Expert Systems With Applications 230 (2023) 120650

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Modeling system dynamics of interacting cruising trains to reduce the impact of power peaks

Alessio Trivella [a,*], Francesco Corman [b]

[a] Industrial Engineering and Business Information Systems, University of Twente, Hallenweg 17, 7522 NH Enschede, The Netherlands
[b] Institute for Transport Planning and Systems, ETH Zürich, Stefano-Franscini-Platz 5, 8093 Zürich, Switzerland

ARTICLE INFO

Keywords:
Intelligent transport systems
Railway traffic dynamics
Automated train operations
Stochastic processes
Energy efficiency
Power peaks

ABSTRACT

Rail operators around the globe are striving to improve the efficiency, automation, safety, and sustainability of railway systems. Despite significant advances in technologies such as artificial intelligence and automated train operations (ATO), achieving these goals is challenging for complex rail networks when accounting for unpredictable factors that alter real-time operations. In this paper, we model railway traffic in a corridor as a string of interacting cruising trains, each subject to random speed variations that are described by a stochastic process. We simulate this dynamic system under assumptions that model human drivers and ATO systems, and compute performance measures focusing on energy consumption and the power peaks arising when multiple trains accelerate simultaneously. Different strategies to smooth these peaks are investigated, including the use of regenerative braking energy, potentially coupled with an electric energy storage, and a rule that uses fixed waiting times before re-acceleration. Our findings shed light on when and why these strategies can be effective at reducing energy consumption and/or shaving the peaks. They also show that employing a well-calibrated ATO controller in which vehicles exchange information about their location improves energy performance compared to a model of a human driven. Finally, a trade-off between energy performance and traffic regularity is exposed, i.e., strategies to reduce power peaks may slow rail traffic down, reducing capacity utilization.

1. Introduction

Rail operators worldwide are eager to make railway systems more efficient and resilient while ensuring safety and high service quality to passengers. Technologies such as automated train operations (ATO) and artificial intelligence (AI) are indeed at the top of the railway innovation agenda in many countries (see examples of Horizon Europe, 2022 and Railtech, 2021) as they offer a way to increase railway capacity without building new track infrastructure. Despite these technologies are being researched and deployed (or planned to), understanding their precise impact and potential on the planning and operations of railway systems is challenging, e.g., due to the complexity of modern rail networks and the presence of unpredictable real-time disturbances that affect operations. Additional barriers include the potentially high initial investments, the complexity of coordinating between various autonomous systems, setting up communication-based train control, legal and regulatory issues as well as cybersecurity concerns (Singh et al., 2021).

Regarding sustainability, transport accounts for a large share of energy consumption and is responsible for about 16% of greenhouse gas emissions globally (OWiD, 2016). Since three quarters of this amount is attributed to road transport, one key direction to promote sustainability and energy efficiency is to switch to collective transport of larger vehicles, rather than using private vehicles or small shipments by trucks. Railway is generally acknowledged to be an energy-efficient mode of transporting passenger and freight. Nonetheless, efforts to reduce its energy footprint are pursued by many transport operators and authorities (UIC, 2017) to cope with skyrocketing energy prices and to meet ambitious climate targets, e.g., set by the European Union for 2030 (EU, 2021).

In fact, energy consumption accounts for a large portion of railway operating costs (Railenergy, 2016). For instance, it is estimated that railway traffic in Germany consumes around 2% of the total electricity demand in the country, resulting in an annual traction electricity bill for Deutsche Bahn AG in the order of 1 billion EUR (Bärmann et al., 2017). The amount of energy needed for moving a train depends on speed and resistances, and results in requirements for energy production. By carefully planning the acceleration and braking processes of trains (i.e., the so-called train speed profiles) or optimizing the timetable,

The code (and data) in this article has been certified as Reproducible by Code Ocean: (https://codeocean.com/). More information on the Reproducibility Badge Initiative is available at https://www.elsevier.com/physical-sciences-and-engineering/computer-science/journals.

* Corresponding author.
E-mail addresses: a.trivella@utwente.nl (A. Trivella), francesco.corman@ivt.baug.ethz.ch (F. Corman).

https://doi.org/10.1016/j.eswa.2023.120650
Received 1 April 2023; Received in revised form 29 May 2023; Accepted 30 May 2023
Available online 2 June 2023

significant energy can be saved (Hansen & Pachl, 2014). For example, the Dutch Railways estimate a saving of about 5% by energy-efficient train driving (Luijt et al., 2017). In electrified railway systems, energy usage is centralized, i.e., all vehicles draw energy at the same time from a distribution network. The total energy required by the system is then the sum of all energies required by all vehicles. Nowadays, many rail electric systems can exploit regenerative braking, i.e., the motor of a decelerating train is turned to a generator that converts mechanical energy to electricity, which can be then used to power nearby trains (Khodaparastan et al., 2019; Lu et al., 2014), potentially in conjunction with an electric energy storage system (De La Torre et al., 2014; González-Gil et al., 2013).

