Assessment of architectures for Automatic Train Operation driving functions

Wang, Ziyulong; Quaglietta, Egidio; Bartholomeus, Maarten G.P.; Goverde, Rob M.P.

DOI
10.1016/j.jrtpm.2022.100352

Publication date
2022

Document Version
Final published version

Published in
Journal of Rail Transport Planning and Management

Citation (APA)
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 and Management*, *24*, Article 100352. https://doi.org/10.1016/j.jrtpm.2022.100352

Journal of Rail Transport Planning & Management 24 (2022) 100352

Contents lists available at ScienceDirect

Journal of Rail Transport Planning & Management

journal homepage: www.elsevier.com/locate/jrtpm

Assessment of architectures for Automatic Train Operation driving functions

Ziyulong Wang [a,*], Egidio Quaglietta [a], Maarten G.P. Bartholomeus [b], Rob M.P. Goverde [a]

[a] Department of Transport and Planning, Delft University of Technology, P.O. Box 5048, 2600 GA Delft, The Netherlands
[b] Department of Automatic Train Operation, ProRail, P.O. Box 2038, 3511 EP Utrecht, The Netherlands

ARTICLE INFO

Keywords:
Automatic Train Operation
Connected Driver Advisory System
Train Path Envelope
Train trajectory generation
ATO-over-ETCS
SWOT

ABSTRACT

Automatic Train Operation (ATO) is well-known in urban railways and gets increasing interest from mainline railways at present to improve capacity and punctuality. A main function of ATO is the train trajectory generation that specifies the speed profile over the given running route considering the timetable and the characteristics of the train and infrastructure. This paper proposes and assesses different possible ATO architecture configurations through allocating the intelligent components on the trackside or onboard. The set of analyzed ATO architecture configurations is based on state-of-the-art architectures proposed in the literature for the related Connected Driver Advisory System (C-DAS). Results of the SWOT analysis highlight that different ATO configurations have diverse advantages or limitations, depending on the type of railway governance and the technological development of the existing railway signaling and communication equipment. In addition, we also use the results to spotlight operational, technological, and business advantages/limitations of the proposed ATO-over-ETCS architecture that is being developed by the European Union Agency for Railways (ERA) and provide a scientific argumentation for it.

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 utilize 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 realized 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 analyzed 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

---

[*] Corresponding author.

E-mail addresses: Z.Wang-19@tudelft.nl (Z. Wang), E.Quaglietta@tudelft.nl (E. Quaglietta), maarten.bartholomeus@prorail.nl (M.G.P. Bartholomeus), R.M.P.Goverde@tudelft.nl (R.M.P. Goverde).

https://doi.org/10.1016/j.jrtpm.2022.100352
Received 6 May 2022; Received in revised form 25 July 2022; Accepted 6 September 2022

advice based on the timetable and the current train delay. 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 et al., 2019). In this way, drivers can anticipate their driving behavior 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 onboard C-DAS subsystem. Each of these three architecture 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 standardizing 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 62290-1, 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 62290-1, 2014; ERA UNISIG, 2017b). 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 (ERA UNISIG, 2017c). 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 – is 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 towards 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 from 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 signaling and communication equipment. Additionally, we also highlight 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 organization 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, 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 3. Next, Section 4 analyzes the ATO driving functions and proposes three design choices for ATO functional architectures. Afterward, we perform the SWOT analysis on these alternatives in Section 5. Lastly, Section 6 concludes this paper. We also include a list of acronyms in Appendix.

2. Literature review

In this section, we first briefly introduce ATO in Section 2.1. Second, we review the ATO-over-ETCS system requirements specification in Section 2.2, followed by the description of the state-of-the-art analogous C-DAS architecture alternatives in Section 2.3.

