studied three types of C-DAS (DAS-Central, DAS-Intermediate, and DAS-Onboard) that differ in the allocation of intelligent functions to either the trackside C-DAS subsystem or the onboard C-DAS subsystem. Each of these three architectural alternatives has proven its usefulness and has been implemented in practice. The International Union of Railways (UIC) developed the Smart Communication For Efficient Rail Activities (SFERA) protocol for standardising the data exchange between TMS and various DAS systems (i.e., S-DAS and C-DAS) across different suppliers to allow interoperability (UIC, 2020).

The automation level of train operation is known as Grade of Automation (GoA), which allocates the responsibility for several basic functions to either onboard staff or automated functions, such as train operation, train speed control, train stopping, train door control and disruption management (IEC, 2014). This concept of GoA was developed for urban passenger railways, while some functions are not relevant for freight trains, for instance, the train door control. C-DAS is regarded as GoA 1, providing support to the manual driving process, while the only automation is the Automatic Train Protection (ATP) in the background. From GoA 2, the train runs automatically with increasing GoA from driver supervision (GoA 2), an attendant in the train who can take over in case of disruptions (GoA 3), and finally fully automated with possible remote control in case of disruptions (GoA 4, which is also known as Fully Automated Operation or Unattended Train Operation) (IEC, 2014; ERTMS Users Group, 2023b). In the remainder of this paper, we refer to ATO for any GoA from GoA 2, while we refer to C-DAS for GoA 1.

To support the interoperability and the rollout of ATO in Europe, the European Union Agency

for Railways (ERA) has been developing a set of technical specifications for ATO over the European Train Control System (ETCS) in mainline railways (ERTMS Users Group, 2023a). In these subsets of ATO-over-ETCS, a specific ATO architecture with the functions and information exchange between the trackside subsystem (ATO-TS) and the onboard subsystem (ATO-OB) has been defined, which will become a mandatory requirement for European railways. Although the Thameslink core in London - the only ATO-equipped mainline railway at present - has been operated in accordance with the ATO-over-ETCS architecture since 2018, the scientific argumentation supporting this architecture is not presented.

Both ATO and C-DAS aim to automate or aid human driving to improve operational efficiency, and the performance of these two relies on the train trajectory generation method and the control strategy. Therefore, C-DAS is a comparable entity to ATO, and C-DAS could be a transition toward ATO. The main difference is that the driver has to follow the speed advice with C-DAS, while in a higher GoA, the ATO-OB sends the brake and traction commands directly to the train. Furthermore, both technologies need the information provided by a TMS that has to be translated into constraints for the trajectory computation, the trajectory generation itself, and its translation into traction/braking commands.

To this end, this paper seeks to investigate and assess different possible ATO architecture configurations by allocating the intelligent components on the trackside or onboard. We take inspiration from the state-of-the-art analogous C-DAS architectures and replace the manual driving functions with corresponding ATO driving functions to establish a set of ATO architectures. Then, we perform an analysis of Strengths, Weaknesses, Opportunities and Threats (SWOT) to identify the advantages and disadvantages of different ATO architectures as well as the resulting limitations to the railway business. The assessment results spotlight that different ATO configurations might have diverse advantages or limitations, depending on the type of railway governance and the technological development of the existing railway signalling and communication equipment. Additionally, we also highlight the operational, technological, and business pros and cons of the ATO-over-ETCS architecture provided by the ERA, which is in line with one of the proposed architectures.

The main contributions of this paper are:

- An analysis of essential functional differences between C-DAS and ATO with respect to manual driving versus automated driving.
- A proposal and investigation of three different ATO architecture configurations based on the cutting-edge C-DAS architectures in the literature.
- A SWOT analysis of possible different ATO architecture configurations to identify their internal and external factors, as well as current and future potentials, in relation to different railway network types and organisation structures.
- A scientific argumentation for the choice of ATO-over-ETCS reference architecture, rendered by the ERA.

The structure of this paper is as follows: we introduce ATO in Section 2.2, along with the driving functions proposed in the ATO-over-ETCS specifications and the state-of-the-art C-DAS literature. Then, we propose the methodology in Section 2.3. Next, Section 2.4 analyses the ATO driving functions and proposes three design choices for ATO functional architectures. Afterwards, we perform the SWOT analysis on these alternatives in Section 2.5. Lastly, Section 2.6 concludes this paper.

2.2 Literature review

In this section, we first briefly introduce ATO in Section 2.2.1. Second, we review the ATO-over-ETCS system requirements specification in Section 2.2.2, followed by the description of the state-of-the-art analogous C-DAS architecture alternatives in Section 2.2.3.

2.2.1 Automatic Train Operation

ATO automates the train driving tasks towards supervisory or autonomous train control. It comprises a trackside subsystem (ATO-TS) and an onboard subsystem (ATO-OB) that are connected by wireless communication (ERA * UNISIG * EEIG ERTMS USERS GROUP, 2023c). The ATO-TS receives the target timetable and infrastructure data from the TMS. In general, the target timetable should be sufficiently robust to be realisable for different train characteristics. This timetable information must be converted into constraints and targets for the train trajectory computation. Based on the operational constraints and speed limits, a train trajectory is computed, specifying the speed profile and associated time-distance profile to the next stop or beyond (Yin et al., 2017). The generated speed profile is used as the reference trajectory to determine the traction or brake commands to the traction and braking systems.

ATO is a non-vital system, and therefore an ATP system is required to supervise the speed and braking curves within a safety envelope, taking into account the Movement Authority (MA) and corresponding track speed limits, e.g., ETCS (ERA * UNISIG * EEIG ERTMS USERS GROUP, 2023a). The TMS also triggers the route setting according to the timetable and the train positions by, for instance, Automatic Route Setting (ARS) (Quaglietta et al., 2016). The actual route is set by the interlocking system, which then reserves the route for the specific train. In the case of ETCS Level 2, the Radio Block Centre (RBC) gets the reserved route information from the interlocking, generates the new MA and sends it to the onboard ETCS. The onboard ETCS extends its dynamic speed profile accordingly until the end of the MA. Thus, it is essential that the train respects the trajectory constraints so that the route setting and train movements are aligned. For that reason, the TMS will plan the time targets or constraints at Timing Points (TPs) carefully, such that train path conflicts are avoided. Specifically, a TP is a location identified in the schedule of a train where a specific time is identified, and this time may be an arrival time, departure time or, in the case of a train not scheduled to stop at that location, the passing time (ERTMS Users Group, 2023b).

2.2.2 ATO-over-ETCS reference architecture

In Europe, a set of ATO-over-ETCS system requirement specifications is being developed, see Figure 2.1. This set includes operational functions, such as speed control, accurate stopping, door operation, and other functionalities that are traditionally the duties of drivers (ERA * UNISIG * EEIG ERTMS USERS GROUP, 2023b).

In this ATO-over-ETCS operational concept, the TMS forwards the infrastructure and timetable information about the route, the targets at TPs (e.g., arrival time, departure time, and minimum dwell time), temporary speed restrictions, and low adhesion (if applicable) to the ATO-TS (ERA * UNISIG * EEIG ERTMS USERS GROUP, 2014). Alternatively, an Infrastructure Manager (IM) may have an integrated TMS and ATO-TS with a similar function division, such that the interface between TMS and ATO-TS is no longer needed. This spatiotemporal information, related to infrastructure and timetable, is transformed into a list of Segment Profiles

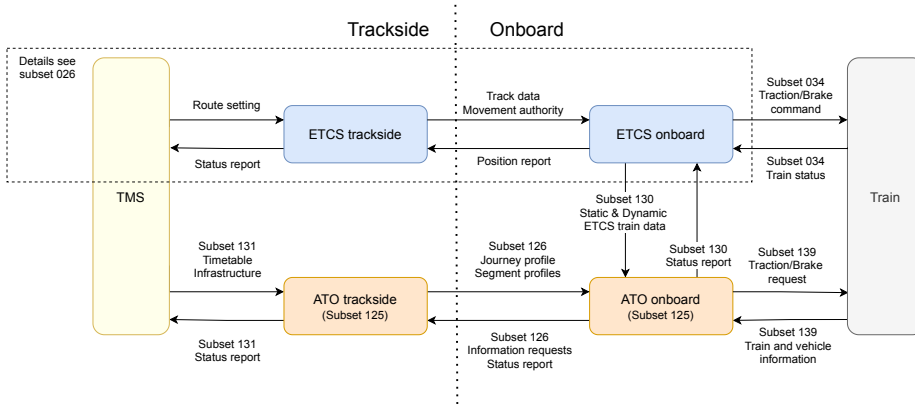


Figure 2.1: ATO-over-ETCS reference architecture

(SPs) and a Journey Profile (JP) at the ATO-TS and then forwarded to empower the ATO-OB driving functions (ERA * UNISIG * EEIG ERTMS USERS GROUP, 2023b,c). The static SP carries the most up-to-date infrastructure details, such as the segment length, the static speed profile, gradient, and curve data. The dynamic JP encloses a list of SPs (route data), the TP constraints (e.g., stopping or passing point and acceptable time allowance to be earlier at the TP), and temporary constraints (e.g., additional speed restrictions and adhesion conditions), representing the current timetable. If the timetable is changed during a journey without re-routing, the timing point information in the JP should be updated correspondingly, while the SPs are maintained. If new routes are given in the rescheduled timetable, then a new JP with a new set of SPs must be provided.

At the heart of the ATO-over-ETCS reference architecture is the ATO-OB subsystem requirements specification with the ATO driving functions as shown in Figure 2.2 (ERA * UNISIG * EEIG ERTMS USERS GROUP, 2023b). The functional features of the ATO-OB driving function consist of four parts, including Timetable Speed Management (TTSM), Supervised Speed Envelope Management (SSEM), Automatic Train Stopping Management (ATSM), and ATO Traction/Brake Control. SSEM computes the maximum speed that the train can run without the intervention of ETCS. ATSM establishes the speed profile to stop the train accurately at the stopping points, and TTSM calculates the optimal speed to meet the arrival time at TPs in the most energy-efficient way. ATO Traction/Brake Control computes the output commands to control the train based on the speeds given by the previous three functions.

The ATO-over-ETCS system architecture indicates that the SP and the JP shall always be determined at the trackside, while the driving functions should be performed at the ATO-OB, similar to the DAS-Onboard architecture, see next subsection. Nonetheless, the scientific reasoning and justification behind this system architecture design are not given in the specification. Besides, the interaction among the ATO-OB functional features proposed by the specification is still ambiguous, as it intends to be a Railway Undertaking (RU) choice for EU countries with a vertically separated railway. Furthermore, research and implementations of C-DAS show that other design choices are also possible (ON-TIME, 2013a). Consequently, the freedom of choice left by the ATO-over-ETCS technical specifications is leading the railway sector worldwide towards the need to identify potential architectures, which could represent the best trade-off of costs and legal responsibilities/burdens with operational and business performances.

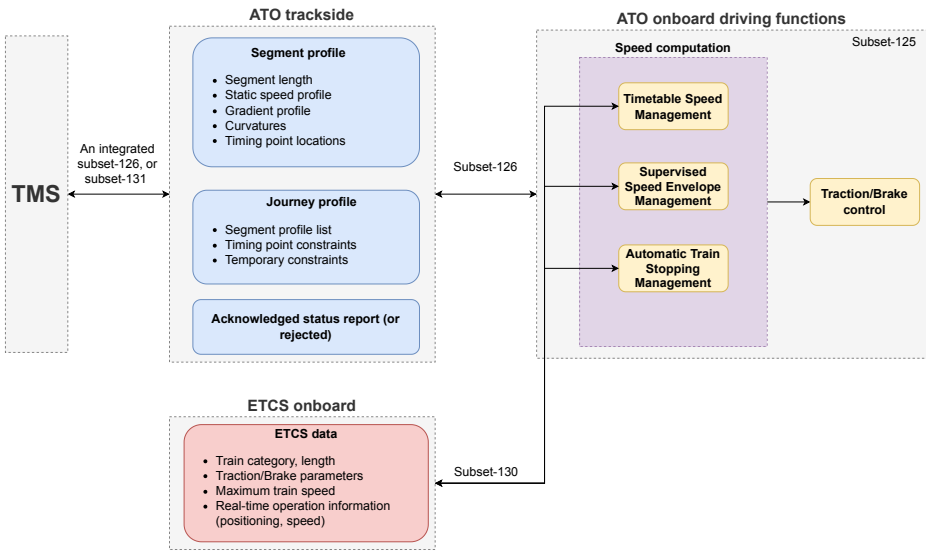


*Figure 2.2: Driving functions of ATO-over-ETCS, based on ERA * UNISIG * EEIG ERTMS USERS GROUP (2023b)*

2.2.3 C-DAS and C-DAS Architectures

Although ATO is less prevalent in mainline railways than in urban railways, C-DAS, as GoA 1, is at the forefront of assisted train operation, with various C-DAS architectures implemented worldwide. Our description of C-DAS follows the functional architecture as proposed by ON-TIME (ON-TIME, 2013a; Quaglietta et al., 2016) and the SFERA standard (UIC, 2020). This architecture is based on an RTTP computed by the TMS. An RTTP specifies routes to trains and the order and times of trains over track sections, which is the detailed output of Conflict Detection and Resolution (CDR) algorithms that keep an up-to-date conflict-free timetable (ON-TIME, 2013b). This RTTP is input to compute a Train Path Envelope (TPE) for each train. A TPE leverages the buffer time between two consecutive trains to build the blocking times on each track detection section and hence is able to identify time and/or speed windows as constraints for energy-efficient train trajectory computations without hindering the operation of neighbouring trains (Quaglietta et al., 2016).

Given the operational constraints of a TPE and the actual train parameters, the train trajectory generation computes an optimal train trajectory that exploits the running time supplements to minimise energy consumption. This train trajectory is then translated into speed advice and displayed to the driver. Typically, the speed advice concerns a sequence of the driving regimes, maximum acceleration to a target speed, maintaining a given cruising speed, coasting without traction, and service braking, with the exact sequence and switching points between regimes depending on the speed limit, gradient profile and the running time supplement. Finally, the driver complies with the given speed advice and controls the train.

Similar to ATO, when using a DAS system, an ATP system independently supervises the train speed with respect to any safety restrictions. Moreover, C-DAS also has a trackside subsystem (DAS-TS) and an onboard subsystem (DAS-OB) that are connected through a communication

channel. ON-TIME (2013a) distinguished three categories of C-DAS systems, depending on the distribution of the C-DAS intelligence between the DAS-TS and the DAS-OB, which is explained by Panou et al. (2013) and Rao et al. (2016). This categorisation is also used in the SFERA standard (UIC, 2020). The C-DAS implementations mainly differ in the communication required between DAS-TS and DAS-OB units:

- DAS-Central (DAS-C): The train trajectory and the corresponding speed advice are computed centrally in the DAS-TS. The only functionality of the DAS-OB is to display the advice. An example is seen in the Admirail/AF used in the Lötschberg base tunnel in Switzerland (ON-TIME, 2013a).
- DAS-Intermediate (DAS-I): Here, the computation is distributed into two parts. The train trajectory is calculated by the DAS-TS and then forwarded to the DAS-OB. The DAS-OB converts the train trajectory into driving advice corresponding to successive regimes and displays it to the driver. It is crucial to ensure the consistency of the parameters used for the computation (in the central unit) of the train trajectory and the reconstruction (in the onboard unit) of the regimes (ON-TIME, 2013a). An instance of implementation is the "Zuglaufregelung" (ZLR, train control) system tested by Deutsche Bahn (Deutsche Bahn, 2020).
- DAS-Onboard (DAS-O): Both the train trajectory computation and the speed advice generation are done in DAS-OB. The TMS only sends the TPE to the DAS-OB. One successful application is the Computer-Aided Train Operation (CATO) system in Sweden (Lagos, 2011).

2.3 Methodology

We develop a three-step method to identify the ATO driving functions, propose different ATO architectures and analyse them. As Figure 2.3 displays, we first show the essential ATO driving functions and the crucial transformations from C-DAS to ATO based on the literature review as a sound basis for the following three steps. Then, we provide a detailed definition of the state-of-the-art ATO driving functions, which is critical to pinpoint the functional intelligence involved in the system architecture. Next, these functions as modules are allocated to either the onboard or the trackside subsystem for establishing three different architecture configurations, representing various opportunities and adversities for the railway sector. Lastly, we perform a SWOT analysis for these three alternatives based on the criteria that were intrinsically presented by them and revealed in the literature to spotlight their advantages and disadvantages.

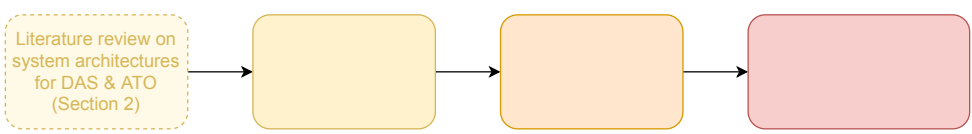


Figure 2.3: Outline of the proposed methodology for assessing different ATO architecture configurations

The first step is to unveil the essential ATO driving functions. We referred to the primary functions of the C-DAS system as ATO and C-DAS share noticeable similarities: 1) C-DAS and ATO are responsible for the train trajectory, which is computed following the goals of the

RUs, such as punctuality, energy consumption, and comfort; 2) They both utilise the (real-time) timetable and infrastructure information from a TMS via trackside subsystems. Yet, C-DAS and ATO also have significant differences. ON-TIME (2013a) identified four primary functions of C-DAS architecture, including TPE generation, train trajectory computation, driving advice derivation and display. Only the train trajectory computation remains consistent in ATO driving, while the other three functions are replaced due to either different interpretations of the input-output pairs or substantial changes from manual to automated driving. Initially, the way of speed tracking is altered. It is manually realised by a driver following the advice from the C-DAS, while ATO has a tracking algorithm inside the traction/brake control since the train operation is automated. Besides, the traction/brake mode in C-DAS is explicitly given to the driver in the form of driving advice. However, it is directly executed in an ATO system and displayed on a Driver Machine Interface (DMI) as supervisory information when appropriate. Besides, the TPE is replaced by a similar concept of JP and SP, although they both signify the spatial-temporal information that a train needs to compute a train trajectory. We conclude and exhibit the comparison between C-DAS and ATO driving functions in Table 2.1, which will be manifested in-depth in the next section.

Table 2.1: Comparison between C-DAS and ATO driving functions

	Spatiotemporal information from the trackside	Train trajectory optimisation	Train driving regime derivation	Propulsion control
Connected Driver Advisory System (C-DAS)	Train Path Envelope (TPE)	Same	Driving advice computation and display via Driver Machine Interface (DMI)	Manual
Automatic Train Operation (ATO)	Journey Profile (JP) and Segment Profile (SP)	Same	Traction or brake mode determination	Automated traction or brake control

Second, we design three ATO architecture alternatives based on the analogue C-DAS architectures as above-mentioned in the literature review and detailed in Panou et al. (2013); ON-TIME (2013a). These three ATO architecture alternatives correspond to their correlative C-DAS architecture options by a distinctive distribution of intelligence between the trackside and onboard subsystems, and thus a different need for information exchange. Regardless of the configuration, the fundamental modules in a structure are always the same: SP and JP construction, train trajectory computation, traction/brake mode determination and traction/brake control.

Third, we use the analysis and the alternatives from the first two steps to perform a SWOT analysis that determines business demands and barriers to these three distinctive ATO architectures. A SWOT analysis is a well-established method for assisting in the formulation of strategies over the past six decades (Learned et al., 1969). The strengths and weaknesses are identified as the internal environment, and the opportunities and threats are related to the external environment (Dyson, 2004). It has been widely applied for developing strategies for numerous emerging transportation topics, such as cooperative perception in self-driving cars (Caillot et al., 2022), multi-modal transportation development in ports (Vasheghani & Abtahi, 2023), market potentials of virtual coupling railway signalling (Aoun et al., 2020), and Mobility as a Service (MaaS) in rural areas (Eckhardt et al., 2018). We start performing the SWOT analysis by naming the strengths and weaknesses of each ATO architecture configuration, namely the intrinsic features that are related to the distribution of intelligence. We begin with an investigation of communication, including communication volume, frequency, and latency requirements. Then,

the impact of the assigned onboard intelligence on train operations and the onboard processing power is analysed. Next, we compare their abilities to cope with driving disturbances and traffic disruptions. Lastly, the prerequisites that the architecture necessitates are concluded. Concerning the environmental factors, opportunities or threats, we check the external factors associated with the developments in the railway industry and the political and legal requirements that affect the railway market, in particular, the influence on the duty division and collaboration between an IM and an RU. Different divided responsibilities would lead to distinct investments and regulations. Finally, we identify the suitable market for each ATO architecture configuration, together with the possibilities of migrating from the existing system architecture and future developments.

2.4 ATO driving functions and ATO architecture configurations

In this section, we first compare the driving function differences in ATO-over-ETCS with C-DAS, particularly JP and TPE (Section 2.4.1). Thereafter, Section 2.4.2 defines three ATO system architecture design choices.

2.4.1 From C-DAS to ATO driving functions

We provide a relation diagram of the different components in the ATO driving system architecture in Figure 2.4, particularly a schematic drawing of the energy-efficient train trajectory with its associated traction/brake command. The upper half of the illustrative diagram shows the speed profile of the energy-efficient train trajectory that comprises optimal cruising speed and coasting points for every train trip. It aims to minimise energy consumption with a given amount of running time between two stops. This train trajectory involves a particular sequence of four driving regimes, namely maximum acceleration, cruising (holding a certain speed with partial traction/brake force, depending on the resistance force), coasting (no traction/brake power), and service braking. The associated time-distance diagram is shown in the lower half of Figure 2.4, which corresponds to the RTTP, i.e., the timing at successive track and switch sections. Information about the route details is represented in the segment profile SP, while timing point TP and temporary constraints for the train trajectory computation are provided in the journey profile JP. The JP is similar to a train path envelope TPE and, in particular, provides time window constraints at (selected) strategic timing points to avoid conflicts with other trains.

From train path envelope to journey profile

Fundamentally, both C-DAS and ATO driving functions should always interact with the TMS through the trackside such that the intent of the TMS can be safeguarded, namely, conflict-free train paths. Both JP and TPE are derived from a route setting plan or RTTP, issued by the TMS. Within the ON-TIME concept, the RTTP replaces a route setting plan and is dynamically maintained by the TMS and strictly followed. Hence, the TMS replaces the local rules of current ARS systems to avoid inconsistent interferences between the TMS CDR function and an autonomous ARS. In other words, the ARS is replaced by a simpler ARS that just executes the RTTP, while the intelligence is moved to the CDR. In this way, both the route setting and the train trajectories have a common source in the RTTP. The specification of a TMS or its output is beyond the scope of ATO-over-ETCS but is essential for conflict-free train movements.

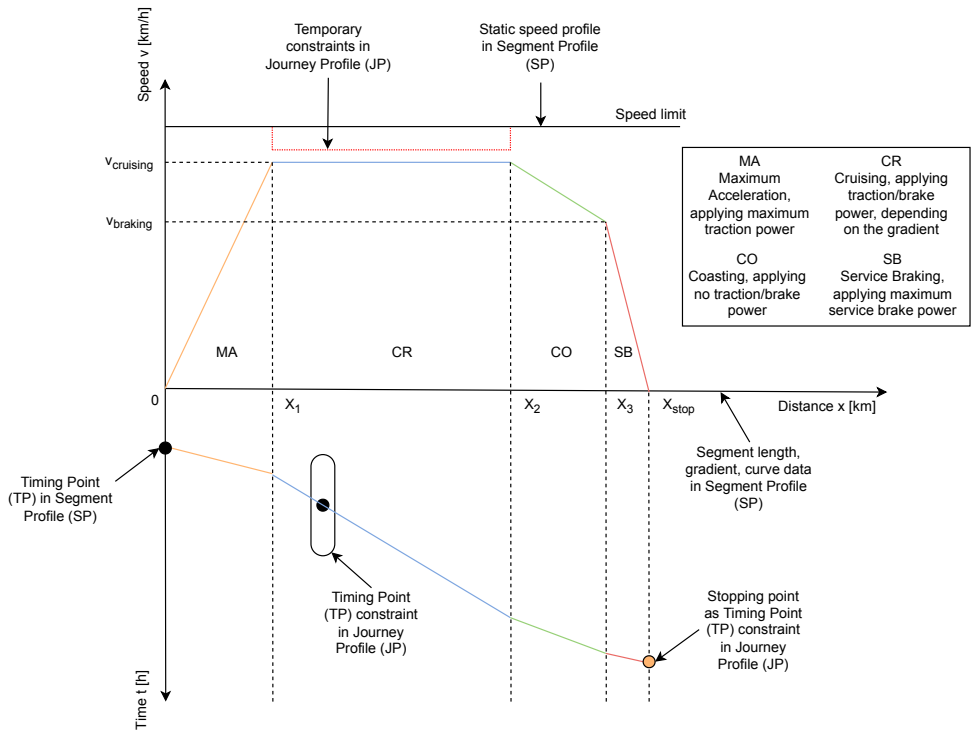


Figure 2.4: Schematic drawing for the relation of different components in the ATO driving system architecture

ON-TIME assumed that the TPE computation should be performed in the TMS. However, this could also be executed as a function of the DAS-TS unless the TMS integrates CDR and train trajectory optimisation as proposed in some recent literature (for example, Luan et al. (2018a)). Similar considerations hold for ATO-over-ETCS, where the JP is defined in the ATO-TS. In general, the interaction between TMS and ATO-TS depends on the algorithms used and the data needed.

The JP and TPE should be conflict-free to avoid operational train path conflicts that will disturb the train traffic with a negative impact on track occupation and schedule adherence. By definition, the JP contains a list of dynamic infrastructure data and operational data (TPs) required by the ATO-OB in order to drive the train, which may be updated during the journey, depending on the scheduled timetable and online traffic regulation (ERTMS Users Group, 2023b). Regarding the TPE, it identifies a sequence of time windows that facilitates train operations in an energy-efficient fashion without conflicting with the neighbouring trains based on the buffer times between them (Quaglietta et al., 2016). The locations and time constraints of TPs in ATO-over-ETCS can be defined flexibly in the JP. This allows the IM to optimise the train trajectories indirectly by either imposing a strict JP for RUs to operate the trains precisely or sending a comparably loose JP for RUs to optimise the train operations within it. These TPs are also used as information points where the train informs the trackside of estimated JP execution

based on the actual operation information. In the recent trajectory optimisation research with the TPE applied, the TPs could be suited at station stops, junction locations, signal positions, and route release points (Wang, P. & Goverde, 2016).

One can regard JP as the output from the TPE computation, similar to the time constraints from the TPE for the train trajectory generation. Moreover, the train trajectory computation is divided into three functional features in ATO-over-ETCS based on the operational constraints from the JP and safety constraints from ETCS, namely the SSEM, the ATSM, and the TTSM. As aforementioned, the SSEM sets speed limits, and the ATSM provides the local speed profile to the stop location. With the given speed limit and stopping point information, the TTSM connects to the local speed profile by providing the speed profile for the whole journey. From an implementation perspective, it is reasonable to decouple the trajectory computation into parts of stations and lines, where the station area optimises track occupation, and the line focuses on energy consumption. Nevertheless, the full train trajectory can be computed at once as explained in the energy-efficient train trajectory optimisation literature (Scheepmaker et al., 2017), particularly the ones which used the concept of a TPE (Wang, P. & Goverde, 2016,b). Besides, both TPE and JP emphasise the need to minimise energy consumption for train trajectory optimisation, meeting the operational and safety constraints. Accordingly, the timetable from the RTTP in ATO-over-ETCS should allow as much flexibility as possible, such that the RU is able to optimise train operation while avoiding conflicting train paths. Lastly, the TPE also allows the option of specifying a speed target or window at the TPs, which can be an extension of the JP functionalities. This is useful to specify, for instance, minimal speeds before uphill slopes or prevent creeping slow train movements in bottleneck areas (Wang, P. et al., 2019).

Therefore, the JP in the ATO-over-ETCS system requirements specification is almost the twin of a TPE applied in the train trajectory optimisation research. Whereas a TPE focuses more on train-centric driving flexibility, a JP offers an IM more functions/flexibility by defining the operational time-space based on the actual or dynamic train parameters and situation. Still, the separation and the task assignment between the IM and the RU for ATO would lead to different ways of developing a JP. In the SFERA standard, as Figure 2.5 shows, the DAS-TS can have two modules that are connected by the SFERA layer, namely one IM DAS-TS module (which can be integrated into the TMS) and one RU DAS-TS module (optional). The output from the TMS is sent to the DAS-TS of the IM first and then communicated through the SFERA layer to the DAS-TS of the RU or directly to the DAS-OB if the DAS-TS of the RU is not present. If the DAS-TS of the RU exists, the output from the DAS-TS of RU is transmitted to the DAS-OB, while the output from the DAS-OB feeds back to the DAS-TS on both sides via the SFERA. The interoperability and compatibility emphasised and strengthened by the SFERA standard shed some light on ATO system architectures, in particular, sufficient buffer times in an RTTP and the driving flexibility in the associated JP should be agreed upon in cooperation with the IM and the RU.

Consistent train trajectory generation

Subject to the information from the JP/TPE, an optimal train trajectory can be computed. The optimal train trajectory in both ATO and C-DAS systems should respect the intermediate time constraints with the objective of, for instance, maximising energy efficiency (between bottlenecks) or minimising track occupation (in bottleneck areas).

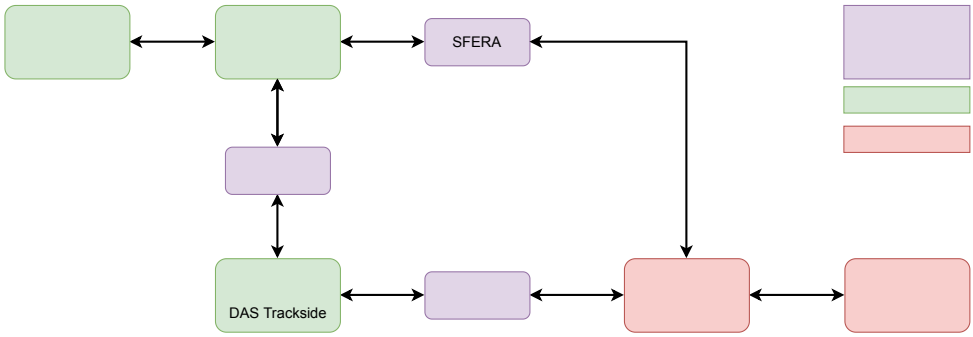


Figure 2.5: Illustrative diagram of the SFERA system architecture, adapted from UIC (2020)

From driving advice derivation to traction/brake mode determination

Normally, a train trajectory comprises a specific sequence of switching points among four traction/brake regimes: acceleration, cruising, coasting, and braking. The speed advice in C-DAS is based on the current and next regime. For ATO, the advice is replaced by information on the current regime and announcing the next. This can be used as supervisory information on the DMI to understand which regime ATO is operating in.

From driving advice display to ATO traction/brake control

In ATO, the traction/brake control determines the corresponding traction/brake command for each mode and sends it to the train traction and braking systems. This traction/brake control takes over the driver function from C-DAS. The advice display is replaced by supervisory information, while automatic driving takes over manual driving. The traction/brake control may be implemented in many ways. One way is regime-based, where the acceleration, coasting, and braking regimes are just defined as applying maximum traction, zero traction/braking, and service braking, respectively. The cruising regime is the most involved one that requires a speed tracking or cruise control algorithm in which the traction and brake commands need to counter variations in the train and line resistance due to, e.g., varying gradients, curves, and wind. Specifically, the cruise control constantly takes the distance and/or speed error as input and determines the traction/brake control to minimise this error. In this way, small deviations from the optimal speed profile will be corrected. If the deviation exceeds a defined error bandwidth, the train trajectory needs to be recalculated. In case the train is not able to be kept manoeuvring within the JP, the TMS needs to be warned and has to generate adjusted time targets for the ATO-TS.

2.4.2 ATO architecture configurations

Hinged on the generic modules that ATO and C-DAS share in common, we propose three functional design alternatives of ATO-TS and ATO-OB. The categorisation is distinguished by the distribution of intelligence shown in Figure 2.6:

- ATO-Central (ATO-C): Here, the ATO-OB is only responsible for the traction/brake control, which inherently includes the speed tracking based on a given traction/brake mode.

The construction of SP and JP, the optimal train trajectory, and the traction/brake mode are computed centrally at the ATO-TS.

- ATO-Intermediate (ATO-I): Here, the construction of SP and JP and the optimal train trajectory are performed at the ATO-TS and sent to the train-borne unit. The traction/brake mode is determined onboard based on the trackside trajectory. Thereafter, the traction/brake control decides the matching tracking commands.
- ATO-Onboard (ATO-O): Here, the ATO-TS only determines the SP and the JP. The ATO-OB is in charge of the optimal train trajectory computation, deciding the traction/brake mode and the corresponding traction/brake control.

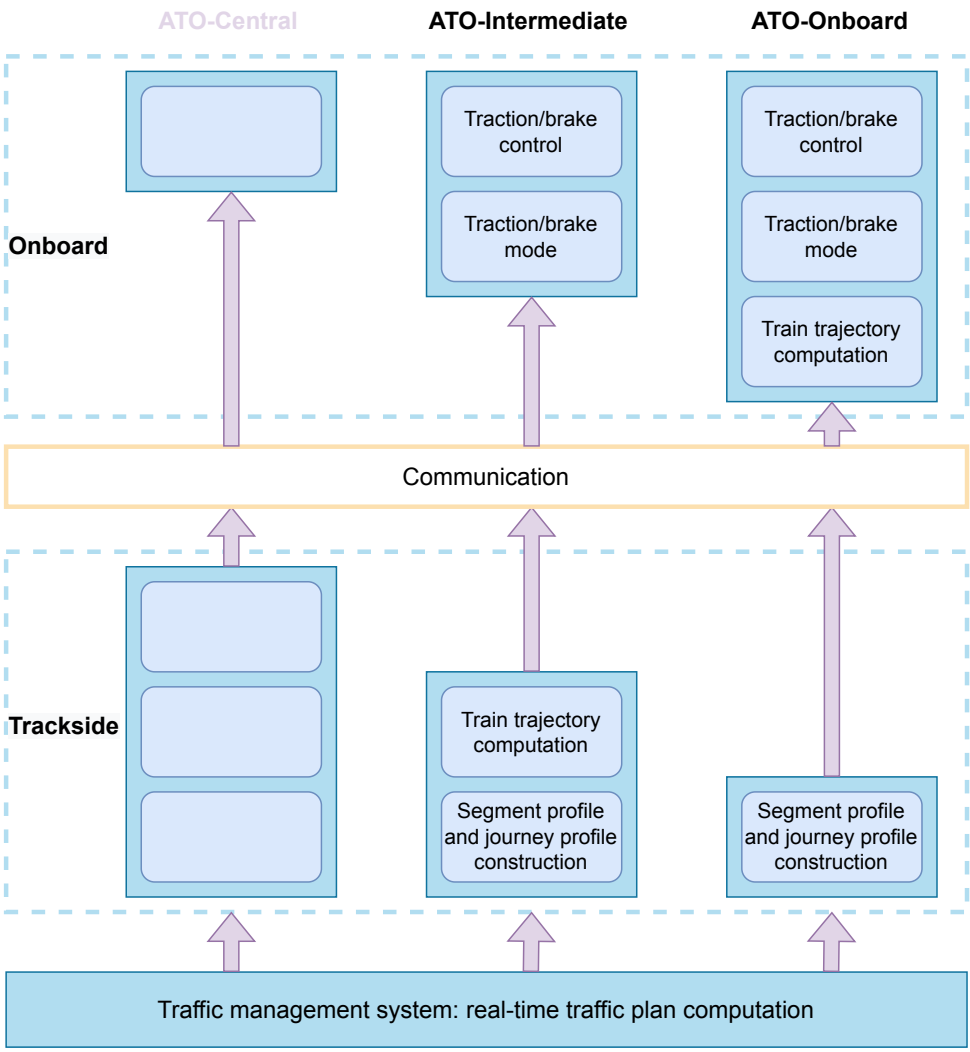


Figure 2.6: Distribution of functions for the ATO architecture alternatives

We exhibit the distribution of ATO intelligence with the input/output information for each design choice in Figure 2.7. In this picture, the existing modules remain the same for all three function divisions, such as the TMS, the onboard and trackside sensors, the vital system (i.e., ATP), and the train propulsion system. Instead, the division of ATO-TS and ATO-OB functions differs from each other. We highlight the ATO-TS of each alternative in a different background colour. The ATO-OB functions of each alternative are listed next to the highlighted trackside part. A detailed discussion follows, focusing on each design choice sequentially.

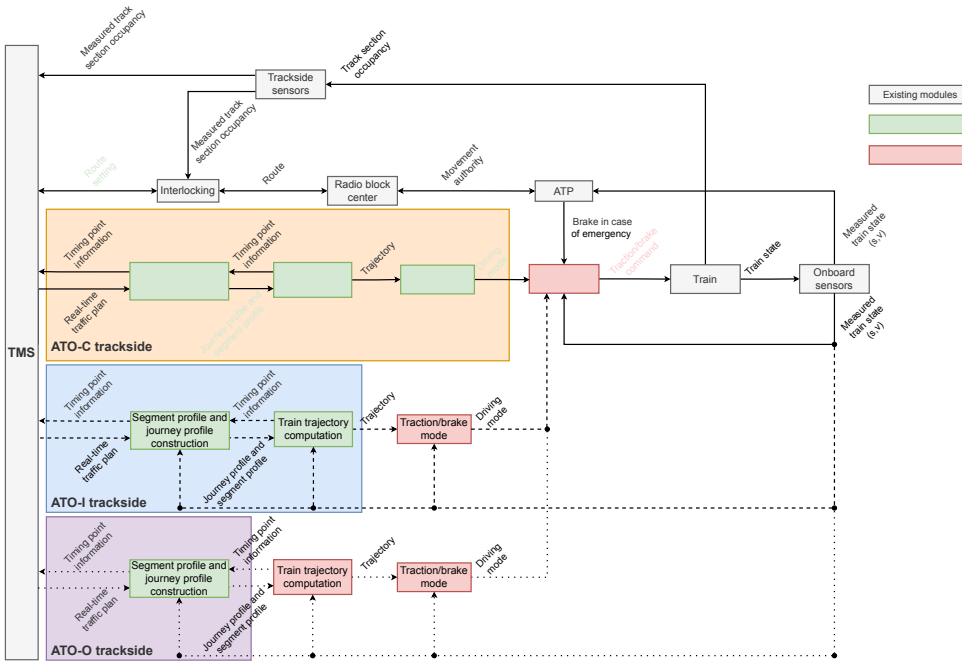


Figure 2.7: Distribution of intelligence for ATO architecture alternatives

ATO-C

In the ATO-C architecture alternative, the train trajectory and the traction/brake mode are computed at the ATO-TS and then sent to the ATO-OB. The ATO-OB receives traction/brake mode and its attributes from the ATO-TS. It also retrieves the current train position and speed from the onboard sensor. Collectively, the ATO-OB determines the traction/brake command to control the train accurately following the computed traction/brake mode.

In particular, the current traction/brake mode is provided along with the target speed. Also, supervisory information about the next mode may be provided with the expected switching point, depending on the train state and the train position of a set route. The traction/brake control translates the current traction/brake mode into traction/brake commands. Notably, the traction/brake commands aim at maintaining the reference cruising speed with respect to resistance forces and disturbances in the cruising regime.

Additionally, the ATO-TS continuously monitors the train driving deviation by comparing the position and speed reported from the ATO-OB with the computed trajectory. It recomputes the

train trajectory when the deviation is larger than the error bandwidth that the ATO traction/brake control allows. Furthermore, when the train is not able to stay within the original JP, the TMS needs to activate the CDR to generate a new RTTP. The updated RTTP will be forwarded to the ATO-TS. It can then be converted into a new JP and a train trajectory. Taking the current train state into account, the new traction/brake mode will be computed, which is later forwarded to the ATO-OB.

Besides, ATO-C allows for an entirely centralised feedback loop for guaranteeing consistent train model parameters during all stages, including RTTP computation, JP/SP construction, and trajectory computation. The onboard sensors obtain the measured train state and feed it back to the central logic, where the train model parameters are adapted, and consistency is ensured.

Communication between ATO-TS and ATO-OB is essential for this process and critical for driving deviation management. Thus, ATO-C highly relies on the quality of the data transmission, the accuracy of the train parameters, and the accuracy of the state measurements. This alternative corresponds to remote control from the trackside, while the train is just following orders as determined from the trackside.

ATO-I

In the ATO-I design choice, the ATO-TS is responsible for computing SP, JP, and the trajectory, based on the RTTP from the TMS. The computed train trajectory is communicated to the ATO-OB, where the traction/brake mode is determined. This mode and its attributes are derived from the reference train trajectory at the current position and time. The train traction/brake control will then decide the control command, according to the traction/brake mode. In the cruising mode, it adaptively responds to the error in speed or time for maintaining the cruising speed with respect to the reference cruising speed, depending on the cruise control algorithm. The trackside subsystem monitors the deviation from the train trajectory and computes a new trajectory if it could mitigate the deviation whilst still fitting in the JP. Otherwise, the TMS must compute a new JP to restart a cycle. In essence, the traction/brake mode translates the trajectory into a sequence of driving regimes as a function of position and time or speed.

For ATO-I, the traction/brake mode needs to be determined on board, including the switching points between the driving regimes. The onboard sensors continuously monitor the train states and send the information to the ATP system, the ATO-OB, and the ATO-TS. The ATO-OB part may include an online dynamic train model parameter calibration system for correcting the trajectory computation to guarantee the performance onboard (Cunillera et al., 2022). The updates of the train parameters should also be fed back to the trackside and shared with the TMS to ensure the consistency of the parameters through the entire loop, including the RTTP and JP/SP construction. In case of inconsistency or a delayed feedback loop to the central system, it may lead to many unnecessary recomputations of train trajectories at the trackside or even JPs by the TMS.

Thus, ATO-I is only applicable in a stable environment with predictable train dynamics and straightforward train trajectories that are easy to track. This alternative also corresponds to remote control from the trackside, with still some intelligence in the train to decide on the traction/braking mode corresponding to the given train trajectory.

ATO-O

Except for the construction of the SP and JP that are calculated at the ATO-TS, all the other three components are realised on board, namely the train trajectory computation, traction/brake mode computation, and traction/brake control.

The RTTP is converted into the SP and JP, which are then entirely communicated to the onboard unit. The ATO-OB algorithms for generating the train trajectory and traction/brake mode have to guarantee that the JP is respected. In case of too large deviations from the train trajectory, the trajectory can be recalculated onboard as long as the JP allows some flexibility. This design allows much shorter control loops that can react quickly to deviations without the need for communication with the trackside. Only when the time targets or TP constraints can not be realised, the trackside is informed that a new JP needs to be calculated, which possibly demands a new RTTP from the TMS.

Regarding the train parameter consistency, the onboard sensors are responsible for monitoring the train states and feeding them to the ATO-OB and the ATO-TS. Similar to ATO-I, an online dynamic train model parameter calibration system is recommended to be deployed for a more accurate real-time train trajectory computation/adjustment. If the monitored parameters are inconsistent with the pre-defined parameters, a timely adjustment is required to be fed back to the JP/SP computation and then RTTP construction in the central system.

This alternative thus corresponds to all intelligence onboard, while the trackside only provides the targets and constraints that guarantee conflict-free train movement.

2.5 SWOT analysis of ATO architecture configurations

Three ATO design choices have been defined in the previous section by allocating the ATO functional components at the trackside or onboard. We conduct a SWOT analysis of these three options in Tables 2.2 and 2.3. The analysis is carried out with a special focus on the following aspects: communication, onboard unit computation power, driving deviation reaction, responsibilities for RUs and IMs, interoperability and potential investment comparison of IM and RU, the typical situations that it might apply to, and the railway system structure that it might suit.

One of the significant distinctions among these three ATO design choices is the requirements of communication delay, i.e., the latency requirements in communication. ATO-C has the highest requirement for communication latency, which is its main weakness. The onboard unit of ATO-C is only in charge of performing traction/brake control, whereas the computations of SP and JP, train trajectory, and traction/brake mode are realised at the ATO-TS. The ATO-TS needs to frequently observe the train deviation, recompute the train trajectory if the deviation is larger than the error bandwidth, and forward the updated traction/brake information to the onboard unit. On the one hand, this results in high demand for the availability of communication channels because frequent contact between the ATO-TS and the ATO-OB is necessary. On the other hand, the communication volume is low as the information communicated only indicates a specific driving regime with its attributes and information about the switching point to the next regime. The small size of the exchange information packet accordingly becomes a strength of ATO-C. For ATO-I, the communication requirements are looser than ATO-C. However, it still needs to constantly monitor the train state and recompute the trajectory if needed. It sends the medium-sized information to the ATO-OB to derive the traction/brake mode from the computed train

Table 2.2: SWOT analysis of ATO architecture configurations (strengths and weaknesses)

Architecture	Strengths	Weaknesses
ATO-C	• Low communication volume • Low onboard processing power for executing traction/brake commands • Predictable train trajectory to IM	• High requirements for communication latency • Frequent communication between trackside and onboard • Limited reaction capability to driving deviations onboard • Trajectory determined with assumed train data • Stable and predictable environments
ATO-I	• Medium communication volume • Medium onboard processing power to determine the driving regime • Accurate traction/brake mode determination due to onboard train state estimation • Predictable train trajectory to IM	• High requirements for communication latency • Medium frequent communication between trackside and onboard • Limited reaction capability to driving deviations onboard • Trajectory determined with assumed train data • Stable and predictable environments
ATO-O	• Low requirements for communication latency • Low communication frequency between trackside and onboard • High reaction capability for driving deviations onboard • Accurate trajectory generation and traction/brake mode determination due to onboard train state and parameter estimation	• High communication volume due to JP and SP transmission • High onboard processing power for train trajectory generation • Train trajectory unknown to IM or requires communication

trajectory. Unlike ATO-C and ATO-I, ATO-O necessities the least communication frequency. It is unnecessary to constantly update the ATO-O unless a JP is changed partially or completely. But it needs a large volume of data (SP and JP) per update to support the onboard computation of train trajectory, traction/brake mode and control when desired.

The three design choices of ATO driving functions have various intelligence allocations, and therefore, they need different computation abilities at the onboard unit, which results in diverse advantages and disadvantages. From ATO-C to ATO-I and ATO-O, the computational power is growing as the complexity and the number of onboard modules are increasing. Correspondingly, it asks for more investments from the RUs. Furthermore, the trajectory optimisation of ATO-O may have to be simplified to allow the low processing power onboard, since the train trajectory generation takes place onboard, which is computationally expensive.