In addition to the total energy consumption, it is important to consider the peak value of the power needed. Power peaks arise when multiple vehicles require large amount of power, especially during acceleration. The power load profiles of railway energy systems are usually highly dynamic in the collective power drawn by the trains in the network. The power load of the Swiss Federal Railways, for instance, is subject to increases and drops of 300 MW daily (50% of maximum load) and of 35 MW within 15-minute intervals (7% of maximum load) (SBB, 2021). In other terms, the load sizing for the infrastructure has to cover, for its half, only variations of power demand; and for the other half, base load. Those peaks threaten grid stability and represent a big concern for operators (Wang et al., 2022). A power distribution network designed for lower peaks might fail when the power demand is very high, causing reduction of delivered energy or, in the worst cases, a blackout. Traffic volume is growing, speed is increasing, vehicles are heavier, resulting in higher chances of high power peaks, but redesigning or upgrading the power distribution requires large costs and long time. Overdimensioning the power network at the design stage is also very costly. In addition to grid stability, another issue with power peaks is that the energy bill paid by rail operators is usually based on both total and maximum consumption (i.e., the highest peak) over the billing period (Albrecht, 2010). Smoothing power peaks has thus a direct impact on operational costs.

Controlling the total and maximum energy consumption of a railway system is challenging because practical railway operations are unavoidably subject to uncertainties affecting running and waiting times of trains. Even though accounting for energy consumption when designing the train timetable is common (Yang et al., 2015), disturbances occurring in real time shift departure and arrival times of trains, altering the planned timetable and the synchronization of acceleration activities. At a microscopic level, a train is subject to random speed variations that are linked with the driver behavior and changes in line voltage and track resistance, among others. Corman et al. (2021) provide empirical evidence of this effect using data from the Swiss railway, and introduce stochastic processes to model a system with two trains. The dynamic characteristics of a follower train are studied given a trajectory of a leader, without looking at energy use.

In this paper, we develop a stochastic model of railway traffic for a more generic string, or platoon, of consecutive trains that interact during cruising. As opposed to car traffic, railway vehicles are subject to a strict safety system and must keep a minimum safety distance. When two trains get too close, a yellow signal is triggered and the follower must decelerate towards a fixed lower speed. This results in extra time lost, in a braking, and a successive re-acceleration to a cruising speed. When multiple trains are considered, a yellow signal may cause a cascade effect on downstream vehicles, forcing more trains to decelerate and re-accelerate, possibly producing a power peak. We are thus particularly interested in examining this behavior for a platoon of cruising trains, its implications, and possible strategies to improve traffic regularity, energy usage, and peak values.

Three models to describe railway traffic dynamics are considered in this paper, namely a deterministic baseline and two stochastic models representing, respectively, a human driver and an ATO controller. Furthermore, different strategies to reduce power peaks are investigated, which are based on technological assumptions and/or train control. Specifically, one strategy implements fixed waiting rules for trains that have triggered a yellow signal. Another measure assumes that the electric railway system can use regenerative braking energy, possibly combined with a track-based electric energy storage. We use simulation for each model and strategy to quantify the emerging properties of the dynamic system, with focus on energy consumption and power peaks. The simulation program is coded in Matlab and is made available online in reproducible capsule on the "Code Ocean" platform.[1]

Some of the key takeaways from this work are the following:

- The model describing an ATO controller that exploits information exchange across vehicles is more efficient than that of a human driver in terms of traffic regularity, consumes less energy, and results in fewer power peaks.
- The considered strategies can shave the power peaks quite effectively. Under the human driver model, for instance, accounting for regenerative braking lowers the overall energy consumption by 3.3%, while coupling this with a storage reduces the peak height by 10%. Combining all strategies further decreases consumption and peaks.
- There is a trade-off between traffic regularity (measured, e.g., as the throughput, i.e., the hourly number of vehicles crossing the corridor) and energy performance (intended as energy consumption and power peaks). Strategies to improve energy performance must be designed and tuned carefully to avoid significant losses in capacity utilization.

Overall, the insights from our work can be valuable for railway operators to better understand how the stochastic dynamics and energy behavior or the system are affected by the use of technologies like ATO, storage systems, and train control algorithms. More specifically, the approach could be used in real-world applications to identify parameters of ATO systems and therefore better design and operate such systems. While the control parameters of ATO are now normatively determined, a comprehensive a-priori evaluation of different ATO specifications in their impact towards traffic fluidity allows a prescriptive determination of the ATO parameters to reach a specific traffic quality. The study also informs rail operators about the potential benefit to combine energy management and train control algorithms.

The rest of the paper is structured as follows. The relevant literature in reviewed in Section 2, the system dynamics and power peak strategies are modeled in Section 3, numerical results and findings are presented in Section 4, and conclusions are drawn in Section 5.

2. Novelty and related work

Four streams of literature are connected with our research and reviewed in the following: improving energy efficiency in railway operations (Section 2.1), dealing with stochasticity in railway models (Section 2.2), the traffic flow theory (Section 2.3), and the management of intelligent railway systems (Section 2.4). A summary of the contributions of this study is eventually provided (Section 2.5).