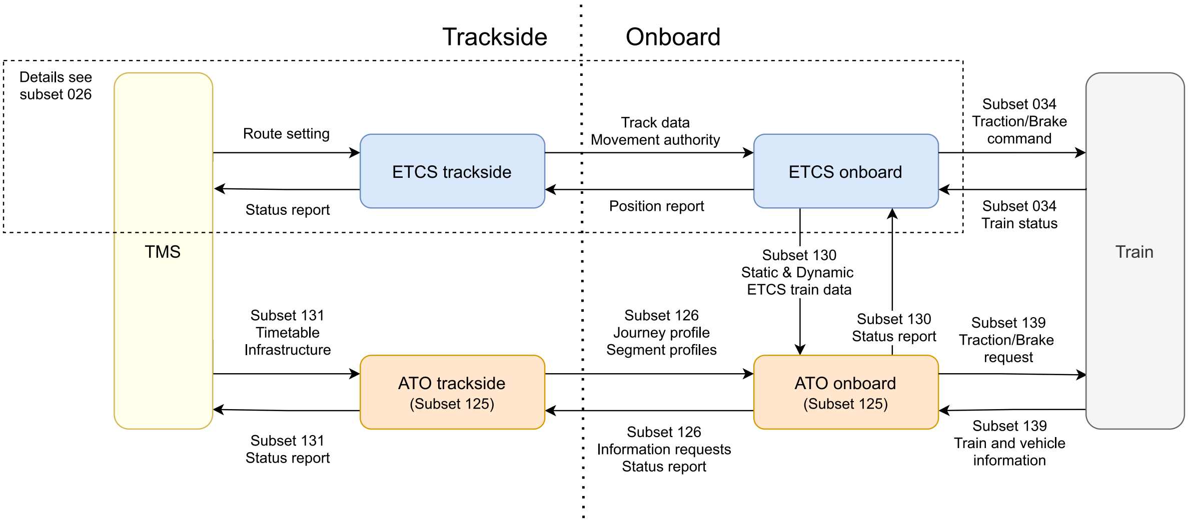

Fig. 1. ATO-over-ETCS reference architecture.

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, 2020). The ATO-TS receives the target timetable and infrastructure data from the TMS. In general, the target timetable should be sufficiently robust to be realizable 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, 2016). The TMS also triggers the route setting according to the timetable and the train positions by e.g., 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. It is thus essential that the train respects the train 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 (ERA UNISIG, 2017a).

2.2. *ATO-over-ETCS reference architecture*

In Europe, a set of ATO-over-ETCS system requirement specifications is being developed, see Fig. 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, 2018).

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, 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 (SPs) and a Journey Profile (JP) at the ATO-TS and then forwarded to empower the ATO-OB driving functions (ERA UNISIG, 2018, 2020). 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 maintain. If new routes are given in the rescheduled timetable, then a new JP with a new set of SPs must be provided.

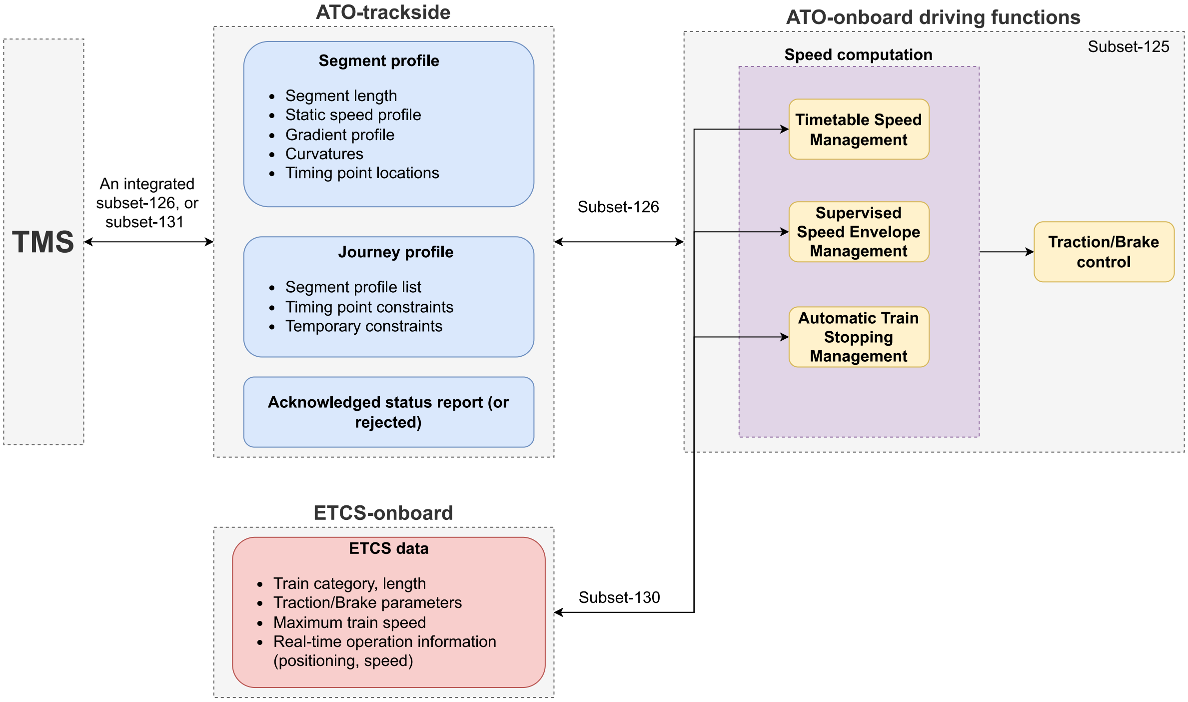

Fig. 2. Driving functions of ATO-over-ETCS, based on ERA UNISIG (2018).

