Real-time railway traffic management under moving-block signalling
A literature review and research agenda

Versluis, Nina D.; Quaglietta, Egidio; Goverde, Rob M.P.; Pellegrini, Paola; Rodriguez, Joaquin

DOI
10.1016/j.trc.2023.104438

Publication date
2023

Document Version
Final published version

Published in
Transportation Research Part C: Emerging Technologies

Transportation Research Part C 158 (2024) 104438

Contents lists available at ScienceDirect

Transportation Research Part C

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

Real-time railway traffic management under moving-block signalling: A literature review and research agenda

Nina D. Versluis [a,*], Egidio Quaglietta [a], Rob M.P. Goverde [a], Paola Pellegrini [b], Joaquin Rodriguez [b]

[a] *Department of Transport and Planning, Delft University of Technology, Stevinweg 1, 2628 CN Delft, The Netherlands*
[b] *COSYS-ESTAS, Université Gustave Eiffel, 20 rue Élisée Reclus, 59650 Villeneuve d'Ascq, France*

ARTICLE INFO

Keywords:
Moving-block signalling
Real-time railway traffic management
Conflict detection and resolution
Rescheduling
Dispatching

ABSTRACT

Railway traffic management is responsible for the detection and resolution of conflicts in case of disturbed operations. To minimise delay propagation, rescheduling decisions are taken by human dispatchers, possibly supported by mathematical models. Existing conflict detection and resolution (CDR) models mostly refer to conventional fixed-block multi-aspect signalling systems, in which minimum train headways are determined based on a preset number of blocks considering worst-case braking distances and number of signal aspects. In moving-block signalling systems, minimum headways are based on absolute braking distances. This paper reviews literature on CDR with the aim to identify gaps and to propose next steps in the research on CDR under moving-block signalling. A research agenda presents various modelling options, for which modelling approaches are proposed based on a comparative analysis.

1. Introduction

Moving-block signalling is an innovative railway signalling solution, offering increased efficiency and capacity compared to fixed-block signalling (UNIFE, 2022). In conventional fixed-block signalling systems, the track is divided into fixed-length block sections, which are protected by trackside signals. These systems rely on trackside train detection for train position and integrity monitoring, with train integrity referring to whether a train has not accidentally split. In these fixed-block signalling systems, train headways, i.e., the minimum safe head-to-head separation distance between two trains, are based on a preset number of block sections considering worst-case braking distances and number of signal aspects. In moving-block signalling systems, the fixed block sections are eliminated. Trackside signalling and train detection are replaced by radio-based cab signalling and onboard train positioning and train integrity monitoring (TIM), respectively. With this, moving-block train headways are continuously based on absolute braking distances to the tail of the preceding train, i.e., the distance a train needs to decelerate to a standstill from its current speed. The continuous braking curve calculation is characteristic to distance-to-go signalling, which is not only featured in moving-block systems but also in advanced fixed-block systems. In fixed-block distance-to-go signalling systems, the fixed-block track division is combined with radio-based cab signalling to enable train headways derived from absolute braking distances to the end of a block.

In moving-block systems, the radio communication system, the train positioning technology and the onboard TIM device are safety-critical components (Martinez and Martin, 2020). So far, TIM is only available for homogeneous, fixed-composition trains

* Corresponding author.
E-mail addresses: n.d.versluis@tudelft.nl (N.D. Versluis), e.quaglietta@tudelft.nl (E. Quaglietta), r.m.p.goverde@tudelft.nl (R.M.P. Goverde), paola.pellegrini@univ-eiffel.fr (P. Pellegrini), joaquin.rodriguez@univ-eiffel.fr (J. Rodriguez).

https://doi.org/10.1016/j.trc.2023.104438
Received 9 May 2023; Received in revised form 11 October 2023; Accepted 24 November 2023
Available online 4 December 2023

and closed networks with low complexity. As a result, moving-block technology is typically deployed on urban railways which have these characteristics (Martinez and Martin, 2020). There, moving-block technology is implemented within Communication-Based Train Control (CBTC) systems. For mainline railways with heterogeneous traffic, variable train compositions and more complex networks, fixed-block signalling is still the standard. In this area, current research focuses on the further development of the safety-critical aspects, rather than on the operational efficiency of the moving-block system (e.g., Himrane et al. (2023) and Lazarescu and Poolad (2021)).

Crucial for the operational efficiency is real-time railway traffic management. It is responsible for the detection and resolution of conflicting operations arising from disturbances, i.e., relatively small delays originating from, for example, variations in rolling stock, dwell times and driver behaviour. In current practice, human dispatchers take rescheduling measures mostly based on experience and preset rules. In the literature, conflict detection and resolution (CDR) models exist to support dispatchers in minimising the impact of disturbances on the network. CDR models generally consider measures such as retiming, i.e., shortening or extending running and/or dwelling times, reordering, i.e., changing passing sequences, and local rerouting, i.e., changing station route and/or platform assignment (Cacchiani et al., 2014). In urban railways, typically only retiming is considered (Pochet et al., 2017).

The literature on the modelling of CDR under moving-block signalling is limited to the preliminary works of Büker et al. (2019), Janssens (2022), Meunier et al. (2023) and Pochet et al. (2016, 2017). The vast majority of CDR models proposed in the literature refers to fixed-block signalling, with modelling approaches relying on the system's inherent infrastructure discretisation and a limited dependence of train headways on speed (e.g., D'Ariano et al. (2007a), Reynolds et al. (2020) and Törnquist and Persson (2007)). A knowledge gap exists regarding the modelling of CDR for moving-block operations in terms of infrastructure representation and speed-headway relation, as well as with respect to the impact of moving-block CDR models on the management of heterogeneous traffic under disturbed conditions.

In this paper, the available literature on CDR models is reviewed with the aim to identify existing gaps and to propose future steps in the research on CDR under moving-block signalling. Specifically, a comparison of modelling approaches considering infrastructure and speed modelling is included. The main contributions of the paper are:

- The identification of gaps and challenges in the research on CDR under moving-block signalling based on a comprehensive review of the existing literature.
- A comparative analysis of modelling approaches for the application to moving-block CDR.
- A research agenda proposing future steps in the development of CDR models for moving-block signalling.

The paper is structured as follows. Section 2 provides background on railway signalling systems. Section 3 presents a literature review on CDR models. In Section 4, gaps in the research on CDR under moving-block signalling are identified and research challenges are formulated. Section 5 presents a research agenda on CDR under moving-block signalling. The paper finalises with concluding remarks in Section 6.

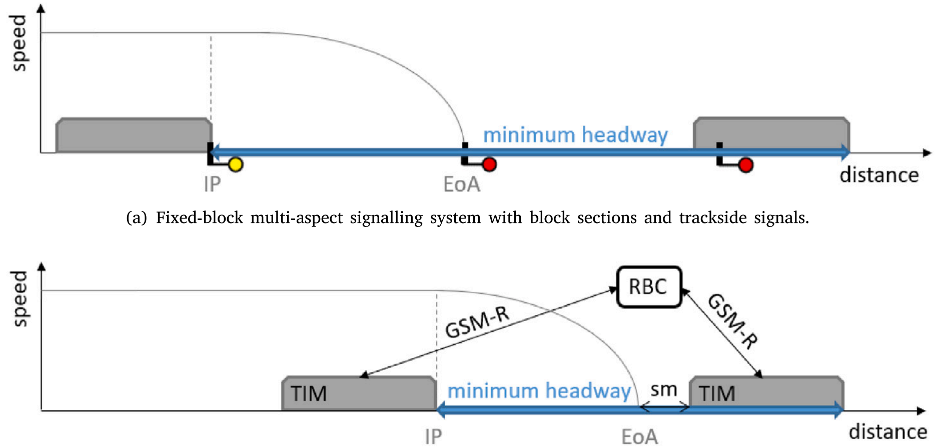
2. Railway signalling systems and blocking time theory

In Section 1, the fixed-block and the moving-block signalling systems have been shortly introduced. Here, the systems are described in more detail and compared to one another in terms of minimum headways. Additionally, blocking time theory is introduced as one of the most frequently used models for minimum headways.

Conventional fixed-block signalling systems are characterised by the division of the track into fixed-length block sections, of which the entry points are protected by trackside signals. These signals use multiple aspects to communicate train movement authorities (MAs), i.e., the permission for a train to move to a specific location, and corresponding speed commands. The colour aspects indicate different instructions to an approaching train. For instance, in the case of three-aspect signalling, they indicate whether the train needs to stop (red), needs to start braking and prepare to stop (yellow) or can proceed without restrictions (green) into the following block section. A train needs to stop if the block section ahead is assigned to another train or otherwise not available. It needs to start braking if it must slow down to be able to halt before the next stop indication. If the system has more than three aspects, additional restrictive aspects exist before a red signal, and hence, increase the number of block sections available to come to a halt.

Fig. 1(a) illustrates the fixed-block multi-aspect system, providing a schematic representation of its main features. These include the brake indication point (IP) at the block entry corresponding to a trackside signal showing yellow, and the end of authority (EoA) at the end of the block right before the first red signal. For the monitoring of train position and integrity, this signalling system relies on trackside train detection devices, such as track circuits or axle counters. To protect against driver errors, multi-aspect block systems are complemented by an automatic train protection (ATP) system, which supervises adherence to the signal aspects. The ATP system can intervene when a train fails to respect restrictive indications.

Moving-block signalling systems feature radio-based cab signalling with a distance-to-go ATP system and onboard train positioning and TIM, while dismissing the concept of fixed block sections (ERA, 2016). The elimination of the fixed block sections, in combination with cab signalling, onboard train positioning and TIM, allows trains to maintain headways based on absolute braking distances. Cab signalling is enabled by bidirectional radio communication, facilitated by, e.g., the Global System for Mobile Communication for Railways (GSM-R), between a radio block centre (RBC) on the trackside and the trains. This bidirectional communication allows a train's MA to continuously be provided up to the tail of the preceding train, or to a movable track element such as a switch, respecting a safety margin. Distance-to-go refers to the onboard computation and supervision of dynamic speed profiles including continuous braking curves to an EoA, enabling the IP to lie anywhere along the track.