Different ATO design choices have dissimilar onboard and trackside control loops when reacting to deviations. Regardless of the design choices, all three options are able to fix the small driving deviations from the reference speed by utilising the control loop onboard, namely the traction/brake control. Additionally, when the train cannot be manoeuvred within the JP, each choice needs to redetermine the JP at the trackside based on an updated RTTP from the TMS,

Table 2.3: SWOT analysis of ATO architecture configurations (opportunities and threats)

Architecture	Opportunities	Threats
ATO-C	• High interoperability between IM and RU • Remote control of the train, such as shunting/stabling or very low-frequency lines for an integrated company without any other railway traffic • Vertical integration applicable • High central unit computation power for cloud platform possibility • Low investment for RUs • Central logic to adapt train model parameters throughout the entire feedback loop	• Control over train operations by IMs, as they are in charge of train trajectory generation and mode determination • Jeopardise free-access rights of RUs due to high dependency on IMs • Vertical separation inapplicable • High investment for IMs • High energy consumption responsibility for trackside
ATO-I	• High interoperability between IM and RU • Vertical integration applicable • Suitable for low complexity networks with homogeneous traffic, corridors with no intersections and shunting/stabling • High central unit computation power for cloud platform possibility • Freedom for RUs to choose supervisory information • Low investment for RUs • Medium responsibility in energy consumption for onboard	• Control over train operations by IMs as they are in charge of train trajectory generation • Jeopardise free-access rights of RUs due to high dependency on IMs • Vertical separation inapplicable • High investment for IMs • Medium responsibility in energy consumption for trackside
ATO-O	• Applicable to all market segments • Both vertical integration and separation are applicable • Freedom for RUs to optimise train trajectories • Freedom for RUs to choose supervisory information • Migration from C-DAS/SFERA possible • Low investment for IMs	• Divergent goals in railway operations of IMs and RUs • Various RUs may develop different trajectory generation solutions • High investment for RUs

which is the trackside control loop. However, the underlying difference lies in the computation location of the train trajectory and traction/brake mode, since the communication between the onboard loop and the trackside loop can be time-consuming. The communication process leads to longer latency and possibly unplanned stops. ATO-O with the train trajectory and traction/mode calculation all on board has the most efficient and powerful onboard control loop, which makes it the fastest deviation-responsive choice. Yet, the generated train trajectory is unknown to the IM, and thus, communication is required to provide this information to the trackside to close the loop with traffic planning. Instead, the train trajectories are predictable to IMs of ATO-C and ATO-I, while the reaction time or reactive capability onboard is a downside. For ATO-C, a deviation out of the manageable bandwidth has to be handled in the trackside control loop, including train

trajectory recalculation and traction/brake mode redetermination. Nevertheless, the trackside has better overall information and could generate the most suitable train trajectory instantly. However, it is the most inefficient one since it demands the trackside control loop considerably. The recalculation of trajectory and traction/brake mode can be frequent, as all kinds of small variances and disturbances need to be adjusted for (such as adhesion conditions if known, traction limitations, headwind, and delays at commercial stops during peak hours). ATO-I is slightly more flexible to smaller disturbances but still needs to rely on assumed train parameters to generate the train trajectories at the trackside, as ATO-C does. Nonetheless, ATO-C has the strength of a fully centralised feedback loop such that the central logic could adapt the train motion parameters throughout the entire feedback loop and ensure the same set of parameters is used in RTTP, JP/SP and train trajectory computation. ATO-O and ATO-I are advised to have an online parameter calibration model onboard where the trajectory is computed or corrected, and include a feedback loop to send the updated train parameters to the trackside when applicable. This train motion parameter feedback loop from the onboard to the trackside is missing in the current ATO-over-ETCS and SFERA standards.

Then, the incentive misalignments between the IM and the RU might lead to different levels of responsibility for RUs and IMs across these three ATO design options. There could be two scenarios in the ATO-TS management. The first scenario is that the ATO-TS is managed by an IM while the ATO-OB is operated by an RU. When both the train trajectory and the traction/brake mode are computed at the trackside, the ATO-OB loses the opportunity to optimise the train trajectory but only performs the traction/brake control. In this way, the costs for IMs are high since an investment in computation is needed. In addition, the responsibility of train operation also partly shifts to the IM side when the train trajectory and traction/brake mode are determined by the IM, although the actual traction/brake control is performed onboard. Most importantly, an IM has to guarantee that the TMS possesses real-time ATP information in order to compute a safe and conflict-free JP. This could bring in extra responsibilities and risks to IMs as they get involved in train operations. On the other hand, if RUs possess the ATO-TS, they can have the freedom to optimise the trajectory inside the JP. Besides, they can reduce their heavy investments in the onboard processors since the trajectory is computed at the trackside. Yet, it will still be costly to construct the trackside assets. Notably, when multiple RUs are involved in the same corridor, the initial investment division and the usage of such trackside equipment can be complex, which implicitly demands an independent party to optimise trajectories over multiple operators or trains by introducing a suitable capacity allocation mechanism and its associated charging principles and tariffs. It is also worth noticing that only when there is no legal and financial split between IMs and RUs, the ATO-TS can be managed by an RU. For ATO-I, the costs and responsibilities for IMs and RUs are similar to those of ATO-C in both scenarios. If ATO-O is selected with the ATO-TS managed by an IM, then the RU receives the basic information from the IM (i.e., JP and SP) and optimises the train trajectory generation and tracking onboard. To this end, it unavoidably brings computation complexity to the onboard unit. Accordingly, the IM has no responsibility for the train operation. When ATO-O trackside is possessed by an RU, the IM can only specify time targets at certain locations. Under this scenario, the RU has the largest freedom to optimise the train trajectory as long as it respects the timetable, but it will induce some extra expenses in the investment.

Regarding the cooperation and the shared responsibilities between IMs and RUs, it should be high for ATO-C and ATO-I since their objectives overlap and hence create an opportunity for collaboration between IMs and RUs. Simultaneously, the reliance of RUs on IMs would

pose a threat to the RUs whose free-access rights and own operations are impaired. Ideally, they should work together to achieve mutually beneficial outcomes. For instance, the train trajectory is calculated at the trackside, and therefore, the IM needs to consider the actual needs of the RUs to generate the reference speed profile. The other way around, the RU needs to closely follow the reference speed profile so that the goals of the IM can be achieved. In case of late braking, the division of responsibility between an RU and an IM also needs to be specified. Hence, bilateral or multilateral agreements between RUs and IMs are necessary to be met. ATO-O has a strict division between ATO-TS and ATO-OB, resulting in a goal and responsibility divergence between the IM and the RU. In particular, various RUs may develop different solution approaches for train trajectory generation, which might cause difficulties in traffic management.

Next, we list the typical conditions as prospective markets where each one of the ATO alternatives could be applied. Both ATO-C and ATO-I have weaknesses in responding to a train driving deviation, which indicates that they are only applicable in stable circumstances where train dynamics and train trajectories are easily predicted. ATO-I is more appropriate to be chosen when trains run in a low-complexity railway network with homogeneous traffic and no intersections with other railway network categories, such as urban railways, dedicated freight lines, or shunting/stabling areas. The shunting/stabling in this paper refers to signalled passenger train operations at yards, which means ATO needs to interact with an ATP, see for instance, Poulus et al. (2018). Even more limited, ATO-C represents the train's remote control since the ATO-OB only follows the order from the trackside. This suggests that it is only applicable for shunting/stabling or running at low speeds on very low-frequency lines without other types of traffic. In contrast, ATO-O generally can be implemented in every circumstance as it has the strongest onboard capability and the lowest communication requirements. It is, therefore, capable of dealing with disturbances, open environments, and mixed and heterogeneous traffic conditions as long as the trajectory can stay inside the JP.

If the railway has already implemented C-DAS, then it can be viewed as a transition technology between manual and ATO driving. Consequently, it is recommended to follow the SFERA standard and opt for ATO-O as SFERA referred to ERA * UNISIG * EEIG ERTMS USERS GROUP (2023c) as the core element for DAS operation and provided additions to it. For this reason, the migration from C-DAS to ATO-O would be a natural step. On the other hand, ATO-C and ATO-I have the power central unit of computing and thus make it easier to develop a cloud computation platform that contains both trackside and onboard intelligence. Accordingly, the system reliability will be enhanced with only sensors and input/output left at the trackside and the train.

Usually, RUs are responsible for the energy payment for railway operations based on either modelled consumption rates or actual metered usage. Since the train trajectory is derived onboard in ATO-O, the goal of sustainability and energy efficiency should remain the duty of RUs. However, the railway traffic becomes highly centralised in the ATO-C and ATO-I system architecture configuration because the onboard subsystem completely executes the train trajectory instructed by the trackside, and hence the contribution from the onboard intelligence is less significant compared to the trackside. Especially in ATO-C, the onboard subsystem simply follows the order from the trackside, which means the responsibility onboard is even more limited. As a result, IMs and RUs should negotiate to meet a bilateral agreement from this commercial aspect.

Lastly, we present the organisational structure that suits every design option the best. Vertical separation means that a country organises its railway into independent IMs and RUs, while

vertical integration means that a country does not separate the railway infrastructure and train operations into entirely different companies (Nash et al., 2014). ATO-C and ATO-I compute the train trajectory at the trackside and thus are in charge of the train operations if the ATO-TS is owned by an IM. This means that they are not suitable for a vertical separation rail business model since the legal and fiscal responsibilities cannot be split between an RU and an IM, particularly ATO-C. On the contrary, ATO-O gives the largest freedom to the ATO-OB such that it supports a vertical separation in the rail organisation. This support could attract more than one competitor to the rail market and contribute to both passenger and freight transport revenue (Huang, W. et al., 2019; Esposito et al., 2020).

2.6 Conclusions

In this paper, we developed a three-step method that first identified the critical functions in an ATO system architecture based on the cutting-edge analogous C-DAS developments in the literature, as well as the key transformations from C-DAS to ATO system architecture. Then, we presented three different ATO architecture alternatives. Lastly, these three design choices are evaluated by performing a SWOT analysis to facilitate the strategy formulation. The goal is to bring the limelight to the recent ATO-over-ETCS system requirements specification from the ERA operationally, technologically and commercially.

The most significant change from C-DAS to ATO is from manual driving to automated driving. The TPE is formalised in the JP of ATO-over-ETCS. The speed advice computation and display are replaced by traction/brake mode determination to feed the DMI with information about the specific driving mode and to feed the traction/brake control. Based on the three main C-DAS architecture alternatives, three functional design alternatives were proposed for ATO: Central, Intermediate and Onboard, which were compared on various criteria.

Our results revealed that ATO-C and ATO-I are suitable in countries with an integrated railway organisation where no legal and financial split exists between the IM and the RU. In contrast, the ATO-O architecture design is appropriate for fully separated organisations. This alternative corresponds to the ATO-over-ETCS system requirements, supporting the vertical separation in EU countries. In ATO-O, the RU has full control over energy consumption, given the constraints set by the trackside. However, the responsibility for the energy consumption becomes limited onboard when ATO-C and ATO-I are deployed, as the train trajectory is derived at the trackside. This means that the trackside should take more responsibility in this regard, and thus, the current energy bill splitting model should be re-examined when the trackside is managed by the IM.

For stable and predictable environments, ATO-C and ATO-I are relevant to be implemented since they are more cost-efficient. In particular, ATO-C corresponds to the train remote control and thus can only be selected at low-frequency lines with homogeneous traffic conditions or shunting/stabling where the onboard unit simply follows the trackside instructions. ATO-I can also be selected for urban railways or dedicated lines without other traffic. On the other hand, ATO-O with the most onboard functions generally can be applied in any circumstances as long as the organisational structure allows, e.g., mainline, high-speed, regional, and freight railways. Moreover, it also shows the ability to effectively react to disruptions by recalculating the train trajectory onboard as long as the JP allows.

Since countries may intend to deploy C-DAS as a transition technology between manual driving and ATO, we suggest that DAS-O is the best option to invest in, as it is in line with the

SFERA standard that uses the ATO-over-ETCS ATO-OB/ATO-TS Form Fit Functional Interface Specification (FFFIS) and thus would provide a no-regret policy regarding a later deployment of ATO-over-ETCS and ease the mixed-use of C-DAS and ATO. Further research could investigate quantitatively how these three different architectures influence capacity by reducing the human reaction time to the signalling system. Moreover, studies that verify the SWOT analysis based on the proposed ATO architecture configurations and their impact could be another research avenue.

Chapter 3

Optimising timing points for effective automatic train operation

The formulation of TPEs as train trajectory generation constraints is essential for enabling ATO-equipped trains to operate conflict-free and energy-efficiently while remaining compliant with a given timetable. This chapter develops a two-stage methodological approach for defining and optimising TPEs to address the mismatch between traffic management, which relies on discrete target event times, and train operation, which allows for multiple feasible driving strategies. In the first stage, a linear programming model derives conflict-free blocking time ranges across multiple feasible driving strategies. In the second stage, a critical-block strategy is proposed to place timing points at signals or stations where blocking time overlaps are most pronounced, and to define their associated time targets or windows. The proposed approach is verified through various controlled experiments and applied to a real-world Dutch railway corridor case study, demonstrating its effectiveness in enhancing network capacity, supporting energy-efficient train control, and adding flexibility in aligning traffic management with train operation.

Apart from minor changes, this chapter has been published as:

Wang, Z., Quaglietta, E., Bartholomeus, M.G.P., Cunillera, A., & Goverde, R.M.P. (2025). Optimising timing points for effective automatic train operation. *Computers & Industrial Engineering*, 206, 111237.

3.1 Introduction

The increasing demand for railway transport has led to a nearly saturated condition in many heavily used railway networks worldwide. Besides deploying advanced signalling systems, the railway sector proposes Automatic Train Operation (ATO) to address this issue (European Commission, 2022). ATO automates train driving partially or fully to reduce the impact of human factors on railway operations, enhancing punctuality, energy efficiency, and capacity (Yin et al., 2017). In particular, ATO can reduce the variations in train running times or dwell times. These processes are now heavily affected by human behaviour, e.g., the driving style of a human train driver or the passenger interactions at platforms when boarding/alighting the trains, respectively (Jansson et al., 2023). With ATO, train driving and dwell times become more predictable.

Hitherto, most literature has focused on developing a railway Traffic Management System (TMS) that generates a real-time traffic plan—defining train routes, reference locations and times, and possible rescheduling strategies—for train operation (Corman & Meng, 2015), or explores the integration of TMS and ATO (Rao et al., 2016). Recent technological developments partition ATO into onboard and trackside components with a standardised communication interface to exchange information dynamically to support or automate train driving (Wang, Z. et al., 2022). To align ATO with timetable-based traffic management, the real-time traffic plan from the TMS must be translated into a sequence of Timing Points (TPs) at successive track locations, with time targets or windows for each connected train to facilitate onboard trajectory generation and tracking. These TPs and their associated time windows or targets are assembled in a Train Path Envelope (TPE) as proposed in the literature (Wang, P. & Goverde, 2016, 2017; Quaglietta et al., 2016). The onboard ATO algorithm uses this TPE as constraints for train trajectory generation. State-of-the-art research mainly focused on developing approaches for punctual and energy-efficient ATO onboard algorithms (Yin et al., 2017; Huang, Y. et al., 2018), whereas a literature gap remains in the definition and modelling of the TPE and its associated TP location and time constraints. In particular, no approach has been developed to optimise the TPs in a TPE for successive trains. Nevertheless, the TP configurations are critical to instruct the onboard algorithms effectively to follow the real-time traffic plan without track occupation conflicts.

Typically, conflict-free timetabling models assume a deterministic time-distance train path (Caimi et al., 2011a; Goverde et al., 2016; Zhang et al., 2019; Wang, P. et al., 2022), as do timetable rescheduling approaches applied in a TMS (Corman et al., 2010; Pellegrini et al., 2014; Quaglietta et al., 2016; Li et al., 2023). In contrast, a train may operate with various train driving strategies that all satisfy scheduled departure and arrival times, such as following a scheduled cruising speed or applying energy-efficient train control (Scheepmaker et al., 2020). These variations lead to a range of train passing times, potentially causing conflicts between successive trains adopting different driving strategies. To this end, the main objective of TPE computation is to provide the ATO onboard with an optimal TP configuration, i.e., a set of TPs with corresponding time constraints, that respects scheduled station arrivals while allowing sufficient flexibility to accommodate uncertainties in train driving strategies and prevent conflicts between successive trains.

Building on this, we consider the following research question: How can TPEs be computed to ensure conflict-free operations of ATO-equipped trains? This question is further broken down into three sub-questions: How can we identify possible misalignment between the TMS real-time traffic plan and ATO speed profile generation to ensure feasibility and conflict-free operations?

How can the locations and time windows of TPs in a TPE be determined effectively while accounting for variability in ATO driving strategies? How does the proposed approach impact railway operations in a real-world case study?

We address these research questions by proposing an innovative TPE computation method that dynamically determines optimised TPs and associated time windows to enable punctual, conflict-free and energy-efficient train operation while allowing flexibility in the adopted ATO onboard driving strategies. Specifically, for each train, we compute a range of blocking times across successive blocks, corresponding to multiple train driving strategies that may be adopted during operation. Critical TPs and time bounds are derived to avoid blocking time conflicts between trains, achieving robustness against uncertainties in train driving strategies and maintaining compliance with the real-time traffic plan. We test and verify the developed model in various experiments and demonstrate its potential in a case study on one of the busiest Dutch railway corridors with heterogeneous traffic. Results from a real-life case study show that the critical-block TP configuration, which forms the TPE, outperforms alternative approaches, such as using only TPs at scheduled stops or successively adding TPs based on the first conflicting block in an iterative manner.

The contributions of our paper are summarised as follows:

- We introduce the concept of a train path envelope by considering multiple train driving strategies to manage uncertainties in train trajectory modelling and enhance alignment between traffic management and train operation.
- We develop a two-stage optimisation-based method to compute train path envelopes, incorporating a critical-block timing point determination approach to guide ATO-driven trains in generating conflict-free, energy-efficient and punctual train trajectories.
- The proposed method is validated on one of the busiest Dutch heterogeneous railway corridors, demonstrating its feasibility and practical applicability.

The proposed method can also be extended to other domains where automated processes must adhere to time targets or windows, ensuring time- and energy-efficient execution of a predefined timetable, such as in aviation, supply chain management, and road transport. The remainder of this paper is organised as follows: Section 3.2 offers an overview of the related works and entails the research gaps, followed by Section 3.3, where the methodological framework is presented. Then, Section 3.4 describes the case study and discusses the results. Section 3.5 provides managerial insights and examines the applicability and limitations of the proposed approach. Finally, Section 3.6 summarises the key findings and outlines future research directions.

3.2 Related works

Rail traffic optimisation can be generally classified into two groups (Rao et al., 2016): On the one hand, the TMS monitors and predicts the status of traffic and infrastructure, detects and resolves train path conflicts, and ultimately computes the real-time traffic plan for train operations. On the other hand, ATO aims to generate and track an optimal train trajectory to respect the real-time traffic plan and provide the resulting real-time traction and brake commands to control the train speed effectively.

Yet, traffic management and train operation have been considered separately for quite a long time in mainline railways (Rao et al., 2016). The concept of a TPE was proposed in the literature to bridge these two based on the input of a (re)scheduled traffic plan from the TMS. A TPE

defines a sequence of time targets or windows at TPs, ensuring conflict-free train operation while enabling train trajectory optimisation and maintaining schedule punctuality (Albrecht, T. et al., 2013; Quaglietta et al., 2016). So far, it has been successfully implemented in train trajectory optimisation (Wang, P. & Goverde, 2016; Wang, P. et al., 2020). Tables 3.1 and 3.2 provide an overview of existing studies on the interaction between traffic management and train operation at the operational planning stage, with a particular focus on the role of TPE in bridging these two domains. The TPE is part of a Journey Profile that is sent from the trackside to the train onboard systems within ATO and Connected Driver Advisory Systems (C-DAS) architectures (Wang, Z. et al., 2022). The ATO onboard then generates a train trajectory satisfying the constraints in the TPE to automatically operate the train.

In essence, a TPE defines time bounds at TPs to ensure sufficient separation between trains (Albrecht, A.R. et al., 2015; Haahr et al., 2017). We refer interested readers to the review paper by

Table 3.1: Overview of the relevant studies on the connections between traffic management and train operation at the operational planning stage (part 1)

Author(s)	**Main focus**	**Railway type**	**Objectives**	**Connections between traffic management and train operation**
Albrecht, T. et al., 2013	Train trajectory optimisation	Mainline	Delay minimisation; Energy efficiency	Target point; Target window
Albrecht, A.R. et al., 2015	Train trajectory optimisation	Mainline	Energy efficiency	Passage point
Quaglietta et al., 2016	Real-time traffic management	Mainline	Conflict detection and resolution; Energy efficiency	Train Path Envelope
Rao et al., 2016	Integrated traffic management and train operation	Mainline	Conflict detection and resolution; Energy efficiency	Main-target point; Sub-target point
Wang P. and Goverde, 2016	Train trajectory optimisation	Mainline	Delay minimisation; Energy efficiency	Train Path Envelope
Haahr et al., 2017	Train trajectory optimisation	Mainline	Energy efficiency	Passage point
Wang P. and Goverde, 2017	Train trajectory optimisation	Mainline	Delay minimisation; Energy efficiency	Timetable constraint set
Luan et al., 2018a	Integrated traffic management and train operation	Mainline	Delay minimisation	Train speed
Luan et al., 2018b	Integrated traffic management and train operation	Mainline	Delay minimisation; Energy efficiency	Train speed
Hou et al., 2019	Integrated traffic management and train operation	Metro	Delay minimisation; Energy efficiency; Demand satisfaction	Speed profile selection
Mo et al., 2019	Integrated traffic management and train operation	Metro	Demand satisfaction; Energy efficiency	Speed profile selection
Wang P. and Goverde, 2019	Train trajectory optimisation; Energy-efficient timetabling	Mainline	Conflict detection and resolution; Energy efficiency	Timetable constraint set
Wang P. et al., 2020	Train trajectory optimisation	Mainline	Parametric uncertainty minimisation; Time and energy cost	Train Path Envelope

Table 3.2: Overview of the relevant studies on the connections between traffic management and train operation at the operational planning stage (part 2)

Author(s)	**Main focus**	**Railway type**	**Objectives**	**Connections between traffic management and train operation**
Li et al., 2021	Integrated traffic management and train operation	Metro	Headway regularity; Energy efficiency	Train headway time; Speed profile selection
Dong et al., 2022	Integrated traffic management and train operation	High-speed	Delay minimisation; Energy efficiency	Train running time; Train trajectory
Wang Y. et al., 2022	Integrated traffic management and train operation	High-speed	Delay minimisation; Energy efficiency	Train running time; Train dwell time
Howlett et al., 2023	Train trajectory optimisation	Mainline	Safe separation; Energy efficiency	Target time
Liu et al., 2023	Integrated traffic management and train operation	Metro	Passenger travel time; Energy efficiency	Train timetable; Speed profile selection
Long et al., 2023	Integrated traffic management and train operation	High-speed	Delay minimisation	Train speed
Ying et al., 2023	Train trajectory optimisation	High-speed	Energy efficiency	Target window
Dong et al., 2024	Integrated traffic management and train operation	High-speed	Delay minimisation; Energy efficiency	Train running time
This paper	Timing Point and Train Path Envelope optimisation	Mainline	Timetable-trajectory alignment; Timetable robustness; Driving flexibility; Delay minimisation; Energy efficiency	Train Path Envelope

Scheepmaker et al. (2017) for the state-of-the-art of energy-efficient train trajectory optimisation. Recent works have further explored trajectory optimisation by using Quadratically Constrained Linear Programming-based (Ying et al., 2023) and analytical solutions (Howlett et al., 2023) to explicitly handle time constraints at TPs. Notwithstanding, few studies focus on optimising the TPE per se. Specifically, the time targets or windows defined according to the scheduled departures and arrivals at stops may not be sufficient to guarantee conflict-free train trajectories during operations when the train trajectories between two stations are modelled deterministically. Hence, Wang, P. & Goverde (2017, 2019) devised the concept of a Timetable Constraint Set to fine-tune the values of time and speed windows based on operability, feasibility, and energy efficiency.

An optimal interaction between traffic management and train operation is required to control the train speed to avoid train path conflicts and minimise the impact of disturbances. Rao et al. (2016) provided an integrated approach that collectively considered the TMS and ATO to generate a set of conflict-free train trajectories with a focus on the speed control of each train. Luan et al. (2018a) modelled the train speeds as decision variables to minimise the total delay and employed three optimisation methods. Among these three approaches, a two-level pre-processing speed profile generation and selection method performed best and was extended

further to energy reduction (Luan et al., 2018b). Along the same research line, Long et al. (2023) proposed a mixed-integer linear programming model to simultaneously determine train orders, routes, departure and arrival times, as well as speed profiles, allowing for a precise representation of the effect and dynamics of temporary speed restrictions on high-speed railways. Continuing the development of integrating traffic management and train operation in high-speed rail, Dong et al. (2022) proposed an online rescheduling framework for real-time disruption management, integrating train rescheduling and speed trajectory optimisation to enhance operational recovery. Dong et al. (2024) introduced a reinforcement learning-based approach that integrates train trajectory optimisation with timetable rescheduling to minimise delays and energy consumption, using rescheduled train running times as the bridge between the two processes. Additionally, Wang, Y. et al. (2022) developed a hierarchical model predictive control method to jointly manage delays and train speed control, where train running and dwell times act as the key linking factors.

Besides, metro systems have widely implemented ATO and mostly adopted this integrated concept (Yin et al., 2017). Usually, the goal is to minimise energy consumption by collectively considering the timetable and the speed profiles. Hou et al. (2019) and Mo et al. (2019) aimed to select the most suitable speed profile out of a pre-generated set for optimising the energy-efficient train operation and scheduling. Furthermore, Li et al. (2021) integrated the dwell time regulation in the timetable with the speed profile selection to enhance the real-time metro system performance. Liu et al. (2023) developed a bi-level model predictive control approach that simultaneously adjusts train schedules and speed profiles based on time-dependent passenger origin–destination demand.

However, ATO development in mainline railways necessitates a distinction between traffic management and train operation due to the vertical separation of railways between infrastructure manager and railway undertaking in European Union member states (Nash et al., 2014). Moreover, the TMS and ATO represent different dynamics and scopes. On the one hand, the TMS reschedules a timetable in case of disturbances and disruptions, focusing on optimal track capacity allocation and travel times on the network level. On the other hand, ATO regulates the trains by computing feasible and energy-efficient speed trajectories over the assigned routes within the margins contained in the real-time traffic plan. Such a discrepancy introduces uncertainty in train running and dwell times, making train timetable (re)scheduling increasingly complex to solve while strongly affecting both the quality and robustness of the timetable (Zhan et al., 2024). Consequently, directly controlling the train from a TMS can not be realised, while determining an optimised set of TPs with temporal constraints along the train route is of utmost necessity to bridge TMS and ATO in the system architecture aforementioned.

Overall, the existing studies are mostly limited to adopting the TPE and similar concepts as constraints for train trajectory optimisation. No existing study provides a method for optimising TP locations with their associated time targets or windows in a TPE to ensure conflict-free train trajectory generation. Nonetheless, this optimisation is crucial for ATO implementation to fully realise its benefits, including enhanced capacity, punctuality, and energy efficiency. The first attempt to compute the TPEs was based on splitting the buffer time between trains at selected locations into two halves and then allocating a half to the time window of each of two successive trains (Wang, P. & Goverde, 2016). Wang, P. & Goverde (2017, 2019) proposed the concept of a Timetable Constraint Set to fine-tune the time and speed windows, although they considered the TPs only at stations while focusing on multi-train trajectory optimisation.

The challenges of determining TPs in a TPE mainly arise from three aspects: First, multiple train driving strategies could be applied for the same running time between two stops, which

leads to uncertainty in train trajectory modelling in the TMS. Second, the computation of conflict-free TPEs for successive trains is still an open problem. Even with a (re)scheduled timetable, operational conflicts may still arise due to variations in train driving behaviour. Therefore, merely following the target time at stops based on an assumed deterministic train time-distance path cannot ensure conflict-free train operations. Third, defining TPs and their associated time windows requires balancing flexibility in train driving strategies with adherence to the real-time traffic plan. A rigid set of TPs may not accommodate variations in train operations, potentially leading to track occupation conflicts, whereas an excessive number of TPs could overly constrain flexibility, reducing ATO adaptability and energy efficiency. As a result, this paper proposes a new method to guide the trains within conflict-free TPEs by identifying an optimal sequence of TPs and associated time windows for each train. Our approach thus contributes to the identified and required future works to interface the TMS, ATO trackside, ATO onboard, and the signalling systems in mainline railways (Goverde et al., 2023) and serves as a cornerstone within the Europe's Rail FP1-MOTIONAL project, which includes efforts to align and enhance the interaction between TMS and ATO (MOTIONAL, 2024).

3.3 Methodology

This section begins with a problem description and key assumptions that form the foundation of the proposed modelling framework (Section 3.3.1). Next, Train Path Slot and TPs are introduced based on blocking time theory and multiple train driving strategies in Section 3.3.2, followed by the complete modelling framework with five modules in Section 3.3.3 to determine an optimised set of TPs. Lastly, the evaluation criteria for quantifying the results are in Section 3.3.4.

3.3.1 Problem description and assumptions

Effective ATO deployment requires seamless alignment between the network-wide real-time traffic plan and real-time train trajectory generation and tracking. The TMS (re)schedules train timetables (times, orders, routes) at the traffic level in response to disturbances and disruptions, whereas the ATO onboard functions at the individual train level, generating and tracking train trajectories (speed over distance) to ensure compliance with operational constraints. However, existing approaches lack a mechanism to dynamically synchronise ATO-driven trajectory generation with timetable-based traffic management, leading to possible misalignments between planned and actual train time-distance trajectories.

This misalignment is particularly pronounced in mainline railways, where the TMS does not directly control train speed, and train trajectory generation depends on dynamic ATO driving strategies. Timetable (re)scheduling models typically determine discrete arrival and departure times at selected timetable points, assuming deterministic (minimum) running times to ensure theoretical conflict-freeness. However, trains may adopt different driving strategies — such as maintaining a target timetable speed or employing energy-efficient control with cruising and coasting regimes — resulting in different continuous speed-distance trajectories between the discrete scheduled timing points, while still adhering to the same computed timetable. This variability results in different time-distance trajectories between stops, introducing operational uncertainties that can cause train path conflicts, particularly for successive trains sharing track resources.

This study proposes a TPE computation method as an interface between the TMS real-time

traffic plan and ATO onboard speed control. This method dynamically determines possible locations where trains may have track occupation conflicts due to incompatible driving strategies and derives extra timing points with associated time constraints to prevent these conflicts. These extra timing points are subsequently collected together with the scheduled event times at the timetable points as an ordered list of time constraints in a TPE for each train. The TPE is then sent to the ATO onboard to guide ATO-driven trains toward conflict-free, punctual, and energy-efficient operations while preserving flexibility in train trajectory generation.

The following key assumptions are used in our approach: (1) The TMS provides a conflict-free real-time traffic plan that specifies scheduled departure and arrival times at designated stops, train routing, and the predefined train precedence order; (2) Trains can adopt different driving strategies (e.g., minimum-time, scheduled-speed, or energy-efficient control), leading to a range of possible passing times between stops; (3) This study assumes a fixed-block signalling system with continuous braking curve supervision (e.g., European Train Control System (ETCS) Level 2) as an example. However, the methodology applies to any signalling system, such as conventional multi-aspect and moving-block signalling systems, with the blocking time computations adapted accordingly; (4) Regenerative braking is not explicitly modelled, but the methodology can be extended to incorporate energy-efficient train control with regenerative braking (Scheepmaker & Goverde, 2020), as any other valid driving strategy, without loss of generality; (5) We assume that data communication follows standard railway telecommunication specifications, such as GSM-R used in ETCS Level 2.

3.3.2 Train Path Slot and timing points

The railway infrastructure can be regarded as a series of block sections that are allocated exclusively to one train at a time, during which it is blocked for other trains (Pachl, 2014). The successive blocking times of each train over a railway corridor in a time-distance graph are the so-called blocking time stairway in which the blocking times can be seen as boxes of time slots over the blocks (Goverde et al., 2013). If on a block section, the blocking times of two successive trains overlap, then a conflict would occur, resulting in unnecessary braking and possibly stopping until the block is released. Otherwise, the white space between two consecutive trains for each block (i.e., the buffer time) can be leveraged to optimise the train trajectory as long as there is no overlap.

Nevertheless, the blocking time stairway could vary for the same departure and arrival time, as the running time in a specific intermediate block section differs among train driving strategies. In particular, the conflict detection and resolution algorithms in a TMS typically rely on a deterministic train path based on a single train trajectory. The inconsistency of the train driving strategies utilised by the traffic management and the train operation leads to different intermediate passing times, which might result in train path conflicts. As a result, we define the concept of a Train Path Slot as an integrated blocking time stairway pattern based on multiple driving strategies, specifying the successive passing windows of each train over a railway corridor. In this paper, we propose three driving strategies: shifted Minimum-Time Train Control (MTTC), Reduced Maximum Speed (RMS, a reduced cruising speed without coasting) and Energy-Efficient Train Control (EETC, combining cruising and coasting). The shifted MTTC considers a late departure tolerance (i.e., an allowed late departure window) equal to the available running time supplement until the next arrival time. It drives the train as fast as possible and respects the imperative target scheduled arrival time at the next stop. The late

departure tolerance represents a further robustness of the scheduled train paths. This strategy provides some flexibility on the departure time due to, for instance, extended boarding. The derived trajectories based on these driving strategies determine the initial lower and upper bounds of the Train Path Slot, i.e., without yet considering the adjacent trains. We present an example of an initial Train Path Slot in Figure 3.1. These preliminary Train Path Slots can reveal potential train path conflicts inherently contained in the real-time traffic plan as they take the time-distance paths from different speed profiles into account.

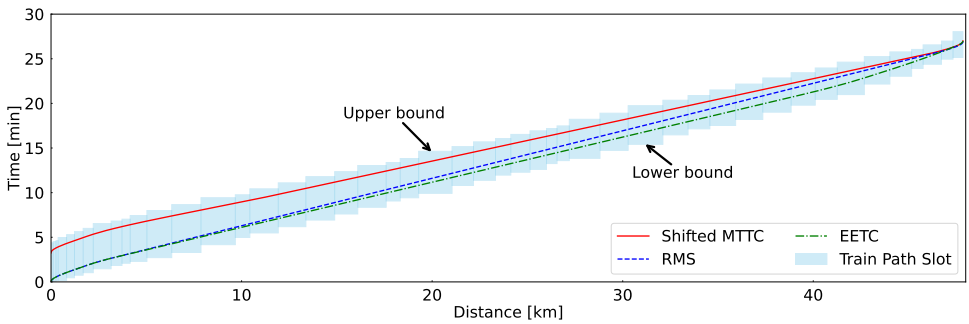


Figure 3.1: Example of a Train Path Slot

Train Path Slots for successive train pairs may overlap and must be resolved to ensure conflict-free operations, forming the basis for an optimised set of TPs. If no overlap occurs, the scheduled arrival time and a departure tolerance up to the running time supplement define the optimised TPs for a train run. Conflict-free Train Path Slots are achieved by considering all adjacent trains and applying Train Path Slot overlap detection and resolution. To resolve conflicts, we formulate a Linear Programming (LP) model that identifies successive conflict-free blocking time windows. This LP model serves as the first stage of a two-step process, where a subsequent computation stage ensures train trajectory feasibility by refining departure tolerances and introducing intermediate TPs. The LP model prioritises adjusting the departure tolerance for the preceding train in a train pair and then imposes constraints on the nominal driving of a following train. Here, nominal driving refers to train operation that adheres to the scheduled departure and arrival times from the plan. The outcome of the LP model is a set of conflict-free Train Path Slots without considering the train dynamics, i.e., the train has to provide a corresponding train trajectory to respect these slots and make them feasible. Therefore, TPs are configured at the entry location of the blocks to ensure that realisable train time-distance profiles are kept within the identified conflict-free Train Path Slots. We propose a critical-block TP configuration that introduces a TP to the block having the largest overlapping blocking time. It is compared with other alternatives of only using TPs at scheduled stops or successively adding TPs based on the first conflicting block in an iterative approach in the case study.

The train driving strategies that meet the optimised lower and upper bounds of a restricted Train Path Slot from the LP model are called a TP response strategy and a departure tolerance response strategy, respectively. These two driving strategies are established to respect the optimisation results and provide feasible event times at TPs as constraints for any onboard train trajectory optimisation approaches, i.e., departure, arrival and passing times. Specifically, the TP response strategy is defined as a combination of RMS and EETC driving strategies with possibly multiple response cruising speeds to avoid Train Path Slot overlaps. The departure tolerance

response strategy relies on the RMS driving strategy. It starts at the adjusted departure tolerance such that the remaining running time supplement is the scheduled running time supplement minus the departure tolerance. This adjusted departure tolerance will always be smaller than the given running time supplement because the initial departure tolerance was defined as the running time supplement for the shifted MTTC to arrive at the scheduled arrival time at the next stop. The discrete TPs with either time targets or windows are eventually assembled into the TPE. The reduction of a restricted Train Path Slot with an optimised set of TPs in a TPE as the output of our modelling framework is visualised on the right-hand side of Figure 3.5.

3.3.3 Modelling framework

The proposed modelling framework uses four types of data, i.e., timetable, route, signalling and rolling stock characteristics. The framework is provided as a flow diagram in Figure 3.2 and consists of five modules, with Train Path Slot overlap resolution following a two-step approach: (1) an LP model that determines conflict-free Train Path Slots, and (2) a response train trajectory generation stage that ensures feasible operations by adjusting departure tolerances and introducing intermediate TPs, ultimately forming the TPE. The five modules are: (1) Train trajectory computations for three relevant train driving strategies, (2) Train Path Slot computation by using blocking time theory, (3) Train Path Slot overlap detection, (4) Train Path Slot overlap resolution through utilising TPs, (5) TPE construction with an optimised set of TPs. Each step feeds into the next, forming the TPE computation process. The following subsections manifest each of the modules, sequentially. The notation used in this paper is summarised in Tables 3.3, 3.4, and 3.5, categorised respectively into: (i) sets, (ii) fixed parameters and variables, and (iii) decision variables and outputs of sequential optimisation.

Table 3.3: Notation of sets

Symbols	Definitions
P, p	P is a set of trains, and p is a train within P
P_2, (p_i, p_j)	$P_2 = \{(p_i, p_j) \mid p_i, p_j \in P, p_i \prec p_j\}$ is a set of successive train pairs with p_i preceding p_j before a possible scheduled overtaking, and (p_i, p_j) is a train pair within P_2
B_p, b	B_p is a set of blocks along the route of train p with an ordering $B_p = \{b_1, \ldots, b_n\}$, and b is a block within B_p
B_p^{parallel}, b	B_p^{parallel} is a set of successive blocks of train p on a track of multiple-track line(s), and b is a block within B_p^{parallel}
$B_{p,b}$, b'	$B_{p,b} \subset B_p$ is a set of blocks that train p has traversed before block b, and b' is a block within $B_{p,b}$
S'_p, s_b	S'_p is a set of all possible Timing Points at the entry location of a block in the running direction of train p, and $s_b \in S'_p$
Γ, γ	Γ is a set of train driving strategies, and γ is a driving strategy within Γ
TPS_p	TPS_p is the Train Path Slot of train p, $\text{TPS}_p = \left\{ \left(b, \underline{t}_{p,b}^{\text{TPS}}, \bar{t}_{p,b}^{\text{TPS}} \right), b \in B_p \right\}$

Train trajectory computation

Train trajectories are computed in two stages: first, using different driving strategies to establish initial Train Path Slots, and later, re-computed as response driving strategies to satisfy the LP

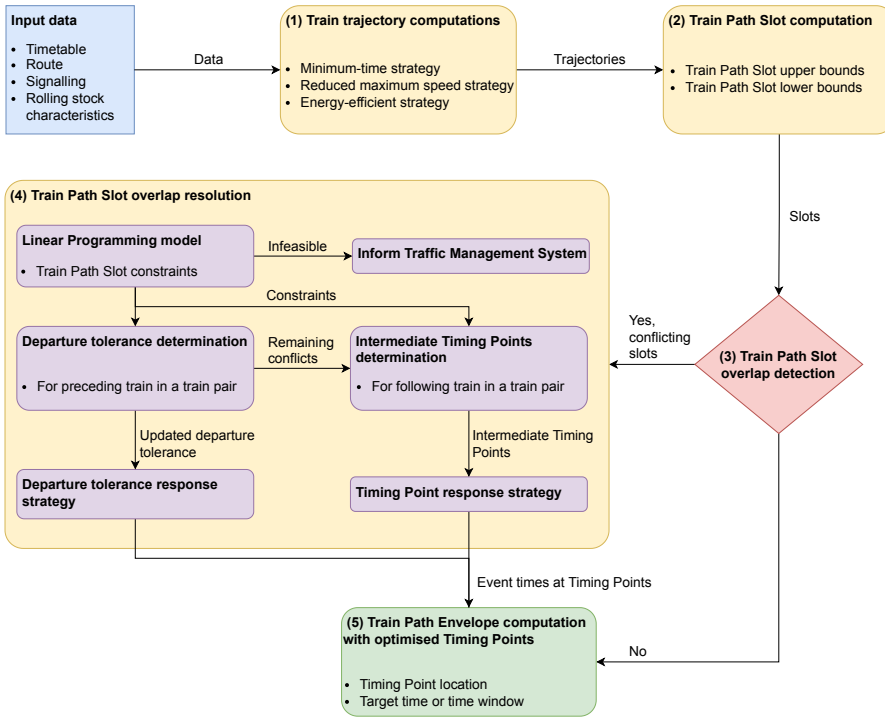


Figure 3.2: Methodological framework of optimised Timing Points for ATO

model's conflict-free Train Path Slot constraints. In this second stage, response driving strategies determine the actual departure tolerances and passing times at intermediate TPs, which together form the TPE. The train trajectory modelling is train-specific and incorporates train-specific parameters, including traction characteristics, braking rates, and mass.

Our train trajectory computation considers three train driving strategies: the shifted MTTC, the RMS and the EETC. The shifted MTTC train driving strategy is based on the MTTC train driving strategy that operates the train as fast as possible, with a departure tolerance. This departure tolerance corresponds to departing late up to the running time supplement. The original MTTC train driving strategy is used in the timetable design to compute the minimum running time, and based on that, determine the additional running time supplement. By applying Pontryagin's Maximum Principle and the Karush-Kuhn-Tucker conditions, we are able to derive the optimal control structure of the MTTC problem as detailed in Albrecht, A.R. et al. (2016a) and Goverde et al. (2021) and then produce the MTTC speed profile. The resulting driving strategy consists of three driving regimes: maximum acceleration of the train until reaching the speed limit, cruising at the speed limit (if reached), and service braking. If there are varying speed limits between two stops, the train has to apply service braking or maximum acceleration to transition between different cruising speeds at these speed limits.

The other two strategies, RMS and EETC, use the full scheduled running time between scheduled departure and arrival times. The EETC driving strategy aims to minimise the total

Table 3.4: Notation of parameters and variables

Symbols	Definitions
Fixed parameters	
r_p	running time supplement of train p between two stops
$t_{p,b}^{\text{setup}}$	setup time of the route for train p in block b
$t_{p,b}^{\text{reaction}}$	sight and reaction time of train p in block b
$t_{p,b}^{\text{release}}$	release time of train p to free up block b
t_p^{buffer}	buffer time in the timetable to prevent delay propagation as part of the Train Path Slot upper bound of train p
$\alpha_{p_i,p_j,b}$	binary parameter to specify whether train p_i precedes p_j in block b. If yes, $\alpha_{p_i,p_j,b}=1$; otherwise, $\alpha_{p_i,p_j,b}=0$
Variables	
$c_{p_i,p_j,b}$	Train Path Slot overlap between trains p_i and p_j in block b
$c_{p_i,p_j}^{\max}$	largest Train Path Slot overlap between trains p_i and p_j in a chosen railway corridor
$\underline{t}_{p,b}^{\text{TPS},\gamma}$	lower bound of the Train Path Slot in block b for train p based on a driving strategy γ
$\bar{t}_{p,b}^{\text{TPS},\gamma}$	upper bound of the Train Path Slot in block b for train p based on a driving strategy γ
$\underline{t}_{p,b}^{\text{TPS}}$	initial lower bound (i.e., start time) of the Train Path Slot in block b for train p
$\bar{t}_{p,b}^{\text{TPS}}$	initial upper bound (i.e., end time) of the Train Path Slot in block b for train p
$t_{p,b}^{\text{approach},\gamma}$	approach time of train p for block b based on a driving strategy γ
$t_{p,b}^{\text{run},\gamma}$	running time of train p in block b based on a driving strategy γ
$t_{p,b}^{\text{clear},\gamma}$	clearing time of train p for block b based on a driving strategy γ

Table 3.5: Notation of decision variables and outputs of sequential optimisation

Symbols	Definitions
Decision variables of the LP model	
$x_{p,b}$	time of the optimised Train Path Slot lower bound of train p in block b
$y_{p,b}$	time of the optimised Train Path Slot upper bound of train p in block b
Sequentially optimised sets and variables	
S_p, s	S_p is a set of Timing Points of train p, and s is a Timing Point within S_p
TPE_p	TPE_p is the Train Path Envelope of train p, $\text{TPE}_p=\left\{\left(s,t_{p,s}^{\min},t_{p,s}^{\max}\right),s\in S_p\right\}$
$\text{TPS}_p^{\text{cons}}$	$\text{TPS}_p^{\text{cons}}$ is the Train Path Slot constraint of train p, $\text{TPS}_p^{\text{cons}}=\{(b,x_{p,b},y_{p,b}),b\in B_p\}$
$t_{p,s}^{\min}$	earliest time for train p to arrive at Timing Point s
$t_{p,s}^{\max}$	latest time for train p to arrive at Timing Point s
τ_p^{D}	departure tolerance of train p
$\hat{\tau}_p^{\text{D}}$	restricted departure tolerance of train p

energy consumption between two stops. Its optimal control structure is well-established in Albrecht, A.R. et al. (2016b) and Goverde et al. (2021). Again, we can obtain the optimal control

structure of the EETC problem by using Pontryagin's Maximum Principle and the Karush-Kuhn-Tucker conditions and then generate the EETC speed profile. The resulting driving strategy comprises four driving regimes: maximum acceleration of the train until reaching the optimal cruising speed, cruising at the optimal cruising speed, coasting without using the engine or brakes of the train, and service braking. The RMS driving strategy is based on a scheduled cruising speed without coasting, which adds a restrictive constraint on top of the EETC control structure to avoid the coasting regime (Scheepmaker et al., 2020). Thus, its resulting driving strategy only has three driving regimes: maximum acceleration of the train until reaching the optimal cruising speed (when no coasting is used), cruising at the optimal cruising speed, and service braking. The RMS essentially determines a timetable speed that trains can maintain to arrive on time, which is a typical running time calculation assumption in timetable planning.