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 Fig. 2 (ERA UNISIG, 2018). 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 of identifying potential architectures, which could represent the best trade-off of costs and legal responsibilities/burdens with operational and business performances.

2.3. C-DAS and C-DAS architectures

Although ATO is not prevailing in mainline railway as that of urban railway, C-DAS as GoA 1 is at the forefront of the assisted train operation and implementations of different C-DAS architectures can be observed throughout the world. 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 neighboring 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 minimize energy consumption. This train trajectory is then translated into speed advice, 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

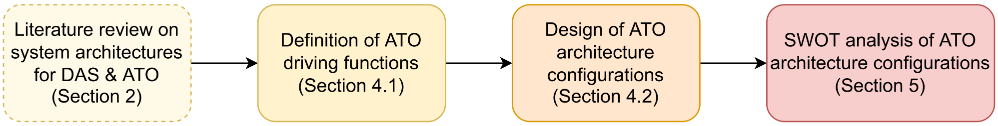

Fig. 3. Outline of the proposed methodology for assessing different ATO architecture configurations.

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 categorization 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 the successive regimes and displays it to the driver. It is crucial to assure 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 (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).

3. Methodology

We develop a three-step method to identify the ATO driving functions, propose different ATO architectures and analyze them. As Fig. 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.

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 utilize 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 the 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 realized 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 1, which will be manifested in-depth in the next section.

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 the

Table 1
Comparison between C-DAS and ATO driving functions.

	Spatial–temporal information from the trackside	Train trajectory optimization	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

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 and Abtahi, 2022), market potentials of virtual coupling railway signaling (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 requirement. Then, the impact of the assigned onboard intelligence on train operations and the onboard processing power is analyzed. 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.

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 4.1). Thereafter, Section 4.2 defines three ATO system architecture design choices.

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 Fig. 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 an optimal cruising speed and coasting points for every train trip. It aims to minimize the 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 Fig. 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.

4.1.1. 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 from 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.

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 optimization as proposed in some recent literature (for example, Luan et al., 2018). 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