2.1. Energy efficiency in railways operations

There is a considerable body of literature dealing with the energy efficiency of railway systems (De Martinis & Corman, 2018). Optimizing the train speed profiles and the timetable are two prominent ways to cut energy consumption without costly investments in new vehicles or electric systems. These are tactical planning problems usually solved days to months before operations.

[1] https://doi.org/10.24433/CO.2417202.v1.

Computing an energy-efficient speed profile for a train running between two stations is commonly known as the train trajectory optimization problem or the energy-efficient train control problem. Here, a speed profile describes the speed of a train as a function of time or distance covered. Determining such a profile means choosing the driving regimes (acceleration, cruising, coasting, and deceleration) and the switching points between them (Howlett, 2000). The goal is finding the profile associated with the lowest energy use, while ensuring a punctual arrival and respecting speed limits and time windows. It is well known that optimizing train control can lead to large energy savings, also above 10% (Hansen & Pachl, 2014). Methods to tackle this problem include the Pontryagin's maximum principle (Albrecht et al., 2016), mathematical programming (Ye & Liu, 2017), and dynamic programming (Trivella et al., 2021). Since we consider a system that is stochastic, dynamic, and comprised of multiple trains, the problem studied in this paper largely deviates from the classical train trajectory optimization problem, despite both problems deal with speed variations and train control at a microscopic scale. Furthermore, train trajectory optimization is usually treated as a deterministic problem that is solved before real-time operations based on the predefined timetable. Instead, this paper deals with a real-time operational problem where unplanned uncertainties affect the ideal speed profile during the cruising phase.

The timetabling problem aims at determining departure and arrival times of trains at stations subject to headway and track capacity constraints (Caprara et al., 2002). The traditional objectives are to minimize operational costs (e.g., number of required vehicles), maximize passenger satisfaction (e.g., by reducing travel time), and minimize deviations from the original timetable in case of a disruption (Binder et al., 2017; Ho et al., 2012). Recently, emphasis has been put on the energy use associated with a timetable. Bärmann et al. (2017) discuss how minimal changes in the planned timetables can impact energy and power of the entire system. Regueiro Sánchez (2021) adopts a similar approach, also limiting the maximum traction power available between stations. Wang et al. (2022) tune a timetable using a mixed-integer linear program combined with a local search algorithm, with the goal of reducing energy consumption and smoothing power peaks. Notice that although this work also considers a system of interacting trains, the problem tackled differs from timetabling for several reasons. First, we are not concerned with setting departure and arrival times as the focus is on a string of trains running in a corridor, effectively considering a platoon of vehicles. Second, reducing power peaks in timetabling implies synchronizing acceleration and deceleration activities at departure and arrival events, whereas our focus is on the dynamics of trains during the cruising phase. Third, similar to train trajectory optimization, timetabling is also a tactical planning problem solved in advance while the one considered in this paper is a real-time operational problem.

We refer to the review by Scheepmaker et al. (2017) for further literature on both the energy-efficient train control and the energy-efficient train timetabling. Some works attempted to combine these two problems by embedding train control between pairs of stations as a subproblem into a timetable optimization framework (Ran et al., 2020; Yang et al., 2015). Moreover, a study on urban railway systems showed that energy consumption could be reduced by 25%–35% by jointly implementing energy-optimized timetables, energy-efficient driving strategies, improved control of comfort functions, and energy storage devices (González-Gil et al., 2014).

Finally, in addition to designing the timetable, real-time train rescheduling models are used to modify the planned timetable and avoid conflicts when disturbances or disruptions occur (Corman & Meng, 2014; Jusup et al., 2021; Luo et al., 2022). The related literature however prioritizes recovering the delays over reducing energy use, with the exceptions of Yin et al. (2016) that account for energy when rescheduling a metro system. Even though what we propose is not a classical train rescheduling problem, it involves altering the control of vehicles to restore the headway between them, which can be considered as a rescheduling measure. To the authors' knowledge, there is no scientific work on real-time railway rescheduling that specifically studies peaks in power consumption and that analyzes the trade-offs between traffic regularity and energy performance.

2.2. Stochasticity in railway models

The majority of the aforementioned works assume that the future is known with certainty at the planning stage. However, railway operations are characterized by uncontrollable stochastic factors that only realize in real time. Those include variable passenger demand, boarding/alighting times, running times, technical failures, and weather to name a few (Trivella & Corman, 2019).

Some recent literature has acknowledged that accounting for uncertainty and counteracting it is critical to improve railway planning in different stages. Thus, uncertainty has been incorporated into train control, timetabling, and rescheduling models (Cacchiani & Toth, 2018; Jusup et al., 2021; Wang et al., 2020; Yang et al., 2016; Yin et al., 2016). For instance, the goal of Cacchiani and Toth (2018) is to compute train timetables that are robust, i.e., that are expected to perform well under random disturbances. Wang et al. (2020) use approximate dynamic programming to tackle train control in the case of randomly varying speed profiles due to uncertain resistance parameters. Yin et al. (2016) use similar methods to reschedule a metro network under varying passenger demands. Yang et al. (2016) propose an integrated stochastic model for optimizing speed profile and timetable in metro systems under uncertain train mass.