We visualise an example of a speed-distance diagram (i.e., speed profile) of these three driving strategies in Figure 3.3. All three driving strategies comprise regimes of maximum acceleration and service braking. The MTTC (in solid red) cruises at the speed limit (in solid black), whereas the RMS (in dashed blue) cruises at a required cruising speed to arrive on time. With the same amount of running time supplement, the EETC (in dash-dotted green) has a higher cruising speed than the RMS due to the following coasting regime. Moreover, the grey area in the figure indicates the track gradient profile.

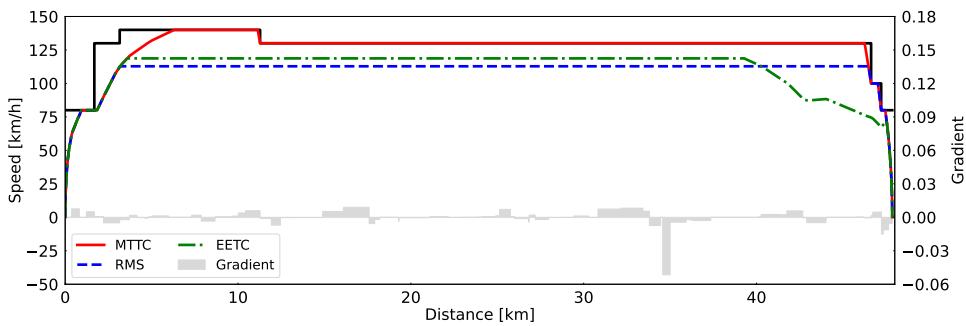


Figure 3.3: Example of the Minimum-Time Train Control (MTTC), Reduced Maximum Speed (RMS) and Energy-Efficient Train Control (EETC) driving strategies

Initial Train Path Slot computation

This subsection determines the lower and upper bounds of a Train Path Slot based on the computed train trajectories, which represent the start and end times of the blocking time stairway diagrams. Specifically, we take the train trajectories with one of the three train driving strategies as input and build the corresponding blocking time stairways. Then, we select the most critical boundaries of the blocking time stairways as the bounds of the Train Path Slot to obtain this integrated blocking time stairway pattern. In general, the shifted MTTC and the EETC driving strategies determine the upper and lower bounds of a Train Path Slot, respectively. Nevertheless, there could be exceptions.

We compute the Train Path Slot lower bound by building the blocking time diagrams based on the shifted MTTC, RMS and EETC driving strategies and selecting the earliest time at any block between two stops. For shifted MTTC driving, we assume that the train exhausts all the

running time supplement r_p before departing between two stops, while the other two depart as scheduled. All three driving strategies follow the scheduled arrival times. Let P be the set of trains. In ETCS Level 2, the approach distance of train $p \in P$ before block b is the absolute braking distance, depending on the train, route and speed. We use the resulting braking time plus a safety margin to indicate the approach time $t_{p,b}^{\text{approach}}$. Hence, the lower bound (denoted by the underscore) of the Train Path Slot $\underline{t}_{p,b}^{\text{TPS}}$ is the earliest reserved time of train p at the entrance point of a block b among all three driving strategies:

$$\underline{t}_{p,b}^{\text{TPS},\gamma} = \sum_{b' \in B_{p,b}} t_{p,b'}^{\text{run},\gamma} - t_{p,b}^{\text{approach},\gamma} - t_{p,b}^{\text{reaction}} - t_{p,b}^{\text{setup}}, \qquad \gamma \in \{\text{RMS, EETC}\}, \tag{3.1}$$

$$\underline{t}_{p,b}^{\text{TPS},\gamma} = r_p + \sum_{b' \in B_{p,b}} t_{p,b'}^{\text{run},\gamma} - t_{p,b}^{\text{approach},\gamma} - t_{p,b}^{\text{reaction}} - t_{p,b}^{\text{setup}}, \qquad \gamma = \text{MTTC}, \tag{3.2}$$

$$\underline{t}_{p,b}^{\text{TPS}} = \min_{\gamma \in \Gamma} \left(\underline{t}_{p,b}^{\text{TPS},\gamma} \right), \tag{3.3}$$

where the set $\Gamma = \{\text{MTTC, RMS, EETC}\}$ represents the set of the three driving strategies and the set $B_{p,b} \subset B_p$ represents the set of blocks before b that train p has traversed. $t_{p,b'}^{\text{run},\gamma}$ denotes the running time of train p in block b', given a certain train driving strategy γ. Hence, the cumulative value of running time over $b' \in B_{p,b}$ is the trip time of train p to the block b. $t_{p,b}^{\text{approach},\gamma}$ is the approach time of train p in block b with a certain train driving strategy γ. Additionally, $t_{p,b}^{\text{reaction}}$ is a constant parameter that represents the sight and reaction time before the braking indication point and $t_{p,b}^{\text{setup}}$ is the setup time of the route in the block that depends on whether the train is on open tracks or within the interlocking areas. For the latter, the $t_{p,b}^{\text{setup}}$ also depends on the number of switches involved in the route setting. While reaction time is a fixed parameter, and setup time is route-dependent, conservative values are chosen to account for stochastic variations, such as the communication delay of the Movement Authority from the Radio Block Centre to the train. As a result, Eq. (3.1) represents the Train Path Slot lower bound based on the nominal train driving with a scheduled departure and arrival time, while Eq. (3.2) hinges on the shifted MTTC driving strategy with a departure tolerance.

We use the shifted MTTC driving strategy to calculate the upper bound of the Train Path Slot $\bar{t}_{p,b}^{\text{TPS}}$. Besides, we add a fixed buffer time t_p^{buffer} after a train as part of the Train Path Slot upper bound. This increases the robustness of the train path since it could (partially) absorb the deviation of the preceding train from the scheduled trajectory and prevent delay propagation to the following train. The value of the buffer time could also be zero, depending on the timetable design. Without loss of generality, this buffer time can also be split into two parts, with one placed before the lower bound and the other after the upper bound. For ease of use in the following equations, we also provide the Train Path Slot upper bounds for the RMS and EETC driving strategies, which correspond to their respective blocking times without the additional buffer time.

$$\bar{t}_{p,b}^{\text{TPS}} = r_p + \sum_{b' \in B_{p,b}} t_{p,b'}^{\text{run},\gamma} + t_{p,b}^{\text{run},\gamma} + t_{p,b}^{\text{clear},\gamma} + t_{p,b}^{\text{release}} + t_p^{\text{buffer}}, \qquad \gamma = \text{MTTC}, \tag{3.4}$$

$$\bar{t}_{p,b}^{\text{TPS},\gamma} = \sum_{b' \in B_{p,b}} t_{p,b'}^{\text{run},\gamma} + t_{p,b}^{\text{run},\gamma} + t_{p,b}^{\text{clear},\gamma} + t_{p,b}^{\text{release}}, \qquad \gamma \in \{\text{RMS, EETC}\}, \tag{3.5}$$

where $t_{p,b}^{\text{run},\gamma}$ is the running time of train p in block b and $t_{p,b}^{\text{clear},\gamma}$ is the time for train p to clear the block b based on a specific train driving strategy γ. $t_{p,b}^{\text{release}}$ is a constant parameter to release the block b.

To represent the blocking time intervals allocated to a train across all blocks it traverses based on multiple driving strategies, we define the initial Train Path Slot TPS_p as a set of blocking time intervals, bounded by $\underline{t}_{p,b}^{\mathrm{TPS}}$ from Eq. (3.3) and $\bar{t}_{p,b}^{\mathrm{TPS}}$ from Eq. (3.4), respectively. This set defines a closed area in time-distance space, bounded by a piecewise constant lower and upper contour corresponding to the most extreme start and end times of the blocking times over successive blocks. This formulation extends the classical blocking time stairway representation by incorporating multiple train driving strategies, thereby capturing a range of possible blocking time intervals rather than a single deterministic trajectory. A visual representation of this concept is provided in Figure 3.1, illustrating how the Train Path Slot defines the train's possible passage times over successive blocks. The formal definition of the initial Train Path Slot is given by:

$$\mathrm{TPS}_p = \left\{ \left(b, \underline{t}_{p,b}^{\mathrm{TPS}}, \bar{t}_{p,b}^{\mathrm{TPS}} \right),\ b \in B_p \right\}. \tag{3.6}$$

Train Path Slot overlap detection

The conflict-free Train Path Slots of all concerned trains without overtaking must satisfy the following conditions:

$$\underline{t}_{p_j,b}^{\mathrm{TPS}} \geq \bar{t}_{p_i,b}^{\mathrm{TPS}} \quad \forall \left(p_i, p_j\right) \in P_2,\ b \in B_{p_i} \cap B_{p_j}, \tag{3.7}$$

where $\left(p_i, p_j\right)$ is a successive train pair in P_2. If an overtaking is scheduled, then the sequence of this train pair is reversed, which we will consider later.

The timetable planning tools of infrastructure managers usually have only target times as TP constraints at stations and/or important junctions. However, train trajectory uncertainty is not accounted for between these TPs and therefore, the Train Path Slots derived from the original timetable may have overlaps, i.e., Eq. (3.7) may not be met for all shared blocks with only the scheduled arrival times at stations. Therefore, we check whether the upper bound of the preceding train p_i is conflicting with the lower bound of the following train p_j, similar to Bešinović et al. (2017). The Train Path Slot overlap $c_{p_i,p_j,b}$ between a successive train pair $\left(p_i, p_j\right)$ at the shared block b is computed as:

$$c_{p_i,p_j,b} = \max \left(0, \left(\bar{t}_{p_i,b}^{\mathrm{TPS}} - \underline{t}_{p_j,b}^{\mathrm{TPS}} \right) \alpha_{p_i,p_j,b} + \left(\bar{t}_{p_j,b}^{\mathrm{TPS}} - \underline{t}_{p_i,b}^{\mathrm{TPS}} \right) \left(1 - \alpha_{p_i,p_j,b}\right) \right)$$
$$\forall \left(p_i, p_j\right) \in P_2,\ b \in B_{p_i} \cap B_{p_j}, \tag{3.8}$$

where the binary parameter $\alpha_{p_i,p_j,b}$ equals 1 if train p_i precedes p_j at block b in the original order as specified in the given timetable while it equals 0 if the timetable indicates that train p_j precedes train p_i due to a scheduled overtaking before block b. If any $c_{p_i,p_j,b} > 0$, then a Train Path Slot overlap exists. Furthermore, we also consider partially parallel routes, such as on partial four-track lines or in stations where a planned overtaking occurs at scheduled stops. Let B_p^{parallel} represent the successive blocks of train p on a track of multiple-track line(s). On these blocks $b \in B_p^{\mathrm{parallel}}$, the Train Path Slot overlaps should not be considered:

$$c_{p_i,p_j,b} = 0 \quad \forall \left(p_i, p_j\right) \in P_2,\ b \in \left(B_{p_i}^{\mathrm{parallel}} \cup B_{p_j}^{\mathrm{parallel}} \right) \setminus \left(B_{p_i}^{\mathrm{parallel}} \cap B_{p_j}^{\mathrm{parallel}} \right). \tag{3.9}$$

If no Train Path Slot overlap has been detected, the current TP configuration—with only target times at stations and/or important junctions—is sufficient to define the TPE and guide the train operation conflict-free. If an overlap is present, a Train Path Slot resolution module is required to optimise intermediate TPs, as detailed in the following subsection.

Train Path Slot overlap resolution

The resolution of Train Path Slot overlaps follows a two-step optimisation process. In the first stage, we formulate an LP model to determine conflict-free Train Path Slots by optimising their lower and upper bounds. The LP model contains two decision variables $x_{p,b}$ and $y_{p,b}$, representing the lower bound and the upper bound of the Train Path Slot in block b. However, since the LP model does not explicitly account for train dynamics, a second-stage sequential optimisation process is required to refine departure tolerances, determine intermediate TPs, and ensure operational feasibility, ultimately deriving the TPEs. For all successive train pairs $\left(p_i, p_j\right) \in P_2$, the LP model is formulated as follows:

$$\text{Maximise} \quad \sum_{p\in P}\sum_{b\in B_p}\left(y_{p,b}-x_{p,b}\right)+\left(\underline{t}^{\text{TPS}}_{p,b}-x_{p,b}\right)=\sum_{p\in P}\sum_{b\in B_p} y_{p,b}+\underline{t}^{\text{TPS}}_{p,b}-2x_{p,b} \tag{3.10}$$

subject to

$$\begin{aligned} c_{p_i,p_j,b} \leq{}& \left[\left(\bar{t}^{\text{TPS}}_{p_i,b}-y_{p_i,b}\right)+\left(x_{p_j,b}-\underline{t}^{\text{TPS}}_{p_j,b}\right)\right]\alpha_{p_i,p_j,b}+\left[\left(\bar{t}^{\text{TPS}}_{p_j,b}-y_{p_j,b}\right)\right. \\ &\left.+\left(x_{p_i,b}-\underline{t}^{\text{TPS}}_{p_i,b}\right)\right](1-\alpha_{p_i,p_j,b}) \quad \forall(p_i,p_j)\in P_2,\ b\in B_{p_i}\cap B_{p_j}, \end{aligned} \tag{3.11}$$

$$y_{p,b}-x_{p,b}\geq \bar{t}^{\text{TPS}}_{p,b}-t^{\text{buffer}}_{p}-\underline{t}^{\text{TPS,MTTC}}_{p,b} \qquad \forall p\in P,\ b\in B_p, \tag{3.12}$$

$$\underline{t}^{\text{TPS}}_{p,b}\leq x_{p,b}\leq \underline{t}^{\text{TPS,MTTC}}_{p,b} \qquad \forall p\in P,\ b\in B_p, \tag{3.13}$$

$$\bar{t}^{\text{TPS},\gamma}_{p,b}\leq y_{p,b}\leq \bar{t}^{\text{TPS}}_{p,b} \qquad \forall p\in P,\ b\in B_p,\ \gamma\in\{\text{RMS, EETC}\}, \tag{3.14}$$

where $\underline{t}^{\text{TPS}}_{p,b}$ and $\bar{t}^{\text{TPS}}_{p,b}$ are provided by the initial Train Path Slot lower and upper bounds.

For a railway corridor, the most flexible conflict-free Train Path Slots for all trains and all blocks they traverse can be found by maximising the terms in the first parentheses of Eq. (3.10), i.e., the difference between the upper and the lower bounds of the Train Path Slots. At the same time, if conflicts exist between a preceding and a following train, the terms in the second parentheses of Eq. (3.10) penalise differences between the optimised and initial lower bounds of the Train Path Slot to minimise impacts to nominal train operations. Essentially, the model primarily adjusts the upper bound of the initial Train Path Slot, which represents the robustness of the train path. This objective function collectively aims to fine-tune the departure tolerance for a preceding train, and only when no departure tolerance remains does it impose constraints on the nominal driving of the following train. In this way, the scheduled departure and arrival times from the real-time traffic plan are respected as much as possible.

Constraint (3.11) guarantees that the adjustments at overlapping shared blocks eliminate all Train Path Slot conflicts. No adjustment is needed if there is no overlap between a pair of successive Train Path Slots.

At any block section, the smallest blocking time is obtained when the train drives as fast as possible, namely by applying an MTTC driving strategy. The right-hand side of constraint (3.12) represents this minimal blocking time, with the buffer time subtracted from the shifted MTTC to derive the plain MTTC upper bound. This constraint ensures that any optimised Train Path Slot will always accommodate at least the MTTC trajectory.

Constraints (3.13) and (3.14) provide the feasible solution space for the decision variables by considering the most critical driving strategy under nominal and delayed train operations. Specifically, the lower bound of a Train Path Slot must be smaller than the blocking time lower

bound derived from the shifted MTTC strategy, while the upper bound must exceed the blocking time upper bound based on the slower of the RMS and EETC strategies.

The LP model provides an optimised, conflict-free set of Train Path Slots, which serves as constraints for subsequent response train trajectory generation. However, since the LP model does not fully capture train dynamics or give TPs, a further computation process is necessary. If the LP model fails to find a feasible solution, this indicates that the (re)scheduled timetable from the TMS contains conflicts that cannot be resolved by introducing additional TPs alone. In such cases, feedback is provided to the TMS, which must adjust the timetable to relax too restrictive headways between conflicting train pairs. Once the timetable is revised, the LP model is re-run in the next iteration to determine necessary TPs.

If the LP model finds a feasible solution, it produces three types of adapted slots as outputs:

1. Unchanged Train Path Slots: These remain intact as there is no overlap. A departure tolerance up to the running time supplement is available, and all three train driving strategies fit within these slots.
2. Adjusted upper bound: The upper bound of the initial Train Path Slot is modified, requiring a fine-tuned departure tolerance at the previous stop.
3. Adjusted lower and upper bounds: The Train Path Slot lower bound of the following train is modified, in addition to the Train Path Slot upper bound of the preceding train. In this case, the preceding train has no departure tolerance and must depart on schedule. The following train requires intermediate TPs with time constraints to resolve remaining Train Path Slot conflicts.

Therefore, the results from the LP model impose a set of Train Path Slot constraints on deriving the response train trajectories as follows:

$$\mathrm{TPS}_p^{\mathrm{cons}} = \left\{ \left(b, x_{p,b}, y_{p,b}\right),\ b \in B_p \right\} \quad \forall p \in P. \tag{3.15}$$

According to the LP result, the departure tolerance is either zero (i.e., the train has to depart on time) or needs to be determined through an iterative process using the departure tolerance response driving strategy until the Train Path Slot constraint is satisfied. The departure tolerance response driving relies on an RMS driving strategy, where the train departs at the adjusted departure tolerance and uses the remaining running time supplement, computed as the difference between the scheduled supplement and the adjusted departure tolerance. Since blocking time overlap is used to fine-tune the departure tolerance, adjusting the running time supplement results in a different train trajectory, which in turn updates the blocking time. Thus, an iterative process is required. Each time, we compute the largest Train Path Slot overlap $c^{\max}_{p_i,p_j}$ of a successive train pair $\left(p_i, p_j\right) \in P_2$ in a chosen railway corridor as follows:

$$c^{\max}_{p_i,p_j} = \max_{b \in B_{p_i} \cap B_{p_j}} \left(c_{p_i,p_j,b}\right). \tag{3.16}$$

This maximum overlap is then used to update the departure tolerance:

$$\hat{\tau}^{\mathrm{D}}_{p_i} \leftarrow \tau^{\mathrm{D}}_{p_i} - c^{\max}_{p_i,p_j}, \tag{3.17}$$

where $\hat{\tau}^{\mathrm{D}}_{p_i}$ is the restricted departure tolerance and $\tau^{\mathrm{D}}_{p_i}$ is the departure tolerance from the last step, which corresponds to the initial departure tolerance in the first step. We present the iterative process described in Algorithm 3.1.

Algorithm 3.1 Description of fine-tuning departure tolerance

Input: Inputs to the modelling framework, the largest Train Path Slot overlap $c^{\max}$, initial departure tolerance τ_p^{D} and Train Path Slot constraint $\mathrm{TPS}_p^{\mathrm{cons}}$

Output: Restricted Train Path Slot, fine-tuned departure tolerance $\hat{\tau}_p^{\mathrm{D}}$

while $c^{\max} > 0$ **do**

Step 1 Update departure tolerance: $\hat{\tau}_p^{\mathrm{D}} = \tau_p^{\mathrm{D}} - c^{\max}$

Step 2 Calculate departure tolerance RMS response train trajectory for $\hat{\tau}_p^{\mathrm{D}}$ to the next scheduled arrival stop

Step 3 Compute the updated Train Path Slot upper bound $\bar{t}_{p,b}^{\mathrm{TPS,RMS}}$ at each block b that train p traverses

Step 4 Set the new largest Train Path Slot overlap: $c^{\max} = \max_{b \in B_p} \left(\bar{t}_{p,b}^{\mathrm{TPS,RMS}} - y_{p,b} \right)$

Step 5 Update the initial departure tolerance: $\tau_p^{\mathrm{D}} = \hat{\tau}_p^{\mathrm{D}}$

end while

We present a time-distance diagram as an example to explain the update of the departure tolerance in Figure 3.4, which corresponds to the second type of adapted slot from the LP model. As shown in the left part of the figure, the initial departure tolerance of the preceding train leads to five Train Path Slot overlaps with the following train. To resolve these conflicts, the largest Train Path Slot overlap is iteratively identified and used to adjust the departure tolerance until a feasible value is found that eliminates all overlaps. Since the departure tolerance response strategy maintains the same total running time supplement, the remaining supplement is given by the initial supplement minus the adjusted departure tolerance. This fine-tuning process leads to a restricted Train Path Slot that meets the Train Path Slot constraint, as the right part of the figure displays.

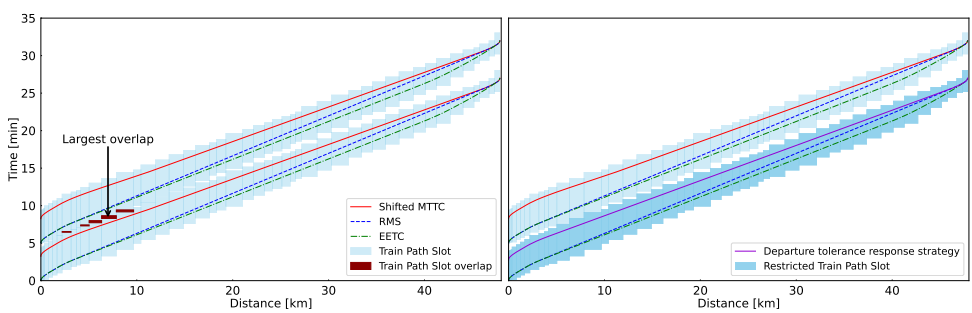


Figure 3.4: Time-distance diagram example of initial conflicting Train Path Slots (left) and optimised Train Path Slots after fine-tuning departure tolerance for the preceding train (right)

The third type of adapted Train Path Slots from the LP model necessitates one or more intermediate TP $s \in S_p$ in case the fine-tuned departure tolerances cannot resolve all the Train Path Slot overlaps. These intermediate TPs are located at the beginning of a block $b \in B_p$ such that the Train Path Slot constraint can be respected. Introducing intermediate TPs triggers a TP response driving strategy that considers the train dynamics and consequently modifies the blocking times. Therefore, each time an intermediate TP is defined, the Train Path Slots must be updated and re-evaluated to determine whether the overlap is resolved or if additional

intermediate TPs are required.

We propose a *critical-block TP configuration*, where an intermediate TP is added to the set S_{p_j} for the following train p_j in a successive train pair $\left(p_i, p_j\right) \in P_2$ at the beginning of the critical block, namely the block with the largest Train Path Slot overlap. Specifically, we add the intermediate TP among all possible TP locations $s_b \in S'_{p_j}$ where the value of the Train Path Slot overlap is maximised:

$$s = s_b \text{ with } b = \underset{b \in B_{p_j}}{\operatorname{argmax}} \left(c_{p_i,p_j,b} \mid \left(p_i, p_j\right) \in P_2\right). \tag{3.18}$$

This TP configuration imposes the most restrictive constraint on train trajectory generation by considering the tightest headway between two trains to ensure conflict-free operations. The concept of the critical block is also used in the International Union of Railways (UIC) timetable compression method to determine the minimum line headway time between two successive trains in (re)scheduling (Goverde et al., 2013; UIC, 2013). However, unlike in timetable compression, where headway times are computed based on single train trajectories, the blocking times here are directly influenced by the fixed scheduled headways from the timetable and the introduction of a TP at the critical block. Since the Train Path Slots account for a range of possible train trajectories, an iterative process is required to adjust train trajectories to fit the requested headways, identify new critical blocks, and introduce additional TPs where necessary.

The TP response strategy combines RMS and EETC driving strategies, using multiple response cruising speeds to meet the Train Path Slot constraints while including a coasting regime between the last intermediate TP and the next stop. First, we consider the previous stop with the scheduled departure time as a starting TP and use the prescribed intermediate TP at the critical block as an endpoint. Aiming for the maximum time window at the intermediate TP, we apply the Golden-section search to find the largest response cruising speed up to the defined intermediate TP. The search terminates once the Train Path Slot constraints at the critical block are met, or there exists only a sufficient buffer time between two trains. If Train Path Slot overlaps persist, additional intermediate TPs are introduced at the next most critical locations. Depending on whether the newly added intermediate TP is upstream or downstream of the previous one, the starting and endpoint of the response trajectory are updated accordingly to determine an optimal response cruising speed. When more than one intermediate TP is defined, maximum traction and service braking are applied to transition between different response cruising speeds. Once all Train Path Slot constraints are satisfied, we refine the response cruising speed between intermediate TPs (if more than one) for the maximum time window. Finally, we compute an energy-efficient train trajectory between the last intermediate TP and the next stop. Depending on the remaining running time supplement, the train either accelerates to an optimal cruising speed before coasting or transitions directly into coasting from the last cruising speed. As the running time supplement is reallocated between stops through intermediate TPs, the train progresses later in its trajectory and ensures sufficient buffer time for the successive train pair after the last intermediate TP.

Train Path Envelope computation with optimised timing points

The final module integrates both stages of Train Path Slot overlap resolution and assembles the optimised TPs with either time windows or target times into TPEs, facilitating conflict-free train operations. The locations and the temporal constraints of the TPs depend on the results from the Train Path Slot overlap detection and resolution modules. Hence, the set of TPs S_p for a train

$p \in P$ could be at stop locations or the beginning of blocks, where constraints must be imposed to train trajectory generation. We formulate the TPE as an ordered set of TP constraints as follows:

$$\text{TPE}_p = \left\{ \left(s, t^{\min}_{p,s}, t^{\max}_{p,s} \right),\ s \in S_p \right\} \quad \forall p \in P, \tag{3.19}$$

where $t^{\min}_{p,s}$ and $t^{\max}_{p,s}$ represent the lower and upper bounds of the departure or passing time window at the TP s based on the driving strategies that define the contours of a conflict-free Train Path Slot. The lower bound of a time window $t^{\min}_{p,s}$ denotes either the passing time of the TP response strategy at an intermediate TP or the scheduled departure time at a stop. The upper bound of a time window $t^{\max}_{p,s}$ corresponds to either the passing time of the shifted MTTC or the departure tolerance response driving strategy at an intermediate TP or the departure tolerance at a stop. If a TP specifies a fixed time (e.g., a scheduled arrival time at a stop), then the time window becomes a target time by having $t^{\min}_{p,s} = t^{\max}_{p,s}$.

We visualise the target time and the time window as two types of TP constraints in Figure 3.5. The left-hand side of Figure 3.5 represents a train run between two stops and has only one target arrival time at the next stop. In contrast, the right-hand side of Figure 3.5 is a restricted Train Path Slot that includes a fine-tuned departure tolerance at the previous station (0 m) and an intermediate passing time window at 26,000 m.

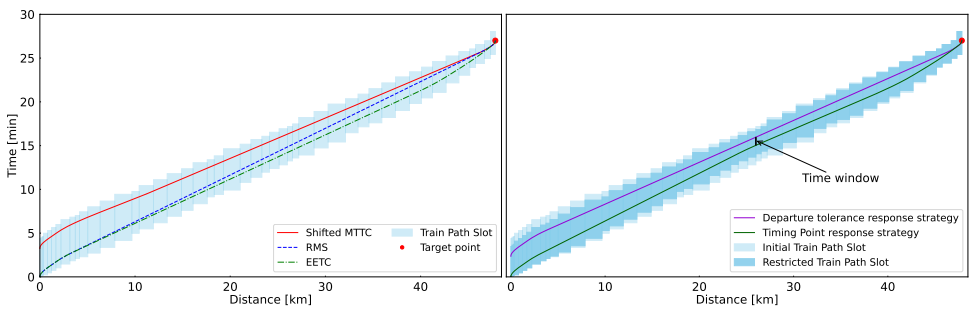


Figure 3.5: Time-distance diagram example of an initial Train Path Slot with just the target time (left) and an optimised Train Path Slot with both target time and time window at Timing Points (right)

3.3.4 Evaluation criteria

Alternative timing point configurations

To assess the effectiveness of the proposed *critical-block TP configuration*, we compare it against two other alternative approaches: the *scheduled TP configuration* and *earliest-conflict TP configuration*. The *scheduled TP configuration* represents the current practice, where only scheduled target departure and arrival times at stops are provided. On the other hand, the *earliest-conflict TP configuration* introduces an intermediate TP to the set S_{p_j} for the following train p_j at the beginning of the first conflicting block. In detail, this TP is placed at the entry location of the block with the lowest topological order among all conflicting blocks where the Train Path Slot overlap is positive:

$$s = s_b \text{ with } b = \underset{b \in B_{p_j}}{\operatorname{argmin}} \left(c_{p_i,p_j,b} \mid \left(p_i, p_j \right) \in P_2 \right). \tag{3.20}$$

Key Performance Indicators

The performance of different TP configurations can be quantified by several Key Performance Indicators (KPIs), considering both timetable-level and train operation-level metrics. From the timetable perspective, the Train Path Slot overlaps under a specific TP configuration are assessed, for instance, those occurring under the *scheduled TP configuration*. Additionally, the infrastructure occupation is quantified based on the UIC timetable compression method (UIC, 2013). From a train operation perspective, we consider the number of added intermediate TPs since this affects the efficiency of the onboard train trajectory generation and the quality thereof, particularly, energy efficiency. We calculate the energy consumption based on the train control problem using the mass-specific applied force and the total distance. Moreover, the number of added intermediate TPs also leads to a different number of possible braking indication points and driving regime changes, both of which have implications for passenger comfort. Note that the number of braking indication points refers to the deceleration action of the ATO onboard to avoid triggering emergency braking because of conflicting block sections. The above concepts are captured in several performance indicators as follows:

- *Train Path Slot overlap time*: The total Train Path Slot conflicting time over the entire railway corridor (in seconds).
- *Infrastructure occupation*: The share of time a given timetable pattern occupies a given infrastructure in an hour (in percentage).
- *Added intermediate Timing Points*: The number of added intermediate Timing Points to resolve the Train Path Slot overlaps.
- *Energy consumption*: The amount of energy consumed by the train operation (in kWh).
- *Braking indications points*: The number of braking indication points to prevent onboard emergency brake intervention.
- *Driving regime changes*: The number of driving regime switches in the computed train trajectory. For instance, the train driving shifting from cruising to braking is counted as one time.

3.4 Analysis and results

This section first presents the model verification with 16 scenarios in Section 3.4.1, demonstrating the effectiveness of fine-tuning departure tolerances and introducing critical-block TPs. The methodology is then applied to a real-world case study in Section 3.4.2, followed by a detailed performance evaluation in Section 3.4.3. Finally, Section 3.4.4 assesses the system improvements achieved by the proposed method.

All computational experiments are conducted on a laptop with an Intel(R) Core(TM) i7-10610U CPU (1.80 GHz) and 16 GB RAM. The LP model for conflict-free Train Path Slot determination is solved to optimality using PuLP, an open-source Python optimisation library, which employs the Revised Simplex Method by default to solve the LP problem. Train trajectory generations are solved using a numerical approach based on train dynamics and optimal control principles. The LP model execution time is typically under 0.1 s, while trajectory computation varies depending on the number of calculations and intermediate TPs required.

3.4.1 Model verification

The proposed method for determining an optimised set of TPs is verified in an experimental setup with 16 scenarios of different train pairs. These scenarios assess the effectiveness of fine-tuning the departure tolerance and the critical-block TP strategy. A 21.5 km route is defined, consisting of 17 block sections and 2 intermediate stations with varying gradients and speed limits. The first and second intermediate stations are 10,187 m and 15,092 m from the origin station, respectively. Both Intercity and Regional trains are operated on this test case with a maximum speed of 140 km/h. The scenarios vary by train order, stopping patterns, and train types. Specifically, an Intercity train may or may not stop at the first intermediate station, while a Regional train may stop at the second intermediate station, both stations, or neither. The service braking rates of Intercity and Regional trains are -0.8 m/s^2 and -0.66 m/s^2, respectively. These values are based on the operational braking characteristics of Dutch passenger trains and are consistent with those used in previous train trajectory optimisation studies. Since only braking rate data was available to describe braking behaviour, the braking force is calculated as the product of braking rate and train mass, plus resistance forces. For a more detailed description of the basic train characteristics used, see Goverde et al. (2021). The blocking time-related parameters in Eq. (3.1)-(3.2) and Eq. (3.4)-(3.5) are set from ETCS Level 2 characteristics and consider the potential influence of ATO, following consultation with the Dutch railway infrastructure manager, ProRail. Specifically, the values are set as follows: $t_{p,b}^{\text{release}} = 2$ s, $t_p^{\text{buffer}} = 30$ s, $t_{p,b}^{\text{reaction}} = 1$ s, and $t_{p,b}^{\text{setup}} = 1$ s for open track. In station areas or junctions, the setup time is defined as $t_{p,b}^{\text{setup}} = 6 \times$ number of switches (in seconds). For ETCS Level 2, the setup time may also include additional processing and communication delays. The 30 s buffer time accounts for these variations and provides sufficient margin for longer setup times when required. The parameter values can be adjusted when better estimates are available without impact on the method. In this test case, the intermediate stations are actual stations with switches.

Table 3.6 presents the computational results for the 16 scenarios, reporting the stop pattern and the train type of both the preceding and following trains, their headway, Train Path Slot overlaps, fine-tuned departure tolerance at stops of the preceding train, intermediate TP locations of the following train, corresponding block indices, time windows at these intermediate TPs, and total computation time for each scenario. The headway values in Table 3.6 are systematically selected to evaluate the effectiveness of the proposed method for TP optimisation under varying traffic conditions. These values distinguish cases where departure tolerances can be fine-tuned to resolve blocking time conflicts from those that require intermediate TPs. For excessively tight headways, the TPE generation module warns the TMS that no feasible conflict-free solution exists. For a certain train pair, we first only utilise fine-tuning the departure tolerance of the preceding train to avoid Train Path Slot overlaps. Then, we illustrate the usefulness of determining intermediate TPs for the following train when the headway is shortened, and no departure tolerance at the previous stop is allowed. For scenarios where the preceding train makes multiple stops, multiple values for departure tolerances are displayed. Similarly, when multiple intermediate TPs are required, their locations and time windows are reported accordingly.

If the departure tolerance fine-tuning cannot fully resolve the Train Path Slot overlaps, intermediate TPs at the entry locations of the most critical blocks provide a further means of resolving overlaps, provided a feasible solution space exists. The results indicate that, in most cases, a single intermediate TP suffices. However, in scenario 4, a lower response cruising speed imposed by an intermediate TP at block 5 successfully resolves the conflict at that location, but causes a new overlap at block 4. In such cases, an additional intermediate TP is necessary at the

Table 3.6: Computational results for all scenarios

Scenario	Preceding train [# stops]	Following train [# stops]	Headway [s]	Train Path Slot overlap [s]	Departure tolerance at stops [s]	Timing Point locations [m]	Block index	Time windows [s]	Computation time [s]
1	Intercity (0)	Intercity (0)	210	80	45	-	-	-	6
2	Intercity (0)	Intercity (0)	180	110	0	14248	block 12	71	7
3	Intercity (0)	Intercity (1)	216	74	78	-	-	-	2
4	Intercity (0)	Intercity (1)	132	158	0	3664, 1929	block 5, block 4	60, 76	12
5	Intercity (1)	Intercity (0)	480	39	full, 48	-	-	-	4
6	Intercity (1)	Intercity (0)	318	81	full, 0	14248	block 12	82	4
7	Regional (0)	Regional (0)	240	76	77	-	-	-	2
8	Regional (0)	Regional (0)	174	142	0	12471	block 11	89	7
9	Regional (0)	Regional (1)	222	94	57	-	-	-	4
10	Regional (0)	Regional (1)	156	160	0	1929	block 4	108	6
11	Regional (1)	Regional (0)	276	40	full, 12	-	-	-	4
12	Regional (1)	Regional (0)	258	58	97, 0	16027	block 14	66	10
13	Regional (1)	Intercity (0)	330	26	full, 28	-	-	-	2
14	Regional (1)	Intercity (0)	300	56	full, 0	16027	block 14	76	4
15	Regional (2)	Intercity (0)	420	37	full, full, 13	-	-	-	2
16	Regional (2)	Intercity (0)	396	61	full, full, 0	14248	block 12	95	6

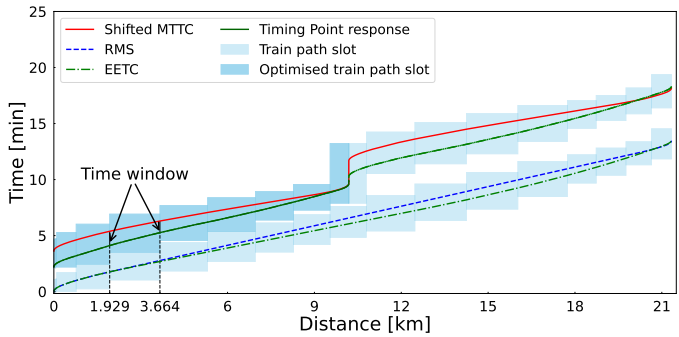
updated most critical block to ensure conflict-free train operations.

For each scenario of a train pair, we observe that the computation time increases when fine-tuning the departure tolerance cannot fully resolve the Train Path Slot overlaps and intermediate TPs are needed with additional calculations. Among the computational steps, generating the TP response train trajectory demands the most computational time. On the one hand, 2 to 6 seconds are required to resolve the Train Path Slot overlaps by fine-tuning the departure tolerances, depending on the number of iterations needed to address the largest overlap in each step. On the other hand, the computation time for determining a feasible TP response train trajectory ranges from 2 to 12 seconds, depending on the number of intermediate TPs introduced. This variation is due to the need to iteratively determine response cruising speeds that satisfy the Train Path Slot constraints.

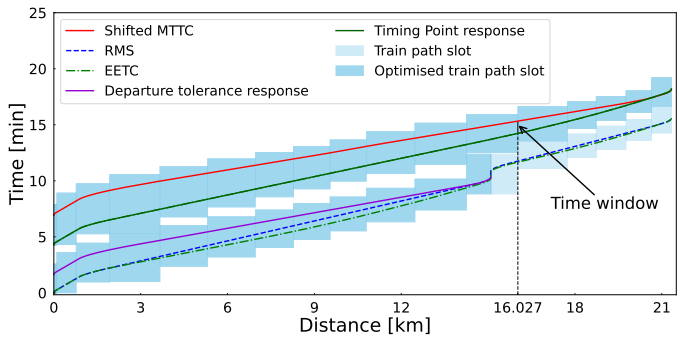
The results further show that the placement of intermediate TPs is influenced by the characteristics of train operations. In homogeneous traffic conditions, the intermediate TP is added mostly halfway between two stops because the energy-efficient train driving deviates gradually from the scheduled train path and the buffer in the timetable is insufficient to absorb this deviation. If the preceding train has multiple stops, the intermediate TPs are mainly specified for the following train after the last stop of the preceding train, where the following train is likely to approach the preceding one. Conversely, if the following train has multiple stops, the intermediate TPs are usually given in the first part of the journey, where the headway is tight. We visualise two examples of the 16 scenarios in Figure 3.6 with scenarios 4 and 12.

3.4.2 Case study description

Our case study takes place on one of the busiest railway traffic corridors in the Netherlands between Utrecht and 's-Hertogenbosch. This fifty-kilometre-long railway corridor includes nine passenger stations: Utrecht (Ut), Utrecht Vaartsche Rijn (Utvr), Utrecht Lunetten (Utln), Houten (Htn), Houten Castellum (Htnc), Culemborg (Cl), Geldermalsen (Gdm), Zaltbommel (Zbm) and



(a) Scenario 4



(b) Scenario 12

Figure 3.6: Optimised Train Path Slots for scenarios 4 and 12

's-Hertogenbosch (Ht). In particular, the Ut-Htnc railway is four-track, and station Gdm has 7 tracks, which permits overtaking. The rest is double-track (one track per direction). We use the periodic timetable of 2022 for the case study corridor (single direction from Ut to Ht). The railway timetable repeats itself every half-hour, featuring three Intercity and three Sprinter (i.e., regional) services per cycle. Nevertheless, the departure intervals of trains are not regular in this alternative Intercity/Sprinter periodic timetable due to the need to plan freight train paths and other constraints (e.g., the arrival of the same rolling stock from the opposite direction or infrastructure limitations). The Intercity stops at the two corridor ends, while the Sprinter stops at every station. As the trains from different directions are independent of each other, we only consider the trains from Ut to Ht within each 30-minute cycle.

The train parameters (e.g., maximum speed, braking rates, and acceleration characteristics) and fixed blocking time components (e.g., reaction times, release times) are consistent with the model verification subsection. Additionally, the speed limit and gradient profile information can be inferred from the speed profile presented in the following subsections. The infrastructure occupation is 62.11% (corresponding to a compressed timetable of 1118 s in 1800 s), calculated following the compression method of the UIC 406 method and the blocking time stairways based on the RMS driving strategy. This is in line with the usual timetable planning approach of infrastructure managers. Note that infrastructure occupation is computed from the first train in a

half-hour cycle to the same train in the next cycle. The infrastructure occupation value is below the recommended value of 75% during peak periods for mixed-traffic corridors.

3.4.3 Results

We start by analysing the Train Path Slots for the case study, as illustrated in Figure 3.7. Generally, the EETC-based train paths occur earlier than the RMS-based ones since the optimal cruising speed is usually higher to allow coasting at a later stage, causing variations in train paths between stops, as only scheduled train event times at stops are specified in the timetable. As a first robustness check and initial Train Path Slot computation, the initial Train Path Slots are calculated using the full departure tolerances, which results in Train Path Slot overlaps. These overlaps indicate potential train path conflicts and serve as input to the LP model, after which departure tolerances are fine-tuned. In total, four Train Path Slot overlaps are detected after the departure of Sprinter trains from stations Htnc, Cl, and Zbm, with a cumulative overlap duration of 1352 s.

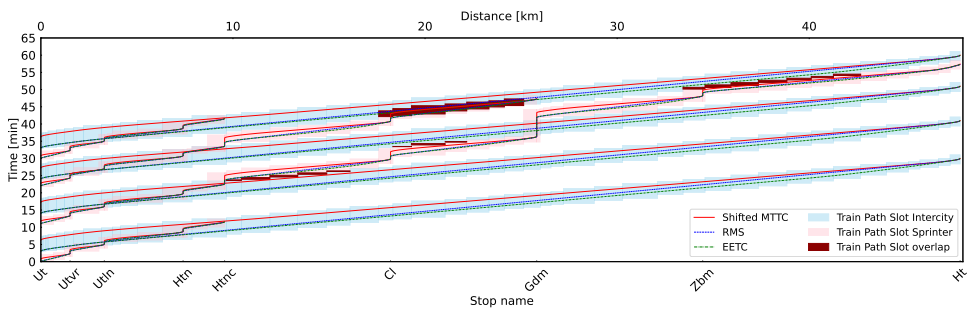


Figure 3.7: Train Path Slots for the original timetable

The proposed LP model is then applied to determine conflict-free Train Path Slots and identify the necessary TP configuration for each train. As shown in Figure 3.8, no Train Path Slot constraints are required for the first Sprinter and third Intercity. Hence, the departure time windows for these trains remain equal to the running time supplement, and the scheduled arrival times at stops correspond to their TP configurations in the TPEs. However, departure tolerance adjustments are required for the second Intercity and second Sprinter at station Cl. Furthermore, the departure tolerance is set to zero for the second Sprinter at station Zbm and the third Sprinter at station Cl. Lastly, intermediate TPs must be introduced for the fourth Intercity train, i.e., the first train in a half-hourly periodic timetable.

Building upon the conflict-free Train Path Slots determined by the LP model, the proposed *critical-block TP configuration* fine-tunes departure tolerances and introduces intermediate TPs at the critical blocks. After applying the fine-tuning departure tolerance algorithm, the departure tolerance of the second Intercity is shortened by 50 s (from 195 s to 145 s) to resolve the Train Path Slot overlaps after the departure of the second Sprinter at station Htnc. For the second Sprinter, its departure tolerance at station Cl is reduced by 37 s (from 76 s to 39 s), and a scheduled departure time is imposed at station Zbm (zero departure tolerance). The third Sprinter also has zero departure tolerance at station Cl. Even with zero departure tolerance, a buffer time of 30 s is still available to accommodate small departure delays. To resolve the remaining overlaps, an intermediate TP is placed at the start of the last block before station Gdm for the

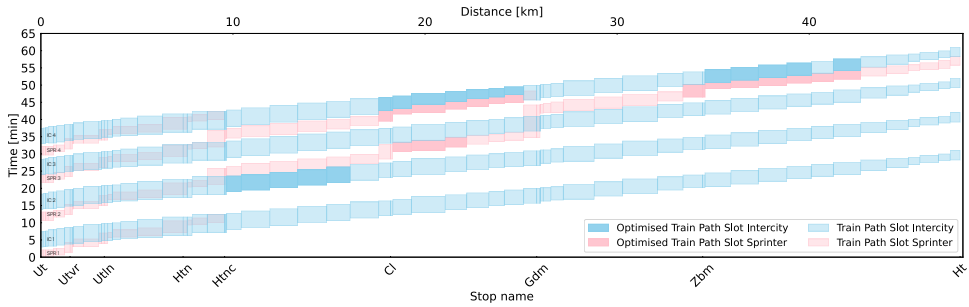


Figure 3.8: Optimised Train Path Slots based on the Linear Programming model

fourth Intercity. This TP is positioned at the most restrictive block, ensuring that all residual conflicts are eliminated. This location of the critical block has been validated against real-life track occupation data provided by the Dutch railway infrastructure manager, ProRail. The results confirm that the conflict detected before station Gdm is observed in 46% of the operated daily schedule in real-world operations.

The TP response strategy uses almost all the running time supplement to meet the Train Path Slot constraint, with approximately 13 s left as the running time supplement for the rest of the journey. The TP response driving strategy avoids any extra braking. The train performs one additional acceleration regime after the newly introduced TP and then attains another cruising speed, followed by coasting and braking before reaching the next stop, as shown in Figure 3.9. Beyond this intermediate TP, no further intermediate TPs are required after station Zbm. This is because the train compensates for its adjusted trajectory by spending additional time in the section between station Ut and the intermediate TP, meeting all Train Path Slot constraints. Hence, the departure tolerance for the conflicted Sprinter train can be fine-tuned again, as illustrated in Figure 3.10.