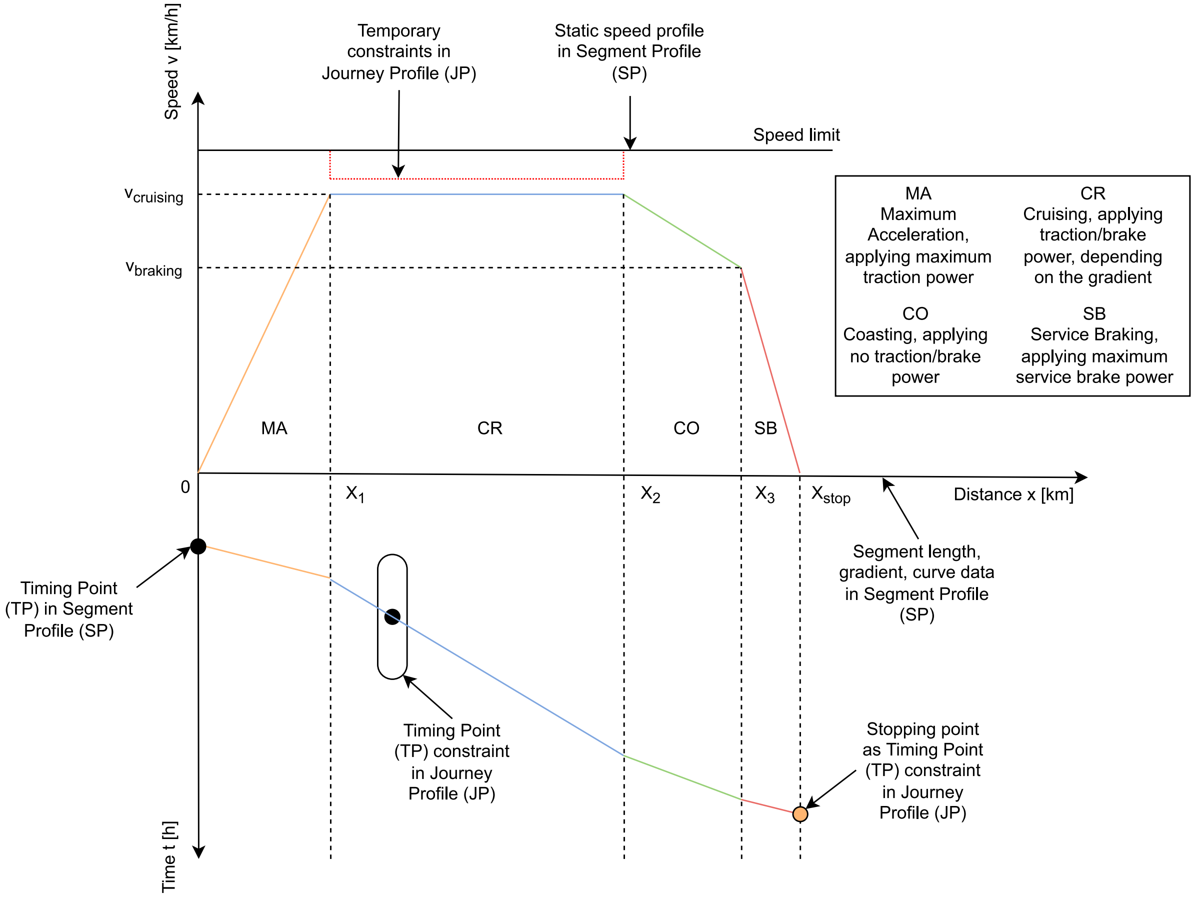

Fig. 4. Schematic drawing for the relation of different components in the ATO driving system architecture.

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 (ERA UNISIG, 2017b). Regarding the TPE, it identifies a sequence of time windows that facilitates train operations in an energy-efficient fashion without conflicting the neighboring 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 optimize 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 optimize the train operations within it. These TPs are also used as information points where the train informs the trackside on estimated JP execution based on the actual operation information. In the recent trajectory optimization research with the TPE applied, the TPs could be suited at station stops, junction locations, signal positions, and route release points (Wang and Goverde, 2016a).

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 optimizes 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 optimization literature (Scheepmaker et al., 2017), particularly the ones which used the concept of a TPE (Wang and Goverde, 2016a,b). Besides, both TPE and JP emphasize the need to minimize energy consumption for the train trajectory optimization, 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 optimize 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 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 optimization research. Whereas a TPE focuses more on train-centric driving flexibility, a JP offers an IM more

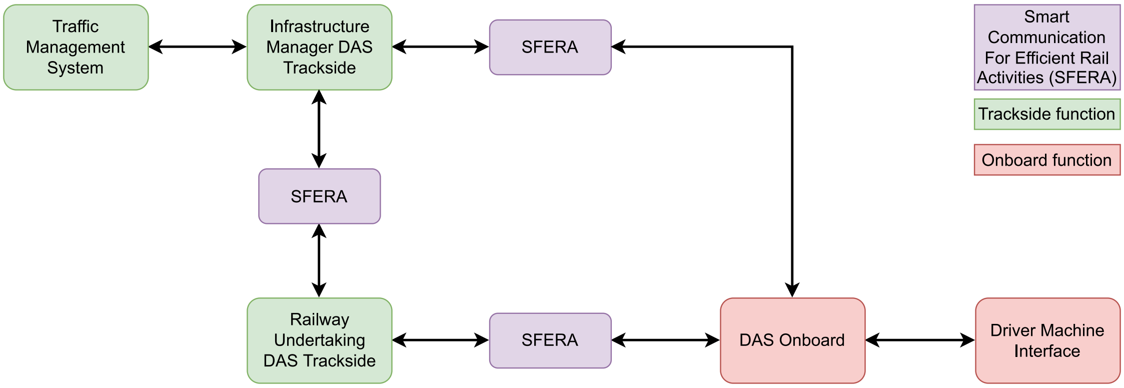

Fig. 5. Illustrative diagram of the SFERA system architecture, adapted from UIC (2020).

functions/flexibility through 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 Fig. 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 emphasized 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 a cooperation of the IM and the RU.

4.1.2. *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, maximizing energy efficiency (between bottlenecks) or minimizing track occupation (in bottleneck areas).

4.1.3. *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.