In this paper, we consider random variations affecting train speed at the microscopic level. While most approaches capture the dynamics of the uncertainties by means of simple independent and identically distributed random variables, we model the uncertainty in train speed using a stochastic process. It refers to a collection of random variables that are function of time, and where the outcome of the process at a time t also depends on its value at time $t-1$. These are more sophisticated models that were shown by Corman et al. (2021) to describe well the train cruising process occurring in real life. When considering a dynamic system that includes a string of trains and/or complex stochastic process models, it is not possible to derive closed-form expressions of key performance indicators using the stochastic differential equations governing the processes. Therefore, a simulation model is developed to study the system dynamics, and derive the relevant performance measures as sample averages over a large number of Monte Carlo sampled trajectories.

2.3. Traffic flow theory

This work builds on the stochastic modeling of the interrelated flow of two successive trains, first introduced by Corman et al. (2021) to extend the vehicular traffic flow theory to railway systems. Whilst Corman et al. (2021) focus on just two consecutive trains (on which only the follower is subject to random effects) and on the regularity of the traffic, we consider in this paper a system of N trains, and examine its behavior in terms of energy consumption and power peaks. Apart from this reference, the literature on traffic flow theory has almost exclusively tackled car traffic, which has similarities with our study but also important differences.

In the case of private vehicles, there is no external safety system which constrains their speeds and distances in-between, which is instead a key feature of railways. When vehicles can be individually controlled, an interesting problem of traffic flow theory is to study under which conditions, a series of coordinated vehicles has impact towards traffic flow stability, i.e., studying string stability. This has been mostly focusing on vehicles with homogeneous, simplified characteristics (Wang, Li et al., 2017).

A second difference is that power distribution in railway traffic is pooled and centralized; hence, there exist issues with simultaneous high

energy maneuvers of different vehicles. In contrast, studies of vehicular traffic do not focus on power peaks because the energy consumption of all vehicles draws from separated energy carriers (such as internal combustion engines, or battery-fed electric motors), although this may change should electric highway systems become more popular in the future. Nevertheless, there is interest in minimizing system properties such as emissions (Qin et al., 2020; Zhang et al., 2020), which have some of their sources in speed and acceleration. Moreover, platooning and coordinated speed control in vehicular road networks have shown to have potential to reduce energy required by 10% (Qin et al., 2020). The current paper thus wants to understand to which extent the implementation of ATO can be an opportunity to minimize system level energy consumption, matching or exceeding the potential for vehicular networks.

2.4. Intelligent railway systems

This paper contributes to the growing literature on intelligent transport systems, and in particular, intelligent railway systems. These are systems that leverage and integrate advanced information and communication technologies such as big data analytics, modern sensors, and automated control and signaling systems to improve the safety, efficiency, and reliability of railway transportation. These technologies fall under the broad sphere of Internet of Things (IoT) and Artificial Intelligence (AI). The reader is referred to Singh et al. (2022) and Tang et al. (2022) for recent reviews on the use of IoT and AI technologies in rail transportation, respectively.

Predictive maintenance is a prominent example of intelligent railway systems. By combining advanced sensors with machine learning algorithms, the condition of rolling stock equipment and tracks, including railroad switches, can be monitored to detect potential failures before they occur. This allows railway operators to proactively repair or replace components, preventing incidents (De Simone et al., 2023; Tashmetov et al., 2022; Wang, Xu et al., 2017). Using drones for visual inspection and monitoring of railway infrastructure has also been proposed (Banić et al., 2019). Another important example is the use of real-time data analytics to manage train traffic. By utilizing data on current train delays, passenger demand, weather conditions, and other factors, intelligent railway systems can adjust train schedules and routes in real-time to reduce delays and improve on-time performance (Corman & Meng, 2014; Sels et al., 2016).

A key role in this picture is played by ATO technologies that can automate various functions of train driving tasks including (depending on the degree of automation), the operation of doors, starting and stopping at stations, and speed control (Quaglietta et al., 2020). A recent review on trends, technologies, advancements, and challenges in the deployment of autonomous trains in rail transportation is provided in Singh et al. (2021). ATO systems are typically fed by timetabling data and data from a traffic management system, and are combined with an automatic train protection system for emergency braking. It is recognized that deploying ATO systems can be valuable as the resulting train control allows for smaller headways, which in turn can lead to a higher track capacity utilization (Poulus et al., 2018). Increased timetable stability, on-time performance, safety, and energy efficiency are all additional potential benefits (Singh et al., 2021), as it is the reduced task-induced fatigue for train driving staff under higher degrees of automation (Brandenburger et al., 2021).

Despite this knowledge, it is difficult to model and accurately quantify these effects, especially due to the complexity of modern rail networks. Moreover, implementing ATO systems may require significant investments; hence, it is crucial to carefully evaluate costs and benefits before deploying such systems. This paper attempts to narrow this knowledge gap by estimating the added traffic regularity and energy performance of ATO, compared to a human driver, in a busy rail corridor where multiple consecutive trains interact with each other following stochastic dynamics. We are also not aware of studies that examine ATO systems from the perspective of reducing power peaks.