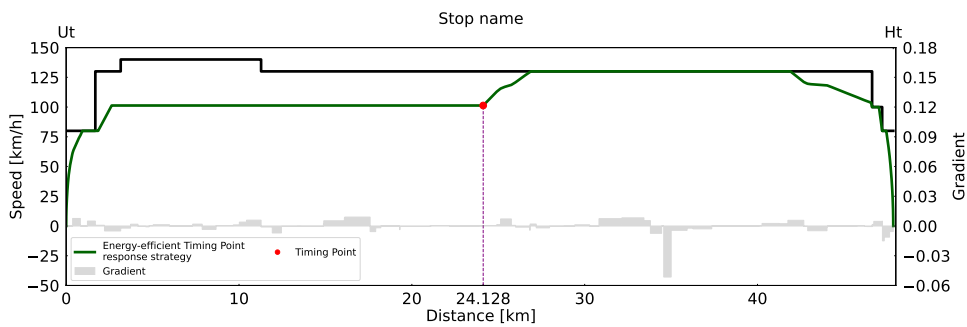


Figure 3.9: Timing Point response speed profile of the conflicted train for the critical-block Timing Point configuration

The total computation time of the optimised ATO TPs for a given timetable depends on the corridor length, the number of conflicting train pairs and the number of intermediate TPs required. It takes around 11 s to generate the initial Train Path Slots for all trains in the case study, including blocking time computation for the given three train trajectories per train, shared track identification, and visualisation of the initial Train Path Slots with departure times added,

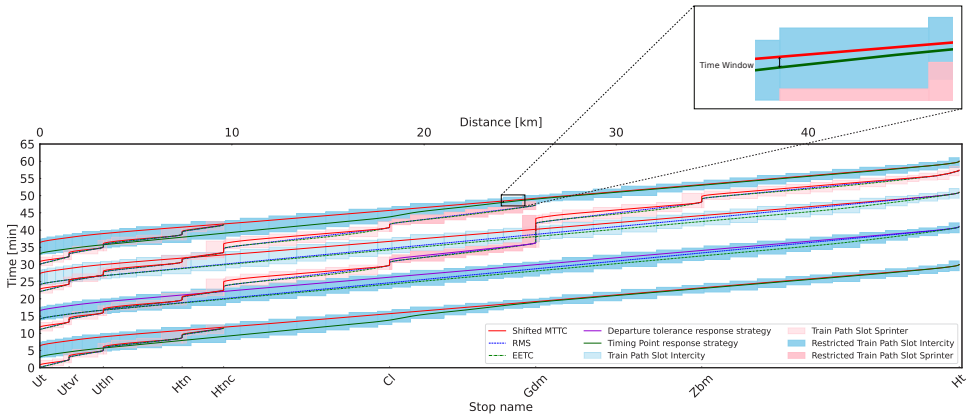


Figure 3.10: Train Path Slots with the critical-block Timing Point configuration

excluding the pre-computed train trajectories used to build the slots. The LP model execution time is 0.1 s, while fine-tuning the departure tolerances of all conflicted trains takes around 12 s. Determining the time window for the critical-block TP configuration needs approximately 6 s to determine the optimal cruising speed of the response speed profile before the intermediate TP. In total, the computation time for the entire case study, involving eight trains with an average departure headway of 5 minutes, is 29 s.

3.4.4 System improvement

To evaluate the system improvement achieved by the proposed *critical-block TP configuration*, we compare its performance with two benchmark cases: the *scheduled TP configuration* and the *earliest-conflict TP configuration*. The *scheduled TP configuration* represents the current practice, where TPs are placed only at scheduled stops. However, as the optimisation results indicate, relying solely on restrictive scheduled departure time windows cannot fully resolve Train Path Slot overlaps, which equals to 623 s between the second and third Sprinters and the fourth Intercity, as shown in Figure 3.11.

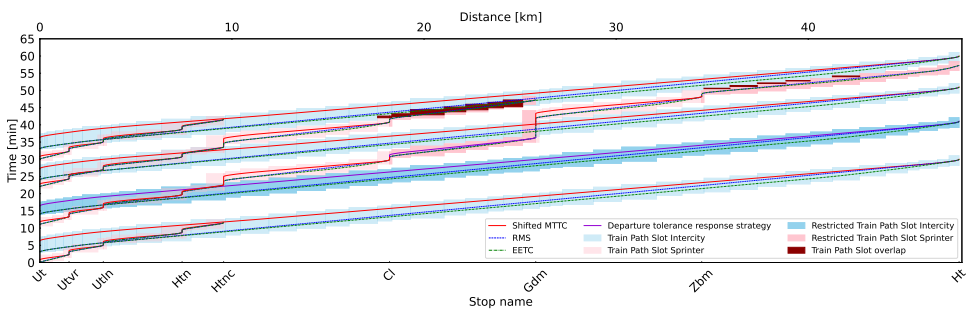


Figure 3.11: Train Path Slots with the scheduled Timing Point configuration

Without adding intermediate TPs, the fourth Intercity must react to the signalling system.

Before reaching conflicting block sections, it follows the optimal cruising speed based on the EETC driving strategy. However, its ATO onboard system then issues a speed reduction command, keeping the train below the braking curve computed by the ETCS onboard system before reaching the conflicting block. The train decelerates accordingly to avoid triggering emergency braking and re-accelerates as soon as its rear has passed the speed reduction section. Yet, this acceleration may trigger another braking curve, as the speed increase could again conflict with the block occupation by the preceding train. The incapability of anticipating the upcoming signals in real-time can lead to a zig-zag movement, as shown in Figure 3.12. Compared to nominal EETC-based operation (without train path conflicts), this Intercity needs to comply with five extra braking indication points and undergoes ten driving regime changes within the Train Path Slot overlapping area. After the conflict area, only 14 s of running time supplement remain, leading the train to adopt an optimised cruising speed below the speed limit to arrive at the next stop punctually. This inefficient response increases energy consumption by 177.99 kWh, representing a 53.2% increase compared to the 334.63 kWh required under nominal EETC operation. The realised infrastructure occupation time is 2455 s per hour, resulting in a 9.79% increase in infrastructure occupation ratio, reaching 68.19%, considering the signalling response driving strategy of the fourth Intercity and the EETC driving strategy for the other trains. In contrast, the *critical-block TP configuration* increases infrastructure occupation time to 2495 s per hour, resulting in an 11.59% increase in infrastructure occupation ratio compared to the baseline of 62.11%, which rises to 69.31%. However, its TP response driving strategy substantially improves energy efficiency, reducing consumption by 152.53 kWh, a 29.75% decrease compared to the signalling response strategy.

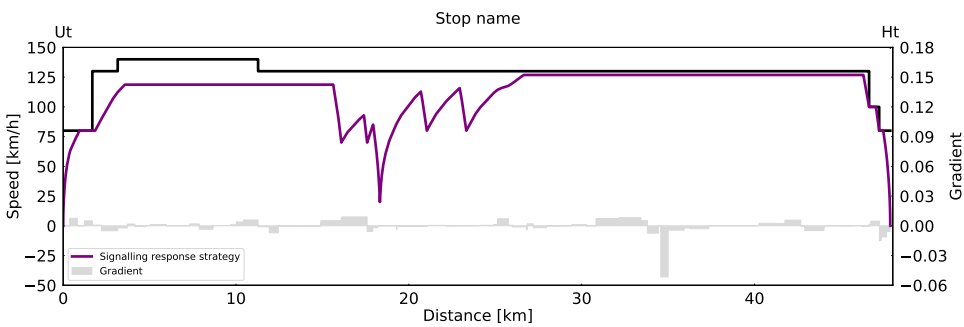


Figure 3.12: Signalling response speed profile of the conflicted train for the scheduled Timing Point configuration

Another benchmark, the *earliest-conflict TP configuration*, fine-tunes departure tolerances and introduces intermediate TPs, starting from the first conflicting block. After an iterative process to meet the conflict-free Train Path Slot constraints from the optimisation model, six additional TPs are required for the fourth Intercity, with the last one corresponding to the critical block TP. The first TP is provided at the beginning of the block before station Cl, while the second and the third TPs are placed at the start of the following two blocks since the Train Path Slot overlap gradually increases across these sections. The same pattern is also seen for the fourth, fifth and sixth TPs, positioned at the start of the last three blocks before station Gdm. Therefore, a larger Train Path Slot overlap in downstream blocks imposes more restrictive constraints for the train trajectory computation, which requires additional TPs to ensure conflict-

free operations. Since the LP model aims to maximise flexibility in blocking time windows, the time windows at intermediate TPs also strive to be as large as possible. To assess sensitivity, we also tested decreasing the window size for the first intermediate TP to make it more restrictive, which resulted in a lower total number of required TPs. However, the TP at the critical block remains essential to achieving effective conflict-free train operations with the given running time supplement. This result reinforces the rationale for adopting the *critical-block TP configuration*. Compared to the signalling response strategy, the TP response strategy for the *earliest-conflict TP configuration* reduces energy consumption by 96.75 kWh (18.87%), requires two fewer braking indication points and results in two fewer driving regime changes. We present the realised Train Path Slots for the earliest-conflict TP configuration in Figure 3.13.

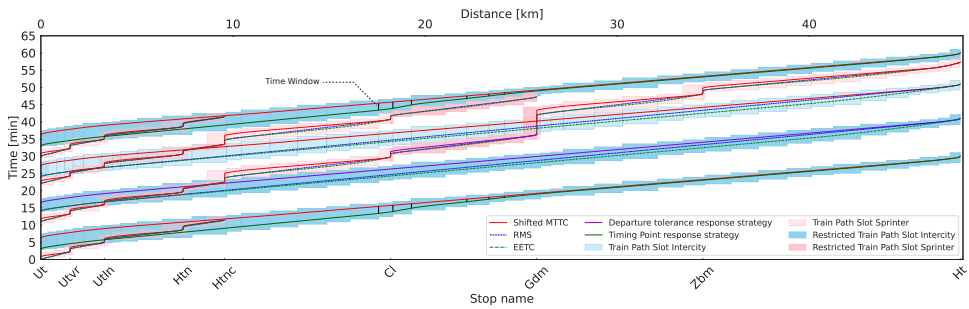


Figure 3.13: Train Path Slots with the earliest-conflict Timing Point configuration

Table 3.7: The results of different Timing Point configurations

	Figure	Slot overlap time [s]	Infrastructure occupation [%]	Added intermediate Timing Points	Energy consumption [kWh]	Extra braking indication points	Extra driving regime changes
Scheduled Timing Point configuration	3.11	623	68.19%	0	512.62	5	10
Critical-block Timing Point configuration	3.10	0	69.31%	1	360.09	0	1
Earliest-conflict Timing Point configuration	3.13	0	69.31%	6	415.87	3	8

We report the results of the three TP configurations in Table 3.7, using the selected KPIs to compare their impact on system performance. In this table, columns regarding added intermediate TPs, energy consumption, extra braking indication points, and extra driving regime changes are quantified based on either the signalling response driving strategy or the TP response strategy of the conflicted train. Extra braking indication points and driving regime changes are benchmarked against the nominal EETC train driving strategy. The *scheduled TP configuration*, representing current practice, demonstrates that TPs at scheduled stops are insufficient, as they fail to resolve the Train Path Slot overlaps. Without intermediate TPs, conflicts persist, leading to inefficient train operations with frequent braking interventions and increased energy consumption. This highlights the necessity of optimising TP configuration to ensure conflict-free train operations.

In contrast, both the *earliest-conflict* and *critical-block TP configurations* eliminate Train Path Slot overlaps and improve operational efficiency. However, the *earliest-conflict TP configuration*

requires six additional TPs to achieve conflict-free operations, introducing multiple constraints that restrict train trajectory generation and tracking flexibility. While it resolves conflicts, it also leads to higher energy consumption and frequent braking interventions as the train must comply with multiple intermediate TPs across its journey. The *critical-block TP configuration* ensures conflict-free operations without imposing excessive trajectory constraints by introducing a single TP at the most restrictive block. Compared to the *earliest-conflict* TP configuration, it significantly reduces energy consumption, braking interventions, and driving regime changes, while maintaining the same infrastructure occupation efficiency. As a result, the *critical-block TP configuration* proves to be the most effective approach, outperforming the alternatives across almost all KPIs.

3.5 Discussion

The proposed method for optimising TPs in TPEs provides practical decision-making benefits for railway infrastructure managers, undertakings, and ongoing standardisation efforts, such as those within the European Rail Traffic Management System (ERTMS) and ATO development. This method establishes a structured interface between traffic management and train operations, ensuring a decoupling of trackside and onboard functions for conflict-free train operations while maintaining driving flexibility. This TPE computation can function as an independent module, either within an ATO trackside or as part of the TMS.

This study advances the formalisation of open issues within ERTMS/ATO, particularly in the requirements specification between TMS and ATO trackside (ERA * UNISIG * EEIG ERTMS USERS GROUP, 2023b), where TP granularity remains undefined for effectively guiding ATO. The proposed critical-block TP configuration introduces a systematic approach to this process, which ensures conflict-free train operations without necessitating major modifications to the existing system architecture. This approach improves traffic fluidity while preserving operational flexibility for railway undertakings and offers a cost-effective alternative to capacity expansion. These findings align with the objectives of Europe's Rail FP1-MOTIONAL project, which aims to enhance traffic management and integrate digital solutions for more resilient and optimised rail operations.

The method applies to both vertically integrated and separated railway structures, aligning timetable-based traffic management with speed-based train operations. Traditional timetable (re)scheduling approaches assume deterministic train trajectories, whereas real-time train operations introduce variability due to different driving strategies and other operational conditions. The proposed Train Path Slot concept bridges this gap by considering multiple driving strategies and incorporating buffer time to accommodate operational uncertainties.

Beyond train path conflict resolution, the model improves train path robustness and energy efficiency. Fine-tuned departure tolerances establish a structured upper bound on allowable departure delays and, therefore, on the latest possible train running time, which enhances the tractability of conflict detection algorithms and the adaptability of train operations. By strategically adjusting TP locations and departure tolerances, trains proactively optimise their trajectories based on the determined TPs and reduce unnecessary driving regime transitions. This approach enables a network-wide green wave, where trains operate at optimised speeds under given headways from the real-time traffic plan without unneeded slowdowns due to restrictive signalling. Once the real-time traffic plan is updated, automated TPE optimisation streamlines downstream train operations, which is applicable in both human-supervised and automated TMS

environments.

While the proposed method optimises TP configurations and generates conflict-free TPEs, certain limitations remain that may affect its scalability and real-time applicability. First, computational efficiency poses challenges for large-scale real-world implementations. Although the LP-based optimisation efficiently resolves conflicts, repeated multiple train trajectory generation requires substantial computation time. This becomes increasingly demanding in networks with diverse train parameters and dynamic operating conditions. Scaling the approach to high-frequency corridors with rapid decision-making requirements remains a challenge.

Second, the selection of driving strategies for the initial Train Path Slot computation follows the most common and logical approaches in train trajectory optimisation, reflecting realistic ATO deployment strategies. These representative strategies define upper and lower bounds for multiple possible train trajectories with a fixed arrival time. Maximal Coasting (MC) (Scheepmaker et al., 2020) could serve as an alternative to EETC, further widening the Train Path Slot by extending its lower bound. Since MC applies maximum acceleration until reaching the speed limit, cruising at the speed limit, followed by a coasting regime, it may lead to earlier conflicts with the preceding train, potentially changing the location of the critical block and affecting the determination of intermediate TPs. The disadvantage is that the wider Train Path Slot could require additional TPs to ensure conflict-free operations. In addition, this study assumes fixed target arrival times. Future research could explore whether allowing arrival time windows may help resolve conflicts in certain cases, provided that it does not compromise punctuality and other operational requirements.

Finally, the current method focuses on TPE computation in a static environment based on a predefined timetable but can be extended to dynamic situations, where TPEs adapt to real-time traffic conditions and operational uncertainties. Future extensions may incorporate continuous updates, allowing TPEs to dynamically adjust based on real-time traffic plan changes from the TMS, status reports from the ATO onboard, and stochastic disturbances or disruptions. This would further contribute to a more adaptive and resilient ATO-driven, TMS-integrated railway digitalisation framework.

3.6 Conclusion

This paper introduced a novel method for optimising Timing Points (TPs) and generating Train Path Envelopes (TPEs), ensuring effective and conflict-free Automatic Train Operation (ATO) for a given real-time traffic plan. The optimised TPs within a TPE serve as constraints for onboard train trajectory generation, thereby aligning speed-based train operation with the timetable-based traffic management.

A key contribution of this work is the introduction of the Train Path Slot concept, which defines a bandwidth of track occupation times corresponding to multiple train driving strategies. Using this approach, we demonstrated that a (re)scheduled timetable may still contain train path conflicts, as conventional (re)scheduling algorithms typically assume a single deterministic train trajectory. Consequently, achieving conflict-free train operations requires intermediate TPs at strategic locations rather than solely relying on scheduled target departure and arrival times at stations.

To identify and avoid these conflicts, we formulated a two-step optimisation process, in which a Linear Programming model was developed to derive conflict-free Train Path Slots, which are subsequently applied as blocking time constraints to compute the TPEs. The proposed

method was verified in various experiments and validated on one of the busiest Dutch railway corridors with a half-hourly periodic timetable. The real-life case study results demonstrated that introducing intermediate TPs at the beginning of blocks with the most significant overlapping blocking times effectively resolved conflicts and outperformed other alternatives. Moreover, the case study confirmed that the proposed TPE generation method was able to detect and eliminate all train path conflicts by adding a minimal number of TPs per train, and thus preventing unnecessary braking and reducing energy consumption by 29.75% for the affected train run.

Future research on TP optimisation will focus on the dynamic adjustment of TPEs based on real-time traffic plan or train status updates. Additionally, alternative train driving strategies, such as incorporating regenerative braking, may be explored as extensions of the proposed method. While these extensions enhance operational efficiency, they do not modify the underlying modelling framework established in this study.

Chapter 4

Sensitivity of train path envelopes for automatic train operation

Prior to closed-loop simulation or integration testing, understanding the responsiveness of TPEs to dynamic railway conditions is essential for their effective implementation. This chapter investigates the sensitivity of TPEs, computed using the approach developed in Chapter 3, to variations in timetable updates from the traffic management system and real-time train status reports. A sensitivity analysis using elementary effects is proposed to quantify the impact of input variations on three key outputs of TPE generation: departure tolerance at stations, time windows at control timing points, and operational tolerance during train runs. A real-life case study on a Dutch railway corridor reveals that control timing points can be introduced into the TPE as headways decrease, to homogenise traffic by better aligning speed profiles and thus resolving conflicts. Timing point locations remain mostly unchanged, while their associated time windows become more sensitive when positioned further along the route. Operational tolerance, which defines the latest conflict-free passing time, becomes more sensitive to headway changes and the distance from the previous stop. The results help identify when speed adjustments based on TPE remain sufficient and when upstream real-time traffic plan updates become necessary to maintain conflict-free and flexible train operation.

Apart from minor changes, this chapter has been published as:

Wang, Z., Quaglietta, E., Bartholomeus, M.G.P., & Goverde, R.M.P. (2025). Sensitivity of train path envelopes for automatic train operation. *IEEE Transactions on Intelligent Transportation Systems*, 26(11), 18921-18933.

4.1 Introduction

Automatic Train Operation (ATO) aims at assisting or automating train driving tasks to improve running time reliability, energy efficiency and punctuality. While ATO is widely implemented in urban and metro railways with predefined train trajectories or coasting strategies (Yin et al., 2017), this study focuses on ATO in mainline railways. In this context, the ATO onboard component generates and tracks optimised train trajectories, adhering to arrival, departure, or passing times at specific locations known as Timing Points (TPs) (ERA * UNISIG * EEIG ERTMS USERS GROUP, 2023b,c). Previous work has proposed to use critical network locations as TPs to determine their time windows, which constitute a Train Path Envelope (TPE) that serves as constraints for generating each train's conflict-free trajectory (Wang, Z. et al., 2022; MOTIONAL, 2024). In this paper, conflict-free operation refers to the absence of blocking time overlaps between successive trains, ensuring that movement authorities can be extended without triggering braking or restrictive signalling (Goverde et al., 2013).

The TPE is aligned with the Real-Time Traffic Plan (RTTP) from the Traffic Management System (TMS), which specifies train routes with reference locations and times and may incorporate rescheduling decisions such as retiming, reordering, and rerouting (Quaglietta et al., 2016; Tschirner et al., 2014). The TMS dynamically maintains a conflict-free RTTP through proactive conflict detection and resolution, which is then translated into train-level trajectory constraints through the TPE computation, while the RTTP is also used for route setting and informing all other actors to guarantee an aligned coordination between train operations and traffic control. Real-time train statuses are also utilised to compute an effective TPE, ensuring its adaptability to evolving conditions.

Dynamic changes in RTTP updates and train status reports can affect the TPE, introducing fluctuations to the train trajectory generation constraints used by ATO onboard algorithms. Such variability may lead to infeasible trajectories or uncomfortable changes in train driving regimes, particularly when planning assumptions no longer hold. However, the sensitivity of TPEs to variations in RTTP updates and train statuses has not been studied yet in the literature.

To address this gap, this paper performs a sensitivity analysis using elementary effects Morris (1991) of the TPE computation under varying operational inputs at the short-term operational planning level, as part of the Europe's Rail FP1-MOTIONAL project (MOTIONAL, 2024). These inputs include RTTP variations, represented by planned headway adjustments, and train status changes, reflected through speed and position updates. The analysis quantifies the importance of these inputs and the extent to which their variations influence the computed TPs and their associated time targets or timing windows along the train route. This form of sensitivity analysis supports early-stage evaluation of operational robustness prior to closed-loop microscopic simulation or integration testing (NASA, 2024), and helps guide the alignment between discrete-event-based traffic management and continuous-speed-based train control. The contributions of this paper are threefold:

- Proposing a sensitivity analysis using elementary effects to quantify the impact of planned headways from the RTTP and updates in train statuses on the locations and time windows of timing points from the TPE generation;
- Applying the proposed method to a real-life case study on a Dutch rail corridor with heterogeneous traffic, demonstrating the effects of headway and status variability on TPE feasibility and responsiveness;

- Identifying when the TPE can effectively resolve conflicts via speed adjustment alone, and when replanning from the TMS is needed to maintain conflict-free operations.

The remainder of this paper is structured as follows: we provide a literature review in Section 4.2 followed by the methodology in Section 4.3. Then, we present and discuss the results in Section 4.4. Finally, Section 4.5 concludes the paper.

4.2 Literature review

4.2.1 Automatic Train Operation

ATO can partially or totally automate train driving functions. Five Grades of Automation (GoA) for ATO are defined in IEC (2014), ranging from manual on-sight driving (GoA 0) to fully unattended driving (GoA 4). From ATO GoA 2 onwards, train traction and braking commands are automated between successive stops for improvements in schedule adherence, capacity, energy efficiency and cost effectiveness compared to manual driving behaviour (Yin et al., 2017). While ATO has been widely adopted in metro systems, its implementation remains largely conceptual for mainline railways with linking to a TMS. This is primarily due to the complexities, such as heterogeneous rail traffic and complex operating environments (Wang, Z. et al., 2022).

To facilitate the implementation of ATO in mainline railways, recent advancements have led to technical specifications of ERTMS/ATO within the European Rail Traffic Management System (ERTMS) (ERA * UNISIG * EEIG ERTMS USERS GROUP, 2023b,c). Figure 4.1 illustrates the ERTMS/ATO reference architecture. Within this architecture, the European Train Control System (ETCS) onboard, operating as a safety-critical component, continuously supervises the ATO onboard using a dynamic speed profile until the end of the movement authority. The ATO trackside generates a Journey Profile, which comprises a list of Segment Profiles (representing route data), along with TP constraints and temporary restrictions (e.g., additional speed limitations and adhesion conditions) (ERA * UNISIG * EEIG ERTMS USERS GROUP, 2023c). This Journey Profile incorporates the TPE, thereby guiding the ATO onboard in generating conflict-free train trajectories while adhering to the current timetable. Furthermore, the ATO onboard provides the train status reports back to the ATO trackside, concerning the position, speed, and estimated arrival time at the upcoming TPs of a train (ERA * UNISIG * EEIG ERTMS USERS GROUP, 2023b).

The current specifications do not standardise the interaction between the TMS and ATO, particularly regarding the role of the ATO trackside. It is not specified which component computes the conflict-free TPEs. The TPEs may either be provided by the TMS, with the ATO trackside taking a passive role in transmitting them to the ATO onboard or actively generated by the ATO trackside itself before being sent to the ATO onboard. Besides, the feedback loops from ATO onboard status reports are also not standardised: the ATO trackside may either forward these reports to the TMS for traffic optimisation or use them to re-optimise the TPEs in a direct loop between the ATO onboard and trackside, enabling faster control responses. Determining the effective interaction between RTTP updates and TPE generation, including the necessary feedback mechanisms, is currently under investigation as part of MOTIONAL (2024).

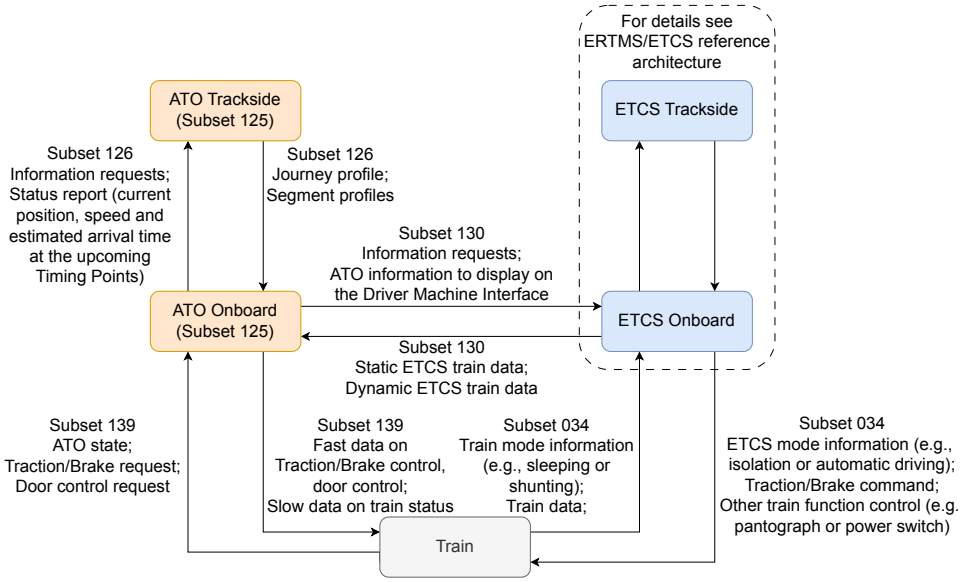


*Figure 4.1: ERTMS/ATO reference architecture, adapted from ERA * UNISIG * EEIG ERTMS USERS GROUP (2023b)*

4.2.2 Related work to train path envelope

Considerable literature exists on the development of TMS functions aimed at dynamically computing up-to-date schedules or RTTPs (Cacchiani et al., 2014; Pellegrini et al., 2015; Fang et al., 2015; Quaglietta et al., 2016). Despite this, automated railway traffic conflict detection and resolution are not yet applied in practice. Consequently, only a few attempts have been made to link TMS and train operation effectively.

On the one hand, traffic management can be integrated with train operation. For example, an integrated approach presented in Rao et al. (2016) computes conflict-free train trajectories at the TMS to derive train trajectory targets for the ATO. Similarly, some approaches utilise approximate discrete train speed profiles as a decision variable to jointly compute dispatching and train control solutions for minimising train delays (Luan et al., 2018a) or energy consumption (Luan et al., 2018b). This integration strategy is also developed in metro operations, where the objective typically revolves around optimising the energy efficiency of both the timetable and train operation (Su et al., 2020; Chen et al., 2023). Additionally, research has explored the selection of a suitable speed profile from a pre-generated set to achieve the same objective (Hou et al., 2019; Li et al., 2021).

On the other hand, the concept of a TPE is introduced to provide constraints in the form of time targets or windows at TPs for the train trajectory generation problem (Albrecht, T. et al., 2013; Quaglietta et al., 2016). This TPE defines an envelope around possible train trajectories, facilitating energy-efficient train operation while meeting timetable punctuality requirements (Wang, P. & Goverde, 2016; Wang, P. et al., 2020). However, TPEs computed for all trains using the given RTTP may not always ensure conflict-free operations. Therefore, a timetable constraint set has been proposed to fine-tune the values of time targets or windows at stations (Wang, P. & Goverde, 2017), which was further developed to achieve conflict-free energy-efficient timetables

(Wang, P. & Goverde, 2019). Moreover, adding control TPs between stops can enforce sufficient separation between trains by imposing time windows that trains must adhere to at these points (Albrecht, A.R. et al., 2015; Haahr et al., 2017; Howlett et al., 2023). To compute conflict-free TPEs for various trains, Wang, Z. et al. (2025a) developed a method to optimise the locations of TPs and their associated time windows, facilitating effective ATO onboard train trajectory generation and execution.

The sensitivity of the TPE with respect to regular updates of the RTTP and the stochastic evolution of rail traffic conditions over time has not been investigated yet. Understanding this sensitivity is crucial for identifying critical thresholds in variations of planned headways within the RTTP, as well as in train speeds and positions, to avoid infeasible trajectory generation or uncomfortable driving regime changes executed by the ATO onboard. This issue is closely tied to practical challenges in the development of ATO technologies.

4.3 Methodology

Based on the identified research gap, this paper investigates the sensitivity of TPEs in relation to variations in headways derived from rescheduled event times in the RTTP at the TMS, as well as dynamic changes in train speeds and positions based on feedback from train status reports provided by ATO onboards. An overview of the methodology is presented in Figure 4.2 and elaborated in the subsequent subsections. Section 4.3.1 introduces the overall problem of TPE generation sensitivity. Section 4.3.2 explains the TPE concept and the key computational steps implemented in the TPE generator, as illustrated in Figure 4.3. These steps include input data extraction, step-wise bandwidth blocking time computation (termed Train Path Slot, TPS) from multiple feasible trajectories, overlapping blocking time detection and resolution by determining control TPs and tolerances, and TPE construction to constrain ATO onboard trajectory generation. If conflicts cannot be resolved, infeasibility is reported to the TMS. Sections 4.3.3 and 4.3.4 describe the influence of RTTP updates and train status reports, respectively. Finally, Section 4.3.5 presents the sensitivity analysis using elementary effects to evaluate output variations.

4.3.1 Problem description and assumptions

This study investigates the sensitivity of TPE computation to variations in two key input sources, namely planned train headways from the RTTP and dynamic train status updates, including train positions and speeds. The analysis focuses on how these variations affect the outputs of the TPE generation, i.e., the computed departure tolerances at scheduled stops, the time windows at control TPs, and the operational tolerances during train operations. Quantifying these sensitivities provides insights into the feasibility and flexibility of the generated TPEs under different operational conditions.

The TPE computation aligns discrete-event-based RTTP updates from the TMS with continuous-speed-based trajectory generation at the ATO onboard system. Specifically, the TMS maintains a conflict-free RTTP through proactive conflict detection and resolution, and the TPE computation translates this plan into trajectory constraints for the ATO onboard system at the short-term operational planning level (Quaglietta et al., 2016). The TPE defines a bounded range of feasible train trajectories that ensure punctual arrival at the next stop, determined by the earliest and latest trajectories under different driving strategies.

To construct the TPE, we generate multiple feasible train trajectories under different driving

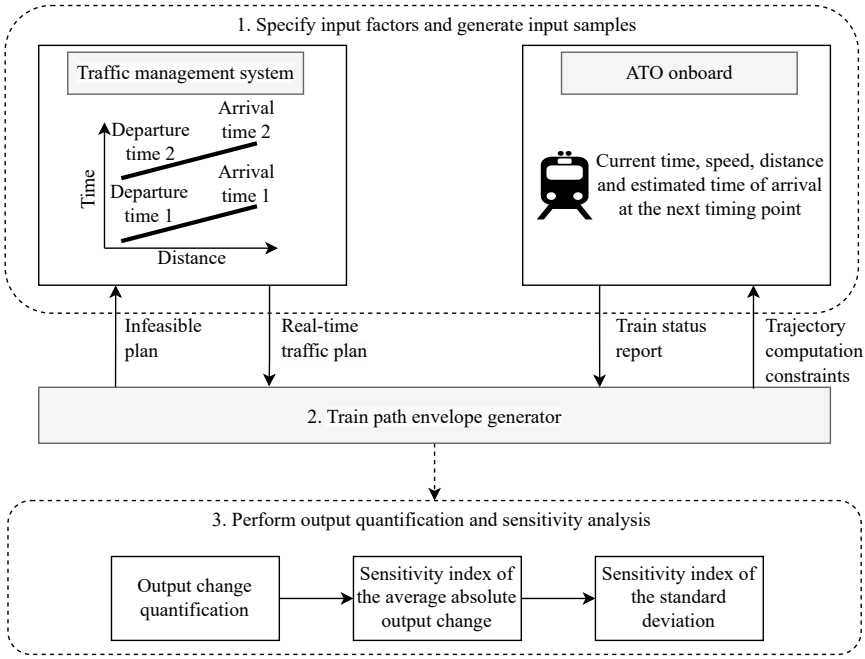


Figure 4.2: Overview of the TPE sensitivity analysis in response to RTTP updates and train status changes

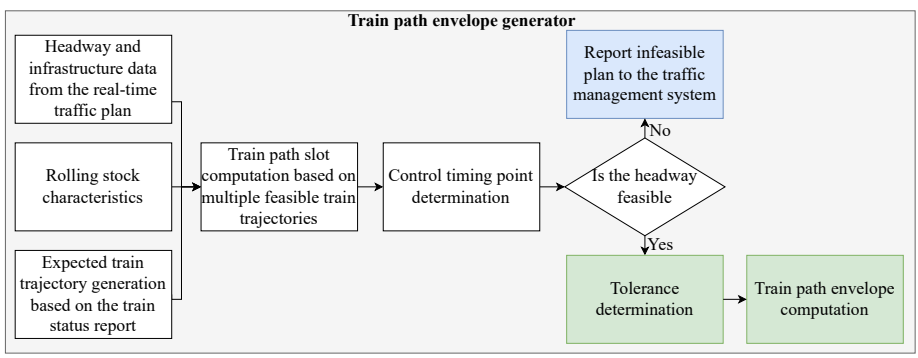


Figure 4.3: Overview of the TPE generator

strategies and derive their corresponding blocking times. The blocking time represents the exclusive track use, which comprises a setup time, sight and reaction time, the approach time to the block section, the running time within the block, the clearing time during which the train vacates the block over its entire length, and the release time of the route (Pachl, 2014). The subsequent blocking times of a train form a blocking time stairway in a time-distance diagram, while the compression of these stairways for successive trains determines the infrastructure occupation

over a corridor (Pachl, 2014). This paper focuses on radio-based fixed-block signalling systems, such as ETCS Level 2 with trackside train detection (Winter, 2009), but the proposed method can also be applied to traditional fixed-block multi-aspect signalling or moving block systems, which all fit the generic blocking time theory.

A conflict occurs when a train reaches its braking curve indication point, meaning it is too close to a block section still occupied by a preceding train and therefore must begin braking (Goverde et al., 2013). Such conflicts are detected through overlapping blocking times, which reflect both interlocking rules and signalling constraints (e.g., ETCS). For train operations to remain conflict-free, the blocking times of successive trains must not overlap within the same block section. This separation allows the following train to receive timely movement authority extensions and proceed without unnecessary braking. Potential blocking time conflicts are resolved by enforcing the earliest passage times at control TPs for the following train and adjusting the departure tolerance of the preceding train, thereby regulating train speed profiles. These adjustments result in a discrete set of TPs with associated time constraints, which together define the TPE and guide conflict-free train trajectory generation. If the TPE computation cannot resolve the conflict, the TMS is notified and must compute an updated RTTP to provide a revised conflict-free plan, e.g., due to insufficient line headway. Recall from Figure 4.1 that ETCS provides safety-critical speed and distance supervision. The corresponding constraints on train separation are modelled in the blocking times.

We assume a static computational setting, where the RTTP and train status reports are available instantaneously. The TPE is computed once per input configuration, representing an updated RTTP and train states. This enables a focused sensitivity analysis on key inputs: line headway, train position, and train speed. Communication between the ATO onboard and the ATO trackside occurs via Train-to-Ground, while communication between the TMS and ATO trackside occurs via Ground-to-Ground. No direct Train-to-Train communication is involved.

4.3.2 Train path slot and train path envelope

The TPE generator in Figure 4.3 implements the computational steps by considering multiple train driving strategies to compute the step-wise bandwidth blocking times (i.e., TPS) and resolving potential blocking time conflicts. This subsection explains the underlying concepts of TPS and TPE computation, control TP determination, and tolerance adjustment, which together enable the construction of a conflict-free TPE.

Typically, TMS algorithms assume a deterministic train trajectory to (re)schedule the timetable and provide an RTTP. This RTTP normally contains only target departure/arrival/passing event times at stations or junctions, which are adequate to ensure conflict-free train operations. The (re)scheduling process often involves presuming a Reduced Maximum Speed (RMS) train driving strategy, where trains cruise at a constant speed below the track speed limit to meet the target arrival time (Scheepmaker et al., 2020). However, a train can follow various trajectories while satisfying the dynamic equations of motion and complying with the same target arrival time at its next stop. One such trajectory is based on the Energy-Efficient Train Control (EETC), incorporating both cruising and coasting regimes (Albrecht, A.R. et al., 2016b; Goverde et al., 2021). Overlap(s) in blocking time between EETC- and RMS-based train paths can occur, requiring additional control TPs beyond the reference locations and times provided by the RTTP to resolve the conflicts. A particularly efficient method to mitigate these overlaps involves optimising TPs at critical blocks associated with the train trajectories (Wang, Z. et al., 2025a).

Specifically, critical block(s) are identified using the timetable compression method and indicate the block with either the shortest time between consecutive blocking times or the largest overlap in case of conflicts (Pachl, 2014).

Let $p \in P$ denote a train, and $(p_i, p_j) \in P_2$ represent an ordered successive pair of trains, where p_i precedes p_j. A control TP for train p_j, denoted as $s \in S_{p_j}$, is assigned at the entry location s_b of a shared critical block section $b \in B_{p_i} \cap B_{p_j}$. This critical block is defined by the largest blocking time overlap, $c^{\max}_{p_i,p_j}$, between these two trains. The passing time of train p_j at this control TP determines the earliest passage time at this critical block to avoid a conflict. In response to the added control TP, a TP Response (TPR) driving strategy adapts the train speed profile by combining RMS and EETC profiles. It employs a lower cruising speed to meet the added control TP and uses the remaining running time supplement for coasting during the later stages, thereby avoiding conflicts with other trains (see Wang, Z. et al. (2025a) for more details).

Furthermore, train departures may experience slight delays within an acceptable tolerance as long as a feasible trajectory exists for the train to reach its destination by the scheduled arrival time. This tolerance, derived from the allocated running time supplement, absorbs minor deviations from the planned schedule by establishing an upper bound on the train trajectories and enhances the robustness of train paths, provided there are no blocking time overlaps. The tolerance reaches its maximum when the train departs late up to the scheduled running time supplement and adopts the Minimum-Time Train Control (MTTC) strategy operating as fast as possible (Goverde et al., 2021; Albrecht, A.R. et al., 2016a), which is termed Shifted MTTC (SMTTC). To determine the tolerance of the preceding train p_i, it is initially set to the running time supplement and fine-tuned until either there are no blocking time overlaps or it reduces to zero. The Departure Tolerance Response (DTR) driving strategy is assumed as an RMS driving strategy that departs at the updated tolerance and arrives on time at the next stopping point.

Wang, Z. et al. (2025a) introduced the TPS as a step-wise bandwidth that includes the blocking time stairways of selected train driving strategies, such as RMS, EETC, and, where feasible, SMTTC, allowing the train to arrive on time at its destination. The TPS spans between the start of the blocking time of the earliest train trajectory (until that block) and the end of the blocking time of the latest train trajectory. We offer an example with three successive RMS-based blocking time stairways as in the TMS (re)scheduling algorithms, alongside the three corresponding initial TPSs integrating RMS and EETC strategies, as shown in Figure 4.4 (a) and (b). This example unfolds along an approximately 50-km-long railway corridor with varying gradients and speed limits. This illustration highlights the TPS overlap between the first and second trains (from bottom to top), requiring tolerance adjustments and possibly adding a control TP for the second train.

Figure 4.4 (c) presents the final three successive TPSs with optimised sets of TPs. It shows that the first train's tolerance is zero, the second is reduced from 195 (full running time supplement) to 141 seconds, and the third remains full. A control TP with a time window is defined at 30.285 km for the second train, and the associated time window is calculated based on the block entry time difference between different driving strategies. Lastly, the TPs are assembled in a TPE to be forwarded to the ATO onboard, which adds flexibility to the ATO onboard train trajectory generation.

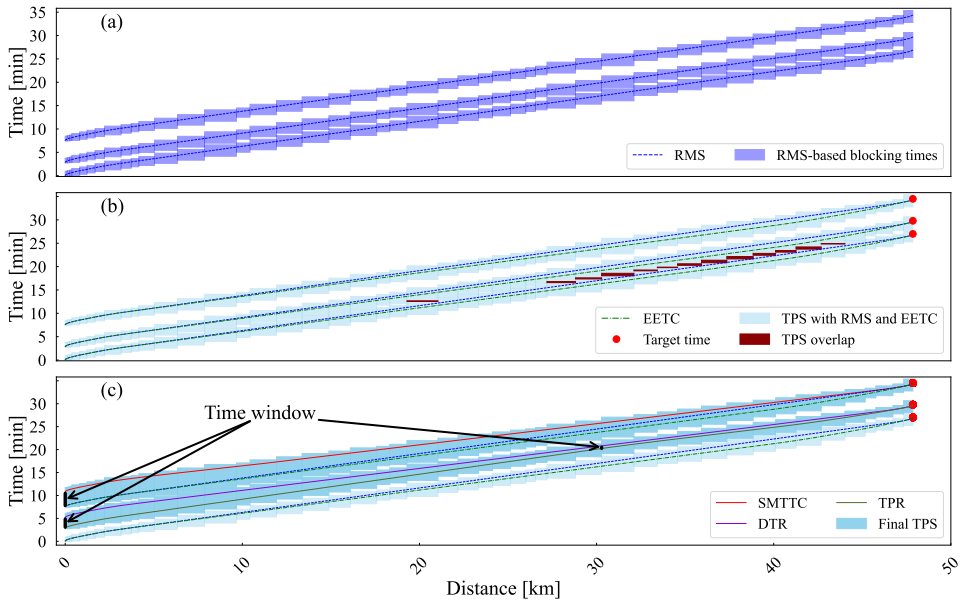


Figure 4.4: (a) Example of three successive conflict-free RMS-based blocking time stairways. (b) Example of three successive initial TPSs with only one target time and TPS overlaps, integrating RMS and EETC driving strategies. (c) Example of the three successive final TPSs with optimised sets of TPs

4.3.3 Train path envelopes and TMS updates

Daily railway operations may deviate from the schedule due to unforeseen events, e.g., extended running or dwell times. These traffic disturbances are monitored and predicted by the TMS, namely, relatively small path deviations without modifying the resources. If conflicts are detected, the TMS resolves them and offers an updated RTTP. Various control measures pertaining to retiming, reordering and rerouting, impact the RTTP by altering scheduled event times, which also implies adjusting train separation distances and times. These altered event times will influence the TPE determination.

In this subsection, line headways corresponding to various (re)scheduled event times within the RTTP serve as critical inputs for determining conflict-free train operations. A line headway is defined as the minimum time interval between successive train departures at the first station of a corridor that ensures conflict-free operations over their routes. These line headways represent potential RTTP variations for successive train paths, categorised into four boundary types, as summarised in Table 4.1 and detailed below.

Let $t_{p,b}^{E,\gamma}$ and $t_{p,b}^{L,\gamma}$ denote the running times to block b for train p according to the earliest and latest train trajectories, which depend on a specific driving strategy γ, where $\gamma \in \Gamma = \{\text{RMS, EETC, TPR, SMTTC, DTR, TPS}\}$. In particular, the TPS driving strategy is a special case representing a bundle of selected driving strategies. Figure 4.5 illustrates the various speed profiles corresponding to the time-distance diagrams presented in Figure 4.4 under identical conditions. The train trajectory modelling is specific to each train, accounting for speed limits, gradients, and train-dependent parameters.

Table 4.1: Different line headways and their assumed driving strategies

Line Headway Type	Meaning	Preceding Train Strategy	Following Train Strategy
TPS headway	Minimum robust line headway allowing all assumed driving strategies	All	All
Nominal driving headway	Minimum conflict-free line headway considering multiple driving strategies that comply with scheduled departure and arrival times, with zero tolerance for the preceding train at departure	RMS & EETC	RMS & EETC
Minimum line headway with one single driving strategy	Minimum conflict-free line headway using the same driving strategy of both trains	RMS | EETC | MTTC	RMS | EETC | MTTC
TPR headway	Minimum line headway with controlled following train, assuming no changes to scheduled departure and arrival events, with the following train reaching the control TP at its earliest permissible time and the preceding train following punctual driving strategies	RMS & EETC	TPR

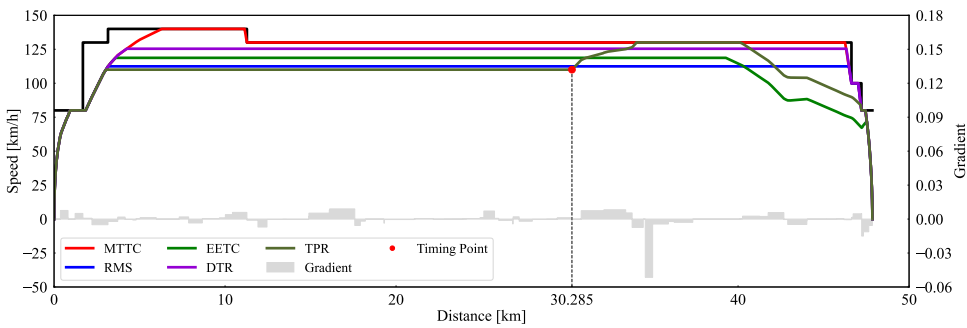


Figure 4.5: Example speed profiles of selected train driving strategies

The corresponding start of the blocking time for a train p at a given block section $b \in B_p$ over a designated railway corridor with multiple stops is represented by:

$$t^{E,\gamma}_{p,b} = t^{E,\gamma}_{p,b} - t^{E,\text{approach},\gamma}_{p,b} - t^{E,\text{reaction}}_{p,b} - t^{E,\text{setup}}_{p,b} - t^{E,\text{buffer}}_{p}, \tag{4.1}$$

where $t^{E,\text{approach},\gamma}_{p,b}$ denotes the approach time of train p to block b for the driving strategy γ, $t^{E,\text{reaction}}_{p,b}$ is a constant parameter representing the ATO reaction time, and $t^{E,\text{setup}}_{p,b}$ signifies the setup time within block b, depending on whether the block is on open tracks or in interlocking areas. Fixed buffer times, denoted as $t^{E,\text{buffer}}_{p}$ at the start and $t^{L,\text{buffer}}_{p}$ at the end of the blocking time, may be included to mitigate delay propagation for timetable robustness and provide tolerance for train trajectory tracking algorithms.

The corresponding end of the blocking time is given by:

$$\bar{t}^{L,\gamma}_{p,b} = t^{L,\gamma}_{p,b} + t^{L,\text{run},\gamma}_{p,b} + t^{L,\text{clear},\gamma}_{p,b} + t^{L,\text{release}}_{p,b} + t^{L,\text{buffer}}_{p}, \tag{4.2}$$

where $t_{p,b}^{L,\mathrm{run},\gamma}$ represents the running time of train p in block b, and $t_{p,b}^{L,\mathrm{clear},\gamma}$ denotes the time for train p to clear the block b based on a specific driving strategy γ. Additionally, $t_{p,b}^{L,\mathrm{release}}$ is a constant parameter used to release block b.