4.1.4. *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 minimize 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 maneuvering within the JP, the TMS needs to be warned and has to generate adjusted time targets for the ATO-TS.

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 categorization is distinguished by the distribution of intelligence shown in Fig. 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 in the ATO-TS.

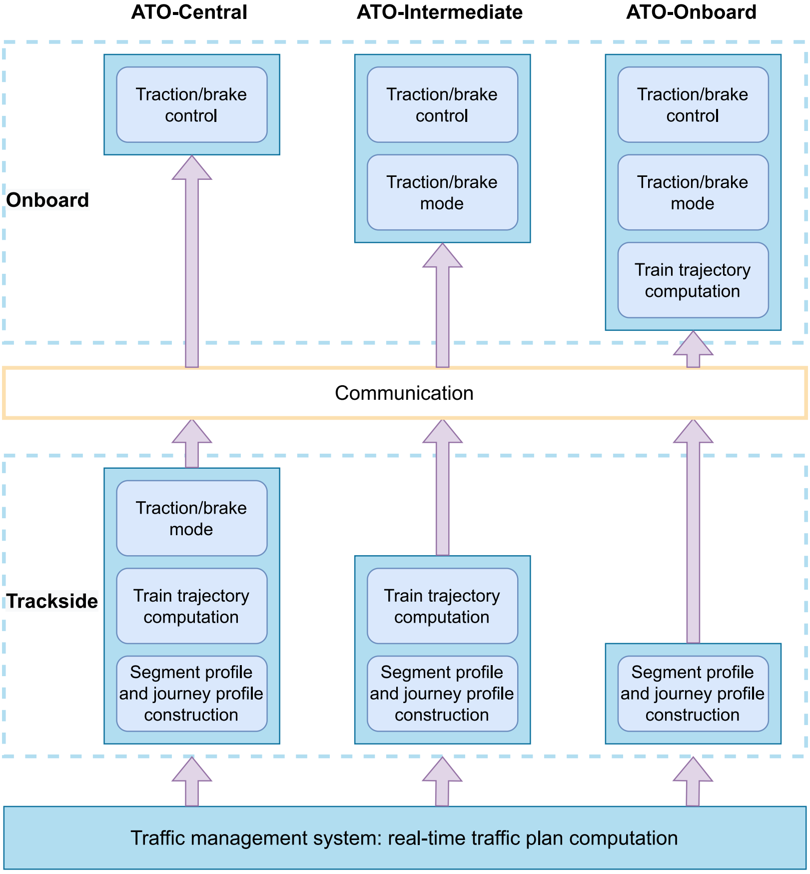

Fig. 6. Distribution of functions for the ATO architecture alternatives.

- ATO-Intermediate (ATO-I): Here the construction of SP and JP and the optimal train trajectory are performed in 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.

We exhibit the distribution of ATO intelligence with the input/output information for each design choice in Fig. 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 varies with each other. We highlight the ATO-TS of each alternative in a different background color. The ATO-OB functions of each alternative are listed next to the highlighted trackside part. Detailed discussions followed are focusing on every design choice, sequentially.