(a) Fixed-block multi-aspect signalling system with block sections and trackside signals.

(b) Moving-block signalling system with onboard train integrity monitoring (TIM), radio block centre (RBC), radio communication (GSM-R) and safety margin (sm).

Fig. 1. Schematic layout of minimum headway between two trains in speed-distance diagram under different signalling systems. The headway is related to the brake indication point (IP) and the end of authority (EoA) depending on the braking curve.

Fig. 1(b) illustrates the moving-block signalling system on an open line, providing a schematic representation of its main features. These include the EoA for the following train at a safety margin behind the tail of the leading train, and the corresponding IP upstream along the track, depending on the train's braking distance.

Moving-block system specifications are defined for urban (IEEE, 2005) and mainline railways (ERA, 2016). For the urban railways, the specifications are in the context of CBTC. For the mainline railways, the moving-block concept is developed within the framework of the European Rail Traffic Management System (ERTMS), consisting of the European Train Control System (ETCS) and GSM-R (UNIFE, 2022). In this context, moving-block signalling falls under ERTMS/ETCS Level 3.

Fixed-block distance-to-go signalling is a signalling solution defined and developed within ERTMS/ETCS as Level 2. In fixed-block distance-to-go signalling systems, the track division into fixed blocks is combined with bidirectional trackside-to-train radio communication. This enables distance-to-go signalling. Hence, the IP can lie anywhere along the track, while the EoA is consistently set at the end of a block section (ERA, 2016). Another distance-to-go solution is ETCS Level 3 fixed virtual block. Under virtual block signalling, the infrastructure is virtually divided into short blocks with fixed lengths. The system relies on onboard train positioning and TIM for train-tail detection to clear a virtual block, so that the EoA lies at the end of the last released virtual block (ERA, 2016; Furness et al., 2017).

The signalling system deployed determines the minimum headway between two trains. Indeed, the minimum headway depends on the EoA and the IP of the following train. In fixed-block (multi-aspect and distance-to-go) systems, the EoA of the following train is located at the end of the block section last released by the leading train. In moving-block systems, the EoA of the following train is consistently maintained at a safety margin behind the tail of the leading train, irrespective of the leading train's position. In fixed-block multi-aspect signalling systems, the IP corresponds to the first block entry upstream ensuring at least the braking distance before reaching the EoA. In signalling systems with distance-to-go ATP, i.e., fixed-block distance-to-go and moving-block systems, the IP is situated precisely the train's safe braking distance further upstream of its EoA. Fig. 1 illustrates the minimum headway in terms of EoA and IP under fixed-block multi-aspect (Fig. 1(a)) and moving-block (Fig. 1(b)) signalling.

In the literature, blocking time theory is used to describe minimum train headways. It is a well-known concept for the detection of track conflicts (Hansen and Pachl, 2014) and it will be used in the following to compare various modelling approaches for CDR. Blocking time is the time a track part, e.g., a block section, is assigned to a train and hence blocked for other trains; it starts when a train requests the track part for its route and ends after the train has traversed and cleared it. The total blocking time of a track part constitutes of the following components:

- Setup time — to request, set and lock the route over the track part;
- Reaction time — to perceive and respond to speed/braking indication;
- Approach time — to run from the IP to the track part;
- Running time — to run over the track part, including possible dwell time;
- Clearing time — to run until the train has left the track part with its full length; and
- Release time — to release the route.

The blocking time and (most of) its components depend on the signalling system (Büker et al., 2019). The difference in setup time between fixed-block multi-aspect and distance-to-go systems relates to the shift from trackside to cab signalling. In the former, the MA is communicated via the trackside signals, while in the latter, the MA is communicated by the RBC through radio communication. Similarly, the reaction time is affected by the shift from trackside to cab signalling. The time to perceive a signal depends on whether the signal is continuously visible (cab signalling) or only from a certain sight distance (trackside signalling). The time to respond to a signal is similar in all systems as the MA information, however received, needs to be translated into an action. The approach time is significantly different in the respective systems due to its dependency on the IP, which is determined differently as explained above. The running time component is the other main change due to the track parts considered: fixed block sections, virtual block sections, or no sections at all in moving block, where the running time component is negligible. The clearing time can be considered the same for the different systems; the time it takes to run over a train length is independent of the signalling system. The release time is influenced by the system's reliance on either onboard TIM or trackside train detection for the track-clear detection because of the different communication systems. Also, whereas in fixed-block multi-aspect systems the signals need to be released, i.e., set to the default value, this subcomponent can be omitted in the systems with distance-to-go signalling. Under moving-block signalling, the blocking times around movable track elements are defined in a similar manner as under fixed-block distance-to-go signalling. For more details on signalling, see Theeg and Vlasenko (2020).

3. Literature review of conflict detection and resolution models

In this section, the existing literature on CDR models is reviewed as a springboard for the identification of gaps in the literature regarding CDR under moving-block signalling. Fig. 2 illustrates that CDR under moving-block signalling is staying behind on the trend of an increasing number of publications on moving-block signalling. A similar trend holds for the number of publications on CDR in general.

To provide a comprehensive analysis, the review extends its focus beyond moving-block signalling, also including the other types of signalling systems described in Section 2. First, the review methodology is described in Section 3.1. Then, the review of CDR models under fixed-block multi-aspect, fixed-block distance-to-go and moving-block signalling are presented in Section 3.2, Section 3.3 and Section 3.4, respectively. Finally, a summary is given in Section 3.5.

3.1. Methodology

Two prominent review papers on CDR for conventional fixed-block operations, i.e., Cacchiani et al. (2014) and Corman and Meng (2015), serve as guideline for the literature to consider until 2015. For the period after that, the literature selection is conducted using forward snowballing, i.e., by identifying follow-up papers or other papers referencing the ones reviewed by Cacchiani et al. (2014) and Corman and Meng (2015), cross-referenced with the search query "rail* AND (dispatching OR rescheduling OR resolution OR 'traffic AND management')" in Scopus (2023). Also, the authors' expertise is used to identify the relevant papers within and beyond the selection procedure.

Foremost, the papers are separated based on the modelled signalling system: fixed-block multi-aspect, fixed-block distance-to-go or moving-block. Within this classification, several (CDR) modelling aspects are considered that relate to the critical importance of train headways in both the modelling of (possibly conflicting) railway operations and moving-block signalling: modelling approach, infrastructure modelling, speed modelling, headway modelling, consideration of rerouting, solution method and objective function.

For CDR, many different modelling approaches have been proposed and applied by many different researchers (Cacchiani et al., 2014).

In the modelling of railway infrastructure, there are two main levels of representation: macroscopic and microscopic. Macroscopic models depict the infrastructure at a lower detail level, often merging parallel tracks into one line and omitting block sections. Microscopic models provide a more detailed representation, including block sections and parallel tracks in junctions and station

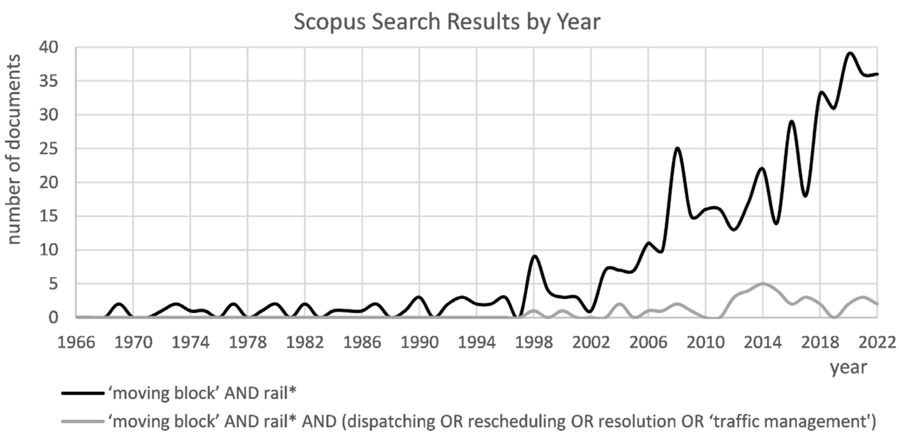


Fig. 2. Available documents in Scopus (2023) by search queries 'moving block' AND rail* and 'moving block' AND rail* AND (dispatching OR rescheduling OR resolution OR 'traffic management').

areas. In the context of CDR, the infrastructure is typically modelled at the microscopic level (Cacchiani et al., 2014). This allows the use of blocking time theory in the modelling of minimum train headways.

While speed modelling is not strictly considered as part of CDR, it plays a crucial role in translating (updated) schedules to train operations (Quaglietta et al., 2016). CDR models can be classified into two types: fixed-speed or variable-speed models. In fixed-speed models, speed variation dynamics resulting from traffic conditions are neglected, and unplanned stops are assumed to lead to immediate stand-stills, without accounting for acceleration and deceleration. In contrast, variable-speed models do incorporate speed dynamics, also in case of unplanned braking.

Headway modelling is heavily influenced by the other aspects. For instance, whether the modelling approach and infrastructure representation level allows for the application of blocking time theory (Cacchiani et al., 2014; Corman and Meng, 2015).

In the context of disturbed railway operations, relevant rescheduling measures are retiming, reordering and local rerouting of trains (Cacchiani et al., 2014). By definition, retiming and reordering measures are considered in CDR models (for mainline railways). The inclusion of rerouting is often explicitly mentioned, e.g., as ‘train routing and scheduling’ (Pellegrini et al., 2014).

The real-time aspect of CDR introduces the need to find a balance between solution quality and computation time. Both commercial solvers and customised heuristics are used for solving CDR problems (Corman and Meng, 2015).

The primary objective of CDR is to minimise the impact of disturbances on the network. Various objective functions can be considered. Typically, they refer to the minimisation of delays (Cacchiani et al., 2014; Corman and Meng, 2015).