Finally, the smart management of energy flows is evermore important to improve sustainability while controlling costs in virtually any industry. In railways, energy storage devices are employed to reuse regenerative braking energy (Ratniyomchai et al., 2014). Although storage devices are traditionally not immediately associated with the concept of "intelligent railway system", they are becoming increasingly popular in electrified railways. Prior work has focused, e.g., on the different characteristics of storage technologies (Liu & Li, 2020) or on optimal sizing of storage assets (De La Torre et al., 2014), while we define an operating policy to distribute regenerative energy in time and space (i.e., across trains) based on quantitative decision making.

2.5. Summary of contribution

This work contributes to the literature by drawing a bridge between the four different active fields of research previously reviewed. The main contribution of this paper is threefold:

- Using stochastic process models, we simulate the dynamics of a platoon of interacting cruising trains and analyze the energy behavior of the resulting system.
- We outline a technique for detecting power peaks arising from the considered system dynamics, and propose three strategies to smooth such peaks based on: (i) train control, (ii) reusing regenerative braking energy, and (iii) managing an energy storage, potentially in combination.
- We provide insights on the impact of different stochastic process models (e.g., describing a human driver vs an ATO system) and different peak reduction strategies towards traffic regularity, energy use, power peaks, and the trade-offs between these potentially conflicting goals.

3. System dynamics and power peaks

The developed methodology is comprised of several modeling steps that are illustrated in Fig. 1. The model of system dynamics for a string of trains is the core element of this methodology and is presented in Section 3.1, including two different stochastic processes and a deterministic benchmark. Section 3.2 describes the key performance indicators (KPIs) of interest related to traffic regularity and energy consumption, and how to obtain them via simulation. To compute some of these KPIs, however, a technique for detecting peaks in energy consumption is developed in Section 3.3. Finally, different strategies to smooth power peaks are considered in Section 3.4, which also affect the dynamics of the system as shown in Fig. 1. The system dynamics module requires as input several simulation and train parameters. However, the values of these parameters are not discussed in this section, but at the beginning of the numerical study when the setting is introduced (Section 4.1).

3.1. Stochastic models of railway traffic

Building on the model for two trains by Corman et al. (2021), we define a general stochastic system of $N \geqslant 2$ consecutive trains running in a railway corridor based on a reference cruising speed v_{CRUISE}. Quantities $v_n(t)$ and $s_n(t)$ denote, respectively, the speed and distance covered by train n at time $t \geqslant 0$, where train $n = 1$ is the leader and trains $n > 1$ are followers. The system dynamics modeled in this section aim to determine the evolution of these quantities (speed and distance covered, for each train) over a predefined time horizon. Thus, the system is comprised of $2N$ state variables.

At $t = 0$, it is assumed that all trains are running at cruising speed $v_n(0) = v_{\text{CRUISE}}$ and have a safe headway from the immediate follower $s_{n-1}(0) - s_n(0) > d_{\text{MIN}}$, where d_{MIN} is the minimum absolute

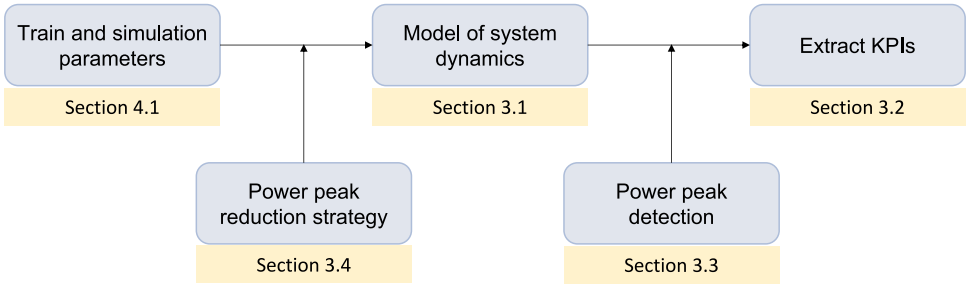

Fig. 1. Modeling steps in the methodology.