The minimum line headway between trains p_i and p_j over this corridor for the given train trajectories is given by the maximum overlap when the successive train paths are scheduled to depart simultaneously:

$$h_{p_i,p_j}^{\gamma_i,\gamma_j} = \max_{b \in B_{p_i} \cap B_{p_j}} \left(\bar{t}_{p_i,b}^{L,\gamma_i} - \underline{t}_{p_j,b}^{E,\gamma_j} \right). \tag{4.3}$$

This minimum line headway time is the minimum time separation between two train paths at the line start, ensuring conflict-free train operations. The critical block for this train pair p_i, p_j where the successive blocking times are closest is the block b_{p_i,p_j}^* where the maximum is achieved:

$$b_{p_i,p_j}^* = \operatorname*{argmax}_{b \in B_{p_i} \cap B_{p_j}} \left(\bar{t}_{p_i,b}^{L,\gamma_i} - \underline{t}_{p_j,b}^{E,\gamma_j} \right). \tag{4.4}$$

The critical block does not have to be unique when multiple blocks are at the same shortest time distance. In the compressed timetable, all blocks other than the critical block(s) have positive buffer time between the successive trains.

The minimum robust line headway $h_{p_i,p_j}^{\mathrm{TPS}}$ that allows all assumed driving strategies is termed a TPS headway. In Eq. (4.1) to (4.4), this corresponds to $\gamma_i = \gamma_j = \mathrm{TPS}$, where TPS includes the RMS, EETC and SMTTC driving strategies. As the line headway decreases, the robustness of the train path declines due to reduced tolerance. When this tolerance is no longer feasible, a second nominal driving headway $h_{p_i,p_j}^{\mathrm{Nom}}$ is determined. This nominal driving headway is defined as the minimum departure interval required for conflict-free operations when a successive train pair may use various driving strategies with punctual departure and arrival. In this case, the SMTTC driving strategy is excluded from the calculations of the previous instance. Third, the minimum line headway for the same driving strategy of both trains h_{p_i,p_j}^{γ} is defined as the interval when a train pair operates under a specific driving strategy without DTR or TPR. This headway is computed from Eq. (4.1) to (4.4) with $\gamma_i = \gamma_j \in \{\mathrm{RMS}, \mathrm{EETC}, \mathrm{MTTC}\}$. For completeness, we include the MTTC driving strategy here as it corresponds to a theoretical minimum line headway based on an assumed deterministic train trajectory.

The three line headway types mentioned earlier are applicable in undisturbed conditions. However, the TPE can include extra control TPs with time windows for potentially conflicting train pairs to avoid conflicts caused by a faster-following train trajectory. This situation can arise when the following train applies the EETC driving strategy and accelerates to a higher cruising speed to optimise coasting phases at later stages. In this case, the preceding train adheres to nominal driving strategies with $\gamma_i = \mathrm{TPS}$, while the following train follows the TPR strategy $\gamma_j = \mathrm{TPR}$ in Eq. (4.1) to (4.4). Without altering the scheduled departure and arrival times of a train pair, the minimum TPR headway $h_{p_i,p_j}^{\mathrm{TPS,\,TPR}}$ is computed by determining the latest permissible passage time of the following train at the critical block to ensure conflict-free operations. If the actual headway is shorter than the TPR headway, indicating that the interval between trains is insufficient to resolve the conflict, then the TMS must provide an updated RTTP. In this case, the TPE computation cannot determine a feasible set of train trajectory generation constraints.

4.3.4 Train path envelopes and train status updates

During operations, the ATO onboard sends train status reports, which are used to optimise or adjust the TPE or the RTTP if the TPE cannot be respected. These reports include the current time t, location $x_{p,t}$, speed $v_{p,t}$, and estimated time of arrival at the next TP, t_{p,σ_p}, where $\sigma_p = \sigma_p(x_{p,t})$ returns the next TP with a time target after the current location $x_{p,t}$. This information is represented as a measurement vector $\hat{a}_{p,t}$:

$$\hat{a}_{p,t} = \left[t, x_{p,t}, v_{p,t}, t_{p,\sigma_p}\right]^T. \tag{4.5}$$

Given the known initial and terminal states of the train, including its start and end times and locations, train trajectories can be computed for both before and after the current position, namely the realised and expected trajectories. The realised train trajectory leads to realised blocking times up to (but excluding) the current block section $b(x_{p,t})$, denoted by superscript r, as $[\underline{t}^{E,r}_{p,b'}, \bar{t}^{L,r}_{p,b'}]$ for blocks $b' \in B^r_p(x_{p,t})$, where $B^r_p(x_{p,t}) = \left\{b' \in B_p \mid s_{b'} \in [0, b(x_{p,t}))\right\}$. Here, $b(x)$ is defined as a function that maps a position x to the corresponding block section containing that position. The expected TPS represents the range of blocking times for blocks in the upcoming route of the train, represented by $[\underline{t}^{E,\mathrm{TPS}}_{p,b}, \bar{t}^{L,\mathrm{TPS}}_{p,b}]$ for blocks $b \in B^e_p(x_{p,t})$, where $B^e_p(x_{p,t}) = \left\{b \in B_p \mid s_b \in (b(x_{p,t}), b(\sigma_p(x_{p,t})))\right\}$. This expected TPS is derived using RMS, EETC, and operational tolerance response driving strategies (described below), reflecting the possible behaviour of the train in the future based on its current position and speed. Upon receiving a measurement within a specific block section, we define the realised blocking time start as the lower contour and the expected TPS end as the upper contour of the current block section, i.e., $[\underline{t}^{E,r}_{p,b(x_{p,t})}, \bar{t}^{L,\mathrm{TPS}}_{p,b(x_{p,t})}]$. This representation captures the infrastructure occupation status, combining the actual movement of the train to its current position with the expected range of blocking times ahead.

To compute the operational tolerance response driving strategy, we assume that the train applies maximum traction promptly upon sending the latest measurement, if applicable, and operates at the maximum feasible speed for the remainder of the journey. As a result of this driving strategy, the train would arrive early, denoted as t'_{p,σ_p}, compared to the originally estimated time of arrival t_{p,σ_p}. The difference between these two arrival times constitutes the operational tolerance, denoted as $d_{p,t} = t_{p,\sigma_p} - t'_{p,\sigma_p}$. This tolerance is used to shift the response driving strategy upon sending the latest measurement, such that the scheduled arrival time can be met. The operational tolerance defines the latest permissible passing time at the moment the preceding train status report is received without inducing blocking time conflicts. It constrains the feasible train trajectories from that point onward and serves as an upper bound for ATO onboard trajectory generation, supporting cruising speed regulation and coasting adjustments.

Information on realised blocking times and expected TPSs within a railway traffic corridor is denoted as $\widetilde{\mathrm{TPS}}_p(t)$ and used for the dynamic generation and adjustment of the TPE. Due to potential asynchronicity in transmitting status reports from each train, previously expected TPSs remain valid until updated. The information $\widetilde{\mathrm{TPS}}_p(t)$ is hence given by:

$$\begin{aligned}\widetilde{\mathrm{TPS}}_p(t) = \Bigg\{ &\left\{ \left[\underline{t}^{E,r}_{p,b'}, \bar{t}^{L,r}_{p,b'}\right],\ b' \in B^r_p(x_{p,t}) \right\}, \left\{ \left[\underline{t}^{E,r}_{p,b(x_{p,t})}, \bar{t}^{L,\mathrm{TPS}}_{p,b(x_{p,t})}\right] \right\}, \\ &\left\{ \left[\underline{t}^{E,\mathrm{TPS}}_{p,b}, \bar{t}^{L,\mathrm{TPS}}_{p,b}\right],\ b \in B^e_p(x_{p,t}) \right\} \Bigg\}. \end{aligned} \tag{4.6}$$

The TPE generation algorithm is re-executed upon receiving a new train status report or at regular intervals, to identify additional required control TPs, fine-tune tolerances or adjust

previously determined TPE to enhance train operations. Any newly added control TPs are positioned downstream after the current block. The associated critical blocks must be assessed to determine if sufficient distance is available to incorporate them and enable the TPR driving strategy. Initially determined TPEs with fine-tuned tolerances or control TPs can be relaxed or adjusted if no or less TPS overlap is detected. This adjustment can improve flexibility in train operations and enhance energy efficiency. For example, cruising at a higher speed and extending coasting can be achieved by expanding or removing the time window constraint at a control TP.

4.3.5 Sensitivity analysis using elementary effects

Tightly (re)scheduled times within the RTTP may not always yield a feasible TPS headway, while variability in train behaviour, as reflected in status reports from the ATO onboard, impacts the flexibility needed to adopt driving strategies that meet the objectives of the railway undertakings, such as passenger comfort and energy efficiency. The dynamic evolution of rescheduling measures at the TMS, combined with realised and expected train trajectories and corresponding track occupations, highlights the need to evaluate their effects on generating conflict-free TPEs.

Our sensitivity analysis uses an approach based on the concept of elementary effects, which capture those changes in an output solely due to changes in a particular input, as introduced in Morris (1991). This method evaluates the sensitivity of the model output to variations in individual input factors by systematically isolating and varying them over their entire feasible range. This approach requires no assumptions on input sparsity, output monotonicity, or smooth functional approximations, making it suitable for the multi-stage, optimisation-based TPE generator.

Consider a function $f(\mathbf{x})$ with n input factors, represented as a vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. The elementary effect E_i of an input factor x_i is calculated by changing x_i with a defined step Δ_i while holding other factors fixed. The method evaluates $n_i + 1$ evenly spaced sample points within the specified range $[\underline{x}_i, \bar{x}_i]$, where the step size is given by $\Delta_i = (\bar{x}_i - \underline{x}_i)/n_i$. These sample points $x_{i,j}$ are generated along the trajectory as $x_{i,j} = \underline{x}_i + (j-1) \cdot \Delta_i$, where $j = 1, 2, \ldots, n_i$, representing the positions of x_i at which its sensitivity is evaluated. The first elementary effect $E_{i,1}$ represents the initial change at $j = 1$, while the general j-th elementary effect $E_{i,j}$ describes the sensitivity at any subsequent j in the parameter space of x_i. The elementary effect $E_{i,j}$ at each point $x_{i,j}$ is calculated as:

$$E_{i,j} = \frac{f(x_1, \ldots, x_{i-1}, x_{i,j} + \Delta_i, x_{i+1}, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, x_{i,j}, x_{i+1}, \ldots, x_n)}{\Delta_i}, \tag{4.7}$$

which represents the gradient of function f at $x_{i,j}$ over the interval defined by the step Δ_i. Since each elementary effect requires two adjacent points, the maximum number of elementary effects for each parameter is constrained to n_i. By evaluating each sample point $x_{i,j}$ along a trajectory, we can obtain a distribution of elementary effects, capturing the variability in the sensitivity of f to changes in x_i. For each input factor x_i, we calculate the mean effect μ_i, and the standard deviation σ_i, which serve as sensitivity indices to quantify the influence of changes in elementary effects E_i:

$$\mu_i = \frac{1}{n_i} \sum_{j=1}^{n_i} E_{i,j}, \tag{4.8}$$

$$\sigma_i = \sqrt{\frac{1}{n_i - 1} \sum_{j=1}^{n_i} \left(E_{i,j} - \mu_i\right)^2}. \tag{4.9}$$

These indices capture the average impact and variability of the influence of each factor on the model output, respectively. A higher mean effect μ_i indicates that an input consistently and strongly influences the output. In contrast, a higher standard deviation σ_i implies that the effect of the input varies across the input space, often due to nonlinearities or interactions. These indices jointly characterise the importance and behaviour of each input with respect to the model output. Algorithm 4.1 summarises the computational steps involved in performing the sensitivity analysis using elementary effects.

Algorithm 4.1 Sensitivity analysis using elementary effects for TPE computation

Input: A scalar-valued TPE model output $f(\mathbf{x})$, input factors $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ with respective bounds $[\underline{x}_i, \bar{x}_i]$

Output: Sensitivity indices (mean effect μ_i, standard deviation σ_i) for each input factor x_i

1: **for** each input factor x_i **do**
2: Generate $n_i + 1$ evenly spaced sample points $x_{i,j}$ in range $[\underline{x}_i, \bar{x}_i]$ with step size $\Delta_i = (\bar{x}_i - \underline{x}_i)/n_i$
3: **for** each point $x_{i,j}$ **do**
4: Compute elementary effect $E_{i,j}$ using Eq. (4.7)
5: **end for**
6: Compute mean sensitivity index μ_i using Eq. (4.8)
7: Compute standard deviation sensitivity index σ_i using Eq. (4.9)
8: **end for**
9: **return** μ_i, σ_i for all i

We begin by analysing the input space of the TPE computation and identifying the most influential factors, i.e., line headway, train position, and train speed. These inputs are then varied in a controlled manner to explore their impact through sensitivity analysis. In our sensitivity analysis, we focus on three TPE model output functions f, corresponding to departure tolerance, the placement of a control TP with its associated time window, and the operational tolerance as defined below. Since a control TP is only generated under sufficiently small headways, these functions are treated separately, each defined over a different headway range.

The first function $f_1(h)$ denotes the tolerance at departure, which depends only on the line headway h. When the headway between two trains is sufficiently large, $f_1(h)$ alone constitutes the TPE in the form of a departure tolerance. However, when $f_1(h)$ is zero, an additional time constraint at a control TP may become necessary. The second function $f_2(h)$ represents the time window at an introduced control TP, which also depends on the line headway h, and is defined over the range of headway values for which $f_1(h) = 0$, i.e., when the departure tolerance is determined to be zero and the blocking time conflict remains, requiring an additional constraint at a control TP. For both $f_1(h)$ and $f_2(h)$, the input factor h is bounded by the minimum and the maximum headway values derived from the RTTP for a specific train pair.

The range between the TPS headway and the nominal driving headway, and the range between the nominal driving headway and the TPR headway, define the feasible adjustment ranges for the departure tolerance of the preceding train and the time window at the control TP for the following train, respectively. The sensitivity indices provide insight into the responsiveness of

the computed departure tolerance and time window to headway changes, particularly in relation to the spatial positioning of TPs. A low mean effect μ_h indicates that headway has limited overall influence on the TPE model outputs, while a low standard deviation σ_h suggests that this influence remains consistent across different headway values.

The third function $f_3(h,x,v)$ represents the operational tolerance during a train run and is a function of line headway h, position x and speed v. Given the interdependence of time, speed, and distance in trajectory generation, speed v is chosen as the input factor to quantify variations in operational tolerance, while h and x define the traffic scenario and the measurement location. Adjustments to speed v are applied within the permissible range set by the original TPE at the selected position x and headway h. These adjustments are constrained by the minimum and maximum speeds of the selected train trajectories within the original TPE.

The sensitivity indices for operational tolerance indicate how variations in speed, headway, or position affect the flexibility available during train operation. A high mean effect μ_v reflects a strong and consistent influence of speed across different positions and headways, while a high standard deviation σ_v suggests that this influence depends on the specific traffic scenario. The results inform where trains may require closer adherence to reference trajectories, or where greater driving flexibility is possible. They also support identifying locations requiring additional buffer times or upstream traffic measures. Although not direct TMS or ATO commands, these indices reflect the dynamic use of the running time supplement.

4.4 Case study and results

4.4.1 Case study description

Our case study is conducted on a Dutch rail corridor, spanning 22 kilometres between Breda (Bd) and Tilburg (Tb) stations. This corridor includes intermediate stations Gilze-Rijen (Gz), Tilburg Reeshof (Tbr), and Tilburg Universiteit (Tbu). The maximum corridor speed limit is 140 km/h.

The case study assumes ETCS Level 2 with fixed blocks, which provides continuous supervision of speed and braking curves Winter (2009). Key parameters are configured as follows: $t_{p,b}^{\text{release}} = 2$ s, $t_{p,b}^{\text{reaction}} = 1$ s, $t_{p,b}^{\text{setup}} = 1$ s (for open track segments), or $t_{p,b}^{\text{setup}} = 6 \times$ number of switches in seconds (in station areas or junctions), and $t_p^{E,\text{buffer}} = t_p^{L,\text{buffer}} = 10$ s (added to the TPS lower and upper contours for the original timetable but not in the following sensitivity analyses).

In a periodic schedule with half an hour cycle time between stations Bd and Tb, two freight (FR) trains and three passenger trains are modelled. Among these, there is one Sprinter (SPR) train (i.e., local train) that stops at every station, alongside two Intercity (IC) trains that only serve stations Bd and Tb. The FR trains pass station Bd at 0.5 min and 14 min, while the SPR train departs from station Bd to Tb at 23 min. Additionally, the two IC trains depart from station Bd at 8 and 20 min. The service braking rates of IC, SPR and FR trains are –0.8 m/s^2, –0.66 m/s^2, and –0.5 m/s^2, respectively.

The initial and optimised TPSs for the original timetable in the Breda-Tilburg corridor are depicted in Figure 4.6 (a) and (b). The TPS overlaps are identified between the IC departing at 20 min and the SPR departing at 23 min from station Bd, with a maximum overlap of 102 s, as highlighted by the red overlaps in Figure 4.6 (a). This overlap can be resolved by adjusting the tolerance of the preceding IC train at station Bd to 50 s without the need for additional control TPs, as shown in Figure 4.6 (b).

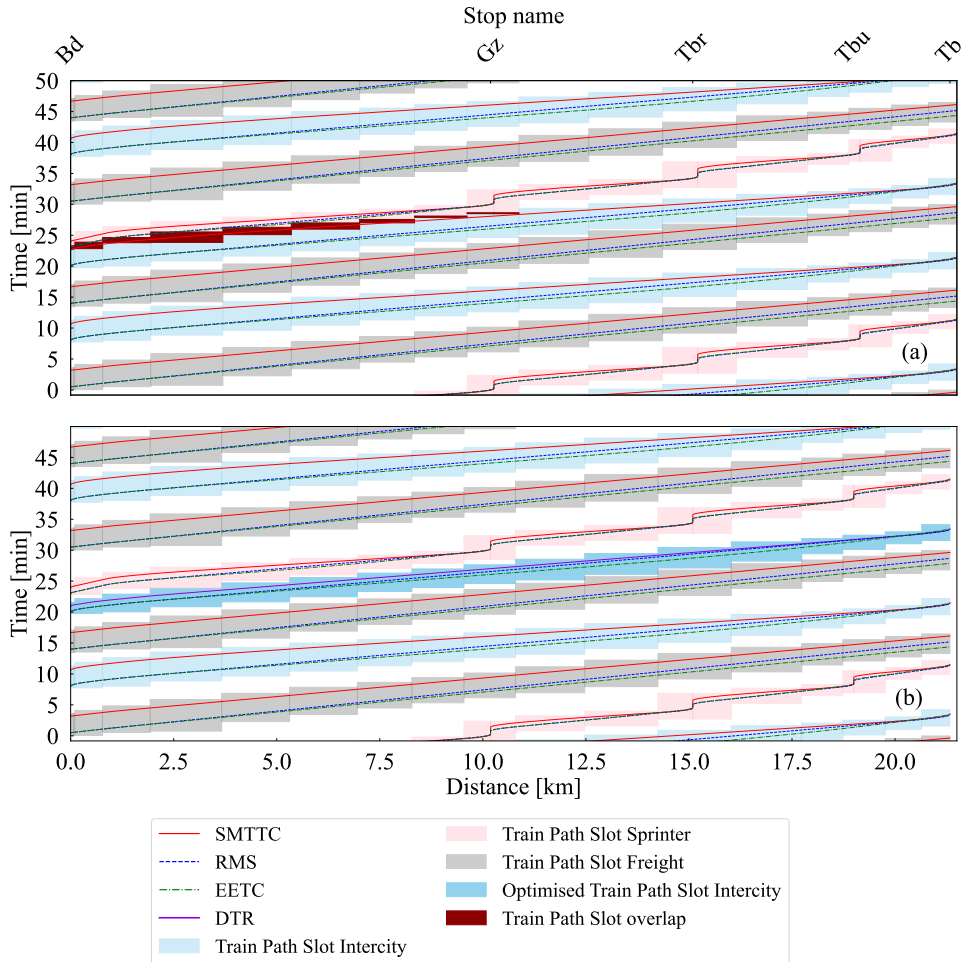


Figure 4.6: (a) Initial (top) and (b) optimised TPSs (bottom) for the original timetable in the Breda-Tilburg corridor

4.4.2 Train path envelope sensitivity to TMS updates

The first sensitivity analysis examines the relationship between the (re)planned line headways at station Bd from a potentially updated RTTP and the TPEs across all possible train pairs in the case study. All permutations of three representative train types are tested, independent of the original timetable sequence, covering distinct line headways for each pair. Headway values are computed as described in the methodology (Table 4.2), and sensitivity is evaluated at departure stations and control TPs. The analysis spans the full headway range in 3 s intervals, and the resulting indices for tolerance and time window adjustments are presented in Table 4.3 and Table 4.4. Cells marked with dashes (–) denote cases where control TPs are inapplicable due to critical blocks occurring at or near the next scheduled stop. For train pairs with two indices, multiple tolerances or time windows are computed, with further explanation provided

below. Figure 4.7 visualises the relationship between headway and the departure tolerance for the preceding train (solid lines) and the time window at the control TP for the following train (dashed lines).

Table 4.2: Computed line headways [s] for all possible train pairs

Train Pair	$h^{\mathbf{TPS}}_{p_i,p_j}$	$h^{\mathbf{Nom}}_{p_i,p_j}$	$h^{\mathbf{EETC}}_{p_i,p_j}$	$h^{\mathbf{RMS}}_{p_i,p_j}$	$h^{\mathbf{MTTC}}_{p_i,p_j}$	$h^{\mathbf{TPS,\ TPR}}_{p_i,p_j}$
IC - IC	261	162	126	132	141	132
IC - FR	309	162	156	156	159	102
IC - SPR	264	117	111	117	114	87
FR - IC	303	237	189	192	243	189
FR - FR	282	153	141	144	138	141
FR - SPR	237	114	96	102	120	114
SPR - IC	447	432	420	432	423	432
SPR - FR	375	369	375	321	258	369
SPR - SPR	231	198	180	180	192	198

Table 4.3: Sensitivity analysis of tolerance at ATO timing points on the headway grid for case study train pairs

Train Pair	**Departure Timing Point**	**Headway Range [s]**		**Sensitivity Indices of Tolerance**	
		Min	**Max**	μ_i	σ_i
IC - IC	Bd	162	261	1.60	0.84
IC - FR	Bd	162	309	1.04	0.15
IC - SPR	Bd	117	264	1.03	0.16
FR - IC	Bd	237	303	2.16	0.56
FR - FR	Bd	153	282	1.11	0.20
FR - SPR	Bd	114	237	1.20	0.33
SPR - IC	Tbu	432	441	2.07	0.28
SPR - FR	Tbu	369	375	5.00	0.24
SPR - SPR	Tbr	198	231	0.94	0.14
	Tbu			1.08	0.14

Considering heterogeneous train pairs, especially where a faster train follows a slower one, we observe that the resulting headway tends to be larger due to the running time differences between trains. This phenomenon is particularly notable when SPR trains precede in a train pair, owing to their extended time-distance train path with multiple intermediate stops. In such cases, the critical block of a train pair occurs at or close to the next scheduled stop, resulting in the TPR headway coinciding with the nominal headway. This is also seen in the FR-SPR train pair, where the critical block is located at station Gz. In these instances, it is not possible to utilise a control TP to resolve the train path conflicts.

Table 4.4: Sensitivity analysis of time window at ATO control timing points on the headway grid for case study train pairs

Train Pair	Timing Point Location [m]	Headway Range [s]		Sensitivity Indices of Time Window	
		Min	Max	μ_i	σ_i
IC - IC	14,248	132	162	1.04	0.27
IC - FR	1,929	102	162	0.81	0.19
IC - SPR	1,929	87	117	0.64	0.20
FR - IC	16,027	189	237	1.06	0.28
FR - FR	6,817 16,027	141	153	1.17 0.61	0.11 1.02
FR - SPR	-	114	114	-	-
SPR - IC	-	432	432	-	-
SPR - FR	-	369	369	-	-
SPR - SPR	-	198	198	-	-

Heterogeneous train pairs, where SPR or FR precedes, exhibit higher mean sensitivity indices (μ_h) for departure tolerance adjustments, indicating that headway has a consistently strong influence on train pairs where slower trains precede. In such cases, the impact on departure tolerance can be twice as large or more compared to the reverse sequence. This effect is particularly pronounced when the conflicting block lies beyond the halfway point of the corridor from the previous stop. For SPR trains, shorter inter-stop distances and limited running time supplements further constrain tolerance adjustments. The passing pattern of FR trains also amplifies this effect, requiring larger adjustments to maintain a conflict-free upper contour in the TPS under the DTR strategy. The SPR–FR pair shows the highest mean effect ($\mu_h = 5.00$), reflecting a steep and consistent increase in departure tolerance adjustments as headway decreases, consistent with the pronounced slope in Figure 4.7 (second orange line from the right).

In contrast, train pairs with an IC preceding typically exhibit lower mean sensitivity indices (μ_h), as conflicts occur earlier in the corridor and the longer scheduled running times allow more flexibility to absorb headway variations through small tolerance changes. Low mean sensitivity indices (μ_h) are also observed for homogeneous train pairs. In particular, in the SPR–SPR pair, reducing the headway from 213 s to 210 s requires fine-tuning tolerances at both stations Tbr and Tbu.

The IC–IC pair, however, is a notable exception, exhibiting a distinct slope change midway through the headway range (around 3.2 min), as shown by the purple line in Figure 4.7. This shift corresponds to a rise in the standard deviation (σ_h), indicating that the influence of headway on the departure tolerance becomes more variable across the headway range. This behaviour arises from the DTR strategy, where reducing cruising speed prolongs running and clearing times while shortening the approach time compared to SMTTC. As a result, the blocking time overlap shifts beyond the midpoint of the corridor (12,471 m), requiring larger adjustments.

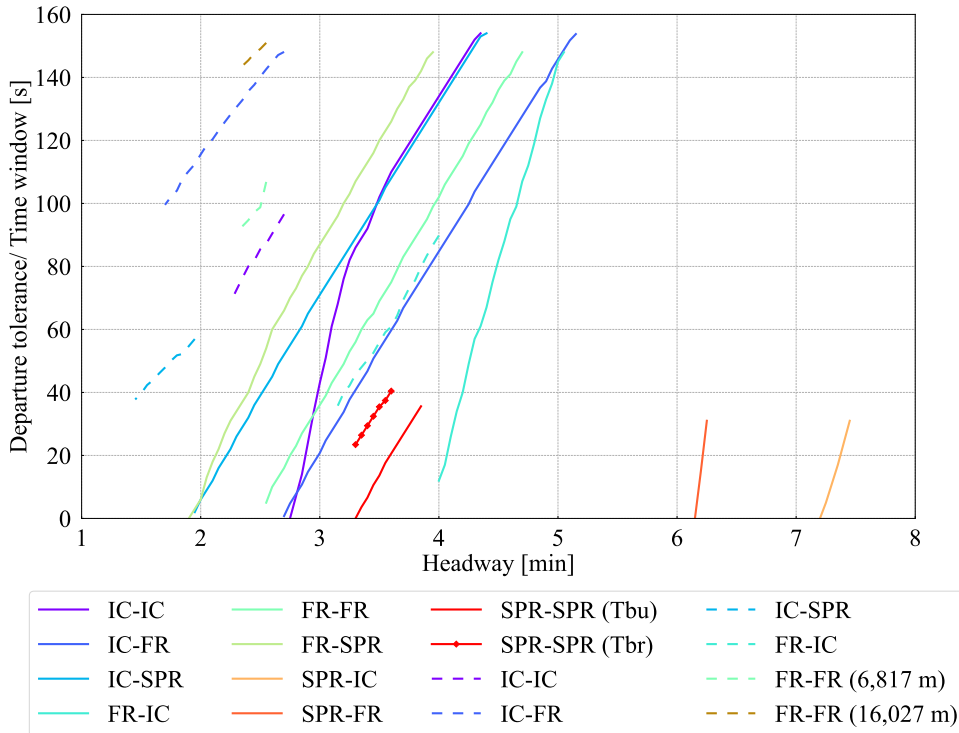


Figure 4.7: Relationship between headway and tolerance for the preceding train (solid line) and headway and time window at the control timing point (dashed line) for the following train across all train pairs in the case study

For the time window sensitivity at control TPs, heterogeneous train pairs, such as IC–SPR and IC–FR, exhibit lower mean sensitivity indices (μ_h) compared to homogeneous pairs. This is primarily because control TPs are introduced earlier along the corridor, allowing more flexibility in the form of running time supplements. For instance, the control TP is located at 1,929 m for the following SPR and FR trains in the IC–SPR and IC–FR pairs, respectively, whereas it appears much later at 14,248 m for the IC–IC pair. Time window sensitivity increases at later TPs due to longer travel distances and less remaining running time supplement. For example, the control TP at 16,027 m exhibits a higher μ_h than those placed earlier. Notably, when the headway is reduced from 153 s to 150 s for the FR–FR train pair, two control TPs are introduced simultaneously to maintain conflict-free operation. When an SPR train leads, the critical block often coincides with a scheduled stop, making it infeasible to define control TPs for resolving conflicts through speed adjustment.

Figure 4.7 highlights a sudden drop in the time window for the FR–FR train pair at 6,817 m (dashed turquoise line). This deviation occurs because the first control TP requires substantial adjustment from the original passing time computed using the EETC strategy. As the train already passes later at the first control TP, the time window adjustment at the second becomes comparatively smaller. However, the standard deviation of the sensitivity index at 16,027 m ($\sigma_h = 1.02$) is significantly higher than at 6,817 m ($\sigma_h = 0.11$), indicating that the effect of headway

on the time window at the second TP is influenced by nonlinearities or interactions between the two control TPs and the degree of headway variation. Consequently, the difference between the original and adjusted time windows at the first and second control points is pronounced.

Overall, adjustments for departure tolerances and time windows show a decreasing trend as headway is reduced. The mean effect μ_h on time window is generally lower than on departure tolerance, reflecting the narrower adjustment range under the TPR strategy. This strategy maintains a steady cruising speed up to the control TP to avoid conflicts, while the remaining running time supplement is used for EETC driving. In contrast, the DTR strategy uses the running time supplement by lowering cruising speeds between stops, enabling broader tolerance adjustments with stronger influence.

4.4.3 Train path envelope sensitivity to train status updates

The second sensitivity analysis investigates a case study featuring the IC train departing at 20 min and the SPR train at 23 min in the original timetable, which showed TPS overlaps (recall Figure 4.6 (a)). The scheduled headway of 180 s is reduced to 150 s and 120 s to evaluate its impact on the TPEs of both trains. The analysis examines various cruising speeds of the IC train within the bounds of the original TPE, considering distances of 5 km, 10 km, and 15 km, based on its scheduled running time of 810 s between stations Bd and Tb. Train speed is incrementally increased by 5 km/h, ranging from 103.48 km/h (RMS) to 134.40 km/h (SMTTC) at 5 km, from 103.48 km/h (RMS) to 140 km/h (SMTTC) at 10 km, and from 101.20 km/h (EETC) to 140 km/h (SMTTC) at 15 km. The resulting sensitivity indices, presented in Table 4.5, quantify the impact of speed changes on the operational tolerance of the IC train at each distance and headway.

Table 4.5: Sensitivity analysis of ATO timing points for the IC-SPR train pair at varying headways and distances

Speed [km/h]	**Operational Tolerance [s] at Various Distances and Headways**								
	5 km			**10 km**			**15 km**		
	120 s	**150 s**	**180 s**	**120 s**	**150 s**	**180 s**	**120 s**	**150 s**	**180 s**
105	22.43	53.86	83.33	40.76	70.85	80.07	40.78	40.78	40.78
110	24.50	57.51	88.15	53.39	84.20	94.89	63.08	63.08	63.08
115	24.50	62.48	92.87	64.22	95.73	107.54	82.97	82.97	82.97
120	24.50	62.48	96.15	64.22	106.05	119.00	100.47	100.47	100.47
125	24.50	62.48	97.14	64.22	114.46	128.09	115.58	115.58	115.58
130	24.50	62.48	97.14	64.22	120.58	136.28	128.64	128.64	128.64
135	-	-	-	64.22	120.58	141.03	137.73	137.73	137.73
140	-	-	-	64.22	120.58	145.60	145.60	145.60	145.60
Sensitivity index μ_i	0.07	0.29	0.46	0.59	1.24	1.64	2.62	2.62	2.62
Sensitivity index σ_i	0.17	0.45	0.45	1.09	1.11	0.98	1.45	1.45	1.45

The results show that when the headway is above the nominal driving headway between these two trains (in this case, 117 s) and the preceding train stays within the original TPE, there is no need to introduce control TPs for the following train. In general, the tolerance at specific points increases with higher cruising speeds to that point, as a greater portion of the running time supplement remains available for operational adjustments. This enables ATO onboard systems to

adopt EETC with coasting while respecting the conflict-free operation constraints. Additionally, larger headways offer a larger buffer between trains, further increasing the robustness.

Nevertheless, TPS overlaps between trains impose a tolerance limit for conflict-free operations. For example, under a 150-second headway, the tolerance stabilises at 62.48 s for speeds from 115 km/h at 5 km and 120.58 s for speeds from 130 km/h at 10 km. Figure 4.8 illustrates that increasing speed beyond certain thresholds does not yield a higher tolerance due to TPS constraints, as shown by the impact of a status report speed of 120 km/h at 10 km under a 150-second headway.

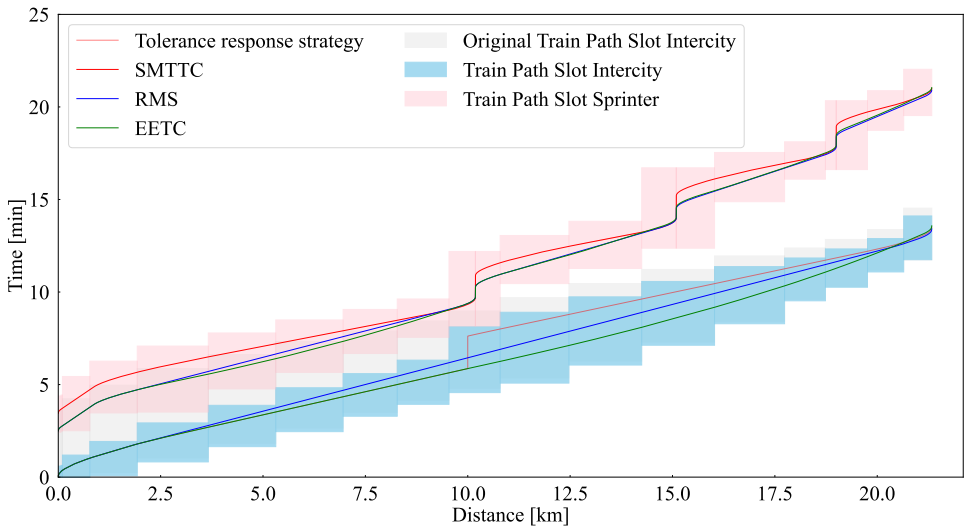


Figure 4.8: Comparison of the original and operational TPSs for the preceding IC train, alongside the original TPS for the following SPR train, with a 150-second headway and a status report sent at 120 km/h at 10 km

In conflict-free scenarios, such as those at 15 km, operational tolerance remains consistent when the running time supplement is fully utilised, with the MTTC strategy starting at the computed tolerance and arriving at the next target time. By contrast, when conflicts are consistently present or absent across headways and the running time supplement is not fully consumed, the difference in operational tolerance corresponds to the headway difference, for instance, at 5 km with a speed of 105 km/h. Slight variations in operational tolerance also arise from applying the tolerance response strategy as an RMS driving strategy, where arrival time deviations in the computation process lead to minor differences in the results.

The sensitivity indices reveal that operational tolerance is less affected by speed variations when the train is closer to its previous stop (e.g., at 5 km), where only a small portion of the running time has elapsed, and speed deviations have limited opportunity to accumulate. This is reflected by low mean (μ_v) and standard deviation (σ_v), indicating limited influence and a consistent response. In such cases, operational tolerance is smaller, and maintaining a speed profile close to the reference trajectory is critical to remain within the TPE. At greater distances (e.g., 10 km and 15 km), speed deviations accumulate more significantly, increasing both μ_v and σ_v, and indicating stronger and more input-dependent effects. These findings confirm and extend the headway-based analysis, where operational tolerance becomes more sensitive at

greater distances and responds more strongly to changes in headway and speed. Although greater tolerance allows for more flexibility in applying energy-efficient strategies such as coasting, it also reflects a higher responsiveness to input variation. In such situations, onboard speed regulation using the TPE may be insufficient, and an RTTP update from the TMS may be required to maintain conflict-free operation.

While the sensitivity indices provide insights into the responsiveness of TPE generation against input variations, their absolute values are mainly meaningful within the same TPE model output and infrastructure segment. The departure tolerance, the time window at control TPs, and the operational tolerance all represent time-based flexibility in relation to the allocated running time supplement along the train route, but serve distinct operational purposes. Consequently, although the sensitivity analysis enables cross-line measurement for a specific model output, comparison across different model outputs requires contextual interpretation.

4.5 Conclusions

This study investigated the sensitivity of the Train Path Envelope (TPE) for Automatic Train Operation (ATO) to variations in real-time traffic plans and train status report updates. The TPE comprises time targets or windows at relevant locations along train routes, serving as constraints for conflict-free train trajectory generation by the ATO onboard. A sensitivity analysis using elementary effects was proposed to evaluate how variations in planned headways within the RTTP and reported train speeds and positions influence the computation of conflict-free TPEs.

The analysis showed that while the spatial structure of the TPE remains unchanged under reduced headways, the associated time windows become increasingly constrained, particularly for heterogeneous train pairs with differing running times and critical blocks positioned further along their routes. For train status updates, operational tolerance becomes more sensitive to speed variations as the train moves further from its previous stop, due to the accumulation of early speed deviations and a reduction in the available running time supplement. These findings support reducing infrastructure occupation by homogenising train trajectories through the addition of a control TP, which enables shorter line headways compared to the unrestricted case where departure tolerance is available. Low sensitivity indices indicate limited input influence and sufficient onboard driving flexibility through TPE adjustments. In contrast, high indices reflect larger and potentially nonlinear changes in TPE model outputs in response to input variations, which may necessitate RTTP updates.

Future research will focus on embedding the TPE generator within a closed-loop microscopic simulation environment to evaluate the dynamic behaviour of TPEs and their interactions with the TMS and ATO under scenarios involving disturbances, disruptions, and human factors.

Chapter 5

Real-time train timetable rescheduling with event time flexibility

This chapter extends the focus from developing TPEs at the corridor level during operations to optimising their network-wide counterpart, event time flexibility. The adherence to a timetable with precise departure and arrival times becomes increasingly challenging in real-world scenarios due to the daily fluctuations in rail traffic, leading to uncertainties that complicate effective real-time traffic management. To address this, we propose a microscopic Train Rescheduling with Flexibility (TRF) model, relying on a Mixed-Integer Linear Programming (MILP) formulation, which jointly determines rescheduled event times and their flexibility to enhance the effectiveness of both ATO-enabled and human-based train operations. Flexibility is modelled as bounded time windows, computed based on timetable allowances and punctuality thresholds, that define the admissible ranges of event times for late departures, early arrivals, and late arrivals, within which the real-time traffic plan remains punctual and conflict-free. The primary objective of the TRF model is to minimise the deviation of event times in the real-time traffic plan from those in the original timetable while simultaneously maximising the size of the associated event time flexibility. A real-life case study on the Dutch railway network with heterogeneous traffic validates the model. The results demonstrate that the proposed model effectively exploits event time flexibility to mitigate disturbances and thus reduce the need for rescheduling interventions. Sensitivity analyses reveal a trade-off between late departure and early arrival flexibility, and indicate a saturation point beyond which increasing the punctuality threshold no longer yields additional flexibility gains.

Apart from minor changes, this chapter is under review as:

Wang, Z., Zhou, R., Correia, G.H.A., Philipsen, E.P., & Goverde, R.M.P. (2026). Real-time train timetable rescheduling with event time flexibility.

5.1 Introduction

Flexibility, as defined by Golden & Powell (2000), denotes a system's inherent capability to adapt, a quality increasingly vital in contemporary society. This adaptability is crucial across diverse domains, including but not limited to information technology (Tafti et al., 2013), economics (Hoberg et al., 2014), power systems (Riaz & Mancarella, 2022) and rail transport (D'Ariano et al., 2008). The ever-increasing rail transport demand has led to rising traffic density, making it more susceptible to perturbations and track occupation conflicts due to disturbances, namely the daily relatively small adverse variations (Cacchiani et al., 2014).

In railway traffic management, a flexible timetable, originally introduced as a tactical planning concept, replaces fixed arrival and departure times with bounded time windows, which improves punctuality without reducing line capacity (D'Ariano et al., 2008). These time windows are derived from predefined timetable allowances, where a larger time window corresponds to a higher degree of flexibility, and the exact arrival and departure times are fixed dynamically in real time according to the network status and current train positions (Schaafsma, 2001). This approach expands timetable feasibility without requiring the execution of algorithms or decision-support tools for slight retiming solutions. Railway infrastructure managers have begun to explore this concept of event time flexibility at departures and arrivals (Caimi et al., 2011b). For passengers, there is always a fixed, published schedule at stations, while traffic controllers work with the so-called flexible timetable, which provides time windows to manage disturbances in real time. The proactive utilisation of time allowances enhances the capability to address disturbances effectively and preventively. Still, the railway timetable rescheduling literature relies predominantly on deterministic models with fixed event times (Zhan et al., 2024). In existing work on flexible timetables, the degree of flexibility has so far been specified exogenously, for instance, through predefined departure and arrival time window sizes, and subsequently used to assess the performance of timetables under different assumed disturbance scenarios (D'Ariano et al., 2008). In contrast, modelling event time flexibility as an endogenous decision variable, where admissible event time bounds are explicitly determined and optimised as part of the train rescheduling process, has received little attention in the literature.

In the past few decades, a plethora of studies have delved into railway traffic management to deal with disturbances. Interested readers can explore comprehensive surveys such as Cacchiani et al. (2014), Corman & Meng (2015) and Fang et al. (2015) for more insights. This stream of research is sometimes referred to as part of the railway timetable rescheduling, which includes both disturbances and disruptions (i.e., relatively large incidents requiring resource modifications). The modelling approaches thereof range from microscopic to macroscopic levels with various methodologies. These include the Alternative Graph (AG) (D'Ariano et al., 2007a; D'Ariano et al., 2007b, 2008; Corman et al., 2010, 2012, 2017), Mixed-Integer Linear Programming (MILP) (Törnquist & Persson, 2007; Krasemann, 2012; Pellegrini et al., 2015, 2019), Model Predictive Control (Caimi et al., 2012; Cavone et al., 2022), reinforcement learning (Zhu et al., 2020) and data-driven models (Huang et al., 2023). In detail, an AG formulation is a typical model used for disturbance management, which is a representation of a job-shop scheduling problem with blocking operations (D'Ariano et al., 2007a; D'Ariano et al., 2007b, 2008; Corman et al., 2010, 2012, 2017). This model incorporated control measures such as retiming, reordering, and rerouting, with the primary aim of minimising the maximum consecutive delay experienced by trains across all visited stations. Differently, Pellegrini et al. (2015, 2019) devised a MILP-based heuristic algorithm that explored all potential train rerouting, reordering, and retiming alternatives

to minimise the maximum consecutive delay at train events. Furthermore, Törnquist & Persson (2007) and Krasemann (2012) developed another MILP model for disturbance management on a railway network comprising single-track and multi-track segments, considering reordering and rerouting measures. Caimi et al. (2012) introduced a model predictive control framework based on solutions derived from a binary linear optimisation model to regulate traffic through retiming, rerouting, and partial speed profile coordination. Incorporating real-life stochastic factors, Huang et al. (2023) utilised recorded traffic control action data to analyse and identify effective actions, for instance, reordering, although their approach was constrained by its dependence on predetermined deterministic parameters. Notwithstanding, most of these models typically assumed deterministic scheduled event times based on service intentions, such as running and dwell times. In such models, time allowances are commonly applied only for delay recovery. When confronted with uncertainties, such as unexpected disturbances, substantial computational efforts are thus required to re-execute these algorithms (Zhan et al., 2024).

In contrast to the aforementioned literature, a flexible timetable relaxes certain pre-planned timetable specifications to allow greater flexibility in real-time traffic management, where detailed event times can be provided in real-time. The concept of flexible timetables was first proposed by Schaafsma & Bartholomeus (2007), in which only a partial order of the trains was determined, and arrival and departure times were provided to traffic controllers only as time windows from the planning stage. D'Ariano et al. (2008) developed an AG model to construct flexible timetables as opposed to "rigid" timetables published to passengers by replacing the scheduled arrival and departure times with maximum arrival and minimum departure times. The results suggested that event time flexibility combined with real-time train scheduling algorithms is promising for increasing railway line capacity and improving train punctuality.

While train rescheduling is commonly categorised as operational planning within the railway transport planning process, a few studies have explored the concept of flexibility at other levels of railway operations. On the tactical planning level, Caimi et al. (2011b) extended the periodic event scheduling problem to generate flexible time windows for departure and arrival times instead of exact times. The enlarged solution space for departure/arrival times on the macroscopic level increases the chance of obtaining feasible train routing subsequently on the microscopic level during operations. For train operations, the concept of a train path envelope was proposed, specifying constraints in the form of time targets or windows at discrete locations, such that trajectories remaining within the envelope are conflict-free and timetable-compliant (Quaglietta et al., 2016; Wang, P. & Goverde, 2016; Wang, Z. et al., 2022). The time windows in a train path envelope allow flexibility to reach conflict-free, energy-efficient and punctual train driving. Although planning successive train path envelopes can represent flexibility at the train trajectory level during operations (Wang, Z. et al., 2025a,b), the arrival times provided to the trajectory generation were kept fixed to preserve punctuality. The direct optimisation of event time flexibility at the traffic management level, where bounded event times are jointly adjusted across trains, remains largely unexplored in the literature.

More broadly, the literature shows that the concept of flexibility within railway planning remains inadequately understood. In particular, event times are commonly treated as fixed values. Timetable allowances embedded in running time supplements for individual train runs are primarily used for delay recovery, whereas buffer times between train paths serve to prevent delay propagation arising from train path deviations, without explicitly representing deviations from scheduled event times during real-time traffic management. In the realm of train rescheduling, this leads to a notable research gap concerning the modelling of event time

flexibility and the determination of optimal event time window bounds. The ON-TIME project proposed a proactive traffic management approach based on the concept of the Real-Time Traffic Plan (RTTP), which specifies the train orders, event times, and routes planned over the operational horizon (Tschirner et al., 2014; Quaglietta et al., 2016). The RTTP is derived from the daily traffic plan, maintained by the traffic controller, and continuously updated to handle disturbances and disruptions in the current traffic situation, and must be followed by all other actors. Incorporating event time flexibility into the RTTP expands the solution space at scheduled event times, allowing variations in running and dwell times to be accommodated, particularly when combined with control measures such as reordering and rerouting. Consequently, minor disturbances can be identified and managed without immediate resolution, thereby handling emergent uncertainties with reduced computational effort and improved efficiency, as the criteria for conflict detection become less restrictive. Moreover, the introduced event time flexibility aims to minimise unnecessary alterations to rescheduling decisions, enabling smooth adjustments to manage disturbances and absorb delays in both automated and conventional train operations.