3.2. Conflict detection and resolution under fixed-block multi-aspect signalling

Three main classes of CDR models for fixed-block multi-aspect signalling are: alternative graph (AG) models (e.g., D’Ariano et al. (2007a) and Mazzarello and Ottaviani (2007)); disjunctive mixed integer linear programming (MILP) models (e.g., Luan et al. (2018), Pellegrini et al. (2015) and Törnquist and Persson (2007)); and time-indexed MILP models (e.g., Bettinelli et al. (2017), Lusby et al. (2013) and Reynolds et al. (2020)).

Other modelling approaches explored in the literature are model predictive control (Caimi et al., 2012), Monte Carlo tree search (Lövétei et al., 2021), stochastic programming (Meng and Zhou, 2011) and constraint programming (Marlière et al., 2023; Rodriguez, 2007). Despite some promising results of these approaches, they have not been further picked up in the research field yet.

3.2.1. Alternative graph models

In AG-based models, the (re)scheduling of railway operations is considered as a no-wait job shop scheduling problem. For this well-known scheduling problem, an AG formulation is developed by Mascis and Pacciarelli (2002). The approach was first applied to railway operations by D’Ariano et al. (2007a) and Mazzarello and Ottaviani (2007). The railway network is represented as an AG consisting of nodes, fixed arcs and alternative arcs, as illustrated in Fig. 3. The nodes correspond to the entry times of trains into block sections, while the weights of the fixed arcs represent the train running times within the block sections and the dwell times at station platform tracks along a route. In D’Ariano et al. (2007a), the running time reflects the time it takes to traverse the block section at the scheduled speed, whereas Mazzarello and Ottaviani (2007) consider the minimum running time based on the maximum line speed. In both cases, the running time is considered fixed for each block section.

The alternative arcs represent conflicting pairs of operations, where the determination of a train passing order and a minimum train headway is required. This is done through selecting one arc of each pair of alternative arcs. The selection process is modelled by disjunctive constraints, formulated as follows:

$$entry_j - entry_i \geq a_{ij} \vee entry_k - entry_h \geq a_{hk}, \ \forall \text{ pairs of alternative arcs } ((i,j),(h,k)),$$

with $entry$ the block entry time of a train and a the weight of the alternative arc between the nodes corresponding to the entries. In the models presented by D’Ariano et al. (2007a) and Mazzarello and Ottaviani (2007), the alternative arcs connect subsequent block entries of pairs of running trains backwards, as shown in Fig. 3. As a result, the running time over the block is directly included into the headway time. The rest of the headway is determined by the weight of the alternative arc. D’Ariano et al. (2007a) include the time between the block exit of the head and of the tail of the train, i.e., the clearing time, supplemented with a default value representing the route setup and release time. In Mazzarello and Ottaviani (2007), the weight consists of a fixed and a variable term corresponding to blocking time components. The latter term depends on the train length and speed, so it includes the clearing time. Note that the time required to traverse the number of sections corresponding to the number of signalling aspects, i.e., the approach

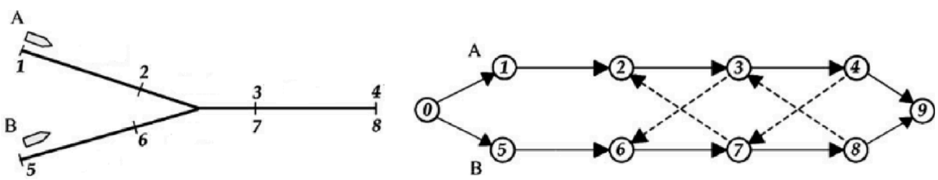


Fig. 3. A simple network (left) and the associated alternative graph (right). The network features trains A and B, and their block section entries 1 to 8. The alternative graph features a source and sink node, 0 and 9 respectively, nodes corresponding to the block entries of trains A and B, fixed arcs (solid) between block entries on a train route, and pairs of alternative arcs (dashed) between conflicting operations. Adapted from Mazzarello and Ottaviani (2007).

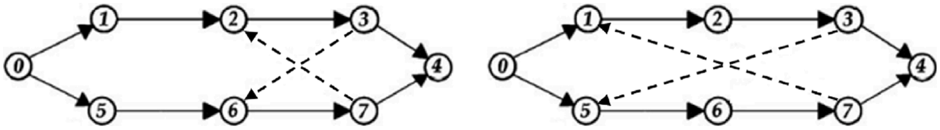


Fig. 4. An alternative graph that does not include approach time in the headway (left), and the same alternative graph that does (right) for three-aspect signalling.

time, is not considered in the headway calculation in these models. In a subsequent version of the AG model, Corman et al. (2009) address the omission by incorporating the approach time in the headway calculation by connecting alternative arcs between two nodes corresponding to block entries that are the number of signal aspects apart. Fig. 4 illustrates the difference in the construction of the alternative graph.

AG-based CDR models require the timetable and the current delays as input, next to the AG representation of the infrastructure. The model objective is to minimise the maximum secondary delay, i.e., the delay trains face due to interaction with trains suffering initial delays. Compliant with this objective, a graph-based solution algorithm is employed to find a selection of arcs that contains source-to-sink paths for all trains, while avoiding the presence of any positive directed cycle and minimising the weight of the critical path. An example of a heuristic solution algorithm is the truncated branch-and-bound algorithm proposed in D'Ariano et al. (2007a). To extend the model to consider objective functions beyond maximum secondary delay, it is necessary to transform the model into a MILP formulation. In that case, the alternative graph serves as an intermediate model to specify the problem and to derive the disjunctive MILP formulation.

The AG approach is particularly suitable for retiming and reordering, though efforts are made to include rerouting (D'Ariano et al., 2008; Mazzarello and Ottaviani, 2007). Other extensions of D'Ariano et al. (2007a) consider the inclusion of speed control. D'Ariano et al. (2007b) propose an iterative approach similar to the method used in Mazzarello and Ottaviani (2007). This iterative approach involves an overarching traffic control algorithm that iterates between the fixed-speed CDR model and an external speed profile optimisation model. Corman et al. (2009) apply the so-called green wave approach, which assumes trains to only stop and wait at stations.

3.2.2. Disjunctive mixed integer linear programming models

In a disjunctive MILP CDR model, the railway operations are described by decision variables and mixed integer linear constraints, complemented with a linear objective function (e.g., Luan et al. (2018), Pellegrini et al. (2014) and Törnquist and Persson (2007)).

The decision variables indicate which train passes which track section at what time. The (re)scheduling decision variables relate to: timing, i.e., at what time does a train pass; ordering, i.e., which one of a pair of trains enters first; and routing, i.e., which route does a train take (Luan et al., 2018; Pellegrini et al., 2014; Törnquist and Persson, 2007). Time variables are typically continuous, indicating entry and exit times of trains on track sections. Binary variables decide on the order in which trains enter sections, as well as on the route trains are assigned.

To obtain linear constraints, disjunctive MILP models rely on the general big-M linearisation method (Bazaraa et al., 2008). The drawback of this method is that it results in a relatively weak linearisation, which makes it hard to solve the model to optimality (Reynolds et al., 2020). For this reason, improved solution methods are proposed in follow-up works of both Pellegrini et al. (2014) and Törnquist and Persson (2007) (Pellegrini et al., 2015, 2019; Törnquist, 2012).

The big-M constraints are used to describe the capacity. More specifically, they are the disjunctive constraints that ensure that possibly conflicting operations (or events) are separated in time. A generic pair of big-M constraints for the CDR problem is represented by the following equations:

$$start_{t',s} - end_{t,s} \geq \Delta_{t,t',s} + M(1 - order_{t,t',s}),\ \forall t, t', s,$$

$$start_{t,s} - end_{t',s} \geq \Delta_{t',t,s} + M order_{t,t',s},\ \forall t, t', s,$$

with t, t' trains, s a track section, $start$ and end the start and ending time, respectively, Δ the minimum separation time, M a sufficiently large number (the big M), and $order$ a binary ordering variable. Depending on whether train t goes before train t' on section s (if $order_{t,t',s} = 1$), a certain minimum separation time ($\Delta_{t,t',s}$ or $\Delta_{t',t,s}$) is to be respected between the two relevant operations, e.g., the occupation of section s by trains t and t'.

In Pellegrini et al. (2014), operations are separated by determining the order in which a track-clear detection section is blocked by a train according to blocking time theory. All blocking time components are included. Based on the number of aspects in the considered multi-aspect signalling system, a reference section is determined for each track-clear detection section. The blocking starting time is linked to the entry time of the reference section. The running time from the reference section to the considered track detection section is based on minimum running times, possibly with an additional component in case of a decision to run slower or to perform an unplanned stop. Only in absence of conflicts and at planned stops the train times are strictly aligned with the scheduled times.

Luan et al. (2018) also consider the full blocking times, but on the level of block sections. The blocking times dynamically depend on train speeds. A mixed integer non-linear programming model is converted into a MILP by approximating the non-linear terms with piece-wise affine functions. In this model, speed is included through predetermined speed profile alternatives per train-block section pair.

In Törnquist and Persson (2007), both station areas and open lines between them are considered in terms of sets of parallel block sections. Being a fixed-speed model, the minimum running time over a block section is considered as the duration of the activity associated with a pair of train and block section. For the separation of events associated with different trains and one block section, however, a fixed time is used, independent of the specific (parallel) block section or train.

Lamorgese and Mannino (2015) address the limitations of big-M formulations by introducing an exact macro/micro decomposition of the CDR problem, inspired by the Benders' decomposition approach. In the master problem, stations are macroscopically represented by nodes. On this macroscopic level, the CDR model is solved exactly, proposing tentative arrival and departure times. In each subproblem, a station is represented microscopically. For each train, a route through the station is searched, to fit the tentative timings. If the subproblem is infeasible, the violations are added to the master problem in an iterative process. The decomposition approach significantly reduces the number of big-M constraints and it has been applied in pilot tests in practice. In follow-up works, the approach is enhanced to speed up the solution process (Lamorgese et al., 2016) and the decomposition is further strengthened by removing the big-M constraints from the master problem (Lamorgese and Mannino, 2019).