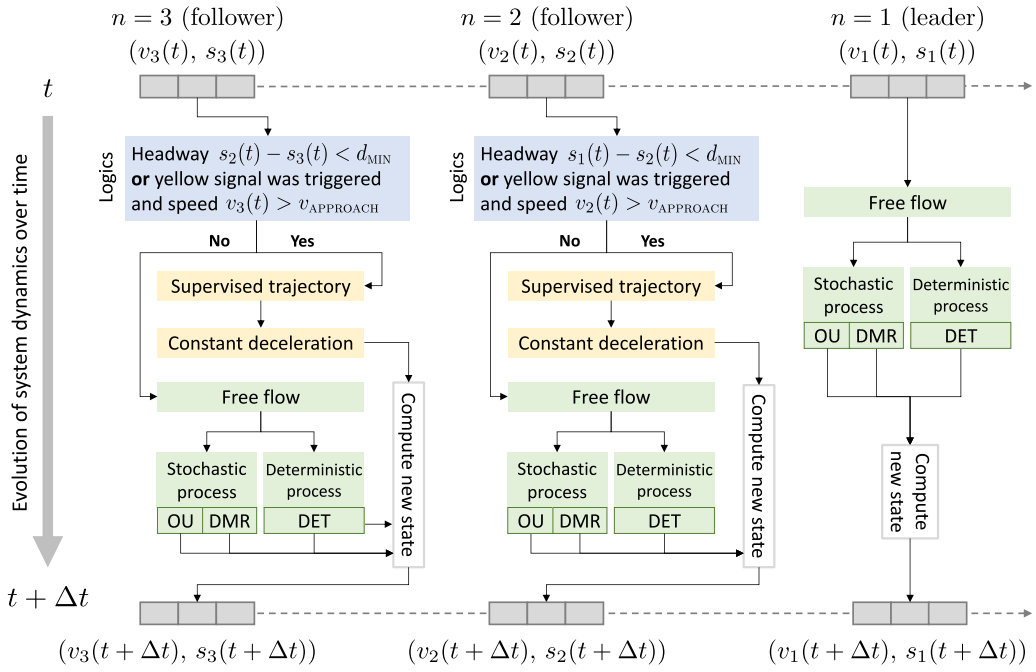

Fig. 2. Overview of train system dynamics.

safety distance between trains. The complete dynamics combine "free flow" and "supervised trajectory" phases as illustrated in Fig. 2 for a three-train example.

Under free flow dynamics, the train speed can follow: (i) a stochastic process modeling a human driver, (ii) a stochastic process modeling an ATO controller, or (iii) a deterministic baseline. Although all trains are initially spaced at a safe distance one from another at $t = 0$, the free flow dynamics may bring a follower train $n > 1$ too close to its preceding train $n - 1$, triggering a yellow signal whenever their headway $s_{n-1}(t) - s_n(t)$ falls below d_{MIN}. This enforces train n to obey to a supervised trajectory, consisting of a deterministic deceleration at constant rate $a_{\text{DET}} < 0$, which continues until an approach speed $v_{\text{APPROACH}} < v_{\text{CRUISE}}$ is reached and headway d_{MIN} is restored.

We now describe in more detail the free flow dynamics. Denote by $W(t)$ a standard Wiener process, which is the simple continuous-time stochastic process that is commonly used as the basis of more advanced ones. Formally, $\{W(t)\}_{t\geqslant 0}$ is defined as the process with increments that are independent and Gaussian-distributed as $W_{t+\Delta t} - W_t \sim N(0, \Delta t)$. The first free-flow model we consider is based on the Ornstein–Uhlenbeck (OU) stochastic process, expressed for train n as:

$$[\text{OU}]: \begin{cases} \mathrm{d}v_n(t) = \beta_n(v_{\text{CRUISE}} - v_n(t))\mathrm{d}t + \sigma_n \mathrm{d}W(t), \\ \mathrm{d}s_n(t) = v_n(t)\mathrm{d}t. \end{cases} \tag{1}$$

with boundary conditions $v_n(0) = v_{\text{CRUISE}}$ and $s_n(0) = 0$, i.e., the distance covered is measured starting at time 0. The stochastic differential Eq. (1) describes the microscopic variations of speed and space over time. Based on this equation, the train speed varies according to a deterministic component that establishes the mean reverting behavior towards v_{CRUISE}, and a stochastic component. At a high level, this model can describe a human train driver that continuously adjusts the train speed to keep it as close as possible to the target value. The parameters defining mean reversion (β_n) and volatility (σ_n) capture the reaction of the driver and of the speed control system.

A more sophisticated model is based on a doubly-mean-reverting (DMR) process, describing an ATO system where trains are aware of the location of the traffic ahead. Under this model, train n accelerates or decelerates to maintain both a target speed and a target headway with train $n - 1$:

$$[\text{DMR}]: \begin{cases} \mathrm{d}v_n(t) = \left[\beta_n(v_{\text{CRUISE}} - v_n(t)) + \alpha_n(s_{n-1}(t) - s_n(t))\right]\mathrm{d}t \\ \qquad\qquad + \hat{\sigma}_n(v_n(t))\,\mathrm{d}W(t), \\ \mathrm{d}s_n(t) = v_n(t)\mathrm{d}t, \end{cases} \tag{2}$$

where $\hat{\sigma}_n(v) := \sigma_n\sqrt{v(v_{\text{MAX}} - v)/[v_{\text{CRUISE}}(v_{\text{MAX}} - v_{\text{CRUISE}})]}$, $v_{\text{MAX}} > v_{\text{CRUISE}}$ is an upper bound on speed, and α_n is a second mean reversion parameter that acts on the headway value. We assume the same boundary conditions to model (1) apply. Let us take a closer look at the new term $\alpha_n(s_{n-1}(t) - s_n(t))$. Since $s_n(t)$ represents the distance covered by train n at time t, a positive difference $s_{n-1}(t) - s_n(t) > 0$ implies that the preceding train $n - 1$ has covered a larger distance than train n, hence the headway between the two vehicles has increased. In this case, $\alpha_n(s_{n-1}(t) - s_n(t)) > 0$ will push the speed $v_n(t)$ up, i.e., the current train

n accelerates to restore the initial headway. Conversely, with negative distance $s_{n-1}(t) - s_n(t) < 0$, indicating that the headway has shrunk, the stochastic process will steer train n to slow down to restore the headway.