4.2.1. 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

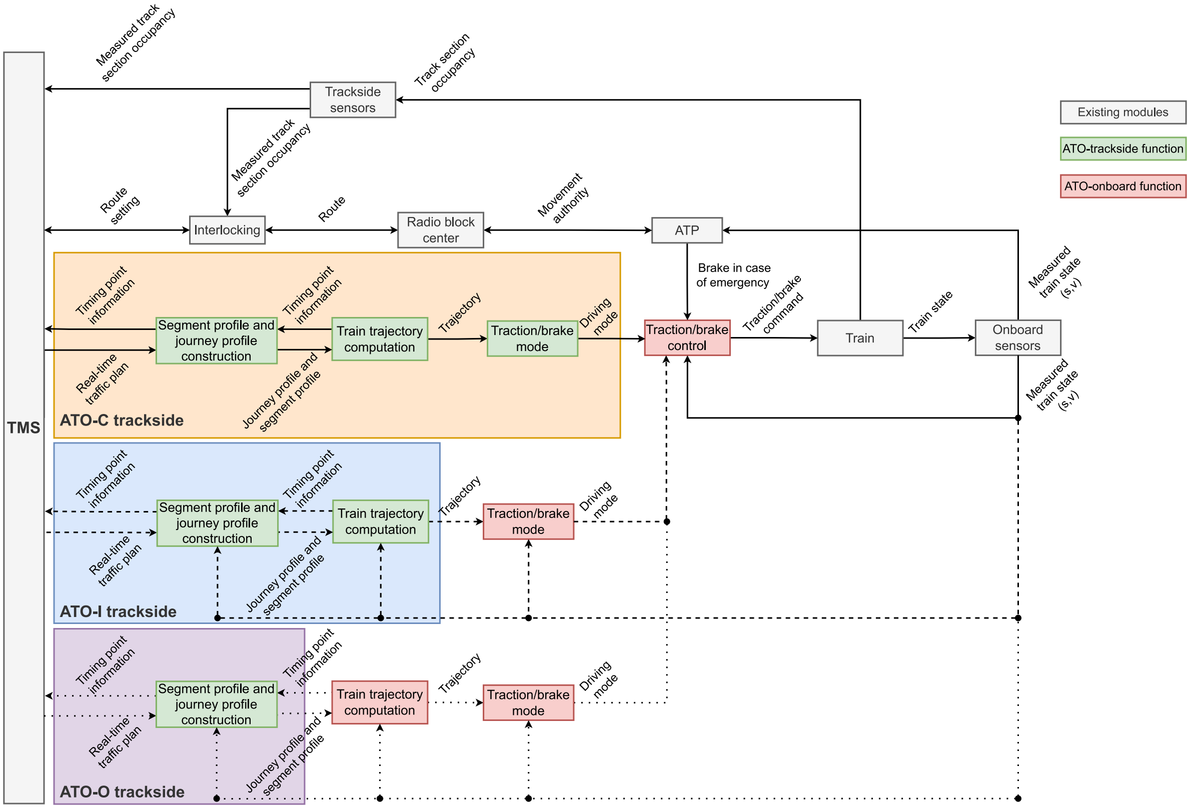

Fig. 7. Distribution of intelligence for ATO architecture alternatives.

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 centralized 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 feedback 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.

4.2.2. 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 onboard, 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 parameter 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.

4.2.3. ATO-O

Except for the construction of the SP and JP that are calculated at the ATO-TS, all other three components are realized onboard, 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 cannot be realized, 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.

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 Table 2. 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 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 realized at the ATO-TS. The ATO-TS needs to frequently observe the train deviation and recomputes the train trajectory if the deviation is larger than the error bandwidth and forwards the updated traction/brake information to the onboard unit. On the one hand, this results in high demand for the availability of the 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. But it still needs to constantly monitor the train state and recomputes the trajectory if needed. It sends the medium-size information to the ATO-OB to derive the traction/brake mode from the computed train 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 computation 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 optimization 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 utilizing the control loop onboard, namely the traction/brake control. Additionally, when the train cannot be maneuvered within the JP, each choice needs to redetermine the JP at the trackside based on an updated RTTP from the TMS, which is the trackside control loop. However, the underlying difference lies in the computation location of 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 onboard 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

Table 2
SWOT analysis of ATO architecture configurations.

ATO 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 • Little reaction capability on driving deviations onboard • Trajectory determined with assumed train data • Stable & 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 and parameter estimation • Predictable train trajectory to IM	• High requirements for communication latency • Medium frequent communication between trackside and onboard • Limited reaction capability on driving deviations onboard • Trajectory determined with assumed train data • Stable & predictable environments
ATO-O	• Low requirements for communication latency • Low communication frequency between trackside and onboard • Rich reaction capability on driving deviations onboard • Accurate trajectory generation and traction/brake mode determination due to onboard train state and parameter estimation	• High communication volume because of the JP and SP transmission • High onboard processing power for the train trajectory generation • Train trajectory unknown to IM or requires communication
ATO 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 the 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 • Jeopardize free-access rights of RUs because of the high dependency of RUs on IMs • Vertical separation inapplicable • High investment for the IMs • High responsibility in energy consumption for trackside
ATO-I	• High interoperability between IM and RU • Vertical integration applicable • Applicable to low complexity networks with homogeneous traffic, corridors with no intersections and shunting/stabling • High central unit computation power for cloud platform possibility • Freedom to RUs to choose supervisory information • Low investment for the RUs • Medium responsibility in energy consumption for onboard	• Control over train operations by IMs as they are in charge of train trajectory generation • Jeopardize free-access rights of RUs because of the high dependency of RUs on IMs • Vertical separation inapplicable • High investment for the IMs • Medium responsibility in energy consumption for trackside
ATO-O	• Applicable to all market segments • Vertical integration & separation applicable • Freedom to RUs to optimize the train trajectory • Freedom to RUs to choose supervisory information • Migration from C-DAS/SFERA • Low investment for the IMs	• Divergent goals in railway operations of IMs and RUs • Various RUs may develop different trajectory generation solutions • High investment for the RUs