The reason to translate an AG model into a MILP, is the flexibility in terms of objective functions (Samà et al., 2015). As long as the objective is formulated as a linear function, all types of objectives can be included into the optimisation model such as a weighted sum of delays, see, e.g., Luan et al. (2018), Pellegrini et al. (2014, 2015) and Törnquist and Persson (2007).

3.2.3. Time-indexed mixed integer linear programming models

Time-indexed MILP models are considered as an alternative to disjunctive MILP formulations as they are known for their strong linearisation and good approximations for scheduling problems (Van den Akker et al., 2000). The CDR problem is formulated as a time-indexed MILP by Bettinelli et al. (2017), Lusby et al. (2013) and Reynolds et al. (2020).

Time-indexed MILP models consider a uniform time discretisation on top of the fixed-block space discretisation. The model formulation builds on resources which correspond to pairs of a time unit and a track part (Bettinelli et al., 2017; Lusby et al., 2013; Reynolds et al., 2020). In Fig. 5, a time-space graph is constructed in which nodes correspond to time-space resources and the arcs to parts of possible train routes indicated by sink-source paths. In general, the time-indexed approach is well-suited for the consideration of rerouting (Bettinelli et al., 2017; Lusby et al., 2013; Reynolds et al., 2020; Zwaneveld et al., 2001).

Typically, binary variables indicate the use of a time-space resource, and capacity is modelled using set packing constraints (Bettinelli et al., 2017; Lusby et al., 2013; Reynolds et al., 2020), such as the following:

$$\sum_{t \in T} x_t^r \leq 1, \ \forall r,$$

with T the set of trains and r a resource. In this generic set packing constraint, the number of trains t assigned to time-space resource r ($x_t^r = 1$) is restricted to at most one. With this, the occurrence of track conflicts is excluded. Resources can also be defined in such a way that they have a capacity greater than 1. For example, when the infrastructure is represented at a macroscopic level in which resources correspond to, e.g., parallel tracks or platforms (Bettinelli et al., 2017).

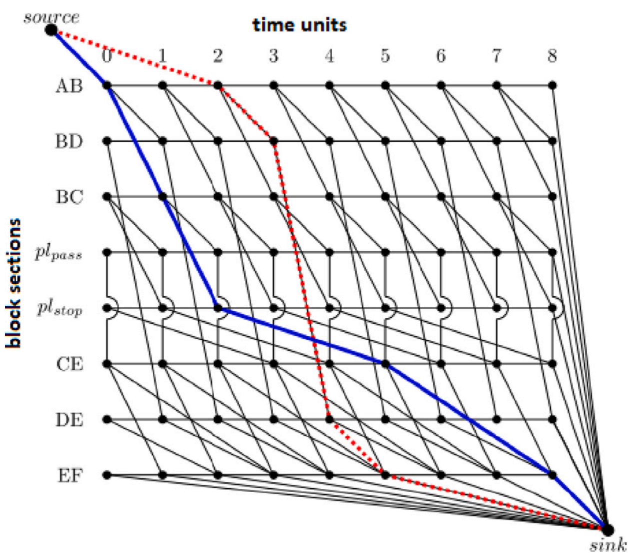


Fig. 5. Example time-space graph with two train routes, i.e., source-sink paths. Adapted from Reynolds et al. (2020).

Time-indexed modelling requires to express the running and headway times with a precision of the time unit. Reynolds et al. (2020) estimate a 'near minimum' running time based on historical real-life data and consider the minimum headway as this running time plus an additional separation time, e.g., 30 s, based on blocking time theory.

The main drawback of the time-indexed approach is the model size (Reynolds et al., 2020; Van den Akker et al., 2000). This can be a restriction in the real-time application of the approach (Reynolds et al., 2020). Several solution methods are proposed to enable the use of the approach for CDR. A validated combination is the use of a heuristic method and a model decomposition (Lusby et al., 2013; Reynolds et al., 2020), while Bettinelli et al. (2017) propose a parallel algorithm based on an iterative greedy approach. The time-indexed formulation allows for various objective functions.

Time-indexed models can consider speed beyond fixed-speed assumptions. Lusby et al. (2013) and Reynolds and Maher (2022) approximate speed profiles by considering the alternatives of running at constant speed and accelerating/decelerating.

3.3. Conflict detection and resolution under fixed-block distance-to-go signalling

The work of Gonzalez et al. (2010) and Mera et al. (2016) on fixed-block distance-to-go signalling nicely illustrates the role of speed on train headways and line capacity at a microscopic level. For the fixed-block distance-to-go signalling system implemented on the Madrid metro line, Gonzalez et al. (2010) describe a line capacity optimisation algorithm. Speed profiles are calculated for trains to come from the scheduled speed (40–90 km/h) to a standstill at a stopping point, i.e., at a station or at the entry of an occupied section. These speed profiles are input for an investigation of track section lengths (typically 100–200 m) and the effects on the capacity parameters of train headway and running times. In Mera et al. (2016), the metro line capacity algorithm is enhanced with the introduction of speed signalling. Based on the number of free sections ahead, a target speed (code) is communicated to the train. This is done per section, obtaining a discrete braking curve.

Literature on CDR under fixed-block distance-to-go signalling is available on the Chinese signalling system, i.e., Level 3 of the Chinese Train Control System (CTCS-3), also called the quasi-moving block signalling system in Chinese literature. Note that the CTCS-3 system is designed for the Chinese high-speed (up to 300 km/h) railway lines. In accordance with the distance-to-go system characteristics (as described in Section 2), CDR models for CTCS-3 consider speed-dependent running times (Liu et al., 2021; Xu et al., 2017). The literature proposes different speed modelling approaches to allow a direct modelling relation between speed and train headway. Fig. 6 illustrates this relation, showing that the speed (level) is translated one-on-one to the number of free block sections needed as minimum headway. The concept of a discrete set of speed levels is included in Xu et al. (2017) and its extension (Xu et al., 2021) to allow speed selection alongside the rescheduling decisions of retiming and reordering (and rerouting). The speed selection and retiming decisions are directly linked via running time constraints. Liu et al. (2021) also consider speed levels, but only in the initial solving phase. In the next step, continuous speed profiles matching the solution are selected from a predefined set. If no speed profile can be found, another iteration of the CDR model is required. Also in Xu et al. (2017), the model is first simplified. To obtain an initial solution for the whole problem, some speed-related variables are fixed according to the scheduled situation, so assuming no effect by the considered disturbances.

The presented CDR models for CTCS-3 rely on a MILP formulation (Liu et al., 2021; Xu et al., 2017). Xu et al. (2017) build upon the AG model introduced in D'Ariano et al. (2007a), resulting in a AG-based MILP. The original model is extended by introducing pairs of alternative arcs related to a possible track conflict for each speed level. For a feasible solution, still only one alternative arc related to this conflict can be selected. As objective function, the total final secondary delay is considered in Xu et al. (2017), in its extension (Xu et al., 2021) and in Liu et al. (2021).

3.4. Conflict detection and resolution under moving-block signalling

The detection and resolution of conflicts under moving-block signalling is barely addressed in the literature. Pochet et al. (2016, 2017) consider CDR for mixed CBTC and non-CBTC operations on suburban railways. A model predictive control approach is presented in a rolling horizon framework, with a genetic algorithm optimising punctuality and regularity indicators. An example of the latter is the deviation from the scheduled headway intervals. Overall, there is no notion of speed. In case of delays up to three minutes, the model can adjust running times and/or dwell times, considering fixed arrival and departure headways. The approach is incorporated in a microscopic simulation tool of the French train operation SNCF. In the practical experiments described, few scenarios in terms of disturbances and networks are considered, and the change of train orders is allowed. Meunier et al. (2023) follow up on the work of Pochet et al. (2017), introducing inter-station running times dependent on traffic conditions, considering

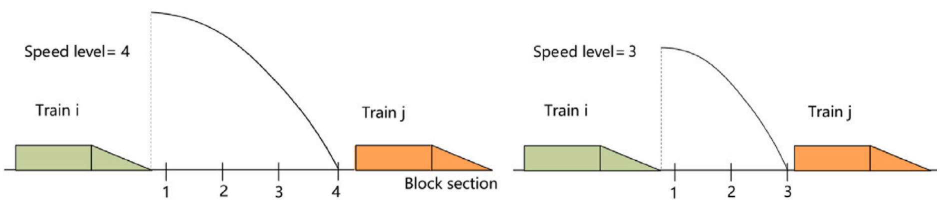


Fig. 6. Example of relation between speed level and train headway in terms of block sections. Adapted from Liu et al. (2021).

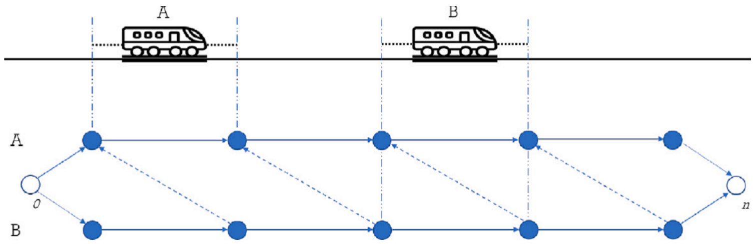


Fig. 7. Alternative graph for two trains on a single track under moving-block signalling. The virtual nodes (blue) represent the discretisation of the line into train lengths. Adapted from Janssens (2022).

disturbances. In their disjunctive MILP model, the running times on single-track lines are related to departure headways based on empirical simulation data.