In addition to the stochastic differential Eq. (2) itself being more sophisticated than (1), notice that there are also key conceptual differences between these two models, especially regarding the way in which trains interact. In the OU model, the stochastic processes governing each train are independent from each other (trains are only linked by the triggering of yellow signals), whereas in the DMR model, the evolution of the state variables for one train (speed, space) also depends on the state variables of the train ahead. More precisely, the latter model enables location information to be exchanged across trains at any point in time. This information is processed and exploited in the form of an improved control systems, which is indeed one of the goals when deploying ATO.

We finally consider a deterministic benchmark that does not account for stochasticity, which is defined by:

$$[\text{DET}]: \begin{cases} \mathrm{d}v_n(t) = 0\mathrm{d}t, \\ \mathrm{d}s_n(t) = v_n(t)\mathrm{d}t. \end{cases} \tag{3}$$

If the deterministic initial speeds are $\bar{v}_n = v_n(0) = v_{\text{CRUISE}}$ for all vehicles $n = 1 \dots, N$, then the system is in an ideal but unrealistic state in which all train pairs will indefinitely preserve the initial headway. Instead, if the speeds are different and $\bar{v}_n > \bar{v}_{n-1}$ for some n, a yellow signal will be triggered at some time $t > 0$ causing train n to decelerate. Once the approach speed is met, this train will then re-accelerate, reaching again $\bar{v}_n > \bar{v}_{n-1}$, trigger a second yellow signal, hence decelerate, and keep cycling over these states. When multiple trains have different initial speeds, the deterministic system will produce much more complex yet cyclic patterns. Notice that (3) can be seen as a special case of a stochastic process with no drift and volatility, e.g., it can be obtained from (1) by setting $\beta_n = \sigma_n = 0$ for all $n = 1 \dots, N$, which results in constant train speeds.

3.2. Performance indicators and their estimation

Key performance indicators of the system related to traffic regularity and energy usage can be computed by drawing scenarios (i.e., sample paths) from the stochastic processes in Monte Carlo simulation. We consider an horizon $[0, T]$, discretize it into steps $\mathcal{T} = \{0, 1, \dots, T\}$ that are equally spaced by Δt, and apply a standard forward Euler scheme to (1), (2), or (3) to obtain speed and space trajectories, i.e., functions describing the evolution of the state variables over time.

To derive train acceleration a_t, traction force f_t, and energy consumption e_t^n at all discrete time steps $\mathcal{T}$, the common dynamic equations for the motion of a railway vehicle are employed.[2] Specifically, the traction force is

$$f_t = a_t \cdot m \cdot \rho + [\gamma_1 + \gamma_2 v_t + \gamma_3 v_t^2], \tag{4}$$

where $a_t = (v_t - v_{t-1})/\Delta t$ is the acceleration, γ_i are the train resistance parameters, m the mass, and ρ the rotating mass factor. The energy consumed at time t (with no regenerative braking) is then

$$e_t^n = \max\{f_t, 0\} \cdot (s_t - s_{t-\Delta t}). \tag{5}$$

We refer to Hansen and Pachl (2014) for more details on train dynamics. Using the described simulation procedure, six KPIs are estimated related to traffic regularity (items 1–3) and energy consumption (items 4–6), and are listed below. Hereafter an *energy profile* $E := \{e_t : t \in \mathcal{T}, e_t = \sum_{n=1}^{N} e_t^n\}$ indicates the joint energy consumption of all trains as a function of time, measured over regular time intervals (e.g., of 30 s) during $[0, T]$.

[2] To ease reading, we drop the dependency on the train index n from a_t, f_t, v_t, and s_t, and only keep index n for the energy consumption e_t^n.

K1. Percentage of scenarios with a triggered yellow signal.
K2. Average first time to yellow (FTTY), which records the first time (s) in which a train triggers a yellow signal (scenarios without signals count as T in the average).
K3. Throughput of the corridor (vehicles/hour) measured as the train speed over headway, where speed and headway are averaged across trains, time steps, and scenarios.
K4. Total energy consumption (kWh) of all N trains over the horizon $[0, T]$, averaged across scenarios.
K5. Maximum value of the energy profile (kWh/30 s) over the horizon $[0, T]$, averaged across scenarios.
K6. Percentage of energy profiles with a power peak, where a *power peak* is defined as the energy (kWh) consumed during a relatively short period of time that considerably exceeds average consumption levels of the system (see Section 3.3 for technical details). The average consumption level refers to the situation where all trains run at cruising speed without disturbances.