To this end, the objective of this paper is to propose a microscopic train rescheduling model that incorporates and optimises event time flexibility during each RTTP update, i.e., operational re-planning. Event time flexibility is modelled as bounded time windows, computed based on timetable allowances and punctuality thresholds, which define admissible ranges for late departures, early arrivals, and late arrivals, within which the RTTP remains punctual and conflict-free. The proposed model is formulated as a MILP problem that considers retiming and reordering as rescheduling strategies, together with blocking time constraints that ensure feasibility at the microscopic signalling level, thereby enabling feasible and efficient RTTP updates under disturbed traffic conditions. In addition, our proposed approach holds the potential for addressing traffic uncertainties arising from both automated and human-based train operations and for extending the triggering range of conflict detection. Consequently, it enhances the efficiency of the conflict detection and resolution module within the traffic management system by broadening the solution space at scheduled events and supporting the alignment between traffic management and train operation. In summary, our contributions to the state-of-the-art are as follows:

- Formalising the concept of event time flexibility as bounded time windows, computed based on timetable allowances and punctuality thresholds, that define admissible ranges for arrival and departure event times.
- Proposing a microscopic Train Rescheduling with Flexibility model that integrates event time flexibility with retiming and reordering decisions and blocking time constraints, enabling more effective and efficient real-time railway disturbance management.
- Demonstrating the potential of event time flexibility to enhance the effectiveness and efficiency of train rescheduling models within the rail traffic management system through a comprehensive real-case study on a part of the Dutch railway network.
- Conducting sensitivity analyses on critical parameters and inputs to validate the robustness and applicability of the proposed approach.

The remainder of this paper is organised as follows: Section 5.2 explains the Train Rescheduling with Flexibility (TRF) problem in more detail, defines event time flexibility within the railway traffic management framework and introduces the proposed mathematical formulation. Section 5.3 presents a real-life case study along with sensitivity analyses. Lastly, Section 5.4 concludes the paper and outlines the future research directions.

5.2 Methodology

In this section, we first describe the problem of introducing event time flexibility into railway traffic management in order to unravel the position of our research. The key ingredient is the proposed TRF model. In the following subsection, we explain the assumptions underlying the model. Next, we detail the MILP formulation for the TRF model, including its objective function, constraints, and the incorporation of a new flexibility arc in the AG model.

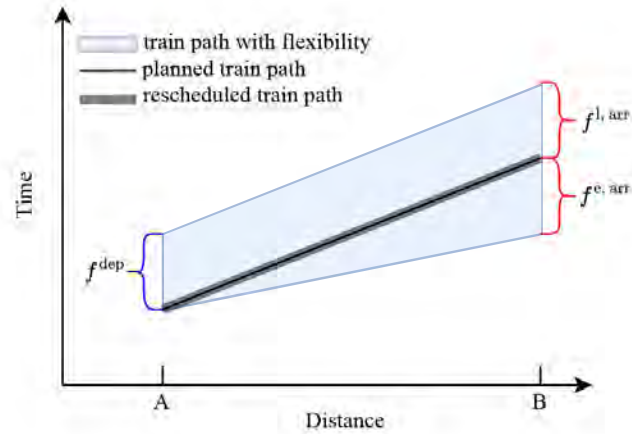
5.2.1 Problem description

A core objective in proactive railway traffic management is to generate and maintain a conflict-free RTTP. This process heavily relies on the efficacy of the predictive conflict detection and resolution algorithm within a rail traffic management system. Typically, these algorithms aim to minimise deviations from the published timetable when train path conflicts or delays are detected, which usually only consider exact departure and arrival times. However, the rigidity of fixed scheduled times poses challenges for trains to adhere to, particularly when confronted with real-time traffic disturbances. These disturbances may arise due to extended passenger boarding/alighting, track adhesion conditions and running time variations. As a result, iterative execution of conflict detection and resolution algorithms may be required to determine a conflict-free RTTP upon detecting conflicts.

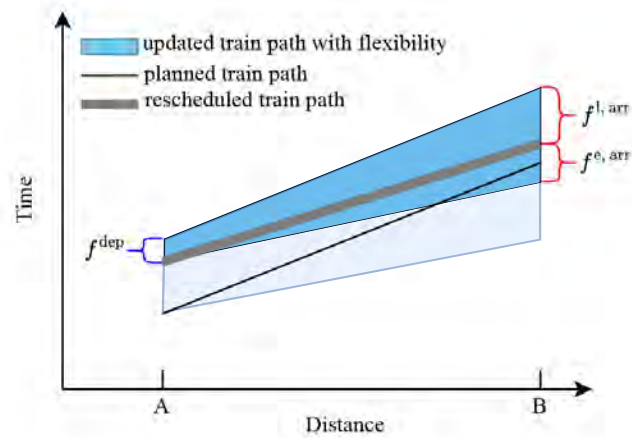
Event time flexibility defines bounded time windows that specify the earliest and latest admissible times at which events may occur without violating conflict-free feasibility or punctuality requirements. This flexibility is computed during each RTTP update when rescheduling, by determining time window bounds for departure and arrival events at critical timetable points, within which the corresponding event times are then selected. Early departures are excluded from the event time flexibility due to their adverse impact on service reliability, potentially leading to missed train trips and extended wait times for passengers (Rietveld et al., 2001). In contrast, late departures are permitted, as the published departure time serves as the earliest allowable departure. Regarding arrival event timings, late arrivals prevail in reality and do not significantly impact train operations as long as they remain within the defined punctuality thresholds. Lastly, while early arrivals are permitted to enhance event time flexibility, they are avoided under energy-efficient train control strategies, because the available running time is fully utilised through coasting rather than arriving ahead of schedule.

Figure 5.1 illustrates two examples of applying a timetable with flexibility in railway traffic management. The left-hand side illustrates event time flexibility under undisturbed conditions, while the right-hand side shows flexibility in the presence of an initial delay, defined here as the delay at the scheduled departure time from station A. A and B are two stations, and the black solid line indicates the planned train path, where the start of this line is the published departure time at station A and the end of this line is the published arrival time at station B. The grey thick line indicates a rescheduled train path, where the two ends of it are rescheduled departure and arrival times. Moreover, the shaded blue area illustrates the train path with flexibility, encompassing late departure, late arrival, and early arrival. The three types of flexibility mentioned above are defined as late departure flexibility (f^{dep}), early arrival flexibility ($f^{\text{e, arr}}$), and late arrival flexibility ($f^{\text{l, arr}}$). Our proposed model will jointly optimise these flexibilities for each arrival and departure such that the maximum delay propagation will not exceed the punctuality threshold.

In Figure 5.1 (a), we first present a scenario where no initial delay occurs. Here, the rescheduled train path coincides with the planned train path, as indicated by the convergence



(a) Event time flexibility without initial delay



(b) Event time flexibility after initial delay

Figure 5.1: Schematic representation of the proposed event time flexibility

of the grey and black lines. The late departure flexibility (f^{dep}) is determined by the difference between the latest permissible and the rescheduled departure time, while the late arrival flexibility ($f^{\text{l, arr}}$) is calculated as the difference between the latest permissible and the rescheduled arrival time. Both of these values must remain within the punctuality threshold. Early arrival flexibility ($f^{\text{e, arr}}$) is defined as the difference between the rescheduled and the earliest possible arrival time, considering the earliest feasible train path at the arrival node without using the running

time supplement. Importantly, the train path with flexibility must not conflict with neighbouring trains. In contrast, Figure 5.1 (b) depicts a scenario where an initial delay is introduced to this train at station A. Here, the rescheduled train path is later than the planned one, but still falls within the defined time bounds of an event time flexibility. As a result, the conflict detection and resolution algorithm is not triggered because the rescheduled event times remain within the flexibility bounds. However, the updated train path with flexibility, shown in the darker shaded blue, also contracts due to the initial delay, necessitating the time windows to be adjusted accordingly.

Therefore, in this paper, we introduce a novel TRF model, aimed at adding and optimising flexibility within the RTTP to support operational traffic re-planning. When no disturbances are present, the optimised flexibility can be regarded as applied to the tactical timetable. This model facilitates the computation of a rescheduling plan that replaces exact event times with bounded time windows without altering the needed resources and track capacity. By defining time windows around scheduled events, the range of conflict detection is broadened from fixed event times to bounded intervals, providing tangible and dynamic bounds to absorb minor train path deviations. Unlike traditional railway (re)scheduling, which primarily relies on predefined buffer times and running time supplements for delay recovery, the TRF model optimises timetable flexibility to dynamically adapt to small variations before actual conflicts arise. Consequently, as long as a train operates within the designated time window, advanced conflict resolution computations can be avoided as the plan remains conflict-free. Furthermore, the added flexibility enhances the robustness of the time-distance diagram by accommodating potential inconsistencies between planned and executed trajectories and variations in train trajectory tracking. In essence, this approach offers the rail traffic management system the flexibility to select arrival and departure times from event time windows to address disturbances, thereby enhancing overall efficiency.

5.2.2 Model assumptions

Common rescheduling measures applied in railway traffic disturbance management pertain to retiming, reordering, and rerouting. Retiming involves adjusting the departure and arrival times of scheduled trains, and reordering features the rearrangement of train sequences to resolve the track occupation conflicts. In this paper, rerouting is not considered due to its computational intensity. Both local and global rerouting alternatives require significant computational resources, which contradicts our objective of swiftly generating an updated timetable with flexibility. Whereas other rescheduling measures, such as cancellation, short-turning, and stop-skipping, are also effective, they are only relevant to manage disruptions characterised by relatively large incidents, e.g., track blockages.

As for the signalling system, a railway network can be regarded as a series of track sections and signals. The signals protect the track sections and regulate rail traffic in the network by enforcing speed restrictions for running trains. Specifically, the track section between two signals is called a block section and can only allow one train in the block section at a time. The minimum safe time interval at which two consecutively scheduled trains can pass the same location is referred to as the train headway. We apply the blocking time theory to calculate this headway time, which is applicable to any signalling system (Pachl, 2014). In this paper, we adopt a three-aspect fixed-block system, where a stop aspect (red) requires trains to stop before the signal, an approach aspect (yellow) means braking and preparing to stop before the signal at danger, and a clear aspect (green) permits trains to proceed at the allowed speed (Goverde et al., 2013).

5.2.3 Train rescheduling with flexibility model

The TRF model is based on the AG model (Mascis & Pacciarelli, 2002), but we reformulate it as a MILP. The AG model offers an intuitive representation that captures the numerous interconnected elements and their associated relationships in the railway rescheduling problem, containing trains, block sections, stations, and operational constraints. A typical AG model aims to minimise the maximum consecutive delay. A MILP reformulation, however, can incorporate multiple objectives, allowing the decision-makers to analyse and explore the trade-offs according to their needs. The AG model provides the structural foundation of our TRF model, serving as the node–arc basis on which the new flexibility arcs are introduced. The notations for sets, indices, parameters, and decision variables are detailed in Tables 5.1 to 5.4, respectively.

Table 5.1: Notation for sets

Symbol	Description
$G = (N, F, A, C)$	graph of representation
N	set of nodes i, j, h, k
F	set of fixed arcs (i, j)
A	set of alternative arcs $((i, j), (h, k))$
$N_a \subset N$	set of arrival nodes
$N_d \subset N$	set of departure nodes
Δ_r	set of superscripts δ for running or dwell time constraints where $\delta \in \Delta_\mathrm{r} = \{\mathrm{e}, \mathrm{r}, \mathrm{l}\}$
Δ_h	set of pairs of superscripts (δ_1, δ_2) for headway time constraints where $(\delta_1, \delta_2) \in \Delta_\mathrm{h} = \{$(e, e), (r, r), (l, l), (e, r), (e, l), (r, e), (r, l), (l, e), (l, r)$\}$

Table 5.2: Notation for indices

Symbol	Description
i, j, h, k	indices for nodes
δ	index for a type of running or dwell time
(δ_1, δ_2)	indices for a type of a pair of headway time

Principles of the Alternative Graph model

In the AG formulation, the model is represented as a graph G with a set of nodes N, fixed arcs F and alternative arcs A (Corman et al., 2012). Each node i represents a signal location, marking either the entrance to a block section, or the arrival or departure at a stop location. Specifically, arrival nodes correspond to the arrival at the stop location on a platform track section, indicating the start of a scheduled dwell, while departure nodes represent the departure from the stop location, denoting the end of the dwell. Each node is associated with a rescheduled time instant t_i^r, representing the start of an operation, such as the traversal of a block section or the dwell at a stop. Only arrival and departure nodes are additionally assigned a planned time instant T_i. An arc (i, j) represents the precedence relationship between nodes i and j. All arcs are directed in an AG formulation, and thus, the order between two events is inherently indicated.

Table 5.3: Notation for parameters

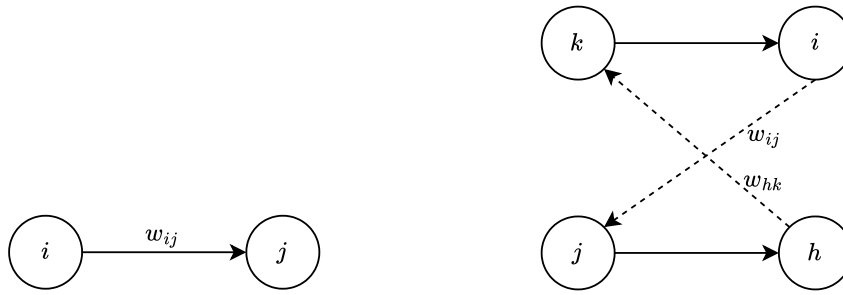
Symbol	Description [unit]
w_{ij}^{f}	arc weight of fixed arc (i,j) [s]
$w_{ij}^{\mathrm{a}}, w_{hk}^{\mathrm{a}}$	arc weight of a pair of alternative arcs (i,j) and (h,k) [s]
T_i	published time at arrival or departure node i [s]
I_i	initial departure delay at departure node i for delayed trains, where $I_i = 0$ for non-delayed trains [s]
p	punctuality threshold [s]
η	weight of rescheduled train event times
$\eta^{\mathrm{f,\,dep}}$	weight of late departure flexibility
M	a sufficiently large value

Table 5.4: Notation for decision variables

Symbol	Type	Description [unit]	Domain
t_i^{r}	continuous	rescheduled time at node i [s]	≥ 0
t_i^{e}	continuous	earliest flexible time at node i [s]	≥ 0
t_i^{l}	continuous	latest flexible time at node i [s]	≥ 0
$f_i^{\mathrm{e,\,arr}}$	continuous	early arrival flexibility at arrival node i [s]	≥ 0
$f_i^{\mathrm{l,\,arr}}$	continuous	late arrival flexibility at arrival node i [s]	≥ 0
f_i^{dep}	continuous	late departure flexibility at departure node i [s]	≥ 0
$\alpha_{((i,j),(h,k))}$	binary	choice of an alternative arc from a pair (equals 0, if arc (i,j) is selected; equals 1, if arc (h,k) is selected)	$\{0,1\}$
$\upsilon_i^{\mathrm{arr}}$	binary	indication whether the difference between the rescheduled arrival time and the published arrival time exceeds the punctuality threshold (equals 0, if not exceeded; equals 1, otherwise)	$\{0,1\}$
$\upsilon_i^{\mathrm{dep}}$	binary	indication whether the difference between the rescheduled departure time and the published departure time exceeds the punctuality threshold (equals 0, if not exceeded; equals 1, otherwise)	$\{0,1\}$

Fixed arcs can be distinguished into processing arcs (indicating the running or dwelling of a train) and timetabling arcs (representing scheduled event times). Figure 5.2 (a) illustrates the former, where w_{ij} denotes the arc weight. Timetabling arcs are required in the AG model to explicitly enforce scheduled event times. In contrast, they are redundant in the MILP formulation, where departure and arrival timing constraints are directly imposed on the nodes/events. Therefore, timetabling arcs are not shown here.

When two trains intend to claim the same block section simultaneously, a pair of alternative arcs is applied to represent possible train orders at these shared resources. Only one alternative arc out of a pair can be chosen, meaning the precedence relationship between two trains entering the shared resource. In Figure 5.2 (b), nodes k and i indicate train A, while nodes j and h signify train B as they enter these shared block sections, respectively. Selecting the alternative arc from nodes i to j implies that train A uses block section 1 before train B. The weight of each alternative arc ensures that the following train enters the shared resource only after the preceding train has fully entered the next block section. This is captured by the weights w_{ij} and w_{hk}, as shown in the figure. The computation of this weight relies on blocking time components, specifically derived



(a) An example of processing arcs *(b) An example of alternative arcs*

Figure 5.2: Representative arcs in the Alternative Graph model

as the minimum headway time between two consecutive trains minus the block running time of the preceding train. In the context of the assumed three-aspect signalling system, as described in this paper, alternative arcs are positioned from a signal located at the end of a block (passed by the preceding train) to the signal situated at the beginning of the block (passed by the following train) (D'Ariano et al., 2007a; Goverde et al., 2013).

Modelling event time flexibility

To incorporate the proposed three types of event time flexibility (i.e., late departure flexibility, early arrival flexibility, and late arrival flexibility) into our model, we introduce the flexibility arcs to the conventional AG model as demonstrated in Figure 5.3. Assuming nodes i and j represent the arrival and departure nodes of a train and w_{ij} is the dwell time at this station, two additional sets of arrival and departure times are introduced, both sharing the same dwell time weight. The grey nodes i and j denote the rescheduled time instants t_i^{r}, the red nodes i' and j' correspond to the latest flexible time instants t_i^{l}, and the blue nodes i'' and j'' serve as the earliest flexible time instants t_i^{e}. Based on the model assumptions, late departure flexibility f^{dep} is specified by the late departure flexibility arc from the rescheduled departure node j to the latest flexible departure node j'. For completeness, we also show the early departure flexibility $f^{\mathrm{e,\,dep}}$ arc from the earliest departure node j'' to the departure node j, which is set to 0 in this study. In the mathematical formulation, the earliest arrival time at node i'' is directly linked to the rescheduled departure time j, so there is no need for a separate early departure flexibility arc. Regarding the arrival nodes, the early arrival flexibility $f^{\mathrm{e,\,arr}}$ arc starts from the earliest arrival node i'' to the rescheduled arrival node i, while the late arrival flexibility $f^{\mathrm{l,\,arr}}$ arc begins from the rescheduled arrival node i to the latest arrival node i'.

All nodes should respect the operational constraints in the AG model, including the nodes on the rescheduled, late and early paths. Although arrival and departure flexibility at stations are the main outputs for dispatchers, each block entry is also associated with the three time instants t_i^{θ}, where $\theta \in \{\mathrm{r}, \mathrm{e}, \mathrm{l}\}$. Therefore, not only the rescheduled train path but also its earliest and latest bounds guarantee conflict-free operations. In particular, as long as train entries at each block section remain within these bounds, the overall solution is conflict-free at the microscopic level.

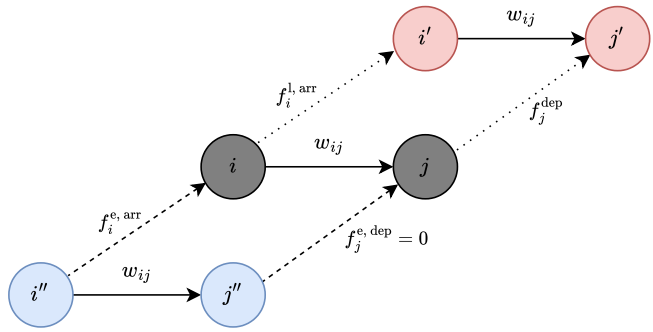


Figure 5.3: Flexibility nodes and arcs in the Alternative Graph model for fictitious arrival and departure nodes i and j

MILP formulation

The proposed MILP formulation for the TRF model is presented in this subsection. Based on the node–arc structure of the AG model, extended with the flexible time instants and flexibility arcs, the MILP reformulation translates this structure into decision variables and constraints that allow the joint optimisation of punctuality and event time flexibility. The objective is to generate a rescheduling plan that minimises timetable deviations caused by initial departure delays of trains, while subsequently maximising event time flexibility at departures and arrival events. This flexibility is incorporated at each RTTP update by allowing admissible variations in departure and arrival times at critical timetable points to be absorbed without triggering additional conflict resolution, thereby improving overall efficiency. By maximising event time flexibility in the objective function, the model identifies the largest feasible admissible ranges for event times without compromising conflict-freeness or punctuality. The objective function is formulated as follows:

$$\text{Minimise} \quad \eta \sum_{i \in N_a \cup N_d} t_i^{\text{r}} - \eta^{\text{f, dep}} \sum_{i \in N_d} f_i^{\text{dep}} - \sum_{i \in N_a} (f_i^{\text{e, arr}} + f_i^{\text{l, arr}}). \tag{5.1}$$

Subject to the following constraints:

Running and dwell time constraints

$$t_j^{\delta} \geq t_i^{\delta} + w_{ij}^{\text{f}} \qquad \forall \delta \in \Delta_{\text{r}}, (i,j) \in F. \tag{5.2}$$

Headway constraints

$$t_j^{\delta_1} \geq t_i^{\delta_2} + w_{ij}^{\text{a}} - M\alpha_{((i,j),(h,k))} \qquad \forall (\delta_1, \delta_2) \in \Delta_{\text{h}}, ((i,j),(h,k)) \in A, \tag{5.3}$$

$$t_k^{\delta_1} \geq t_h^{\delta_2} + w_{hk}^{\text{a}} - M(1 - \alpha_{((i,j),(h,k))}) \qquad \forall (\delta_1, \delta_2) \in \Delta_{\text{h}}, ((i,j),(h,k)) \in A. \tag{5.4}$$

Timetable constraints

$$t_i^{\text{r}} \geq T_i + I_i \qquad \forall i \in N_a \cup N_d. \tag{5.5}$$

Departure flexibility constraints

$$t_i^{\mathrm{r}} - T_i - p \leq M\upsilon_i^{\mathrm{dep}} \qquad \forall i \in N_d, \tag{5.6}$$

$$t_i^{\mathrm{r}} - T_i - p \geq -M(1-\upsilon_i^{\mathrm{dep}}) \qquad \forall i \in N_d, \tag{5.7}$$

$$t_i^{\mathrm{l}} - T_i - p \leq M\upsilon_i^{\mathrm{dep}} \qquad \forall i \in N_d, \tag{5.8}$$

$$f_i^{\mathrm{dep}} \leq t_i^{\mathrm{l}} - t_i^{\mathrm{r}} \qquad \forall i \in N_d, \tag{5.9}$$

$$f_i^{\mathrm{dep}} \leq p(1-\upsilon_i^{\mathrm{dep}}) \qquad \forall i \in N_d. \tag{5.10}$$

Earliest departure time constraints

$$t_i^{\mathrm{e}} = t_i^{\mathrm{r}} \qquad \forall i \in N_d. \tag{5.11}$$

Arrival flexibility constraints

$$t_i^{\mathrm{r}} - T_i - p \leq M\upsilon_i^{\mathrm{arr}} \qquad \forall i \in N_a, \tag{5.12}$$

$$t_i^{\mathrm{r}} - T_i - p \geq -M(1-\upsilon_i^{\mathrm{arr}}) \qquad \forall i \in N_a, \tag{5.13}$$

$$t_i^{\mathrm{l}} - T_i - p \leq M\upsilon_i^{\mathrm{arr}} \qquad \forall i \in N_a, \tag{5.14}$$

$$f_i^{\mathrm{l,\,arr}} \leq t_i^{\mathrm{l}} - t_i^{\mathrm{r}} \qquad \forall i \in N_a, \tag{5.15}$$

$$f_i^{\mathrm{l,\,arr}} \leq p(1-\upsilon_i^{\mathrm{arr}}) \qquad \forall i \in N_a, \tag{5.16}$$

$$f_i^{\mathrm{e,\,arr}} \leq t_i^{\mathrm{r}} - t_i^{\mathrm{e}} \qquad \forall i \in N_a. \tag{5.17}$$

Train path relation constraints

$$t_i^{\mathrm{l}} \geq t_i^{\mathrm{r}} \geq t_i^{\mathrm{e}} \qquad \forall i \in N \setminus (N_a \cup N_d). \tag{5.18}$$

The first term of the objective function (5.1) aims to obtain the rescheduling plan with the smallest deviation from the published timetable by minimising the rescheduled times at arrival and departure nodes. We do not consider the deviation of the rescheduled times at intermediate (block section) nodes in the objective function, as the published times at these nodes are not given. The second and third terms of the objective function seek to maximise the possible flexibility of the rescheduled plan. We set a sufficiently large number of η as the weight of the rescheduled train event times in the objective function at arrival and departure nodes in the objective function to represent its primary goal. Then, we introduce the weight of late departure flexibility $\eta^{\mathrm{f,\,dep}}$ to investigate whether preference for the distribution of flexibility between arrivals and departures would affect the model performance and to see its implications. When $\eta^{\mathrm{f,\,dep}}$ equals 1, it indicates that there is no preference for the distribution of flexibility. A larger $\eta^{\mathrm{f,\,dep}}$ than 1 means that more late departure flexibility is preferred, while a smaller $\eta^{\mathrm{f,\,dep}}$ denotes that more arrival flexibility is favoured. Since departures allow only late departure flexibility, whereas arrivals

may allow both early and late arrival flexibility, the two cases are not fully symmetric. Therefore, this imbalance is examined in the sensitivity analysis.

The rescheduled times at nodes must satisfy the running or dwell time constraint given by constraint (5.2). w_{ij}^{f} indicates the minimum running time through a block section or the dwell time at a station, which is part of the defined fixed arcs. These constraints thus guarantee microscopic feasibility. Constraint (5.2) should be applied to all event times attributed to a node, as expressed by the set Δ_{r}, including the rescheduled time, the earliest flexible time, and the latest flexible time. Constraint (5.3) and Constraint (5.4) are headway constraints that regulate the minimum required headway time between two trains. Only one of the alternative arcs out of a pair can be chosen, which determines the order of the two trains using $\alpha_{((i,j),(h,k))}$. w_{ij}^{a} is the weight of alternative arcs, calculated as the minimum headway time between trains minus the block running time of the preceding train. To ensure that no conflicts occur between each pair of train paths (i.e., early, late and rescheduled), the headway constraints should be respected between rescheduled times, between flexible times, as well as between rescheduled times and flexible times, which is denoted by a set Δ_{h}.

The rescheduling plan aims to minimise deviations from the published times. Specifically, trains should arrive at stations and depart as early as possible in case of delays, while adhering as closely as possible to the published times when no initial delays are present or when trains arrive early. To enforce this, constraint (5.5) ensures that the rescheduled event time t_i^{r} is not earlier than the published event time T_i plus any given initial departure delay I_i. The delay term I_i is defined only for the initial departure event of a train and equals zero for all other events. As a result, for arrival events and non-delayed departure events, the constraint reduces to $t_i^{\mathrm{r}} \geq T_i$, while for an initially delayed departure, it requires $t_i^{\mathrm{r}} \geq T_i + I_i$. This constraint applies exclusively to the rescheduled event times and does not restrict the admissible flexibility bounds, such that early arrivals may still be permitted through the event time flexibility limits.

Event time flexibility is added when the difference between the rescheduled times and the published times does not exceed the punctuality threshold. Regarding departure flexibility, constraints (5.6) and (5.7) are used to determine $\upsilon_i^{\mathrm{dep}}$, indicating whether the punctuality threshold is exceeded for departure events. Constraint (5.8) is introduced to ensure that the difference between the latest and the published departure times is less than or equal to the punctuality threshold. Furthermore, the late departure flexibility is calculated as the difference between the latest flexible and the rescheduled departure times, as given by constraint (5.9). In addition, constraint (5.10) regulates that late departure flexibility can only be added when the punctuality threshold is not exceeded, namely $\upsilon_i^{\mathrm{dep}}$ equals 0. Otherwise, late departure flexibility equals 0, and the latest flexible departure times equal the rescheduled departure times. Lastly, constraint (5.11) ensures that the earliest event time equals the rescheduled event time at all departure nodes. This prevents early departure flexibility, which is not intended in the model.

A similar logic is applied to the late arrival flexibility. It can only be added when the difference between the rescheduled arrival times and the published times does not exceed the punctuality threshold, represented by constraints (5.12) and (5.13). The late arrival flexibility is computed as the difference between the latest flexible arrival time and the rescheduled arrival time, as given by constraint (5.15). Constraint (5.16) specifies that late arrival flexibility can only be added when the punctuality threshold is not exceeded, namely $\upsilon_i^{\mathrm{arr}}$ equals 0. Otherwise, late arrival flexibility equals 0, and the latest flexible arrival times equal the rescheduled arrival times. In our study, early arrival flexibility is defined as the difference between the rescheduled arrival time and the earliest flexible arrival time, as designated by constraint (5.17). Since the

punctuality threshold is only defined for lateness, the upper bound of early arrival flexibility is not limited in the formulation, which means very early arrivals are also allowed, although they are typically exploited for energy-efficient driving. Finally, constraint (5.18) ensures that the rescheduled event time at each intermediate node of a rescheduled train path lies within the corresponding time window defined by the earliest and latest feasible times of the train path with flexibility.

The proposed TRF model can be readily extended to incorporate passenger transfers. In the original AG formulation, this is achieved by introducing connection arcs between the arrival of a feeder train and the departure of a waiting train, which are kept or discarded together (Corman et al., 2012). In the MILP formulation, three sets of binary variables can be added: (i) to maintain or discard a connection arc, (ii) to check whether the feeder train exceeds the maximum acceptable arrival delay, and (iii) to check whether the waiting train exceeds its latest departure time. To focus on the event time flexibility formulation, we exclude them in the current setting.

5.3 Application of the model to a real-world case study

This section applies the proposed TRF model to a part of the Dutch railway network. We first describe the case study and detail the parameter settings. Then, the model is applied to a total of 1000 scenarios featuring random departure delays, along with an additional undisturbed scenario to showcase its effectiveness. Specifically, we evaluate the benefits of added event time flexibility by the computed flexibility values, the number of rescheduling measures applied or avoided, and the computational time, all analysed in relation to the magnitude of the initial delays. Next, we conduct two sensitivity analyses: the first focuses on the distribution of flexibility across departure and arrival events, while the second assesses the impact of varying punctuality thresholds.

5.3.1 Case study description

The selected dispatching area for the case study is located in the southern part of the Dutch railway network. It consists of a mix of converging, diverging and crossing tracks and twelve stations, including 's-Hertogenbosch (Ht), Tilburg (Tb), Breda (Bd), Eindhoven (Ehv), Eindhoven Philips Strijp (Ehs), Best (Bet), Boxtel (Btl), Oisterwijk (Ot), Tilburg Reeshof (Tbr), Tilburg Universiteit (Tbu), Gilze Rijen (Gz), and Vught (Vg) as shown in Figure 5.4. The total network length is approximately 138 km with 978 block sections and 25 platform tracks. The busiest corridor operates with a mean headway of around 4.28 minutes between stations Ehv and Bd.

This case study area contains rail traffic heterogeneity with two train types: Intercity (IC, which only serves major stations) and Sprinter (SPR, which stops at every station, also known as local trains). An overview of the selected trains operating under nominal conditions in one direction is provided in Table 5.5, where each train line follows a half-hourly cycle and repeats periodically throughout the service hours. Platform track allocations are indicated in parentheses alongside the station names. The platform track numbers refer to the platform designations, with an 'a' or 'b' at station Ht indicating separate platform phases, which can be accessed or exited via a double slip switch located halfway along the platform. As shown in Figure 5.4, IC train service 21400 and SPR services 4400 and 6400 share common tracks between stations Ehv and Bet. Similarly, IC train services 800, 1100, 3500, and 3900 operate on the other shared tracks in

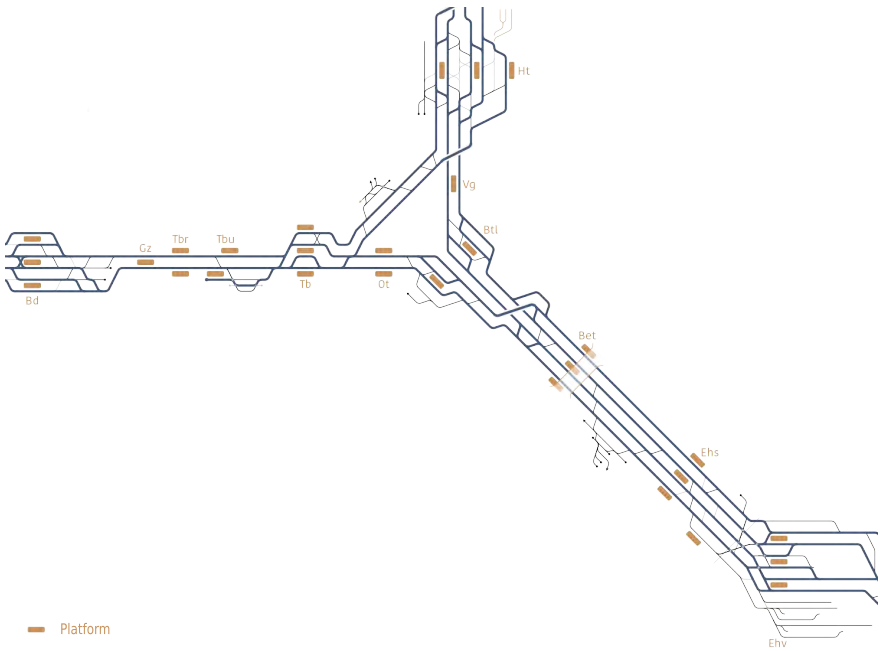


Figure 5.4: Track layout of the case study area

this four-track corridor. Before reaching station Btl, IC train services 21400 and 1100, together with SPR train service 6400, enter the corridor from stations Btl to Tb, while the remaining trains proceed towards station Ht. Upon departing station Tb, IC train services 21400 and 1100 diverge from SPR train service 6400, which continues its journey towards its terminal station Tbu. In contrast, the other trains merge and run along the same tracks in the corridor between stations Tb and Bd. In the opposite direction, all trains merge at a switch and run along the same tracks from stations Ht to Tb after departing the platform area at station Ht.

Table 5.5: Overview of selected trains in the case study area under nominal conditions

Train ID	Train stops (platform)	Train type	Departure time
21400	Ehv(6)-Tb(2)-Bd(7)	IC	0
800	Ehv(4)-Ht(3a)	IC	6
3600	Ht(7b)-Tb(3)-Bd(8)	IC	12
1100	Ehv(5)-Tb(2)-Bd(7)	IC	14
3500	Ehv(6)-Ht(3a)	IC	17
3900	Ehv(5)-Ht(3a)	IC	27
6600	Ht(7b)-Tb(3)-Tbu(1)-Tbr(1)-Gz(3)-Bd(7)	SPR	3
4400	Ehv(6)-Ehs(1)-Bet(1)-Btl(2)-Vg(1)-Ht(4b)	SPR	6
6400	Ehv(4)-Ehs(1)-Bet(1)-Btl(6)-Ot(2)-Tb(2)-Tbu(3)	SPR	21

Table 5.5 also reports the scheduled departure times of each train from their respective origin departure stations. A total time window of 2.5 hours is considered in the case study, where

the first hour serves as a warm-up period for the model, followed by a representative hour during which initial delays are introduced and performance is evaluated, and a final 30-minute cool-down period to allow trains to exit the corridor. Figure 5.5 presents the time-distance diagram for the three corridors, accompanied by a schematic illustration of the corresponding track layout. In the figure, blue lines denote IC train services and red lines represent SPR train services. Different shades of each colour are used to distinguish the same train as it traverses multiple corridors. For instance, IC 800, departing from station Ehv at 6 minutes, leaves the Ehv–Bd corridor at station Btl on the leftmost side of the diagram and, at that moment, appears in the Btl–Ht corridor on the rightmost side.

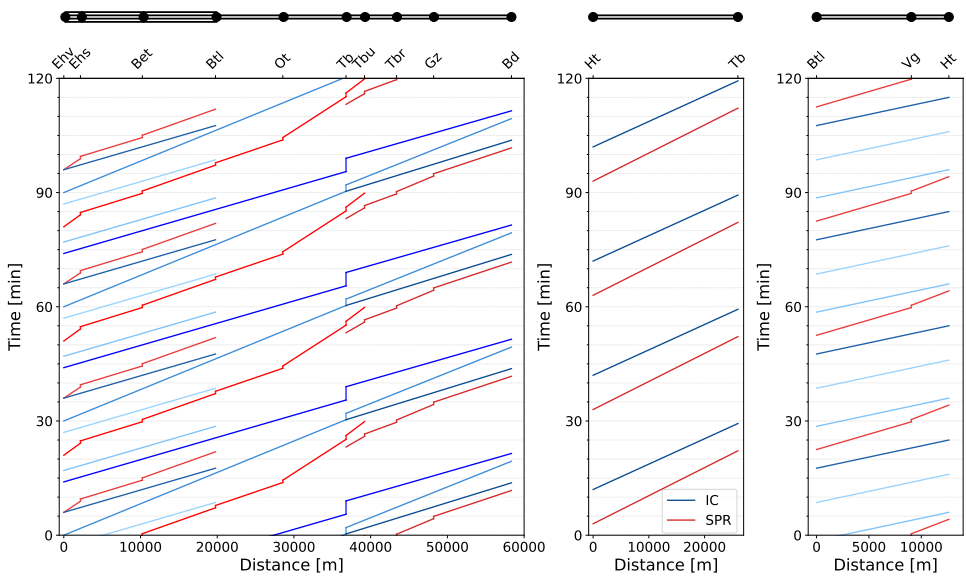


Figure 5.5: Time-distance diagram for the case study network

The weights assigned to running arcs and alternative arcs are derived from input data, determined by the minimum running time of a train through a block section and the minimum headway time (minus the block running time) between two trains at specific locations where alternative arcs are established, respectively. The model also accounts for short block sections with progressive speed signalling, represented by alternative arcs between successive signals and detailed in Goverde et al. (2013) for the Dutch NS'54 three-aspect fixed-block system. The minimum dwell time is set to 60 s at major stations and 36 s at other stations. The weight of the rescheduled time in the objective function is set to $\eta = 100$, i.e., 100 times larger than the unit weight of event time flexibility, indicating that the primary goal is to obtain rescheduled times with the least deviation from the published times. Then, event time flexibility is added and maximised on top of this rescheduling plan. We first set the weight $\eta^{\mathrm{f,\,dep}} = 1$, denoting no preference for departure and arrival flexibility, and consider a punctuality threshold of $p = 180$ s, which is further explored in the subsequent sensitivity analyses. The big-M constant is set to 18,000 s (i.e., 5 hours), which is sufficiently large to cover trains entering before and leaving after the focused hour, thereby avoiding boundary effects in the optimisation. All parameter values are determined in consultation with the Dutch railway infrastructure manager ProRail and

may vary depending on country-specific operational standards and infrastructure characteristics.

With all the relevant information included, the optimisation model comprises 14,576 continuous variables, 25,815 binary variables, and 373,370 constraints. Computational scenarios were solved to optimality (with a gap less than 0.00001%) using the optimisation software GUROBI release 12.0 on a laptop with an Intel(R) Core(TM) i7-10610U CPU (1.80 GHz) and 16 GB RAM. The corresponding computation times are reported in the next subsection on results.

5.3.2 Application of the model under undisturbed and disturbed scenarios

To test the effectiveness of the proposed model, we use a Gamma distribution to randomly generate 1000 scenarios, denoted as S1 to S1000. Each scenario introduces initial departure delays for all trains at the boundary stations (i.e., the first node of each train within the modelled area, such as Ehv and Ht) during the representative hour within the 2.5-hour planning horizon. Based on real-life operational data from the second quarter of 2019 provided by ProRail, the average departure delay and standard deviation are 35.4 s and 84.6 s for IC trains, and 32.4 s and 84 s for SPR trains at these boundary stations, averaged across the boundary stations Ehv and Ht, where the values are similar. The parameters of the Gamma distribution are computed as shape $= (\mu/\sigma)^2$ and scale $= \sigma^2/\mu$, where μ and σ denote the empirical mean and standard deviation, respectively. Applying these relations results in shape and scale values of 0.18 and 202.18 for IC trains, and 0.15 and 217.78 for SPR trains. Figure 5.6 illustrates the fitted Gamma probability density functions, where large departure delays are observed with decreasing probability.

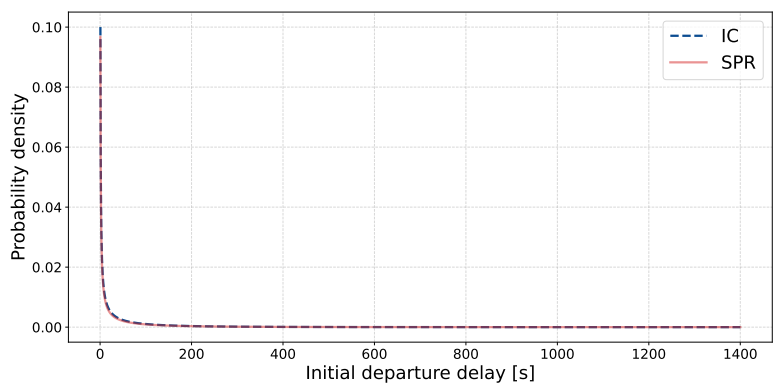


Figure 5.6: Probability density of the fitted Gamma distributions for IC and SPR trains

Table 5.6 presents aggregated statistics across the 1000 disturbed scenarios. The table begins with the disturbance context as reflected by the delay. The average initial delay ranges from 4.31 s in scenario S243 to 111.75 s in scenario S101, computed across all 18 trains in the focused hour. Scenario S243 also shows the smallest maximum delay and the lowest standard deviation, whereas the most severe disturbance is observed in scenario S298, where the maximum delay reaches 1332.84 s (over 22 minutes) for the second SPR train service 4400. In contrast, the smallest delays are below 1 s, observed across multiple trains in different scenarios. The objective values remain close to the undisturbed scenario S0 (251,701,146) as delay minimisation dominates the optimisation, while maximisation of flexibility has only a secondary influence. Across the disturbed scenarios, the minimum objective is 251,710,627 in scenario S243, where

delays are minimal, while the maximum is 253,042,577 in scenario S161. Because scenario S161 combines high delays (average 101.42 s, maximum 1271.69 s, standard deviation 269.40 s) with the lowest total flexibility (13,526.19 s, or 49.01 s per event in the focused hour), it leads to the largest objective value within the disturbed set.

Table 5.6: Aggregated statistics of rescheduling performance across 1000 disturbed scenarios

Metric	Minimum	Mean	Maximum
Average initial delay [s]	4.31	34.38	111.75
Maximum initial delay [s]	33.10	288.20	1,332.84
Standard deviation of initial delay [s]	8.20	70.40	278.78
Objective value	251,710,627	251,833,415	253,042,577
Flexibility per event (focused hour) [s]	49.01	64.92	67.01
Retiming average at station events [s]	0.48	6.90	85.38
Retiming maximum at station events [s]	85.67	307.84	1,332.84
Reordering count (binary variable $\alpha_{((i,j),(h,k))}$ choices)	0	70.16	2,733
Computation time [s]	30.96	72.45	898.93

Flexibility per event remains relatively consistent across the disturbed scenarios, averaging about one minute. However, local event time deviations can still be substantial. Retiming of station events averages 6.90 s, but in extreme scenarios, individual events are retimed by more than 1300 s in scenario S390. Reordering is modelled by binary variables $\alpha_{((i,j),(h,k))}$, which represent alternative precedence choices between trains at specific locations. On average, about 70 of the 25,445 binary variables change per scenario. Many scenarios require no reordering at all, indicating that the original train order can often be preserved, whereas in extreme scenarios, such as scenario S998, more than 2700 order changes are determined. Computation times average 72.46 s, ranging from the shortest of 30.96 s in scenario S314 to the longest of 898.93 s in scenario S562. Further analysis of these patterns follows.

We present the computational results in Figure 5.7, where the computed late departure flexibility, early arrival flexibility, and late arrival flexibility are plotted against the average initial delay, maximum initial delay, and standard deviation of the initial delay, respectively, in the left-hand column of subfigures. The right-hand column shows the corresponding computation time for each scenario. The base scenario S0, without any initial delays, is included as the leftmost point in each subfigure for reference. The shaded contours show the interquartile range (25th–75th percentiles), indicating the variability of the data points. The central curve is a locally weighted smoothing (LOESS) that traces the overall trend without assuming a predefined functional form. At the high-delay end, only a few distinct values remain, which prevents reliable quartile estimation. The shaded contour therefore terminates, while the smoothing curve continues as it is fitted to all available points. The average and median values across all 1001 scenarios are 7625.08 s and 7673.63 s for late departure flexibility, 3810.27 s and 3803.10 s for early arrival flexibility, and 6482.83 s and 6604.62 s for late arrival flexibility.