Janssens (2022) presents a model concept based on the fixed-block CDR model described in D'Ariano et al. (2007a). In the development of a moving-block CDR model, the possibility is explored to model station areas and the open line, i.e., the track between stations, distinctively. The fixed block sections in the station areas are modelled as in the AG model described in D'Ariano et al. (2007a). For the moving block sections, the model is extended with virtual nodes and corresponding fixed arcs and pairs of alternative arcs. These model components are visualised in Fig. 7, showing an alternative graph representation of two trains on a single track under moving-block signalling. Note that, in the figure, the alternative arcs are already selected according to train ordering. The virtual nodes correspond to grid points resulting from a (fixed) discretisation of the line based on train lengths. The virtual nodes are connected by fixed arcs whose weights correspond to minimum train clearing times, i.e., the time it takes the train to traverse its length when running at maximum speed. The alternative arcs connect subsequent nodes of two trains, indicating the order of the trains and the minimum headway between them. The moving-block minimum headway is derived from the fixed-block blocking times, leaving out the (fixed) block traversing time. For a small case study, the approach provides a valid result within reasonable time using a commercial solver, with the objective to minimise the maximum secondary delay. However, the model size significantly increases because of the additional components. Moreover, the model does not consider rerouting and adopts the fixed-speed assumption from D'Ariano et al. (2007a).

The general moving-block literature showcases the relevance of including speed modelling beyond fixed-speed assumptions, with the speed-dependency of the moving-block headway as a main point of focus (Gao et al., 2020; Liu, 2016; Wang et al., 2014; Xu et al., 2014). Gao et al. (2020) point out that under moving-block signalling, the minimum headway is negatively correlated with the running time, which corresponds to the positive correlation between braking distance and speed: the lower the speed, the shorter the braking distance and therefore the headway. The balance between the capacity, as result of the minimum headway, and effective operations, related to the running time or speed, is addressed by, e.g., Liu (2016) and Xu et al. (2014). Liu (2016) proposes a dynamic system approach to address the dual control problem of finding the optimal speed and a safe following distance. Xu et al. (2014) dynamically optimise the balance between running times and headways, e.g., by introducing speed limits to reduce braking distances and, hence, headways.

Dynamic system approaches are well-represented in moving-block literature (e.g., Gao et al. (2020), Liu (2016), Wang et al. (2014) and Xu et al. (2014)). Dynamic systems are characterised by the use of differential equations, and they are widely applied in optimal train control and train trajectory models (Gao et al., 2020; Liu, 2016; Wang and Goverde, 2019). In the optimal control models of Gao et al. (2020) and Liu (2016), the focus is on speed profile optimisation depending on train dynamics and constrained by track characteristics. This train trajectory optimisation is done for single trains. The modelling of moving-block headways would require cross-terms between the state trajectories of successive trains resulting in complex multi-train trajectory optimisation models (Wang and Goverde, 2019). Wang and Goverde (2019) apply multi-train trajectory optimisation in the context of fixed-block timetabling, considering timing, ordering and routing options.

Wang et al. (2014) implement the problem of train separation within the context of trajectory optimisation using the pseudospectral method and by formulating a MILP, both solved with commercial solvers. The two methods show similar performance with respect to the balance between solution quality and computation time, even though the linearisation necessary for the MILP results in a discrete-space model. (Disjunctive) MILP modelling is also used in timetable optimisation for moving-block operations (Schlechte et al., 2022). Train orders and routes are scheduled with headways based on emergency braking curves consistent with a predefined set of possible speeds.

Besides the speed-dependency and the infrastructure modelling, also moving-block as a system needs to be considered for the modelling of moving-block headways. In the context of ERTMS/ETCS Level 3, Büker et al. (2019) incorporate the system characteristics such as the bidirectional radio communication and the onboard train positioning and integrity into the blocking time components.

3.5. Summary

Table 1 provides a summary of the reviewed papers specifically on CDR modelling. The table classifies a representative selection of papers for fixed-block multi-aspect, fixed-block distance-to-go and moving-block signalling in terms of the modelling aspects

Table 1
Overview of literature on conflict detection and resolution under fixed-block multi-aspect, fixed-block distance-to-go and moving-block signalling.

Publication	Modelling approach	Infra modelling	Speed modelling	Headway modelling	Re-routing	Solution method	Objective function
Fixed-block multi-aspect signalling							
Bettinelli et al. (2017)	TI-MILP	BS	Fixed	Blocking	Yes	Iterated greedy	Total arrival and departure
Corman et al. (2009)	AG	BS	Green wave	Blocking	No	Heuristic	Max secondary
D'Ariano et al. (2007a)	AG	BS	Fixed	Occupation	No	B&B	Max secondary
D'Ariano et al. (2007b)	AG	BS	Iterative	Occupation	No	B&B	Max secondary
D'Ariano et al. (2008)	AG	BS	Fixed	Occupation	Yes	B&B	Max secondary
Lamorgese and Mannino (2015)	D-MILP	BS	Fixed	Default	Yes	Macro/micro decomposition	Total arrival delay cost
Luan et al. (2018)	D-MILP	BS	Levels	Blocking	No	Two-level	Total mean absolute
Lusby et al. (2013)	TI-MILP	TDS	Variable	Occupation	Yes	B&P	Weighted total
Mazzarello and Ottaviani (2007)	AG	BS	Iterative	Occupation	Yes	Two-level	Max secondary
Pellegrini et al. (2014)	D-MILP	TDS	Fixed	Blocking	Yes	Solver	Max/total secondary
Pellegrini et al. (2015)	D-MILP	TDS	Fixed	Blocking	Yes	Heuristic	(Weighted) total final
Reynolds et al. (2020)	TI-MILP	TDS	Fixed	Blocking	Yes	B&P	Custom utility
Reynolds and Maher (2022)	TI-MILP	TDS	Levels	Blocking	Yes	B&P	Custom utility
Samà et al. (2015)	AG-MILP	BS	Fixed	Occupation	No	Solver	Multi-criteria
Törnquist and Persson (2007)	D-MILP	BS	Fixed	Default	Yes	Solver	Total/weighted final
Törnquist (2012)	D-MILP	BS	Fixed	Default	Yes	Greedy depth-first	Total final
Fixed-block distance-to-go signalling							
Liu et al. (2021)	MILP	BS	Iterative	Default	No	Bi-level	Total final secondary
Xu et al. (2017)	AG-MILP	BS	Levels	Default	No	Two-step	Total final secondary
Xu et al. (2021)	AG-MILP	BS	Levels	Default	Yes	Two-step	Total final secondary
Moving-block signalling							
Janssens (2022)	AG	(M)BS	Fixed	Occupation	No	Solver	Max secondary
Meunier et al. (2023)	AG-MILP	Micro	Variable	Blocking	No	Solver	Punctuality/ regularity
Pochet et al. (2016)	MPC	Micro	Fixed	Default	No	Genetic algorithm	Punctuality/ regularity

AG: alternative graph, D-MILP: disjunctive mixed integer linear programming, TI-MILP: time-indexed mixed integer linear programming, AG-MILP: AG-based mixed integer linear programming, MPC: model predictive control, (M)BS: (moving) block section, TDS: track-clear detection section, B&B: branch-and-bound, B&P: branch-and-price, Max: maximum.

introduced in Section 3.1. Publications are deemed representative if appointed as such in the general literature, e.g., in the review papers of Cacchiani et al. (2014) and Corman and Meng (2015), and/or if they are unique in terms of the considered modelling aspects. In this way, Table 1 provides a concise overview of the type of models existing for CDR. In case of multiple publications by the same authors that do not differ much in terms of aspects, the first paper is included.

It emerges that the most commonly applied modelling approaches are AG, disjunctive MILP (D-MILP) and time-indexed MILP (TI-MILP), as also found in the literature (Reynolds et al., 2020). Note that AG-based MILP (AG-MILP) is an example of a disjunctive MILP approach. An exception is the model predictive control (MPC) model applied by Pochet et al. (2016) in a suburban context. Each approach has its strengths and weaknesses in terms of, e.g., inclusion of speed modelling, suitability for rerouting, computational performance and flexibility in objective functions. For example, the restriction to minimise the maximum (secondary) delay as objective for AG models.

Practically all CDR models describe the infrastructure at a microscopic level. Within the microscopic level, a distinction can be made based on whether or not the block sections (BS) are represented as a sequence of track-clear detection sections (TDS). The main advantage of considering the infrastructure in terms of TDSs is the possibility to model sectional route release in interlocking areas, which can provide capacity benefits (Hansen and Pachl, 2014). Nevertheless, the TDSs are not often modelled.

Conventionally, CDR models are fixed-speed models. While fixed-speed models capture most essential characteristics and provide highly practical solutions in terms of rescheduling decisions (Pellegrini et al., 2014), they may lead to infeasible solutions in terms of exact timings due to their reliance on minimum or scheduled running times in case of disturbances (Reynolds and Maher, 2022). As alternative to the fixed-speed assumptions, the modelling of speed levels is introduced. This allows to better capture the speed-dependency of the headway and use the speed variability to influence the headway. Specifically for the modelling of fixed-block distance-to-go operations, there is the tendency to move away from fixed-speed assumptions.

The modelling of train headways is clearly interwoven with other modelling aspects such as the modelling approach and the speed modelling. Independent of other aspects, the options are to determine minimum train headways based on the physical occupation or the longer blocking times of trains, or to fix them at default values. This generally follows from the foreseen model application. For instance, the default (arrival) headway between homogeneous suburban trains (Pochet et al., 2016).

The rescheduling measures of retiming and reordering are always considered, even for the CBTC application on suburban lines (Pochet et al., 2016). Rerouting decisions, however, are only included in part of the models. If rerouting is considered, it is regularly in follow-up work. This is, for example, the case with the rerouting model for fixed-block distance-to-go, i.e., Xu et al. (2021). Notable is the lack of rerouting in moving-block applications. The general reason is twofold. Firstly, in practice, disturbances resulting in delays of a couple of minutes are typically resolved with retiming and reordering alone. Secondly, incorporating rerouting presents modelling challenges, primarily related to computational performance (Pellegrini et al., 2015).