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. Nonetheless, ATO-C has the strength of a fully centralized 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 are 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 optimize 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 the real-time ATP information in order to compute a safe and conflict-free JP. This could bring in extra responsibilities and meanwhile as risks to IMs as they get involved in train operations. Oppositely, if RUs possess the ATO-TS, they can have the freedom for optimizing 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 optimize 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 that 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 optimizes 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 optimize 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 make 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 need 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 the 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 a stable circumstance 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 signaled 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 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 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 (2020) 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 them easier to develop 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 the railway operations based on either modeled consumption rates or actual metered usage. Since the train trajectory is derived onboard in ATO-O, the goal of sustainability and the energy-efficiency should remain the duty of RUs. However, the railway traffic becomes highly centralized in 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 organizational structure that suits every design option the best. A vertical separation means that a country separates its railway into totally separated IMs and RUs, while a 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 organization. This support could attract more than one competitor to the rail market and contribute to both passenger and freight transport revenue (Huang et al., 2019; Esposito et al., 2020).

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. Eventually, 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 formalized 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 organization 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 organizations. 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 reexamined 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-efficiently. 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 organizational 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 a mixed-use of C-DAS and ATO. Further research could investigate quantitatively how these three different architectures influence the capacity by reducing the human reaction time to the signaling system. Moreover, studies that verify the SWOT analysis based on the proposed ATO architecture configurations and the impact of them could be another research avenue.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was financially supported by the Dutch railway infrastructure manager ProRail. The authors would like to thank ProRail for making this research possible.

Appendix. List of acronyms

ARS	Automatic Route Setting
ATO	Automatic Train Operation
ATO-C	ATO-Central
ATO-I	ATO-Intermediate
ATO-O	ATO-Onboard
ATO-OB	Automatic Train Operation Onboard subsystem
ATO-TS	Automatic Train Operation Trackside subsystem
ATP	Automatic Train Protection
ATSM	Automatic Train Stopping Management
CATO	Computer-Aided Train Operation
C-DAS	Connected Driver Advisory System

CDR	Conflict Detection and Resolution
DAS	Driver Advisory System
DAS-C	DAS-Central
DAS-I	DAS-Intermediate
DAS-O	DAS-Onboard
DAS-OB	Driver Advisory System Onboard subsystem
DAS-TS	Driver Advisory System Trackside subsystem
DMI	Driver Machine Interface
ERA	European Union Agency for Railways
ETCS	European Train Control System
FFFIS	Form Fit Functional Interface Specification
GoA	Grade of Automation
IM	Infrastructure Manager
JP	Journey Profile
MA	Movement Authority
MaaS	Mobility as a Service
RBC	Radio Block Centre
RTTP	Real-Time Traffic Plan
RU	Railway Undertaking
S-DAS	Standalone Driver Advisory System
SFERA	Smart Communication For Efficient Rail Activities
SP	Segment Profile
SSEM	Supervised Speed Envelope Management
SWOT	Strengths, Weaknesses, Opportunities and Threats
TMS	Traffic Management System
TP	Timing Point
TPE	Train Path Envelope
TTSM	Timetable Speed Management
UIC	International Union of Railways

References

Aoun, J., Quaglietta, E., Goverde, R.M.P., 2020. Investigating market potentials and operational scenarios of virtual coupling railway signaling. Transp. Res. Rec. 2674 (8), 799–812.

Caillot, A., Ouerghi, S., Vasseur, P., Boutteau, R., Dupuis, Y., 2022. Survey on cooperative perception in an automotive context. IEEE Trans. Intell. Transp. Syst. http://dx.doi.org/10.1109/TITS.2022.3153815.

Cunillera, A., Bešinović, N., Van Oort, N., Goverde, R.M.P., 2022. Real-time train motion parameter estimation using an unscented Kalman filter. Transp. Res. C 143, 103794.

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

Dyson, R.G., 2004. Strategic development and SWOT analysis at the university of warwick. European J. Oper. Res. 152 (3), 631–640.

Eckhardt, J., Nykänen, L., Aapaoja, A., Niemi, P., 2018. Maas in rural areas - case Finland. Res. Transp. Bus. Manag. 27, 75–83.