Overall, as initial delays increase, all three types of event time flexibility decline. The total flexibility decreases from 18,499.19 s in the undisturbed scenario S0 to 13,526.19 s in scenario S161, which combines substantial delays with the lowest overall flexibility. Late departure flexibility shows a consistent downwards trend with increasing average delay, dropping from

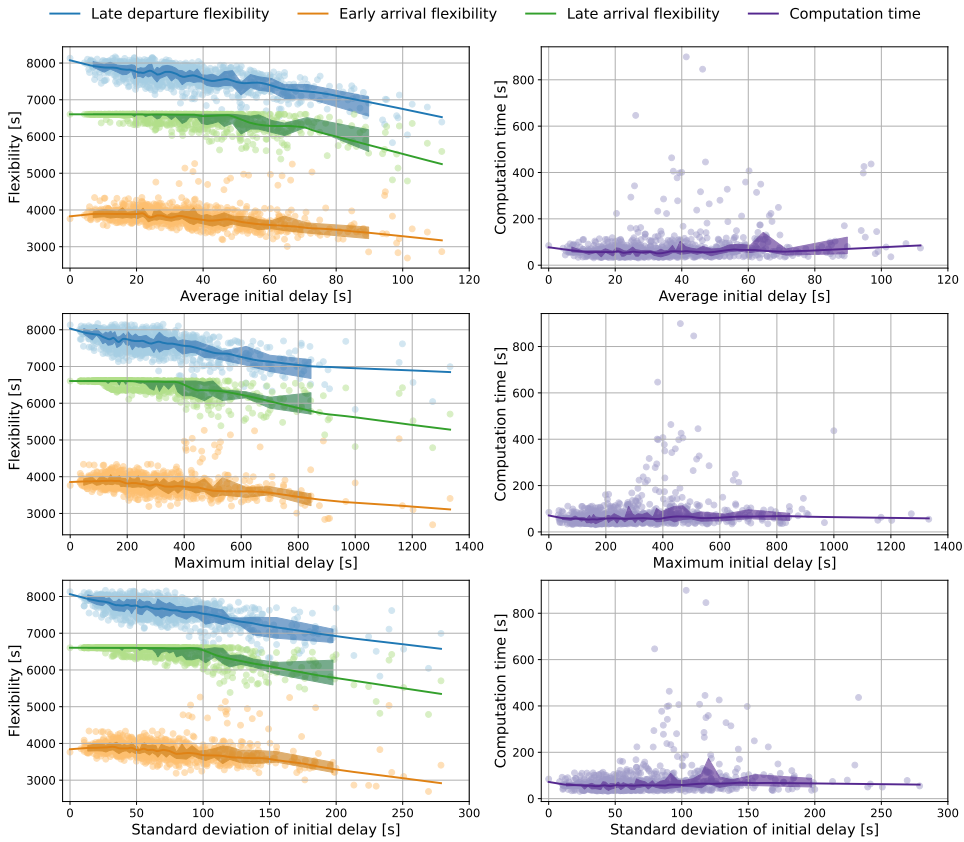


Figure 5.7: Event time flexibility and computation time across the undisturbed and 1000 disturbed scenarios

8,167.98 s in scenario S710 (average initial delay of 13.57 s) to 5,831.83 s in scenario S390 (average initial delay of 96.92 s, where the most extensive retiming takes place). Particularly, scenario S710 results in greater late departure flexibility than scenario S0 (8,132.45 s), due to train reordering and headway interactions that create additional leeway in the timetable. Late arrival flexibility also declines, with a sharp reduction beyond an average initial delay of 60 s, indicating that most timetable allowances are exhausted beyond this point. Among the three flexibility types, early arrival flexibility shows the highest variability since late departures naturally reduce the available time window for early arrivals.

Compared to average delay, event time flexibility is less sensitive to the maximum initial delay or its standard deviation. For instance, scenario S298, which simultaneously has the highest maximum initial delay (1332.84 s) and the highest standard deviation (278.78 s), still maintains moderate flexibility, including 6,995.32 s for late departure, 3,407.85 s for early arrival, and 5,704.62 s for late arrival. Likewise, scenario S130, despite a high standard deviation of 239.74 s, an average delay of 72.73 s, and a maximum delay of 1,154.15 s, retains 18,410.28 s of total flexibility, which is close to the undisturbed scenario.

Furthermore, computation time peaks at 898.93 s (scenario S562), where the average initial

delay is 41.36 s, the maximum initial delay is 461.34 s, and the standard deviation of the initial delay is 103.34 s. The second-highest computation time of 845.89 s occurs in scenario S933, with an average delay of 46.26 s, a maximum of 508.27 s, and a standard deviation of 118.17 s. These scenarios indicate that moderate delays spread across many trains, particularly in densely scheduled corridors, expand the feasible solution space and require the solver to evaluate more combinations, leading to deeper branching. In contrast, both low-delay and high-delay extremes tend to offer more constrained solution structures, resulting in reduced computation times. For example, scenario S314 has the shortest computation time of 30.96 s, with an average delay of 26.93 s, a maximum of 163.21 s, and a standard deviation of 44.44 s. A similar pattern is observed in other scenarios: scenario S101, which has the highest average initial delay, solves in just 74.24 s, while scenario S298, which has the highest maximum delay and standard deviation, completes in 54.88 s.

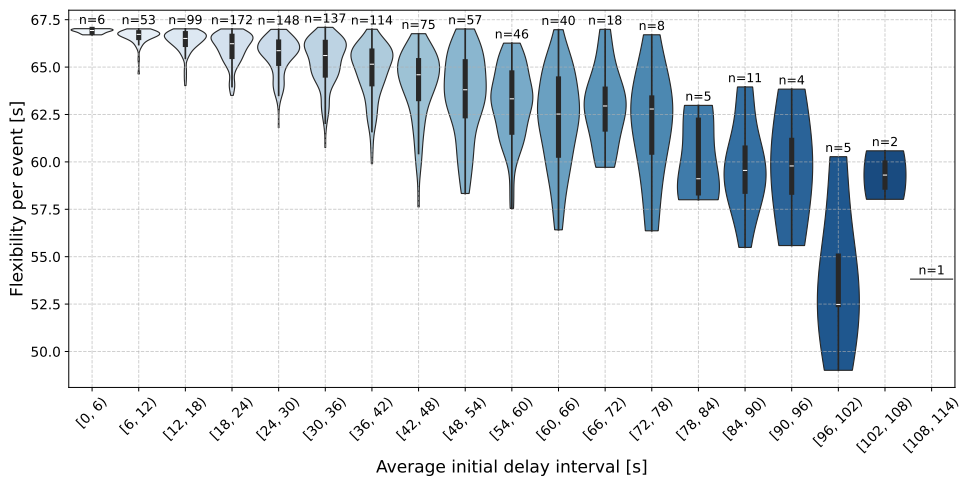


Figure 5.8: Flexibility per event across 6-second bins of scenario-level average initial delay

To analyse how event time flexibility varies with delay severity, the 1001 scenarios are grouped into 6-second bins (i.e., $[(n-1)\cdot 6, n\cdot 6)$), and the flexibility per event during the representative hour is visualised using a violin plot in Figure 5.8. The shade of blue indicates the delay severity, while the sample size within each bin is shown above the violins. The white marker denotes the median, the thick black bar shows the interquartile range, and the thin black line indicates the distribution excluding outliers (defined as values beyond 1.5 times the interquartile range from the 25th and 75th percentiles). The median flexibility per event decreases from 66.93 s in the (almost) undisturbed case (0–6 s) to 52.48 s in the bin with average delays between 96 and 102 s. In mid-range bins (30 to 60 s), the median remains around 66.31 s before declining more noticeably. As delay severity increases, the variability in flexibility also tends to grow, as shown by the broader interquartile ranges and more dispersed distribution shapes. This reflects increased variation in how flexibility is reallocated across events due to differences in train interactions. Scenarios with low flexibility correspond to cases where available flexibility is used to absorb delays, while high-flexibility scenarios indicate potential train reordering that minimises delay and introduces additional flexibility.

Complementing the trend analysis at the aggregated level, we provide a visualisation of the

train paths with flexibility under the undisturbed scenario (S0) for the representative corridor from Ehv to Bd (see Figure 5.9). Following the same styling and colour scheme as in the time–distance diagram, the solid lines indicate the scheduled train paths, with blue for IC and red for SPR services. Each train path is constructed by connecting the scheduled departure and arrival times at consecutive stations. Event time flexibilities are derived from the earliest and latest feasible departure and arrival times computed by the model and shown as shaded bands around the scheduled paths. Visual overlaps between flexibility bands may occur due to the drawing method, which does not distinguish between parallel tracks or separate platform tracks at stations, and these overlaps do not indicate track occupation conflicts. The underlying optimisation results, however, incorporate microscopic blocking times at the block-section level, ensuring that the timetable is conflict-free at the microscopic scale. When no initial delays are added to trains, the flexibility per event reaches a maximum of 67.03 s.

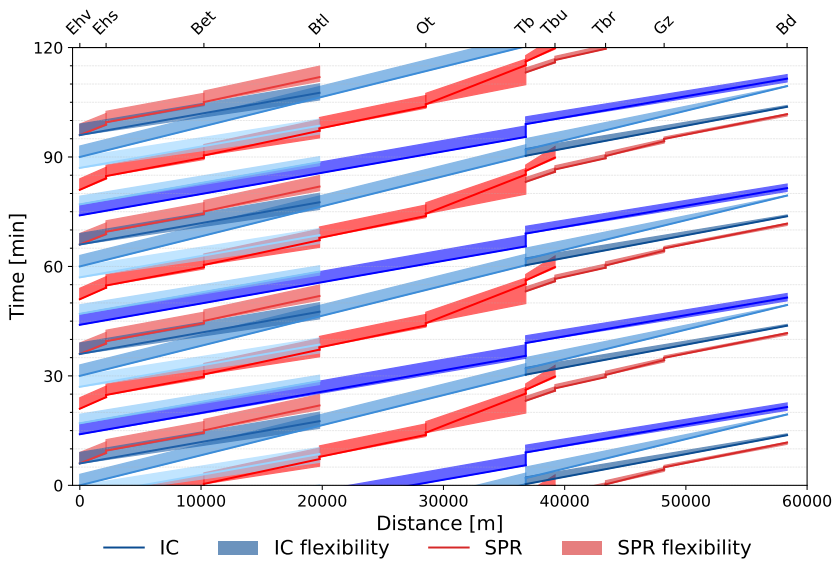


Figure 5.9: Train paths with flexibility for the corridor Ehv–Bd under the undisturbed scenario (S0)

The train paths under the disturbed scenario with the highest average initial delay (scenario S101) are shown in Figure 5.10, with the originally scheduled train paths indicated by grey dashed lines. Compared to the undisturbed scenario, train paths exhibit varying degrees of retiming depending on the severity of their initial delays. For instance, IC train service 1100, departing at 14 minutes with an initial delay of 87.36 s, is highlighted by the green circle and loses all late departure flexibility at station Ehv, decreasing from the full value of 180 s to 0 s. Correspondingly, its early arrival flexibility at station Tb reduces from 39.32 s to 0 s, as the running time supplement is used to recover from the delay. In a more extreme case, SPR train 6400, scheduled to depart at 51 minutes with an initial delay of 909.88 s, is marked by the purple circle. As a result, all three types of flexibility drop to 0 s across intermediate stops until the train departs from Tb, after which full flexibility is restored. Early arrival flexibility is only regained before the final stop at Tbu, where it reaches approximately 15 s. The train order between SPR

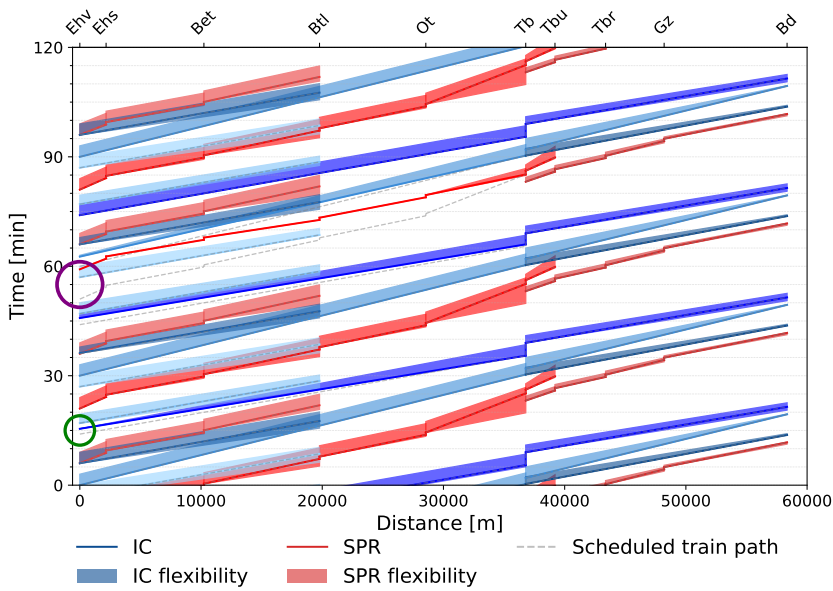


Figure 5.10: Train paths with flexibility for the corridor Ehv–Bd under the disturbed scenario (S101)

6400 and its originally following IC 3900 (scheduled at 57 minutes) is reversed, indicating a reordering to minimise delay and maximise flexibility. This reordering benefits the preceding IC 3500 (scheduled to depart at 47 minutes and with an initial delay of nearly 0 s), which gains additional buffer as the delayed SPR 6400 no longer restricts the passing time at the shared switch. Its late departure flexibility increases from 148.57 s to 180 s at station Ehv.

Lastly, we present the ratios of rescheduling decisions across all scenarios in Figure 5.11, classified into three categories: (1) rescheduled event times that remain within the flexibility bounds computed in the undisturbed scenario (S0), (2) rescheduled event times that fall outside those bounds, and (3) minor delays of 6 s or less, which are considered negligible and expected to be absorbed using existing running time supplements. Scenarios are grouped into 6-second bins based on the average initial delay, consistent with the previous violin plot. For each scenario, 138 station event times (including both arrivals and departures) are evaluated, and the proportional distribution of event times across the three categories is calculated. These ratios are visualised as box plots, showing the distribution across scenarios within each delay bin. Each box depicts the median (black horizontal line), the interquartile range (box edges), and outliers, defined in the same manner as the violin plot.

Across all delay bins, a substantial share of rescheduled event times falls into the category of absorbed minor delays, reflecting operational practice where small fluctuations in train paths are common and typically handled by running time supplements. As the average initial delay increases, the ratios of events rescheduled either within or beyond the originally computed flexibility bounds grow steadily. When the average delay is below 30 s, approximately 90% of event time deviations remain minor and absorbable, while only around 2% require adjustment within the flexibility bounds. For scenarios with larger disturbances, the proportion of within-

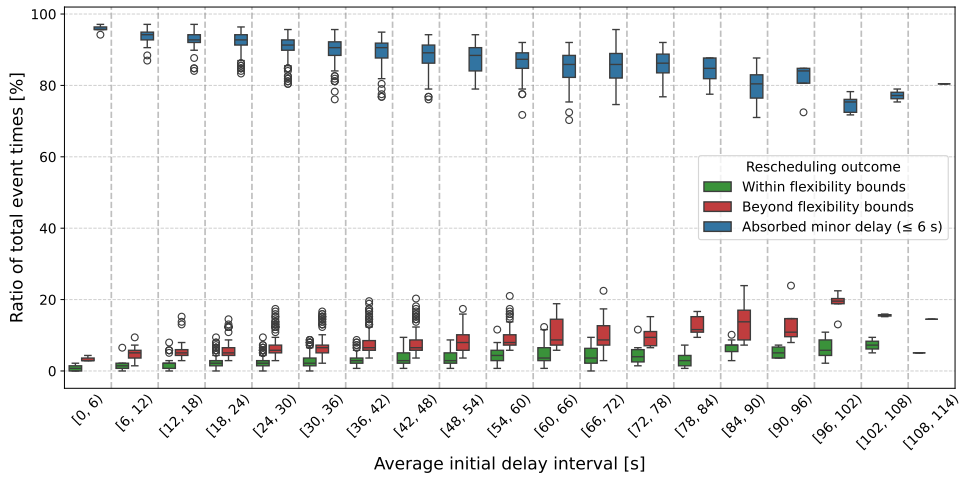


Figure 5.11: Ratios of station event times per scenario falling within flexibility bounds, beyond flexibility bounds, or absorbed as minor delays (≤ 6 s), grouped by average initial delay

bound rescheduling increases from 2.17% at 30 s to 7.26% at both the 84 s and 102 s average delay bins, suggesting that event time flexibility is more actively utilised under heavier delays. Beyond-bound rescheduling adjustments also escalate, indicating a rising need for significant retiming when delays exceed the available time allowances. While event time flexibility cannot fully eliminate the need for rescheduling, these results demonstrate its value in enabling a controlled adaptation to delays without relying on full-scale, network-wide optimisation. It reduces computational burden, eases the operational workload for traffic controllers and dispatchers, and provides bounded time intervals for station events that can be regarded as train path envelopes, serving as trajectory generation constraints for train operations. For both ATO-enabled and human-based train operations, such flexibility maintains consistency between real-time train operations and the RTTP under both nominal and disturbed conditions, thereby improving overall performance.

5.3.3 Sensitivity analysis on flexibility weight in the objective function

In this sensitivity analysis, we adjust the value of the late departure flexibility weight $\eta^{\text{f, dep}}$ to assess its impact on the distribution of event time flexibility and the trade-off between departure and arrival flexibilities. We vary $\eta^{\text{f, dep}}$ across values of 0.9, 1.0, and 1.1 for a selected subset of 101 scenarios, consisting of the undisturbed base scenario (S0) and the first 100 disturbed scenarios (S1 to S100), as this analysis focuses on internal flexibility balance rather than model behaviour under varying disturbance levels. The weight of rescheduled times remains fixed at $\eta = 100$, maintaining our primary objective of minimising deviation from the originally scheduled timetable. We also tested a wider range of $\eta^{\text{f, dep}}$ values, from 0.5 to 2.0. The resulting event time flexibility did not change, as the model consistently prioritises one side of the trade-off once a slight preference is introduced. For values below 0.9, the flexibility allocation remains identical to that of $\eta^{\text{f, dep}} = 0.9$, while for values above 1.1, it remains the same as for

$\eta^{f,\,dep} = 1.1$, except for a few cases when $\eta^{f,\,dep} = 0.5$, where changes remain marginal (i.e., less than 1 s per event). In all such cases, only the objective value and computational time vary, while the flexibility allocation remains stable.

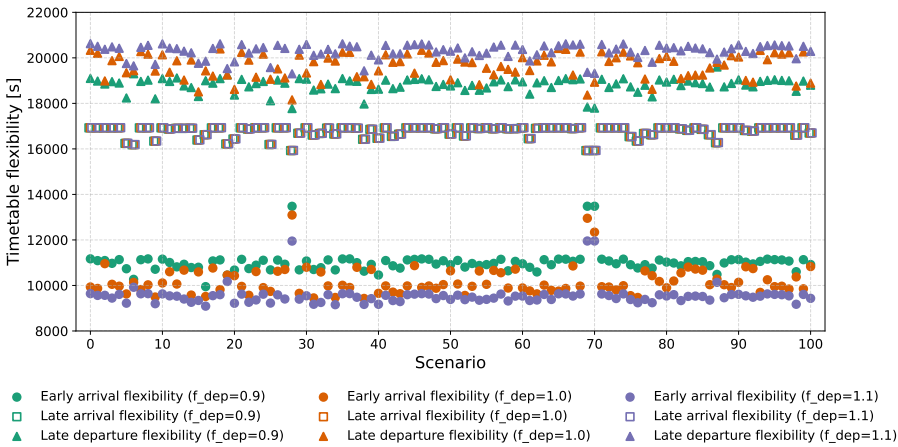


Figure 5.12: Event time flexibility by type across 101 scenarios for $\eta^{f,\,dep} = 0.9$, 1.0, *and* 1.1

We present the results of this analysis in Figure 5.12. The plot shows a consistent monotonic trend as $\eta^{f,\,dep}$ varies from 0.9 to 1.1. Across all 101 scenarios, late arrival flexibility remains unchanged for all three parameter values. In contrast, early arrival flexibility decreases while late departure flexibility increases, reflecting the intended shift in relative importance introduced by the weight variation.

A trade-off among different event time flexibility types exists only between late departure and early arrival flexibilities. The platform tracks are protected by entry signals at a station, and hence, the minimum headway time must be strictly respected between the departure of a preceding train and the arrival of a following train. The successive arrival and departure of the preceding train are additionally constrained by the minimum dwell time. When the late departure flexibility of the preceding train decreases, the following train benefits from the additional margin, increasing its early arrival flexibility. However, this reduction in late departure flexibility also limits the late arrival flexibility of the same train, as the required dwell time must still be satisfied. As a result, late arrival flexibility remains unchanged across all tested weights, since any further shift would violate the minimum dwell or headway constraints.

To provide a clearer illustration of the results, we present the train paths with flexibility under the undisturbed scenario (S0) in Figure 5.13 and Figure 5.14, comparing the two late departure flexibility weights $\eta^{f,\,dep} = 0.9$ and 1.1. IC service 1100 departs from station Ehv at 14 minutes in each cycle, with an intermediate stop at station Tb before continuing to station Bd. As $\eta^{f,\,dep}$ increases from 0.9 to 1.1, its late departure flexibility at station Ehv rises from 59.79 s to 180 s, while its early arrival flexibility at station Tb drops from 106.25 s to 0 s, as indicated by the orange circles. Similarly, IC service 800 and SPR service 4400 depart simultaneously at 6 minutes and are subject to a headway constraint before reaching station Btl. When the late departure flexibility weight is adjusted from 0.9 to 1.1, the late departure flexibility of IC service 800 increases from 111.55 s to 180 s, while the early arrival flexibility of SPR service 4400 decreases from 136.20 s to 67.76 s at station Btl, as shown by the black circles.

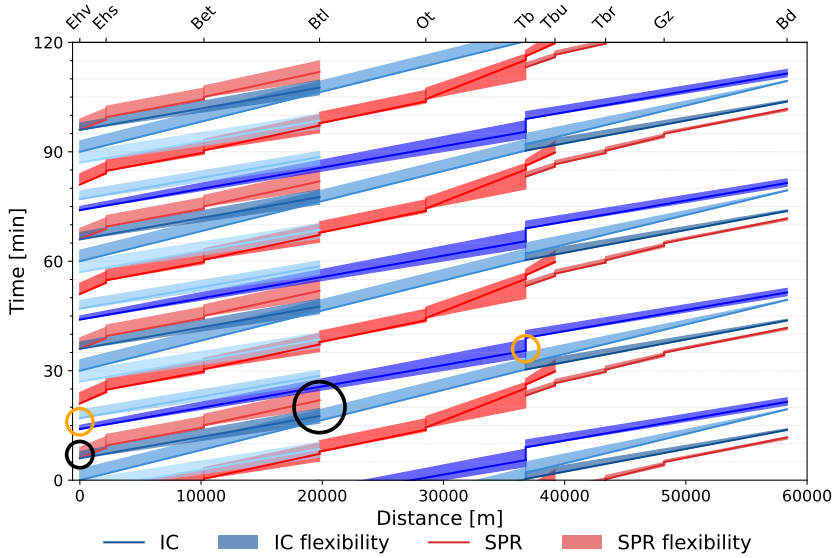


Figure 5.13: Train paths with flexibility for the corridor Ehv–Bd under the undisturbed scenario (S0) with $\eta^{\mathrm{f,dep}} = 0.9$

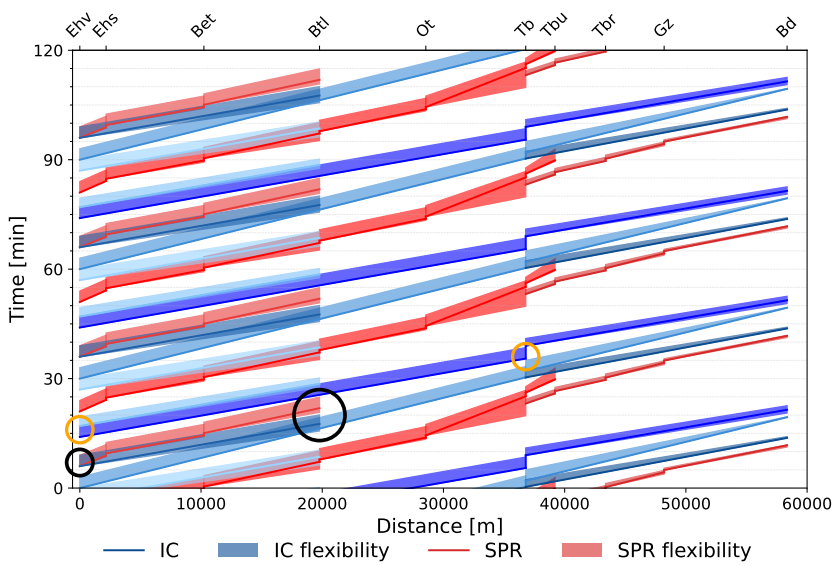


Figure 5.14: Train paths with flexibility for the corridor Ehv–Bd under the undisturbed scenario (S0) with $\eta^{\mathrm{f,dep}} = 1.1$

5.3.4 Sensitivity analysis on punctuality threshold

In the previous results, the punctuality threshold p was fixed at 180 s, specifying the maximum allowable deviation between the latest and published departure or arrival times. This sensitivity analysis investigates the impact of varying p on the resulting event time flexibility. We gradually increase the value of the punctuality threshold from 90 s to 900 s with an increment of 90 s, and present the corresponding flexibility per event in Figure 5.15. This analysis is conducted on the undisturbed scenario (i.e., without any initial delays) and focuses on the representative evaluation hour. It can be observed that flexibility per event increases with a more permissive punctuality threshold, as trains are allowed to arrive and depart later relative to their scheduled times, provided that headway and trip time constraints are respected.

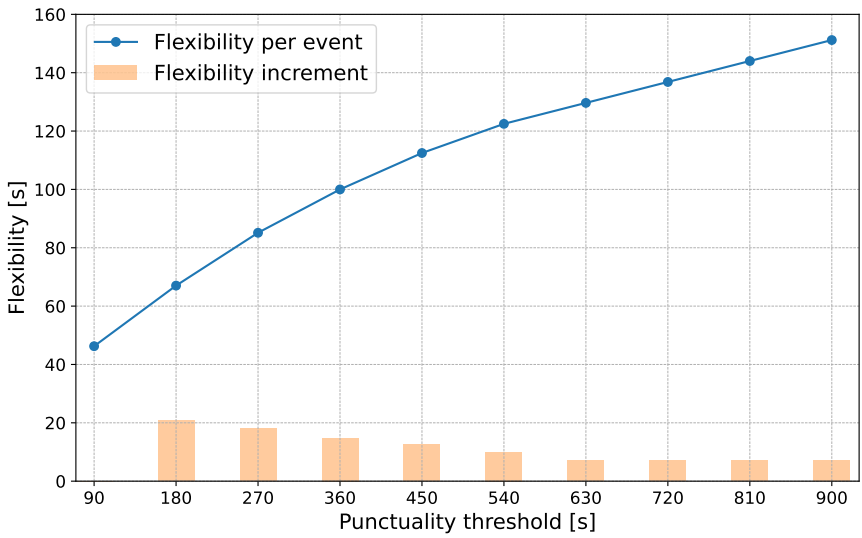


Figure 5.15: Flexibility per event and its increments for varying punctuality thresholds p

We further examine the flexibility increment, defined as the difference in flexibility per event between adjacent values of the punctuality threshold. The results show a decreasing trend, where the increment drops from 20.77 s (from 90 s to 180 s) to 7.18 s beyond 630 s. This indicates a saturation point, beyond which increasing the punctuality threshold does not lead to additional flexibility gains. This phenomenon results from the interaction of multiple constraints. The punctuality threshold directly regulates the latest allowable departure and arrival times, while the event time flexibility is also bounded by running and dwell time constraints, as well as headway constraints. Additional flexibility gains are mainly attributed to increased departure and late arrival margins, particularly for trains operating on four-track sections with sufficient headway, as well as in other parts of the network where buffer times are available.

5.4 Conclusion

Daily operational disturbances challenge the fixed event times of the “rigid” planned timetable, prompting the need for explicitly bounded flexibility in rescheduling. This paper proposed

a train rescheduling model that incorporates event time flexibility to optimise the updated timetable contained in the Real-Time Traffic Plan (RTTP) used in proactive railway traffic management. Event time flexibility is modelled as bounded time windows, computed based on timetable allowances and punctuality thresholds, that define admissible ranges of event times for late departures, early arrivals, and late arrivals. Within these flexibility bounds, the RTTP remains punctual and conflict-free, thereby reducing unnecessary rescheduling interventions. The microscopic rescheduling model provides reordering and retiming decisions that jointly minimise the timetable deviations and maximise flexibility, while guaranteeing conflict-freeness at the block signalling level.

The proposed model was validated using a real-life case study on the Dutch railway network. The validation included one undisturbed scenario and 1,000 disturbed scenarios with random initial delays applied to each train. The results demonstrated that the model effectively minimises delay while introducing event time flexibility. Although event time flexibility generally decreases with increasing initial delays, interactions such as train reordering or additional leeways created when a following train is delayed can locally increase flexibility. The introduced event time flexibility helps absorb minor delays and reduces the need for active rescheduling interventions, as long as the delay remains within the computed flexibility bounds. Such controlled flexibility can be used in train path envelopes that constrain train trajectory generation for both ATO-enabled and human-based train operations. A trade-off is observed only between late departure and early arrival flexibility, due to constraints imposed by minimum headway and dwell times. In addition, the analysis reveals that flexibility per event stabilises beyond a certain punctuality threshold, indicating a practical limit to the benefits of further relaxing punctuality thresholds.

Future research will explore a dynamic simulation environment coupled with an RTTP with event time flexibility to demonstrate its effectiveness in real-world scenarios. Additional rescheduling measures beyond retiming and reordering could also be considered. For example, incorporating rerouting may further enhance flexibility, although this would expand the model dimensionality. Lastly, the concept of event time flexibility can be integrated into other types of rescheduling models, enabling them to better cope with operational uncertainties in their mathematical formulations.

Chapter 6

Conclusions

This thesis aims to enhance the interaction between traffic management and ATO to support a railway system that utilises capacity and energy more effectively, improves alignment between planning and operations, and responds more effectively to real-time conditions. Several research questions were proposed to achieve this objective, which were addressed throughout Chapters 2 to 5. In this chapter, Section 6.1 presents the main findings. Sections 6.2 and 6.3 provide recommendations for practice and outline directions for future research, respectively.

6.1 Main findings

The main research question of this PhD thesis is: *"How can ATO be aligned effectively with the traffic plan to achieve conflict-free, energy-efficient, and flexible railway operations?"* To address this overarching question, four sub-questions were formulated and answered throughout Chapters 2 to 5, as summarised below.

What are the functional and operational requirements for ATO to interact seamlessly with TMS?

Effective alignment of ATO and traffic management requires a clear understanding of how driving functions are allocated and coordinated across the system architecture. Chapter 2 investigates these architectural requirements, with a particular focus on the definition and distribution of ATO driving functions between onboard and trackside subsystems. The analysis is situated in the context of mainline railway networks and ongoing standardisation efforts under the ERTMS/ATO specifications. Drawing on state-of-the-art architectures proposed in the literature for C-DAS, this chapter proposes and evaluates multiple ATO driving function configurations by varying the distribution of automation-related functionalities, including TPE generation, train trajectory computation, traction or brake mode selection, and traction or brake control. To assess the trade-offs among the proposed architectures, a SWOT analysis is conducted, considering differences in governance, signalling systems, and communication capabilities across railway networks. Results show that the ATO-O architecture design is well-suited to fully separated organisational structures. This configuration aligns with the ERTMS/ATO system requirements, where train trajectories are computed and tracked onboard, while the trackside is responsible for determining trajectory generation constraints. This supports the vertical separation organisational model of EU railway systems. For more stable and predictable operating environments, ATO-C, where only traction or brake control is performed onboard, and ATO-I, where the traction or brake mode is

also determined onboard in addition to control, remain relevant. These configurations offer more cost-efficient implementations under an integrated organisational model, particularly on low-frequency, low-speed lines with homogeneous traffic. The findings also clarify the relationship between established C-DAS architectures and proposed ATO configurations, showing that DAS-O corresponds closely to ATO-O in function and offers a no-regret option for future deployment of ATO-over-ETCS. This compatibility facilitates mixed use of C-DAS and ATO in transitional or hybrid contexts. Lastly, this chapter establishes the position of TPE computation at the trackside, serving as an interface between TMS and ATO, to facilitate the interaction between onboard train control and network-level traffic planning.

How can ATO account for train driving variations while ensuring feasibility, energy efficiency, and punctuality?

Chapter 3 addresses the mismatch between the discrete target event times at stations or junctions determined at the TMS level, which typically rely on fixed speed models or lack explicit speed profiles, and the continuous speed–distance trajectory optimisation applied at the ATO level. This misalignment can lead to deviations between planned and actual train movements, particularly when trains adopt different feasible driving strategies, and may result in track occupation conflicts. To address this, the chapter proposes an optimisation-based method to compute TPEs that define and optimise the train trajectory generation constraints for ATO-equipped trains. By incorporating multiple train driving strategies, the method improves the robustness of the resulting TPEs and accounts for operational variability. A two-stage approach is proposed: First, a linear programming model determines conflict-free blocking time ranges across multiple feasible driving strategies. Second, a structured optimisation process establishes operationally valid TPEs by determining departure tolerances at stations and intermediate timing points with their associated time windows along the train route. A critical-block strategy is introduced to place intermediate timing points at signal positions where blocking time overlaps are maximal, with corresponding time windows derived to ensure conflict-freeness while allowing flexibility in train driving. The approach is verified through controlled experiments and applied to a real-life case study on a Dutch corridor with heterogeneous traffic. The results confirm that optimised TPEs improve network performance in terms of capacity, energy use, and timetable adherence, while also strengthening the alignment between train operation and traffic management.

How sensitive is ATO to real-time variations in train statuses and traffic plan?

Chapter 4 investigates how TPEs respond to variations in inputs, including updated RTTPs from the TMS and train status reports fed back from the ATO onboard. A sensitivity analysis using elementary effects is proposed to quantify such variations across three key outputs of TPE generation: departure tolerance at stations, time windows at control timing points, and operational tolerance during operations (i.e., the latest permissible passing time). The analysis first defines the boundaries of input variation and then separately evaluates the effects of headway changes, representing timetable updates at the RTTP level, and train status changes, representing real-time operational dynamics. A real-life case study on a Dutch railway corridor reveals that departure tolerance is particularly sensitive to shorter headways, especially with heterogeneous train pairs. Time windows at control timing points remain spatially stable but show increasing sensitivity further downstream along the route, particularly under tight headway conditions. Operational tolerance is more affected by speed variations at lower speeds and over longer

distances, due to the cumulative effect of prolonged cruising deviations. These findings serve as an initial screening of the TPE generation method proposed in Chapter 3, prior to closed-loop simulation and integration testing. They also provide insight into the boundary conditions under which speed regulation based on TPE adjustment suffices, and when upstream RTTP updates with rescheduling decisions from the TMS become necessary to preserve conflict-freeness and flexibility. As a result, this chapter contributes to a more operationally informed understanding of how TPEs facilitate alignment between real-time traffic management and train operation.

How can event time flexibility enhance the alignment between network-wide rescheduling and ATO-enabled train operations?

While Chapters 3 and 4 develop the computation and examine the sensitivity of TPEs as train-level trajectory generation constraints, Chapter 5 extends this concept to the network level by incorporating event time flexibility into timetable rescheduling. A Train Rescheduling with Flexibility (TRF) model is introduced, formulated as a MILP problem, in which departure, arrival, and passage events are no longer modelled as fixed variables but are jointly optimised together with their associated time windows. These time windows are derived from timetable allowances (such as running time supplements and buffer times) and punctuality thresholds, which define the maximum permissible deviation from scheduled departure or arrival times. By allowing scheduled event times to be determined within these optimised windows, the model defines tangible and dynamic bounds that can absorb minor train path deviations without triggering immediate train rescheduling actions. These flexible event times serve as network-wide counterparts to TPEs, enabling bounded train driving flexibility without compromising conflict-freeness. A case study on a part of the Dutch railway network demonstrates that incorporating event time flexibility transforms the deterministic time–distance train paths into bandwidths and enables effective absorption of initial train delays. The optimisation simultaneously determines time window sizes and rescheduling decisions such as re-timing and re-ordering, and is evaluated against a benchmark without event time flexibility and a large set of randomly disturbed scenarios. A sensitivity analysis of the weights in the objective function, which reflect the preference for event time flexibility, shows that trade-offs primarily arise between early arrival flexibility and late departure flexibility. Furthermore, increasing the punctuality threshold reduces the incremental benefit of additional event time flexibility, with no gain once the timetable becomes saturated. The results demonstrate the potential of integrating event time flexibility into traffic management to enhance network-level adaptability and reduce delays arising from train operation.

In response to the main research question, this thesis demonstrates that effective alignment between ATO and the timetable can be achieved by introducing controlled flexibility across planning and operational levels through enhanced interaction. At the system architecture level, a well-considered allocation of ATO driving functions to either the trackside or onboard supports the integration of TPEs as train trajectory generation constraints derived from the timetable, within diverse governance, signalling, and communication settings. At the operational level, optimised TPEs accommodate multiple feasible driving strategies, offering robustness to operational variability while ensuring conflict-freeness and allowing for energy-efficient train driving. The sensitivity analysis helps identify how TPEs respond to input variations, indicating when onboard speed regulation via TPE adjustments may suffice and when upstream rescheduling by the TMS becomes necessary. At the network level, event time flexibility functions as a network-wide counterpart to TPEs, providing bandwidth-defined train paths that help reduce rescheduling

interventions while maintaining punctuality under disturbances. Moreover, such flexibility reduces alterations to the TPEs forwarded from the ATO trackside to the ATO onboard, leading to more stable ATO trajectories and fewer required updates. Together, these findings establish a comprehensive multi-level approach for aligning ATO with traffic management through the use of trajectory-based and event-time-based flexibility, supporting scalable, conflict-free, flexible and energy-efficient railway operations.

6.2 Recommendations for practice

This thesis has led to seven practical recommendations that aim to support the effective interaction between TMS and ATO. These recommendations are intended for IMs, RUs, and system developers involved in the specification, implementation, and operation of TMS and ATO. They are as follows:

1. Railway organisations should adopt the ATO-O architecture defined in this thesis as a reference for assigning responsibilities between onboard and trackside subsystems. This architecture aligns with the ERTMS/ATO specifications and enables a functional separation, where (re)scheduling and the generation of train trajectory constraints are handled at the trackside, while trajectory optimisation and tracking are performed onboard. Such a structure supports scalable ATO deployment across both vertically separated and vertically integrated railway organisations, and facilitates an effective interface between traffic management and train operation.
2. IMs shall consider the TPE generation method developed in this thesis, either within the TMS or at an active ATO trackside subsystem, as a means to generate conflict-free yet flexible trajectory generation constraints that accommodate multiple feasible driving strategies. This addition does not require significant changes to the existing system architecture but enhances operational performance and strengthens the coordination between traffic management and train operation.
3. IMs and system developers shall consider pre-computing the headway boundary conditions identified in this thesis, where departure tolerance adjustments are sufficient, control timing points are required, or timetable rescheduling is needed. They are also advised to explore the proposed sensitivity analysis to evaluate the responsiveness of TPEs in highly utilised corridors. This proactive assessment can help identify conditions under which TPEs may be less robust, informing the need for closer supervision by the TMS and potential intervention from traffic controllers or signallers.
4. IMs are recommended to assess the potential of the TRF model introduced in this thesis to enhance real-time train timetable rescheduling with event time flexibility so as to improve ATO effectiveness. By jointly optimising event time windows and rescheduling decisions such as train ordering and timing, the model enables more adaptive responses to disturbances. This built-in flexibility, based on timetable allowances and punctuality thresholds, reduces the need for frequent re-planning, alleviates the workload of traffic controllers, and supports improved punctuality.
5. RUs and IMs should align their planning and operational objectives to determine optimal TPEs and flexible timetables. RUs can utilise the flexibility embedded in TPEs and time–window–based timetables to implement energy-efficient driving strategies. IMs, in turn, should take into account the operational needs of RUs by allowing sufficient

headways or time windows where possible, while still ensuring effective use of capacity.

6. RUs and IMs shall communicate their preferred driving strategies, as these are critical for developing conflict-free and realistic timetables that enable energy-efficient and capacity-effective train operation under both nominal and disturbed conditions. To support this coordination, system developers can implement the methods proposed in this thesis to detect inconsistencies between planned and actual operations and to enhance the interaction between TMS and ATO.
7. It is beneficial for IMs and RUs to engage in structured dialogues to improve alignment between planning and operations.

6.3 Recommendations for future research

This thesis has taken a first step towards enhancing the interaction between TMS and ATO by introducing the concepts of controlled trajectory-based and event-time-based flexibility. Nevertheless, the developed models are subject to several limitations and assumptions that open new avenues for further research.

Train operation. The train trajectory optimisation approach adopted in this thesis uses numerical integration based on an adaptive third-order Runge–Kutta method, constrained by control laws and boundary conditions derived from Pontryagin's Maximum Principle (PMP) to solve the optimal train control problem. While this method is computationally efficient for stop-to-stop trajectory computation, it is less efficient under free-speed conditions where the speed is not predetermined, for example, at a control timing point location. To address this, PMP could be used to derive the necessary optimality conditions for such intermediate constraints, offering a more structured basis for either analytical solutions or guided numerical computation, depending on whether the resulting equations are explicit or implicit. The search algorithm for determining the optimal cruising speed can be further refined, including for the timing point response and energy-efficient train control driving strategies, especially in cases with multiple stops. Such refinements may support faster train trajectory computation under multiple timing point constraints and improve TPE generation, particularly in C-DAS and ATO implementations.

Furthermore, current ERTMS/ATO specifications lack standardised mechanisms for feeding real-time train parameters, such as train mass, traction, and braking characteristics, back to the trackside ATO subsystem. This information can be used to refine trajectory computation and provide updated running time estimates, thereby improving alignment between operational planning and real-time control. Harmonising analytical train trajectory models with tabular traction representations and ensuring consistent modelling of ETCS braking curves would further improve the fidelity of train operation modelling. This would also support more accurate estimation of energy use under varying traction and braking profiles, especially in the presence of fluctuating power supply conditions arising from multiple operational and infrastructural factors, including substation spacing and characteristics, voltage variations, and regenerative braking. Finally, future research should consider trajectory computation models for trains that do not rely on external electricity, such as those powered by diesel or batteries.

TPE optimisation. Beyond static computation, dynamic re-computation of TPEs during operations could be investigated to reflect evolving traffic conditions, including disturbances, disruptions, or early arrivals. This would extend the use of TPEs from short-term operational planning into real-time applications. Advanced control methods, such as event-triggered model

predictive control or rolling horizon optimisation, could be applied to enable such functionality. Moreover, future research could extend the current TPE generation method from individual corridors to networks of interconnected corridors with merging and diverging train lines.

The proposed methods should also be tested in closed-loop simulation environments that include interactions between ATO or C-DAS, TMS, and human agents such as signallers, traffic controllers, and train drivers. Such simulations should be conducted in realistic, large-scale network setups that include complex infrastructure layouts, different rolling stock types, and a range of disturbance or disruption scenarios. This would provide a more comprehensive understanding of system-level performance and potential challenges associated with implementing TPE generation in practice. In parallel, simulation-based validation would help determine the most suitable position of TPE generation to inform ongoing standardisation efforts, i.e., whether within the TMS or in the ATO/C-DAS trackside subsystem.

Additionally, coordinated train control presents a promising direction for refining the definition and formulation of TPEs. By allowing timing points to be placed flexibly and dynamically along the route, rather than being limited to predefined signal or station locations, it becomes possible to accommodate shorter headways and to support more adaptive train driving strategies. Such developments would enable joint adjustment of TPEs among successive trains, allowing both preceding and following trains to coordinate their trajectories in response to real-time traffic states. It may also be beneficial to embed blocking time constraints directly into the trajectory optimisation framework to improve consistency between train control and infrastructure occupation modelling, potentially enhancing the computation of conflict-free TPEs. Lastly, the interaction between network-level event time flexibility and train-level TPE optimisation could be further explored to determine how train driving bandwidth can be allocated efficiently and effectively across successive services.

Traffic management. Enhancing real-time traffic management by introducing event time flexibility offers a valuable avenue for future research. The TRF model developed in this thesis enables event times to be selected within bounded, conflict-free time windows, improving adaptability to disturbances. Future research could explore scaling up the model to large networks. The model could also be extended to support cross-border and long-distance services by explicitly modelling time windows at border stations and harmonising local and international timetable constraints. Such improvements would facilitate smoother integration and reduce the risk of cascading delays across national boundaries. Integrating passenger-related factors, such as route choice behaviour and transfer preferences, would further increase the practical relevance of the TRF model. This would enable assessment of passenger-perceived delays and allow for rescheduling decisions that are better aligned with traveller needs.

To enhance the capability of the model for managing disruptions, future work could incorporate additional dimensions of flexibility beyond timing and train ordering. These may include station-level spatial flexibility, such as local re-routing through platform adjustments, and service-level rescheduling strategies, such as stop-skipping, short-turning, or global re-routing. To support effective decision-making under these conditions, it is necessary to quantify the marginal impact of each strategy, both in isolation and in combination, under different disruption scenarios that include both stations (nodes) and track segments (links). This quantification would assist IMs in identifying capacity bottlenecks, selecting the most effective rescheduling measures, and avoiding disproportionate or counterproductive interventions.

Advancing the solution approach is also essential for operational deployment of the TRF model. Research into decomposition-based algorithms, tailored heuristics, and hybrid meth-

ods that apply combinatorial optimisation with machine learning could significantly reduce computational time and support real-time deployment under complex, dynamic conditions.

Bibliography

Albrecht, A.R., P. G. Howlett, P. J. Pudney, X. Vu, P. Zhou (2015) Energy-efficient train control: The two-train separation problem on level track, *Journal of Rail Transport Planning & Management*, 5(3), pp. 163–182.

Albrecht, A.R., P. G. Howlett, P. J. Pudney, X. Vu, P. Zhou (2016a) The key principles of optimal train control—Part 1: Formulation of the model, strategies of optimal type, evolutionary lines, location of optimal switching points, *Transportation Research Part B: Methodological*, 94, pp. 482–508.

Albrecht, A.R., P. G. Howlett, P. J. Pudney, X. Vu, P. Zhou (2016b) The key principles of optimal train control—Part 2: Existence of an optimal strategy, the local energy minimization principle, uniqueness, computational techniques, *Transportation Research Part B: Methodological*, 94, pp. 509–538.

Albrecht, T., A. Binder, C. Gassel (2013) Applications of real-time speed control in rail-bound public transportation systems, *IET Intelligent Transport Systems*, 7(3), pp. 305–314.

Aoun, J., E. Quaglietta, R. M. P. Goverde (2020) Investigating market potentials and operational scenarios of virtual coupling railway signaling, *Transportation Research Record*, 2674(8), pp. 799–812.

Bešinović, N., R. M. P. Goverde, E. Quaglietta (2017) Microscopic models and network transformations for automated railway traffic planning, *Computer-Aided Civil and Infrastructure Engineering*, 32(2), pp. 89–106.

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

Caillot, A., S. Ouerghi, P. Vasseur, R. Boutteau, Y. Dupuis (2022) Survey on cooperative perception in an automotive context, *IEEE Transactions on Intelligent Transportation Systems*, 23(9), pp. 14204–14223.

Caimi, G., F. Chudak, M. Fuchsberger, M. Laumanns, R. Zenklusen (2011a) A new resource-constrained multicommodity flow model for conflict-free train routing and scheduling, *Transportation Science*, 45(2), pp. 212–227.

Caimi, G., M. Fuchsberger, M. Laumanns, M. Lüthi (2012) A model predictive control approach for discrete-time rescheduling in complex central railway station areas, *Computers & Operations Research*, 39(11), pp. 2578–2593.

Caimi, G., M. Fuchsberger, M. Laumanns, K. Schüpbach (2011b) Periodic railway timetabling with event flexibility, *Networks*, 57(1), pp. 3–18.

Cavone, G., T. Van den Boom, L. Blenkers, M. Dotoli, C. Seatzu, B. De Schutter (2022) An MPC-based rescheduling algorithm for disruptions and disturbances in large-scale railway networks, *IEEE Transactions on Automation Science and Engineering*, 19(1), pp. 99–112.