The literature on CDR is dominated by heuristic solution methods (Cacchiani et al., 2014; Corman and Meng, 2015). Indeed, various heuristics method to solve CDR problems are proposed and applied. Even the commercial solvers typically rely on heuristics when taking into account a limited computation time. The two favoured streams of heuristic methods are tailored branch-and-bound (B&B), among which truncated branch-and-price (B&P) algorithms, and multi-step optimisation methods.

The objective of minimising the impact of disturbances on the network is generally considered by minimising the propagation of train delays. In line with that, practically all model objectives are a type of delay minimisation. While the AG models are limited to the minimisation of the maximum (secondary) delay, the MILP models include various (linear) objective functions, e.g., total mean, weighted total or final secondary delay. The exception to delay minimisation is the maximisation of punctuality and regularity in a suburban context by Meunier et al. (2023) and Pochet et al. (2016). Note that not only are different objectives not compatible with all modelling approaches, but also that different objectives can lead to different (optimal) solutions for the same model.

4. Research gaps and challenges

This section is dedicated to the identification of research gaps and the related challenges regarding the modelling of CDR under moving-block signalling. The literature review in Section 3 shows that the majority of existing CDR models refers to fixed-block signalling systems whereas only little research addresses CDR under moving-block signalling, leaving a clear gap to fill in future research.

The main characteristic for the modelling of railway traffic under moving-block signalling is the minimum train headway based on absolute braking distance. For minimum headways, default values can be determined per section, per train and/or per speed level. The alternative is to determine headways using blocking time theory such that they rely on the time a track part is assigned to a train. The blocking time includes the occupation time, during which a train is physically present on the considered block section, plus the other blocking time components including in particular the approach time corresponding to the braking behaviour before the block section. Hence, to model signalling conflicts, and not only physical conflicts, minimum headways should include all blocking time components.

Overall, the gap to bridge in the research on CDR under moving-block signalling is the development of a model that describes railway traffic operating at moving-block characteristic headways. This gap gives rise to two main challenges to be faced in future research.

The first challenge is to adapt CDR models such that the EoA can be assigned to any point on the open line instead of only at block entries as customary under fixed-block signalling. This yields the need of a continuous modelling of the infrastructure rather than the conventional discretisation into track sections. A continuous representation of the infrastructure would also accommodate IPs anywhere on the track, including in station areas where fixed sections remain in place to protect switches in interlocking areas.

The second challenge relates to the modelling of the headways as a continuous function of train speed. While in fixed-block multi-aspect signalling system the role of speed is limited, in signalling systems with distance-to-go ATP the impact of speed on rescheduling decisions increases. However, the extent of the impact of modelling a fully continuous speed-headway relation is unclear. The effect on rescheduling decisions, the main output of the model, may be limited; these decisions are generally taken in interlocking areas where speed limits and protective blocks are in place, possibly diminishing the effects of continuous speed-headway relation.

5. Research agenda

A research agenda is drafted based on the gaps and challenges identified in Section 4. Four main research steps towards CDR under moving-block signalling are: (1) an analysis of infrastructure modelling options, (2) an analysis of speed modelling options, (3) the identification of modelling options for moving-block CDR, and (4) an assessment of modelling approaches for the identified modelling options.

Fig. 8 presents the research steps and their connections. In the following, the research steps are addressed to provide stepping stones in the research on CDR under moving-block signalling.

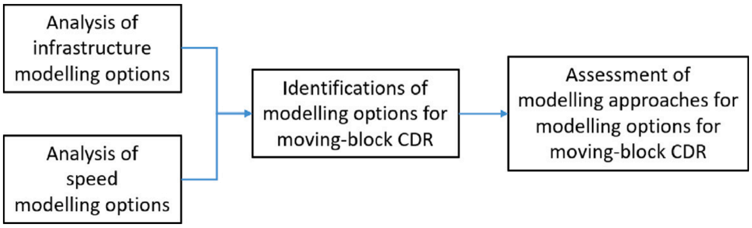
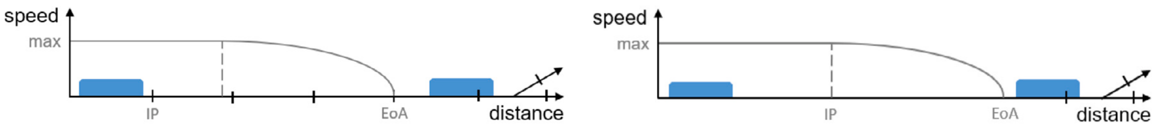


Fig. 8. Flow chart of the proposed research agenda.



(a) Discrete infrastructure. (b) Continuous infrastructure.

Fig. 9. Illustrative examples of infrastructure modelling options: (a) discrete and (b) continuous. The end of authority (EoA) and the brake indication point (IP) for maximum speed are indicated.

5.1. Infrastructure modelling options

An analysis of infrastructure modelling options can help to assess the impact of infrastructure modelling on CDR models in terms of, e.g., delay propagation minimisation, rescheduling decisions and computational performance. Various modelling options, discrete and continuous, can be considered and compared.

In case of a discrete representation of the infrastructure, the open line is divided into sections by a discretisation grid. Because of the infrastructure discretisation, an MA up to the position of the tail of the preceding train cannot be modelled. Instead, the EoA lies at the grid point that is released last by the preceding train. Similar to the EoA, the IP, i.e., where an approaching train receives the indication to start braking to stop at the EoA, has to be related to a discrete grid point. The effect of a discrete representation of the infrastructure on the EoA and the IP is illustrated in Fig. 9(a). The impact of discretising the infrastructure is expected to strongly depend on the type and fineness of the discretisation.

There are roughly two types of discretisation that can be applied: equidistant discretisation or variable discretisation. With equidistant discretisation, the interval length between grid points is set equal to an overall fixed value. This value can be, e.g., derived from train lengths or from the inherent moving-block system discretisation due to discontinuous communication. Based on this, 200 metres can be considered as starting point in the search for an appropriate discretisation interval (Furness et al., 2017). With variable discretisation, the interval lengths can vary over the infrastructure. For example, shorter intervals can be considered closer to critical points such as in station areas, or interval lengths can be related to the (maximum) track speed (the lower, the shorter).

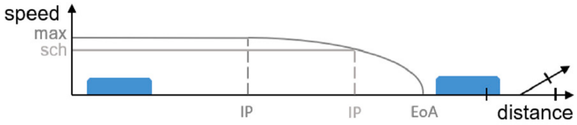
The consideration of the open line in continuous space directly affects the EoA and the IP. In continuous space, the EoA can be set at a safety margin behind the tail of the preceding train. The approaching train can then run freely up to the 'actual' braking point, i.e., where the train should start braking. Fig. 9(b) illustrates the effect of a continuous representation of the infrastructure on the EoA and the IP.

5.2. Speed modelling options

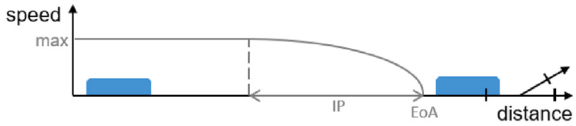
As for infrastructure modelling, an analysis of the speed modelling options can help to assess their impact on CDR models in term of, e.g., delay propagation minimisation, rescheduling decisions and computational performance. Various options to incorporate speed dynamics, discrete and continuous, can be considered and compared. Key is to keep in mind the possibly limited effect due to speed limits and protection zones/blocks around decision points at switches.

With speed modelled discretely, braking distances can be (pre)calculated for each of the discrete speed alternatives. The speed alternatives can correspond to, e.g., speed levels or speed profiles. Speed levels can be defined to align with different track speed (limits), while speed profiles can be derived from the scheduled and/or minimum running times. In both cases, the impact of discretising speed strongly depends on the fineness of the discretisation, i.e., the number of alternatives. Fig. 10(a) relates the discrete modelling of speed to the EoA and the IP; the EoA is independent of speed, but the IP can be different for each speed alternative.

With continuous modelling of speed, the minimum headway can be better approximated by computing the braking distance based on the actual speed. Hence, the IP cannot be precomputed. As illustrated in Fig. 10(b), the IP can lie anywhere between the start of the braking curve when running at maximum speed and the EoA.

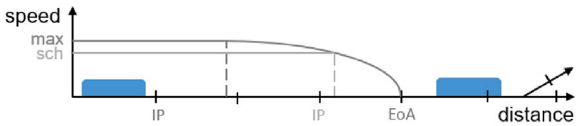


(a) Discrete speed.

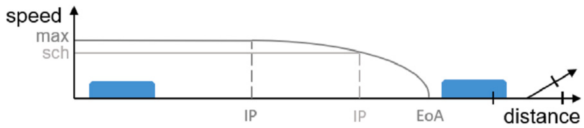


(b) Continuous speed.

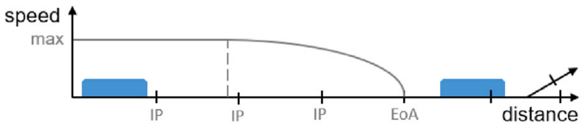
Fig. 10. Illustrative examples of speed modelling options: (a) discrete and (b) continuous. The end of authority (EoA) and the possible brake indication points (IP) are indicated based on maximum (max) and scheduled (sch) speed.



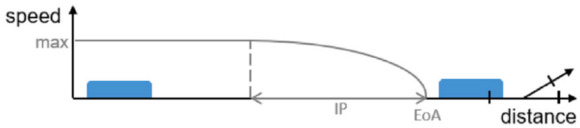
(a) Discrete infrastructure and discrete speed.



(c) Continuous infrastructure and discrete speed.



(b) Discrete infrastructure and continuous speed.



(d) Continuous infrastructure and continuous speed.

Fig. 11. Illustrative examples of moving-block modelling options in terms of the modelling of the infrastructure (discrete on the left versus continuous on the right) and speed (discrete on the top versus continuous on the bottom). The end of authority (EoA) and the possible brake indication points (IP) are indicated.