ERA UNISIG, 2014. ATO-TS / TMS Interface Specification. REF: SUBSET-131, ISSUE: 0.0.3.

ERA UNISIG, 2016. ERTMS/ETCS System Requirements Specification. REF: SUBSET-026, ISSUE: 3.6.0.

ERA UNISIG, 2017a. ATO over ETCS Glossary. REF: 13E154, Version: 1.4.

ERA UNISIG, 2017b. ATO over ETCS Glossary. REF: 13E154, Version: 1.6.

ERA UNISIG, 2017c. ATO over ETCS Operational Principles. REF: 12E108, Version: 1.4.

ERA UNISIG, 2018. ATO over ETCS System Requirements Specification. REF: SUBSET-125, ISSUE: 0.1.0.

ERA UNISIG, 2020. ATO-OB / ATO-TS FFFIS Application Layer. REF: SUBSET-126, ISSUE: 0.0.17.

Esposito, G., Cicatiello, L., Ercolano, S., 2020. Reforming railways in the EU: An empirical assessment of liberalisation policies in the European rail freight market. Transp. Res. A Policy Pract. 132, 606–613.

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.

Huang, W., Zhang, Y., Shuai, B., Xu, M., Xiao, W., Zhang, R., Xu, Y., 2019. China railway industry reform evolution approach: Based on the vertical separation model. Transp. Res. A Policy Pract. 130, 546–556.

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

Lagos, M., 2011. CATO offers energy savings. Railw. Gaz. Int. 167 (5), 50–52.

Learned, E.P., Andrews, K.R., Christensen, C.R., Guth, W.D., 1969. Business Policy: Text and Cases. RD Irwin.

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

Luijt, R.S., Van den Berge, M.P.F., Willeboordse, H.Y., Hoogenraad, J.H., 2017. 5 years of dutch eco-driving: Managing behavioural change. Transp. Res. A Policy Practice 98, 46–63.

Nash, C.A., Smith, A.S.J., Van de Velde, D., Mizutani, F., Uranishi, S., 2014. Structural reforms in the railways: Incentive misalignment and cost implications. Res. Transp. Econ. 48, 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 D4.1: Functional and technical requirements specification for perturbation management. URL: https://www.ontime-project.eu/deliverables.aspx.

Panou, K., Tzieropoulos, P., Emery, D., 2013. Railway driver advice systems: Evaluation of methods, tools and systems. J. Rail Transp. Plan. Manag. 3 (4), 150–162.

Poulus, R.W., Van Kempen, E.A., Van Meijeren, J.C., 2018. Automatic Train Operation. Driving the Future of Rail Transport. Technical Report 842373, Netherlands Organisation for Applied Scientific Research, The Hague, URL: http://resolver.tudelft.nl/uuid:5294ea46-7bd7-47b9-8f45-8f9ae2b30a52.

Quaglietta, E., Pellegrini, P., Goverde, R.M.P., Albrecht, T., Jaekel, B., Marlière, G., Rodriguez, J., Dollevoet, T., Ambrogio, B., Carcasole, D., Giarolig, M., Nicholsonh, G., 2016. The ON-TIME real-time railway traffic management framework: A proof-of-concept using a scalable standardised data communication architecture. Transp. Res. C 63, 23–50.

Rao, X., Montigel, M., Weidmann, U., 2016. A new rail optimisation model by integration of traffic management and train automation. Transp. Res. C 71, 382–405.

Scheepmaker, G.M., Goverde, R.M.P., Kroon, L.G., 2017. Review of energy-efficient train control and timetabling. European J. Oper. Res. 257 (2), 355–376.

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

Vasheghani, M., Abtahi, M., 2022. Strategic planning for multimodal transportation in ports. Marit. Policy Manag. http://dx.doi.org/10.1080/03088839.2022.2061060.

Wang, P., Goverde, R.M.P., 2016a. Multiple-phase train trajectory optimization with signalling and operational constraints. Transp. Res. C 69, 255–275.

Wang, P., Goverde, R.M.P., 2016b. Two-train trajectory optimization with a green-wave policy. Transp. Res. Rec. 2546 (1), 112–120.

Wang, P., Goverde, R.M.P., Van Luipen, J., 2019. A connected driver advisory system framework for merging freight trains. Transp. Res. C 105, 203–221.

Yin, J., Tang, T., Yang, L., Xun, J., Huang, Y., Gao, Z., 2017. Research and development of automatic train operation for railway transportation systems: A survey. Transp. Res. C 85, 548–572.