Chen, M., Q. Wang, P. Sun, X. Feng (2023) Train control and schedule integrated optimization with reversible substations, *IEEE Transactions on Vehicular Technology*, 72(2), pp. 1586–1600.

Corman, F., A. D'Ariano, A. D. Marra, D. Pacciarelli, M. Samà (2017) Integrating train scheduling and delay management in real-time railway traffic control, *Transportation Research Part E: Logistics and Transportation Review*, 105, pp. 213–239.

Corman, F., A. D'Ariano, D. Pacciarelli, M. Pranzo (2010) A tabu search algorithm for rerouting trains during rail operations, *Transportation Research Part B: Methodological*, 44(1), pp. 175–192.

Corman, F., A. D'Ariano, D. Pacciarelli, M. Pranzo (2012) Bi-objective conflict detection and resolution in railway traffic management, *Transportation Research Part C: Emerging Technologies*, 20(1), pp. 79–94.

Corman, F., L. Meng (2015) A review of online dynamic models and algorithms for railway traffic management, *IEEE Transactions on Intelligent Transportation Systems*, 16(3), pp. 1274–1284.

Cunillera, A., N. Bešinović, N. Van Oort, R. M. P. Goverde (2022) Real-time train motion parameter estimation using an Unscented Kalman Filter, *Transportation Research Part C: Emerging Technologies*, 143, p. 103794.

D'Ariano, A., F. Corman, D. Pacciarelli, M. Pranzo (2008) Reordering and local rerouting strategies to manage train traffic in real time, *Transportation Science*, 42(4), pp. 405–419.

D'Ariano, A., M. Pranzo, I. A. Hansen (2007b) Conflict resolution and train speed coordination for solving real-time timetable perturbations, *IEEE Transactions on Intelligent Transportation Systems*, 8(2), pp. 208–222.

Deutsche Bahn (2020) Produktbeschreibung Grüne Funktionen der Zuglaufregelung, URL https://gruen.deutschebahn.com/de/massnahmen/zuglaufregelung.

Dong, H., X. Liu, M. Zhou, W. Zheng, J. Xun, S. Gao, H. Song, Y. Li, F.-Y. Wang (2022) Integration of train control and online rescheduling for high-speed railways in case of emergencies, *IEEE Transactions on Computational Social Systems*, 9(5), pp. 1574–1582.

Dong, H., L. Ning, M. Zhou, H. Song, W. Bai (2024) Deep reinforcement learning for integration of train trajectory optimization and timetable rescheduling under disturbances, IEEE Transactions on Neural Networks and Learning Systems, Early Access, available: https://doi.org/10.1109/TNNLS.2024.3357494.

Dyson, R. G. (2004) Strategic development and SWOT analysis at the University of Warwick, *European Journal of Operational Research*, 152(3), pp. 631–640.

D'Ariano, A., D. Pacciarelli, M. Pranzo (2007a) A branch and bound algorithm for scheduling trains in a railway network, *European Journal of Operational Research*, 183(2), pp. 643–657.

D'Ariano, A., D. Pacciarelli, M. Pranzo (2008) Assessment of flexible timetables in real-time traffic management of a railway bottleneck, *Transportation Research Part C: Emerging Technologies*, 16(2), pp. 232–245.

Eckhardt, J., L. Nykänen, A. Aapaoja, P. Niemi (2018) MaaS in rural areas - case Finland, *Research in Transportation Business & Management*, 27, pp. 75–83.

ERA (2018) Command and Control 4.0, URL https://www.era.europa.eu/content/command-and-control-40_en.

ERA (2024) A compelling vision for the target railway system, URL https://www.era.europa.eu/system/files/2022-11/Compelling%20vision%20for%20a%20target%20rail%20system%20(1).pdf, ISBN 978-92-9477-468-2, DOI 10.2821/59548.

ERA * UNISIG * EEIG ERTMS USERS GROUP (2014) ATO-TS / TMS Interface Specification, SUBSET-131, ISSUE: 0.0.3.

ERA * UNISIG * EEIG ERTMS USERS GROUP (2023a) ERTMS/ETCS - System Requirements Specification, SUBSET-026.

ERA * UNISIG * EEIG ERTMS USERS GROUP (2023b) ERTMS/ATO - System Requirements Specification, SUBSET-125.

ERA * UNISIG * EEIG ERTMS USERS GROUP (2023c) ERTMS/ATO - ATO-OB / ATO-TS FFFIS Application Layer, SUBSET-126.

ERTMS Users Group (2023a) ERTMS/ATO Operational Principles, EUG Reference: 12E108.

ERTMS Users Group (2023b) ERTMS/ATO Glossary, EUG Reference: 13E154.

Esposito, G., L. Cicatiello, S. Ercolano (2020) Reforming railways in the EU: An empirical assessment of liberalisation policies in the european rail freight market, *Transportation Research Part A: Policy and Practice*, 132, pp. 606–613.

European Commission (2022) Europe's Rail Joint Undertaking Master Plan, URL https://rail-research.europa.eu/wp-content/uploads/2022/03/EURAIL_Master-Plan.pdf, draft version; Brussels, 17 Feb 2022.

Fang, W., S. Yang, X. Yao (2015) A survey on problem models and solution approaches to rescheduling in railway networks, *IEEE Transactions on Intelligent Transportation Systems*, 16(6), pp. 2997–3016.

Federal Railroad Administration (2022) Automated Train Operation Interface Requirements Specification Development Summary Report, https://railroads.dot.gov/elibrary/automated-train-operation-interface-requirements-specification-development-summary-report.

Golden, W., P. Powell (2000) Towards a definition of flexibility: in search of the holy grail?, *Omega*, 28(4), pp. 373–384.

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

Goverde, R. M. P., F. Corman, A. D'Ariano (2013) Railway line capacity consumption of different railway signalling systems under scheduled and disturbed conditions, *Journal of Rail Transport Planning & Management*, 3(3), pp. 78–94.

Goverde, R. M. P., G. M. Scheepmaker, P. Wang (2021) Pseudospectral optimal train control, *European Journal of Operational Research*, 292(1), pp. 353–375.

Goverde, R. M. P., S. Su, Z. Tian (2023) Energy-Efficient Train Operation: Conclusions and Future Work, in: *Energy-Efficient Train Operation: A System Approach for Railway Networks*, Springer, Cham, pp. 229–234.

Government of the Netherlands (2019) Public Transport in 2040: Outlines of a vision for the future, URL https://www.government.nl/documents/publications/2019/06/13/public-transport-in-2040-outlines-of-a-vision-for-the-future.

Haahr, J. H., D. Pisinger, M. Sabbaghian (2017) A dynamic programming approach for optimizing train speed profiles with speed restrictions and passage points, *Transportation Research Part B: Methodological*, 99, pp. 167–182.

Hoberg, G., G. Phillips, N. Prabhala (2014) Product market threats, payouts, and financial flexibility, *The Journal of Finance*, 69(1), pp. 293–324.

Hou, Z., H. Dong, S. Gao, G. Nicholson, L. Chen, C. Roberts (2019) Energy-saving metro train timetable rescheduling model considering ATO profiles and dynamic passenger flow, *IEEE Transactions on Intelligent Transportation Systems*, 20(7), pp. 2774–2785.

Howlett, P., P. Pudney, A. Albrecht (2023) Optimal driving strategies for a fleet of trains on level track with prescribed intermediate signal times and safe separation, *Transportation Science*, 57(2), pp. 399–423.

Huang, P., Z. Li, Y. Zhu, C. Wen, F. Corman (2023) Train traffic control in merging stations: A data-driven approach, *Transportation Research Part C: Emerging Technologies*, 152, p. 104155.

Huang, W., Y. Zhang, B. Shuai, M. Xu, W. Xiao, R. Zhang, Y. Xu (2019) China railway industry reform evolution approach: Based on the Vertical Separation Model, *Transportation Research Part A: Policy and Practice*, 130, pp. 546–556.

Huang, Y., L. Yang, T. Tang, Z. Gao, F. Cao, K. Li (2018) Train speed profile optimization with on-board energy storage devices: A dynamic programming based approach, *Computers & Industrial Engineering*, 126, pp. 149–164.

IEC (2014) Railway applications - Urban guided transport management and command/control systems - Part 1: System principles and fundamental concepts, 62290-1.

Jansson, E., N. O. Olsson, O. Fröidh (2023) Challenges of replacing train drivers in driverless and unattended railway mainline systems—A Swedish case study on delay logs descriptions, *Transportation Research Interdisciplinary Perspectives*, 21, p. 100875.

Krasemann, J. T. (2012) Design of an effective algorithm for fast response to the re-scheduling of railway traffic during disturbances, *Transportation Research Part C: Emerging Technologies*, 20(1), pp. 62–78.

Lagos, M. (2011) CATO offers energy savings, *Railway Gazette International*, 167(5), pp. 50–52.

Learned, E. P., K. R. Andrews, C. R. Christensen, W. D. Guth (1969) *Business policy: Text and cases*, RD Irwin.

Li, S., R. Liu, Z. Gao, L. Yang (2021) Integrated train dwell time regulation and train speed profile generation for automatic train operations on high-density metro lines: A distributed optimal control method, *Transportation Research Part B: Methodological*, 148, pp. 82–105.

Li, Z., J. Yin, S. Chai, T. Tang, L. Yang (2023) Optimization of system resilience in urban rail systems: Train rescheduling considering congestions of stations, *Computers & Industrial Engineering*, 185, p. 109657.

Liu, X., A. Dabiri, J. Xun, B. De Schutter (2023) Bi-level model predictive control for metro networks: Integration of timetables, passenger flows, and train speed profiles, *Transportation Research Part E: Logistics and Transportation Review*, 180, p. 103339.

Long, S., L. Meng, Y. Wang, J. Miao, X. Luan, F. Corman (2023) Integrated speed modeling and traffic management to precisely model the effect and dynamics of temporary speed restrictions to high-speed railway traffic, *Transportation Research Part C: Emerging Technologies*, 152, p. 104148.

Lu, C., J. Liu, Y. Liu, Y. Liu (2019) Intelligent construction technology of railway engineering in China, *Frontiers of Engineering Management*, 6(4), pp. 503–516.

Luan, X., Y. Wang, B. De Schutter, L. Meng, G. Lodewijks, F. Corman (2018a) Integration of real-time traffic management and train control for rail networks - Part 1: Optimization problems and solution approaches, *Transportation Research Part B: Methodological*, 115, pp. 41–71.

Luan, X., Y. Wang, B. De Schutter, L. Meng, G. Lodewijks, F. Corman (2018b) Integration of real-time traffic management and train control for rail networks - Part 2: Extensions towards energy-efficient train operations, *Transportation Research Part B: Methodological*, 115, pp. 72–94.

Luijt, R. S., M. P. F. Van den Berge, H. Y. Willeboordse, J. H. Hoogenraad (2017) 5 years of Dutch eco-driving: Managing behavioural change, *Transportation Research Part A: Policy and Practice*, 98, pp. 46–63.

Mascis, A., D. Pacciarelli (2002) Job-shop scheduling with blocking and no-wait constraints, *European Journal of Operational Research*, 143(3), pp. 498–517.

Mo, P., L. Yang, Z. Gao (2019) Energy-efficient train operation strategy with speed profiles selection for an urban metro line, *Transportation Research Record*, 2673(4), pp. 348–360.

Morris, M. D. (1991) Factorial sampling plans for preliminary computational experiments, *Technometrics*, 33(2), pp. 161–174.

MOTIONAL (2024) TMS and ATO/C-DAS Timetable Test & Simulation Environment, Deliverable D15.2, FP1-MOTIONAL, EU-Rail.

NASA (2024) Standard for Models and Simulations, NASA-STD-7009B.

Nash, C. A., A. S. J. Smith, D. Van de Velde, F. Mizutani, S. Uranishi (2014) Structural reforms in the railways: Incentive misalignment and cost implications, *Research in Transportation Economics*, 48, pp. 16–23.

ON-TIME (2013a) Deliverable 6.1: Specification of a driving advisory system (DAS) data format. (Revision 2.2), URL https://www.ontime-project.eu/deliverables.aspx.

ON-TIME (2013b) Deliverable 4.1: Functional and technical requirements specification for perturbation management, URL https://www.ontime-project.eu/deliverables.aspx.

Pachl, J. (2014) Timetable Design Principles, in: Hansen, I. A., J. Pachl, eds., *Railway Timetabling & Operations: Analysis, Modelling, Optimisation, Simulation, Performance Evaluation*, Eurailpress, Hamburg, pp. 13–46.

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

Pellegrini, P., G. Marlière, R. Pesenti, J. Rodriguez (2015) RECIFE-MILP: An effective MILP-based heuristic for the real-time railway traffic management problem, *IEEE Transactions on Intelligent Transportation Systems*, 16(5), pp. 2609–2619.

Pellegrini, P., G. Marlière, J. Rodriguez (2014) Optimal train routing and scheduling for managing traffic perturbations in complex junctions, *Transportation Research Part B: Methodological*, 59, pp. 58–80.

Pellegrini, P., R. Pesenti, J. Rodriguez (2019) Efficient train re-routing and rescheduling: Valid inequalities and reformulation of RECIFE-MILP, *Transportation Research Part B: Methodological*, 120, pp. 33–48.

Poulus, R. W., E. A. Van Kempen, J. C. Van Meijeren (2018) Automatic train operation. Driving the future of rail transport, Tech. Rep. 842373, Netherlands Organisation for Applied Scientific Research, The Hague.

Quaglietta, E., P. Pellegrini, R. M. P. Goverde, T. Albrecht, B. Jaekel, G. Marlière, J. Rodriguez, T. Dollevoet, B. Ambrogio, D. Carcasole, M. Giaroli, G. Nicholson (2016) The ON-TIME real-time railway traffic management framework: A proof-of-concept using a scalable standardised data communication architecture, *Transportation Research Part C: Emerging Technologies*, 63, pp. 23–50.

Rao, X., M. Montigel, U. Weidmann (2016) A new rail optimisation model by integration of traffic management and train automation, *Transportation Research Part C: Emerging Technologies*, 71, pp. 382–405.

Riaz, S., P. Mancarella (2022) Modelling and characterisation of flexibility from distributed energy resources, *IEEE Transactions on Power Systems*, 37(1), pp. 38–50.

Rietveld, P., F. Bruinsma, D. van Vuuren (2001) Coping with unreliability in public transport chains: A case study for netherlands, *Transportation Research Part A: Policy and Practice*, 35(6), pp. 539–559.

Schaafsma, A. A. M. (2001) *Dynamisch Railverkeersmanagement besturingsconcept voor railverkeer op basis van het Lagenmodel Verkeer en Vervoer*, Ph.D. thesis, Delft University of Technology, Delft, The Netherlands, in Dutch, English summary.

Schaafsma, A. A. M., M. G. P. Bartholomeus (2007) Dynamic traffic management in the Schiphol bottleneck, in: Hansen, I. A., A. Radtke, J. Pachl, E. Wendler, eds., *Proceedings of the Second International Seminar on Railway Operations Modelling and Analysis*, pp. 1–18.

Scheepmaker, G. M., R. M. P. Goverde (2020) Energy-efficient train control using nonlinear bounded regenerative braking, *Transportation Research Part C: Emerging Technologies*, 121, p. 102852.

Scheepmaker, G. M., R. M. P. Goverde, L. G. Kroon (2017) Review of energy-efficient train control and timetabling, *European Journal of Operational Research*, 257(2), pp. 355–376.

Scheepmaker, G. M., H. Y. Willeboordse, J. H. Hoogenraad, R. S. Luijt, R. M. P. Goverde (2020) Comparing train driving strategies on multiple key performance indicators, *Journal of Rail Transport Planning & Management*, 13, p. 100163.

Su, S., X. Wang, Y. Cao, J. Yin (2020) An energy-efficient train operation approach by integrating the metro timetabling and eco-driving, *IEEE Transactions on Intelligent Transportation Systems*, 21(10), pp. 4252–4268.

Tafti, A., S. Mithas, M. S. Krishnan (2013) The effect of information technology–enabled flexibility on formation and market value of alliances, *Management Science*, 59(1), pp. 207–225.

Tschirner, S., B. Sandblad, A. W. Andersson (2014) Solutions to the problem of inconsistent plans in railway traffic operation, *Journal of Rail Transport Planning & Management*, 4(4), pp. 87–97.

Törnquist, J., J. A. Persson (2007) N-tracked railway traffic re-scheduling during disturbances, *Transportation Research Part B: Methodological*, 41(3), pp. 342–362.

UIC (2013) Code 406 – Capacity (second edition), International Union of Railway, Paris.

UIC (2020) Data exchange with Driver Advisory Systems (DAS) following the Smart Communications For Efficient Rail Activities (SFERA) protocol, IRS-90940.

Vasheghani, M., M. Abtahi (2023) Strategic planning for multimodal transportation in ports, *Maritime Policy & Management*, 50(7), pp. 957–979.

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

Wang, P., R. M. P. Goverde (2016b) Two-train trajectory optimization with a green-wave policy, *Transportation Research Record*, 2546(1), pp. 112–120.

Wang, P., R. M. P. Goverde (2017) Multi-train trajectory optimization for energy efficiency and delay recovery on single-track railway lines, *Transportation Research Part B: Methodological*, 105, pp. 340–361.

Wang, P., R. M. P. Goverde (2019) Multi-train trajectory optimization for energy-efficient timetabling, *European Journal of Operational Research*, 272(2), pp. 621–635.

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

Wang, P., A. Trivella, R. M. P. Goverde, F. Corman (2020) Train trajectory optimization for improved on-time arrival under parametric uncertainty, *Transportation Research Part C: Emerging Technologies*, 119, p. 102680.

Wang, P., Y. Zhu, F. Corman (2022) Passenger-centric periodic timetable adjustment problem for the utilization of regenerative energy, *Computers & Industrial Engineering*, 172, p. 108578.

Wang, Y., S. Zhu, S. Li, L. Yang, B. De Schutter (2022) Hierarchical model predictive control for on-line high-speed railway delay management and train control in a dynamic operations environment, *IEEE Transactions on Control Systems Technology*, 30(6), pp. 2344–2359.

Wang, Z., J. Aoun, C. Szymula, N. Bešinović (2023) Promising solutions for railway operations to cope with future challenges — Tackling COVID and beyond, *Journal of Rail Transport Planning & Management*, 28, p. 100405.

Wang, Z., E. Quaglietta, M. G. P. Bartholomeus, A. Cunillera, R. M. P. Goverde (2025a) Optimising timing points for effective automatic train operation, *Computers & Industrial Engineering*, 206, p. 111237.

Wang, Z., E. Quaglietta, M. G. P. Bartholomeus, R. M. P. Goverde (2022) Assessment of architectures for automatic train operation driving functions, *Journal of Rail Transport Planning & Management*, 24, p. 100352.

Wang, Z., E. Quaglietta, M. G. P. Bartholomeus, R. M. P. Goverde (2025b) Sensitivity of train path envelopes for automatic train operation, *IEEE Transactions on Intelligent Transportation Systems*, 26(11), pp. 18921–18933.

Winter, P. (2009) Potential for benefits with ERTMS, in: Winter, P., ed., *Compendium on ERTMS: European Rail Traffic Management System*, Eurailpress, pp. 199–232.

Yin, J., T. Tang, L. Yang, J. Xun, Y. Huang, Z. Gao (2017) Research and development of automatic train operation for railway transportation systems: A survey, *Transportation Research Part C: Emerging Technologies*, 85, pp. 548–572.

Ying, P., X. Zeng, A. D'Ariano, D. Pacciarelli, H. Song, T. Shen (2023) Quadratically Constrained Linear Programming-based energy-efficient driving for High-speed Trains with neutral zone and time window, *Transportation Research Part C: Emerging Technologies*, 154, p. 104202.

Zhan, S., J. Xie, S. Wong, Y. Zhu, F. Corman (2024) Handling uncertainty in train timetable rescheduling: A review of the literature and future research directions, *Transportation Research Part E: Logistics and Transportation Review*, 183, p. 103429.

Zhang, Y., A. D'Ariano, B. He, Q. Peng (2019) Microscopic optimization model and algorithm for integrating train timetabling and track maintenance task scheduling, *Transportation Research Part B: Methodological*, 127, pp. 237–278.

Zhu, Y., H. Wang, R. M. P. Goverde (2020) Reinforcement learning in railway timetable rescheduling, in: *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, pp. 1–6.

Summary

Railway operations are undergoing a digital transformation, driven by the adoption of an advanced Traffic Management System (TMS) and the deployment of Automatic Train Operation (ATO) to meet growing demands for responsiveness, reliability, and resilience. A critical challenge in this context is ensuring effective interaction and service alignment between TMS and ATO, particularly on mainline railways. A TMS typically assumes deterministic running times between scheduled stops and target event times at designated timetable points, while ATO adaptively selects driving strategies based on real-time conditions and service objectives, ranging from minimum-time running under disturbances to energy-efficient control using coasting regimes under nominal conditions. Differences in scope among tactical planning, operational planning, and real-time train operation can lead to mismatches that result in conflicts in track occupation, thereby compromising punctuality, energy efficiency, and network capacity.

This thesis investigates, designs, and optimises how service event-time-based and train trajectory-based flexibility can be systematically defined, computed, and integrated to enable effective TMS–ATO alignment and improve their interaction across different planning levels and real-time operations. To this end, a multi-level framework is developed around the concept of the Train Path Envelope (TPE), which defines a set of timing points at discrete locations along a train's route, each associated with time constraints. These constraints, formulated in line with current ATO standardisation requirements, guide trains to operate in a conflict-free manner by avoiding blocking time overlaps along the route, to remain punctual by adhering to scheduled times and minimising delays under disturbances, and to improve energy efficiency by leveraging embedded flexibility that promotes energy-efficient train control.

The research starts by reviewing completed and ongoing standardisation activities and exploring the architectural design of ATO connected to TMS, with particular attention to the allocation of automated functions. Based on the state-of-the-art Connected Driver Advisory System (C-DAS) architectures proposed in the literature, three alternative ATO system designs are introduced, each varying in the allocation of four key functions: TPE generation, train trajectory computation, traction or brake mode selection, and traction or brake control. These architectures are systematically compared across governance, communication, and signalling dimensions through a Strengths, Weaknesses, Opportunities, and Threats (SWOT) analysis. Building on this system architectural investigation, the concept of the TPE is introduced as an effective interface to align traffic management and train operation, with its appropriate position within current standardisation specifications identified as the foundation for the modelling approaches developed in the subsequent chapters.

To compute conflict-free TPEs for successive trains in a railway corridor, a two-stage optimisation framework is proposed. The first stage evaluates multiple punctual train driving strategies and determines conflict-free blocking time constraints across scheduled train pairs. The second stage applies a strategy based on critical blocks to identify the placement of timing points,

define associated time windows, and fine-tune departure tolerances to ensure conflict-freeness while adhering to the timetable and preserving flexibility for energy-efficient driving. Controlled experiments and a real-world application on a Dutch mainline corridor demonstrate that the approach effectively translates scheduled or rescheduled timetables into structured, conflict-free time windows suitable for guiding ATO- or C-DAS-equipped trains.

The responsiveness of the TPE generation method is further investigated through a sensitivity analysis using elementary effects, supporting its integration into closed-loop simulation environments and real-world systems. This analysis quantifies the impact of input variations such as timetable updates from the TMS and real-time train status reports from the ATO onboard. A case study on a Dutch railway corridor with various rolling stock types shows that while the spatial placement of timing points remains largely unchanged, their associated time constraints require adjustment to maintain conflict-freeness and ensure sufficient driving flexibility. Time window size and operational tolerance are more sensitive for heterogeneous train pairs that have progressed further along their routes and operate at low speeds.

Alongside train-level TPE generation, this thesis also introduces a traffic-level counterpart by proposing a Train Rescheduling with Flexibility (TRF) model. A Mixed-Integer Linear Programming (MILP) problem formulation is developed to determine an optimal timetable under disturbances, where fixed event times are relaxed into time windows. Late departure flexibility, as well as early and late arrival flexibility, define a tractable bandwidth for train operation, derived from available timetable allowances (e.g., running time supplements and buffer times) and from given punctuality thresholds (i.e., permissible deviations from scheduled times). As long as trains remain within this bandwidth, no immediate rescheduling interventions are required. Results on a part of the Dutch network indicate that the model effectively embeds event time flexibility, outperforms a no-flexibility benchmark, and shows promise in handling disturbances.

Together, this thesis presents a multi-level methodology for enhancing the interaction between TMS and ATO by incorporating operational flexibility into train-level and traffic-level planning, thereby contributing to ongoing railway digitalisation and standardisation efforts.

Samenvatting

Railverkeerssystemen ondergaan een digitale transformatie, gedreven door de invoering van geavanceerde Traffic Management Systemen (TMS) en de implementatie van Automatic Train Operation (ATO) om te voldoen aan de toenemende vraag naar adaptieve, betrouwbare en veerkrachtige dienstverlening. Een belangrijke uitdaging daarbij is het waarborgen van een effectieve interactie en afstemming tussen TMS en ATO, vooral op hoofdspoorlijnen. TMS gaat doorgaans uit van deterministische rijtijden en vaste dienstregelingstijden op dienstregelingspunten, terwijl ATO adaptief rijstrategieën kiest op basis van realtime omstandigheden en operationele doelen, van minimale rijtijd bij verstoringen tot energiezuinig rijden met uitrollen onder nominale omstandigheden. Deze mismatch tussen planning en uitvoering kan leiden tot rijwegconflicten, wat ten koste gaat van punctualiteit, energie-efficiëntie en capaciteit.

Dit proefschrift onderzoekt, ontwerpt en optimaliseert hoe flexibiliteit in gebeurtenissen en de bijbehorende tijden en snelheidsprofielen systematisch kan worden gedefinieerd, berekend en geïntegreerd om effectieve afstemming tussen TMS en ATO mogelijk te maken en hun interactie op verschillende planningsniveaus en in de realtime operatie te verbeteren. Een meerlagig raamwerk wordt ontwikkeld rond het concept van de Train Path Envelope (TPE), een verzameling timingpunten met bijbehorende tijdvensters langs het traject. Deze beperkingen, opgesteld volgens de ATO-standaardisatie-eisen, sturen treinen aan om conflictvrij te rijden door overlappende bloktijden te vermijden, punctueel te blijven door geplande tijden op de dienstregelingspunten aan te houden, en energiezuinig te rijden met behulp van de ingebedde bandbreedtes.

Het onderzoek begint met een analyse van afgeronde en lopende standaardisatie-initiatieven en een verkenning van de systeemarchitecturen van ATO gekoppeld aan TMS, met bijzondere aandacht voor de toewijzing van Automatiseringsfuncties. Op basis van de meest recente architecturen van Connected Driver Advisory Systems (C-DAS) uit de literatuur worden drie alternatieve ATO systeemontwerpen geïntroduceerd, die variëren in de toewijzing van vier kernfuncties: TPE-generatie, berekening van snelheidsprofielen, keuze van tractie- of remmodus, en tractie- of remsturing. Deze architecturen worden systematisch vergeleken op governance, communicatie en beveiliging via een SWOT-analyse (Strengths, Weaknesses, Opportunities, and Threats). Op basis van deze analyse wordt het TPE-concept geïntroduceerd als een effectieve interface om verkeersmanagement en treinbesturing op elkaar af te stemmen, waarbij de juiste positie binnen de huidige standaardisatiespecificaties wordt vastgesteld als basis voor de in de volgende hoofdstukken ontwikkelde modelleringsbenaderingen.

Voor het berekenen van conflictvrije TPE's voor opeenvolgende treinen in een corridor wordt een tweestaps optimalisatiemethodiek voorgesteld. In de eerste stap worden meerdere punctuele rijstrategieën berekend en worden conflictvrije bloktijden bepaald tussen geplande treinen. De tweede stap past een strategie toe gebaseerd op kritieke blokken om de plaatsing van timingpunten, de bijbehorende tijdvensters en de vertrektoleranties vast te leggen, zodat

conflictvrijheid wordt gegarandeerd terwijl de dienstregeling kan worden gevolgd en flexibiliteit voor energiezuinig rijden behouden blijft. Gecontroleerde experimenten en een toepassing op een Nederlandse hoofdspoorcorridor tonen aan dat deze aanpak geplande of aangepaste dienstregelingen effectief vertaalt naar gestructureerde, conflictvrije tijdvensters die geschikt zijn voor ATO- of C-DAS-aangestuurde treinen.

De adaptiviteit van de TPE-methode wordt verder onderzocht via een gevoeligheidsanalyse met behulp van elementaire effecten, ter ondersteuning van een integratie in simulatieomgevigen en daadwerkelijke systemen. Deze analyse kwantificeert de impact van inputvariaties, zoals dienstregelingsupdates van het TMS en realtime treinstatusrapporten van de ATO-aan boord. Een casestudy op een Nederlandse spoorcorridor met verschillend rollend materieel toont aan dat de ruimtelijke plaatsing van timingpunten grotendeels stabiel blijft, maar dat de bijbehorende tijdvensters moeten worden aangepast om conflictvrijheid en voldoende rijflexibiliteit te waarborgen. De grootte van het tijdvenster en de operationele tolerantie blijken gevoeliger voor heterogene treinparen die verder op hun traject zijn gevorderd en met lage snelheden rijden.

Naast de TPE-berekening op trein-niveau introduceert dit proefschrift ook een tegenhanger op verkeersniveau in de vorm van het Train Rescheduling with Flexibility (TRF)-model. Een gemengd geheeltallig lineair optimaliseringsprobleem wordt ontwikkeld om een optimale dienstregeling onder verstoringen te bepalen, waarbij vaste dienstregelingstijden worden verruimd tot tijdvensters. Vertrekflexibiliteit en vroege en late aankomstflexibiliteit definiëren een beheersbare bandbreedte voor treinoperaties, afgeleid van beschikbare rijtijdspelingen, buffertijden en vastgestelde punctualiteitsdrempels. Zolang treinen binnen deze bandbreedte blijven, zijn directe herplanningsacties niet noodzakelijk. Resultaten op een deel van het Nederlandse netwerk geven aan dat het model flexibiliteit in aankomst- en vertrektijden effectief integreert, beter presteert dan een model zonder flexibiliteit, en potentie biedt voor het effectiever omgaan met verstoringen.

Gezamenlijk presenteert dit proefschrift een meerlagige methodologie om de interactie tussen TMS en ATO te verbeteren door operationele flexibiliteit te integreren in zowel trein- als verkeersniveauplanning, en levert daarmee een bijdrage aan de verdere digitalisering en standaardisatie van het spoorwegsysteem.

About the author

Ziyulong Wang was born in 1996 in Beijing, China. He obtained his Bachelor's degree in Traffic and Transportation Engineering from Beijing Jiaotong University (BJTU) in 2018 and his Master's degree in Civil Engineering from TU Delft in 2020 (cum laude). During 2020-2025, he pursued a PhD at TU Delft's Department of Transport & Planning, focusing on aligning railway traffic management with automatic train operation on mainline networks, in collaboration with the Dutch infrastructure manager, ProRail. His research interests include railway operations, traffic management, public transport networks, and data analytics.

He has contributed to multiple railway research and industrial projects, including RouteLint, RAILS, and Europe's Rail FP1 MOTIONAL and FP2 R2DATO. Within Europe's Rail, he played an important role in developing the Train Path Envelope Generator, which was integrated with the FRISO simulation platform to support real-time railway applications and human-in-the-loop simulation.

He has also been actively involved in education, delivering over 30 lectures on railway and public transport planning and operations at TU Delft and in the BJTU-TUD collaborative BSc programme in Traffic and Transportation, receiving consistently positive student feedback. He has supervised five MSc and four BSc students, whose theses have contributed to ongoing research and innovation.

Dr. Wang aspires to bridge research, education, and practice in the fields of railway and public transport. He is an associate member of the Institution of Railway Signal Engineers (IRSE) and a member of the International Association of Railway Operations Research (IAROR).

Publications

Journal papers

1. **Wang, Z.**, Zhou, R., Correia, G.H.A., Philipsen, E.P., Goverde, R.M.P., 2026. Real-time train timetable rescheduling with event time flexibility, *under review*.
2. Liu, Y., **Wang, Z.**, Cats, O., Pei, X., Shang, P., 2026. Semi-flexible transit service optimization considering scenario-based demand fluctuations. *Transportation Research Part E: Logistics and Transportation Review*, 206, 104589.
3. Yu, C., Li, H., **Wang, Z.**, Ma, W., Goverde, R.M.P., Cats, O., 2025. Why do they travel backwards? Understanding the passenger behaviour in congested urban public transport systems. *Transportation Research Part A: Policy and Practice*, 200, 104630.
4. **Wang, Z.**, Quaglietta, E., Bartholomeus, M.G.P., Goverde, R.M.P., 2025. Sensitivity of train path envelopes for automatic train operation. *IEEE Transactions on Intelligent Transportation Systems*, 26(11), 18921-18933.
5. **Wang, Z.**, Quaglietta, E., Bartholomeus, M.G.P., Cunillera, A., Goverde, R.M.P., 2025. Optimising timing points for effective automatic train operation. *Computers & Industrial Engineering*, 206, 111237.
6. **Wang, Z.**, Huang, K., Massobrio, R., Bombelli, A., Cats, O., 2024. Quantification and comparison of unimodal public transport network hierarchies. *Physica A: Statistical Mechanics and its Applications*, 634, 129479.
7. **Wang, Z.**, Aoun, J., Szymula, C., Bešinović, N., 2023. Promising solutions for railway operations to cope with future challenges — Tackling COVID and beyond. *Journal of Rail Transport Planning & Management*, 28, 100405.
8. **Wang, Z.**, Quaglietta, E., Bartholomeus, M.G.P., Goverde, R.M.P., 2022. Assessment of architectures for Automatic Train Operation driving functions. *Journal of Rail Transport Planning & Management*, 24, 100352.
9. **Wang, Z.**, Pel, A.J., Verma T., Krishnakumari, P., Van Brakel, P., Van Oort N., 2022. Effectiveness of trip planner data in predicting short-term bus ridership. *Transportation Research Part C: Emerging Technologies*, 142, 103790.
10. Tang, R., De Donato, L., Bešinović, N., Flammini, F., Goverde, R.M.P., Lin, Z., Liu, R., Tang, T., Vittorini, V., **Wang, Z.**, 2022. A literature review of Artificial Intelligence applications in railway systems. *Transportation Research Part C: Emerging Technologies*, 140, 103679.
11. Tan, Y., Xu, W., Jiang, Z., **Wang, Z.**, Sun, B., 2021. Inserting extra train services on high-speed railway. *Periodica Polytechnica Transportation Engineering*, 49(1), 16 - 24.
12. Xu, W., Tan, Y., Sharma, B., **Wang, Z.**, 2020. Cyclic timetable scheduling problem on high-speed railway line. *Periodica Polytechnica Transportation Engineering*, 48(1), 31 - 38.

Conference contributions

1. Keskin, K., **Wang, Z.**, Goverde, R.M.P., 2024. Microscopic railway simulation of train path envelopes, in: *The 11th International Conference on Railway Operations Modelling and Analysis (RailDresden)*, Dresden, Germany.
2. **Wang, Z.**, Zhou, R., Correia, G.H.A., Philipsen, E.P., Goverde, R.M.P., 2024. Adding timetable flexibility to real-time railway traffic management, in: *The 11th International Conference on Railway Operations Modelling and Analysis (RailDresden)*, Dresden, Germany.
3. **Wang, Z.**, Huang, K., Massobrio, R., Bombelli, A., Cats, O., 2023. Quantification and comparison of unimodal public transport network hierarchies, in: *The 103rd Transportation Research Board Annual Meeting (TRBAM)*, Washington DC, USA.
4. **Wang, Z.**, Quaglietta, E., Bartholomeus, M.G.P., Cunillera, A., Goverde, R.M.P., 2023. Conflict-free train path planning using ATO timing points, in: *The 10th International Conference on Railway Operations Modelling and Analysis (RailBelgrade)*, Belgrade, Serbia.
5. **Wang, Z.**, Pel, A.J., Verma T., Krishnakumari, P., Van Brakel, P., Van Oort N., 2022. Effectiveness of trip planner data in predicting short-term bus ridership, in: *The 15th International Conference on Advanced Systems in Public Transport (CASPT)*, Tel-Aviv, Israel.
6. **Wang, Z.**, Quaglietta, E., Goverde, R.M.P., 2021. Evaluating different architectures of automatic train operation, in: *The 9th International Conference on Railway Operations Modelling and Analysis (RailBeijing)*, Beijing, China.
7. Tang, R., De Donato, L., Bešinović, N., Flammini, F., Lin, Z., Liu, R., Tang, T., Vittorini, V., **Wang, Z.**, 2021. Artificial Intelligence in railway transport: The state-of-the-art survey and future research agenda, in: *The 31st European Conference on Operational Research (EURO)*, Athens, Greece.
8. **Wang, Z.**, Luo, D., Cats, O., Verma, T., 2020. Unravelling the hierarchy of public transport networks, in: *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, Rhodes, Greece.
9. Xu, W., **Wang, Z.**, Sharma, B., Tan, Y., 2018. The reasonable cycle length for periodic train line planning problem, in: *2018 8th International Conference on Logistics, Informatics and Service Sciences (LISS)*, Toronto, Canada.

TRAIL Thesis Series

The following list contains the most recent dissertations in the TRAIL Thesis Series. For a complete overview of more than 400 titles, see the TRAIL website: www.rsTRAIL.nl.

The TRAIL Thesis Series is a series of the Netherlands TRAIL Research School on transport, infrastructure and logistics.

Wang, Z., *Optimising Performance of Automatic Train Operation on Railway Networks*, T2026/6, March 2026, TRAIL Thesis Series, the Netherlands

Hadi, A.H., *DEM Modelling of Multi-Component Segregation in the Blast Furnace Charging System*, T2026/5, February 2026, TRAIL Thesis Series, the Netherlands

Farhani, M., *Demand Management Strategies for Operations of Shared Mobility Services*, T2026/4, February 2026, TRAIL Thesis Series, the Netherlands

Yao, X., *Driving Heterogeneity in Traffic Flow Theory: An action-based framework for identification, modelling, and simulation*, T2026/3, January 2026, TRAIL Thesis Series, the Netherlands

Versluis, N.D., *Optimising Railway Traffic Management under Radio-Based Distance-to-Go Signalling*, T2026/2, January 2026, TRAIL Thesis Series, the Netherlands

Jiao, Y., *Proactive Collision Risk Quantification in Multi-directional Traffic Interactions*, T2026/1, January 2026, TRAIL Thesis Series, the Netherlands

Asadi, M., *Accessibility and Road Safety: Integration of road safety in accessibility evaluation*, T2025/20, November 2025, TRAIL Thesis Series, the Netherlands

Akse, R., *Understanding and untangling the uncertainty knot: How to catalyse decision-making in mobility innovations*, T2025/19, November 2025, TRAIL Thesis Series, the Netherlands

Führer, K., *Participatory Decision-making under Deep Uncertainty: Modelling mobility transitions*, T2025/18, November 2025, TRAIL Thesis Series, the Netherlands

Picco, A., *Monitoring and Feedback in Driving*, T2025/17, October 2025, TRAIL Thesis Series, the Netherlands

Cebeci, M.S., *Behaviour of Prosumers in Last-mile Logistics: The case of crowdshipping*, T2025/16, September 2025, TRAIL Thesis Series, the Netherlands

Kuijpers, A., *Enabling Inter-Organizational Collaboration Through Platforms: The role of trust*, T2025/15, September 2025, TRAIL Thesis Series, the Netherlands

Song, R., *Human-MASS Interaction in Decision-Making for Safety and Efficiency in Mixed Waterborne Transport Systems*, T2025/14, June 2025, TRAIL Thesis Series, the Netherlands

Destyanto, A.R., *A Method for Evaluating Port Resilience in an Archipelago*, T2025/13, June 2025, TRAIL Thesis Series, the Netherlands

Karademir, C., *Synchronized Two-echelon Routing Problems: Exact and approximate methods for multimodal city logistics*, T2025/12, May 2025, TRAIL Thesis Series, the Netherlands

Vial, A., *Eyes in Motion: A new traffic sensing paradigm for pedestrians and cyclists*, T2025/11, May 2025, TRAIL Thesis Series, the Netherlands

Chen, Q., *Towards Mechanical Intelligence in Soft Robotics: Model-based design of mechanically intelligent structures*, T2025/10, April 2025, TRAIL Thesis Series, the Netherlands

Eftekhar, Z., *Exploring the Spatial and Temporal Patterns in Travel Demand: A data-driven approach*, T2025/9, June 2025, TRAIL Thesis Series, the Netherlands

Reddy, N., *Human Driving Behavior when Interacting with Automated Vehicles and the Implications on Traffic Efficiency*, T2025/8, May 2025, TRAIL Thesis Series, the Netherlands

Durand, A., *Lost in Digitalisation? Navigating public transport in the digital era*, T2025/7, May 2025, TRAIL Thesis Series, the Netherlands

Dong, Y., *Safe, Efficient, and Socially Compliant Automated Driving in Mixed Traffic: Sensing, Anomaly Detection, Planning and Control*, T2025/6, May 2025, TRAIL Thesis Series, the Netherlands

Droffelaar, I.S. van, *Simulation-optimization for Fugitive Interception*, T2025/5, May 2025, TRAIL Thesis Series, the Netherlands

Fan, Q., *Fleet Management Optimisation for Ride-hailing Services: from mixed traffic to fully automated environments*, T2025/4, April 2025, TRAIL Thesis Series, the Netherlands

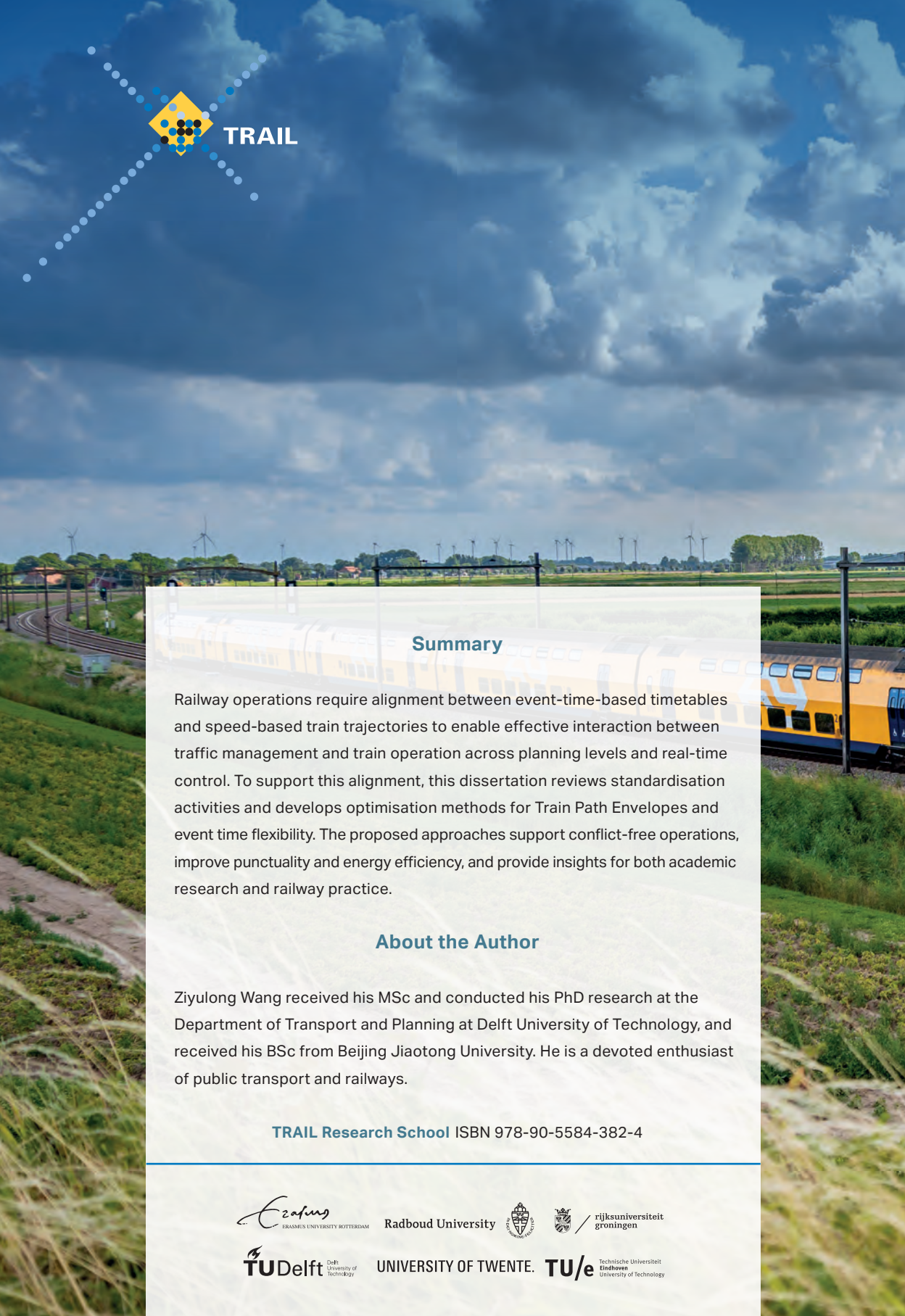
Hagen, L. van der, *Machine Learning for Time Slot Management in Grocery Delivery*, T2025/3, March 2025, TRAIL Thesis Series, the Netherlands

Schilt, I.M. van, *Reconstructing Illicit Supply Chains with Sparse Data: a simulation approach*, T2025/2, January 2025, TRAIL Thesis Series, the Netherlands

Ruijter, A.J.F. de, *Two-Sided Dynamics in Ridesourcing Markets*, T2025/1, January 2025, TRAIL Thesis Series, the Netherlands

Fang, P., *Development of an Effective Modelling Method for the Local Mechanical Analysis of Submarine Power Cables*, T2024/17, December 2024, TRAIL Thesis Series, the Netherlands

Zattoni Scroccaro, P., *Inverse Optimization Theory and Applications to Routing Problems*, T2024/16, October 2024, TRAIL Thesis Series, the Netherlands



TRAIL

Summary

Railway operations require alignment between event-time-based timetables and speed-based train trajectories to enable effective interaction between traffic management and train operation across planning levels and real-time control. To support this alignment, this dissertation reviews standardisation activities and develops optimisation methods for Train Path Envelopes and event time flexibility. The proposed approaches support conflict-free operations, improve punctuality and energy efficiency, and provide insights for both academic research and railway practice.

About the Author

Ziyulong Wang received his MSc and conducted his PhD research at the Department of Transport and Planning at Delft University of Technology, and received his BSc from Beijing Jiaotong University. He is a devoted enthusiast of public transport and railways.

TRAIL Research School ISBN 978-90-5584-382-4

Erasmus University Rotterdam · Radboud University · rijksuniversiteit groningen · TU Delft Delft University of Technology · University of Twente. · TU/e Technische Universiteit Eindhoven University of Technology