5.3. Modelling options for moving-block conflict detection and resolution

Following from the infrastructure and speed modelling options, the modelling options to consider in the development of a CDR model for moving-block signalling are: (1) discrete infrastructure and discrete speed, (2) discrete infrastructure and continuous speed, (3) continuous infrastructure and discrete speed, and (4) continuous infrastructure and continuous speed. The four options are shortly described in the following and illustrated in Fig. 11.

Fig. 11(a) features an equidistant infrastructure discretisation and two speed profile alternatives (maximum and scheduled), to exemplify the modelling option of discrete infrastructure and discrete speed. Indeed, the EoA lies at the grid point that is last released by the preceding train and the speed-dependent IP at the grid point before the start of the relevant braking curve.

Fig. 11(b) features an equidistant infrastructure discretisation and continuous speed, to exemplify the modelling option of discrete infrastructure and continuous speed. Indeed, the EoA lies at the grid point that is last released by the preceding train and the speed-dependent IP at the grid point before the start of the braking curve of the actual train speed.

Fig. 11(c) features a continuous infrastructure and two speed profile alternatives (maximum and scheduled), to exemplify the modelling option of continuous infrastructure and discrete speed. Indeed, the EoA lies at a safety margin behind the tail of the preceding train and the speed-dependent IP at the start of the relevant braking curve.

Fig. 11(d) exemplifies the fourth modelling option by featuring a continuous infrastructure and continuous speed. Indeed, the EoA lies at a safety margin behind the tail of the preceding train and the speed-dependent IP at the start of the braking curve of the actual train speed, which can be anywhere on the track between the IP related to the maximum speed profile and the EoA.

The modelling options can be further specified based on the first two research steps. Within the underlying infrastructure and speed modelling options, the possibilities may be narrowed down and/or a candidate possibility may be put forward. Crucial for any option is the balance between computational effort and solution quality, both in terms of the approximation of the moving-block headway and the optimality of the rescheduling decisions. This balance does not only depend on the modelling options but also on the modelling approach. Hence, modelling approaches are analysed with respect to the presented modelling options in Section 5.4.

5.4. Modelling approaches for moving-block conflict detection and resolution

For the four modelling options, the CDR modelling approaches of AG, disjunctive MILP and time-indexed MILP, as well as the moving-block dynamic system approach are analysed. The comparative analysis is based on the advantages and disadvantages of the approaches, which are summarised in Table 2.

The three CDR approaches inherently rely on a discretisation of the infrastructure for the modelling of decisions. In all three cases, a finer discretisation grid will increase the model size significantly, whether in number of nodes and arcs (AG) or variables (MILP). For time-indexed MILP, the model size is already a restriction as it considers time-space resources as a consequence of its inherent time discretisation. In this also lies its strength, because it facilitates the strong model linearisation based on set packing constraints for the modelling of capacity. The set packing constraints require resources with a fixed capacity, which can restrict the

Table 2
Advantages, disadvantages and proposed moving-block modelling options of modelling approaches.

Modelling approach	Advantages (+) / Disadvantages (−)	Proposed modelling option(s)
Alternative graph	(+) Continuous time (+) Model extension for moving block (−) Iterative speed modelling (−) Rerouting requires meta-heuristic (−) Decision points from infra discretisation (−) Principally headway based on occupation (−) Single objective function	Discrete infrastructure and discrete speed Discrete infrastructure and continuous speed
Disjunctive MILP	(+) Continuous time (+) Speed level applications (+) Suitable for rerouting (+) Headway based on blocking time (+) General objective function (−) Decision points from infra discretisation (−) Weak linearisation due to big-M (−) Linear objective and constraints	Discrete infrastructure and discrete speed Discrete infrastructure and continuous speed
Time-indexed MILP	(+) Strong linearisation (+) Potential for speed modelling (+) Suitable for rerouting (+) General objective function (−) Imposed time discretisation (−) Model size (−) Set packing capacity constraint	Discrete infrastructure and discrete speed
Dynamic system	(+) Applied to moving-block operations (+) Continuous infrastructure (+) Dynamic speed-headway relation (+) Applied to timetabling (−) Not used in rescheduling (−) Complex multi-train models	Continuous infrastructure and discrete speed Continuous infrastructure and continuous speed

model in its space and time discretisation. As the headway is expressed in space and/or time units, the headway is also limited in its flexibility.

The time-continuous disjunctive MILP does not have this restriction as its linearisation is based on the big-M method. The big-M constraints allow for a (more) flexible headway definition. However, it comes at a price: a weak linearisation that makes it harder to solve the model to optimality. The necessary linearisation for MILP models does not only apply to the model constraints, but also to the objective function. Besides that, MILP objective functions are very general, in contrary to the single possible objective in AG models.

Two other aspects in which MILP models seem to outperform AG models are the suitability for rerouting and the potential to include speed dynamics. Both disjunctive MILP and time-indexed MILP models have considered speed alternatives and initially include rerouting, while the inclusion of speed modelling and rerouting in AG models requires external speed profile optimisation models and meta-heuristics. On a different note, only the AG approach has so far been applied to moving-block CDR.

Despite the difference between the approaches, they all build on the important aspect of headway modelling. The MILP models derive headways from train blocking times, while initial AG models only consider physical occupation times. This is not necessarily a restriction, because the missing approach times can be included by a slight reformulation. Anyway, modelling moving-block headways will require a reconfiguration, certainly in an attempt to include speed beyond fixed-speed assumptions or speed levels. For the AG approach, continuous speed would require to rebuild the graph after every changed speed decision because of the graph structure in which alternative arcs define headways by connecting infrastructure points. In theory, MILP models can include continuous speed variables. However, in time-indexed MILP the effects would be downscaled because of the expression of time in fixed units. Moreover, continuous speed would mean that running and clearing times will become variable as they can no longer be precalculated. With this, the requirement of linear constraints complicates the model formulation. For example in combination with route decisions because whether a route is used or not co-determines the running and clearing times.

In the end, only the dynamic system approach has proven to be able to describe the dynamic relation between speed and headway. The approach has been applied to general moving-block models, but not to CDR. Retiming decisions are included through the continuous speed modelling, but reordering and rerouting decisions have no clear connection to the dynamic system. Also of influence is the need to consider switches as discrete sections. In general, the consideration of multiple trains result in complex models which are hardly considered in literature.

To conclude, proposed modelling approaches per modelling option are (also see Table 2):

1. Discrete infrastructure and discrete speed: AG, disjunctive MILP or time-indexed MILP.
2. Discrete infrastructure and continuous speed: AG or disjunctive MILP.
3. Continuous infrastructure and discrete speed: dynamic system.
4. Continuous infrastructure and continuous speed: dynamic system.

6. Conclusions

Conflict detection and resolution (CDR) under moving-block signalling remains underexposed in the literature, in comparison to moving-block signalling and CDR in general. This paper reviewed the literature on CDR models for fixed-block multi-aspect, fixed-block distance-to-go and moving-block signalling with the aim to identify research gaps and to propose future steps towards the development of moving-block CDR models.

The main gap is identified to be the lack of CDR models that accurately describe railway operations considering moving-block characteristic headways, based on absolute braking distances. Therefore, two main research challenges are the modelling of the infrastructure in continuous space and the inclusion of speed dynamics.

It is proposed to address the identified research gaps by an analysis of various options for infrastructure and speed modelling, followed by an assessment of modelling approaches for the different options. The CDR approaches of alternative graph (AG) and (disjunctive) mixed integer linear programming (MILP) are proposed for modelling options including a discrete infrastructure representation, while a shift to dynamic system approaches is suggested for the modelling of the infrastructure in continuous space.

Consistent with the proposal, future research should aim at (further) investigating the possibilities of the existing approaches of AG and disjunctive MILP in terms of infrastructure discretisation and speed alternatives. This has been taken up in the work of Versluis et al. (2023); a CDR model approximating moving-block operations is obtained by enhancing the disjunctive MILP model of Pellegrini et al. (2015). Furthermore, a comparison of the different modelling options is needed to determine the extent to which continuous modelling of infrastructure and speed brings advantages over discretised models in terms of solution quality and computational efficiency. Such a comparison will also improve the understanding of the context in which accurate modelling of moving-block operations does imply different rescheduling decisions compared to existing fixed-block CDR models. These insights are invaluable for the development of effective moving-block CDR models.

CRediT authorship contribution statement

Nina D. Versluis: Conceptualization, Visualization, Writing – original draft, Writing – review & editing. **Egidio Quaglietta:** Conceptualization, Writing – review & editing, Supervision. **Rob M.P. Goverde:** Writing – review & editing, Supervision. **Paola Pellegrini:** Writing – review & editing. **Joaquin Rodriguez:** Writing – review & editing.

Acknowledgements

The research in this paper has received funding from the Shift2Rail Joint Undertaking (JU) under the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101015416 (PERFORMINGRAIL). The JU receives support from the European Union's Horizon 2020 research and innovation programme and the Shift2Rail JU members other than the Union.

References

Bazaraa, M.S., Jarvis, J.J., Sherali, H.D., 2008. Linear programming and network flows. John Wiley & Sons.

Bettinelli, A., Santini, A., Vigo, D., 2017. A real-time conflict solution algorithm for the train rescheduling problem. Transp. Res. B 106, 237–265.

Büker, T., Graffagnino, T., Hennig, E., Kuckelberg, A., 2019. Enhancement of blocking-time theory to represent future interlocking architectures. In: RailNorrköping 2019, 8th International Conference on Railway Operations Modelling and Analysis, No. 069. ICROMA, Linköping University Electronic Press, pp. 219–240.

Cacchiani, V., Huisman, D., Kidd, M., Kroon, L., Toth, P., Veelenturf, L., Wagenaar, J., 2014. An overview of recovery models and algorithms for real-time railway rescheduling. Transp. Res. B 63, 15–37.

Caimi, G., Fuchsberger, M., Laumanns, M., Lüthi, M., 2012. A model predictive control approach for discrete-time rescheduling in complex central railway station areas. Comput. Oper. Res. (39), 2578–2593.

Corman, F., D'Ariano, A., Pacciarelli, D., Pranzo, M., 2009. Evaluation of green wave policy in real-time railway traffic management. Transp. Res. C 17 (6), 607–616.

Corman, F., Meng, L., 2015. A review of online dynamic models and algorithms for railway traffic management. IEEE Trans. Intell. Transp. Syst. 16 (3), 1274–1284.

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

D'Ariano, A., Pacciarelli, D., Pranzo, M., 2007a. A branch and bound algorithm for scheduling trains in a railway network. European J. Oper. Res. 183 (2), 643–657.

D'Ariano, A., Pranzo, M., Hansen, I.A., 2007b. Conflict resolution and train speed coordination for solving real-time timetable perturbations. IEEE Trans. Intell. Transp. Syst. 8 (2), 208–222.

ERA, 2016. ERTMS/ETCS System Requirements Specification (Subset-026). Tech. Rep., European Union Agency for Railways.

Furness, N., Van Houten, H., Arenas, L., Bartholomeus, M., 2017. ERTMS level 3: the game-changer. IRSE News 232, 2–9.

Gao, H., Zhang, Y., Guo, J., 2020. Calculation and optimization of minimum headway in moving block system. In: 2020 IEEE 5th International Conference on Intelligent Transportation Engineering. ICITE, IEEE, pp. 482–486.

Gonzalez, J., Rodriguez, C., Blanquer, J., Mera, J.M., Castellote, E., Santos, R., 2010. Increase of metro line capacity by optimisation of track circuit length and location: In a distance to go system. J. Adv. Transp. 44 (2), 53–71.

Hansen, I.A., Pachl, J. (Eds.), 2014. Railway Timetabling & Operations: Analysis, Modelling, Optimisation, Simulation, Performance Evaluation. Eurail Press, Hamburg.

Himrane, O., Beugin, J., Ghazel, M., 2023. Implementation of a model-oriented approach for supporting safe integration of GNSS-based virtual balises in ERTMS/ETCS level 3. IEEE Open J. Intell. Transp. Syst.

IEEE, 2005. IEEE Std 1474.1-2004, IEEE Standard for Communications-Based Train Control (CBTC) Performance and Functional Requirements. Tech. Rep., Institute of Electrical and Electronics Engineers.

Janssens, M.L., 2022. Multi Machine Approaches for Conflict Resolution Under Moving Block Signalling (Unpublished master's thesis). Delft University of Technology, The Netherlands.
Lamorgese, L., Mannino, C., 2015. An exact decomposition approach for the reail-time train dispatching problem. Oper. Res. 63 (1), 48–64.
Lamorgese, L., Mannino, C., 2019. A noncompact formulation for job-shop scheduling problems in traffic management. Oper. Res. 67 (5), 1586–1609.
Lamorgese, L., Mannino, C., Piacentini, M., 2016. Optimal train dispatching by benders'-like reformulation. Tranp. Sci. 50 (3), 910–925.
Lazarescu, M.T., Poolad, P., 2021. Asynchronous resilient wireless sensor network for train integrity monitoring. IEEE Internet Things J. 8 (5), 3939–3954.
Liu, R., 2016. Simulation model of speed control for the moving-block systems under ERTMS level 3. In: 2016 IEEE International Conference on Intelligent Rail Transportation. ICIRT, IEEE, pp. 322–327.
Liu, F., Xun, J., Liu, R., Yin, J., Dong, H., 2021. A real-time rescheduling approach by using loop iteration for high-speed railway traffic. IEEE Intell. Transp. Syst. Mag. 15 (1), 318–332.
Lövétei, I.F., Kővári, B., Bécsi, T., 2021. MCTS based approach for solving real-time railway rescheduling problem. Period. Polytech. Transp. Eng. 49 (3), 283–291.
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.
Lusby, R.M., Larsen, J., Ehrgott, M., Ryan, D.M., 2013. A set packing inspired method for real-time junction train routing. Comput. Oper. Res. 40 (3), 713–724.
Marlière, G., Sobieraj Richard, S., Pellegrini, P., Rodriguez, J., 2023. A conditional time-intervals formulation of the real-time railway traffic management problem. Control Eng. Pract. 54 (2), 187–194.
Martinez, L., Martin, U., 2020. Terminology, differences, and challenges of communications-based train control and European train control systems. WIT Trans. Built Environ. 199, 15–26.
Mascis, A., Pacciarelli, D., 2002. Job-shop scheduling with blocking and no-wait constraints. European J. Oper. Res. 143 (3), 498–517.
Mazzarello, M., Ottaviani, E., 2007. A traffic management system for real-time traffic optimisation in railways. Transp. Res. B 41 (2), 246–274.
Meng, L., Zhou, X., 2011. Robust single-track train dispatching model under a dynamic and stochastic environment: A scenario-based rolling horizon solution approach. Transp. Res. B 45 (7), 1080–1102.
Mera, J.M., Carabano, E., Soler, M., Castellote, E., 2016. Increasing metro line capacity by optimisation of track circuit in a speed code automatic train protection system. Proc. Inst. Mech. Eng. F 230 (1), 165–180.
Meunier, H., Baro, S., Borodin, V., Dauzére-Pérès, S., Pochet, J., 2023. A microscopic modeling approach for real-time train retiming under disturbances for a CBTC suburban railway line. Available at SSRN: https://ssrn.com/abstract=4333829.
Pellegrini, P., Marlière, G., Pesenti, R., Rodriguez, J., 2015. RECIFE-MILP: An effective MILP-based heuristic for the real-time railway traffic management problem. IEEE Trans. Intell. Transp. Syst. 16 (5), 2609–2619.
Pellegrini, P., Marlière, G., Rodriguez, J., 2014. Optimal train routing and scheduling for managing traffic perturbations in complex junctions. Transp. Res. B 59, 58–80.
Pellegrini, P., Pesenti, R., Rodriguez, J., 2019. Efficient train re-routing and rescheduling: Valid inequalities and reformulation of RECIFE-MILP. Transp. Res. B 120, 33–48.
Pochet, J., Baro, S., Sandou, G., 2016. Supervision and rescheduling of a mixed CBTC traffic on a suburban railway line. In: 2016 IEEE International Conference on Intelligent Rail Transportation. ICIRT.
Pochet, J., Baro, S., Sandou, G., 2017. Automatic train supervision for a CBTC suburban railway line using multiobjective optimization. In: 2017 IEEE International Conference on Intelligent Transportation Systems. ITSC.
Quaglietta, E., Pellegrini, P., Goverde, R.M.P., Albrecht, T., Jaekel, B., Marlière, G., Rodriguez, J., Dollevoet, T., Ambrogio, B., Carcasole, D., Giaroli, M., Nicholson, 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.
Reynolds, E., Ehrgott, M., Maher, S.J., Patman, A., Wang, J.Y., 2020. A multicommodity flow model for rerouting and retiming trains in real-time to reduce reactionary delay in complex station areas. Optimization Online.
Reynolds, E., Maher, S.J., 2022. A data-driven, variable-speed model for the train timetable rescheduling problem. Comput. Oper. Res. 142, 105719.
Rodriguez, J., 2007. A constraint programming model for real-time train scheduling at junctions. Transp. Res. B 41 (2), 231–245.
Samà, M., Meloni, C., D'Ariano, A., Corman, F., 2015. A multi-criteria decision support methodology for real-time train scheduling. J. Rail Transp. Plan. Manag. 5 (3), 146–162.
Schlechte, T., Borndörfer, R., Deniß en, J., Heller, S., Klug, T., Küpper, M., Lindner, N., Reuther, M., Söhlke, A., Steadman, W., 2022. Timetable optimization for a moving block system. J. Rail Transp. Plan. Manag. 22, 100315.
Scopus, 2023. Scopus. https://www.scopus.com.
Theeg, G., Vlasenko, S. (Eds.), 2020. Railway Signalling and Interlocking. PMC Media.
Törnquist, J., 2012. Design of an effective algorithm for fast response to the rescheduling of railway traffic during disturbances. Transp. Res. C 20, 62–78.
Törnquist, J., Persson, J., 2007. N-tracked railway traffic re-scheduling during disturbances. Transp. Res. B 41 (3), 342–362.
UNIFE, 2022. ERTMS. https://www.ertms.net/.
Van den Akker, J.M., Hurkens, C.A., Savelsbergh, M.W., 2000. Time-indexed formulations for machine scheduling problems: Column generation. INFORMS J. Comput. 12 (2), 111–124.
Versluis, N.D., Pellegrini, P., Quaglietta, E., Goverde, R.M.P., Rodriguez, J., 2023. An approximate conflict detection and resolution model for moving-block signalling by enhancing RECIFE-MILP. In: Presented at RailBelgrade 2023, 10th International Conference on Railway Operations Modelling and Analysis. ICROMA, Belgrade, Serbia, 25-28 April.
Wang, Y., De Schutter, B., Van Den Boom, T.J., Ning, B., 2014. Optimal trajectory planning for trains under fixed and moving signaling systems using mixed integer linear programming. Control Eng. Pract. 22 (1), 44–56.
Wang, P., Goverde, R.M.P., 2019. Multi-train trajectory optimization for energy-efficient timetabling. European J. Oper. Res. 272, 621–635.
Xu, P., Corman, F., Peng, Q., Luan, X., 2017. A train rescheduling model integrating speed management during disruptions of high-speed traffic under a quasi-moving block system. Transp. Res. B 104, 638–666.
Xu, P., Zhang, D., Guo, J., Liu, D., Peng, H., 2021. Integrated train rescheduling and rerouting during multidisturbances under a quasi-moving block system. J. Adv. Transp..
Xu, L., Zhao, X., Tao, Y., Zhang, Q., Liu, X., 2014. Optimization of train headway in moving block based on a particle swarm optimization algorithm. In: 13th International Conference on Control, Automation, Robotics and Vision. ICARCV, pp. 931–935.
Zwaneveld, P.J., Kroon, L.G., Van Hoesel, S.P., 2001. Routing trains through a railway station based on a node packing model. European J. Oper. Res. 128 (1), 14–33.