Notice that Monte Carlo simulation is the only viable option to estimate the aforementioned KPIs. In fact, deriving these KPIs in closed form using the stochastic differential Eqs. (1)–(2) is intractable for the considered system. Previous research had already highlighted this challenge: even for the simplest stochastic process and a single train subject to random speed variations, only one KPI (the FTTY) could be computed analytically (Corman et al., 2021). Computing KPIs for (3) is also not possible when considering the complete dynamics that combine free flow and supervised trajectory phases for multiple interacting trains (see Fig. 2).

3.3. Power peak detection

Since thousands of simulations are needed to obtain statistically relevant averages, hence reliable KPIs, a method to detect power peaks in a simulated energy profile is needed. Peaks are relative to the railway system considered, e.g., number of trains in the system, their cruising speed, etc. Therefore, a non-parametric procedure that requires no assumptions on the input data and energy profiles is developed to identify peaks in consumption.

Algorithm 1: Power peak detection

Input: Energy profile $E = \{e_t : t \in \mathcal{T}\}$; Parameters $(w, \theta_1, \lambda_1, \theta_2, \lambda_2)$, with $\theta_1 \geqslant \theta_2$ and $\lambda_1 \geqslant \lambda_2$; Set of peak points $P \leftarrow \infty$; Variable $FullPeakFound \leftarrow FALSE$

Step 1. Apply Gaussian filter to E with window w to get a smoothed energy profile $\hat{E} = \{\hat{e}_t : t \in \mathcal{T}\}$

Step 2. Compute mean (z) and standard deviation (g) of $\hat{E}$

Step 3. Initialize peak points $P \leftarrow \{t \in \mathcal{T} : \hat{e}_t \geqslant \theta_1 z + \lambda_1 g\}$

while $FullPeakFound = FALSE$ **do**

Step 4. Try to extend peak points set:

$P^N \leftarrow \{t \in \mathcal{T} \setminus P : \hat{e}_t \geqslant \theta_2 z + \lambda_2 g \wedge (\hat{e}_{t-1} \in P \vee \hat{e}_{t+1} \in P)\}$

if $P^N = \infty$ **then**

$FullPeakFound \leftarrow TRUE$

else

$P \leftarrow P \cup P^N$

Output: Peak points $P \subset \mathcal{T}$

The proposed method is outlined in Algorithm 1. The algorithm takes as input the energy profile as well as additional parameters that regulate the detection mechanism. It starts by applying a Gaussian-weighted moving average filter to obtain a smoothed profile, so that

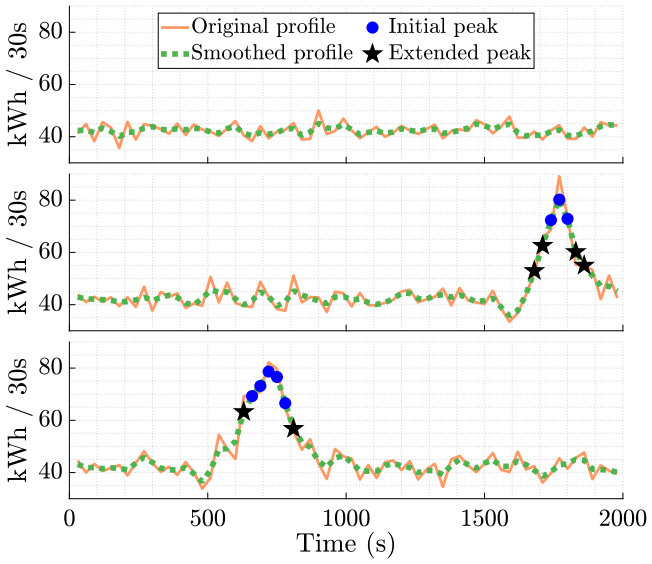

Fig. 3. Examples of simulated energy profiles and detected peaks.

points with high value due to the background stochastic speed fluctuations can be excluded (Step 1). In other words, the effect of this filtering step is to smooth out the fluctuations resulting from the stochastic process while preserving the major peaks that arise when multiple trains accelerate. Then, it defines as peaks the points marked as outliers according to a metric based on mean and standard deviation of the overall profile; specifically, the points that lie above $\theta_1 z + \lambda_1 g$ (Steps 2–3). Finally, in the *while* loop, the initially detected sets of peak points are extended to reconstruct the entire peak. This is done by iteratively incorporating neighboring peak points that fulfill a relaxed outlier condition, i.e., points that are higher than $\theta_2 z + \lambda_2 g$ and with a neighbor already in the peak set (Step 4).

Examples of the described peak detection procedure are illustrated in Fig. 3. The figure shows three simulated energy profiles for a system with six trains governed by OU processes. The dashed line, round markers, and star-shaped markers represent the output, respectively, of Step 1, Steps 2–3, and Steps 4 of Algorithm 1. Notice that the panels in the figure do not represent the energy consumption of an individual train, but represent instead the joint (total) energy consumption of the entire set of six trains over the given time horizon (each panel is a different simulation run of the dynamic system). We verified that our method is accurate as the identified peaks correspond indeed, in almost all cases, to multiple trains accelerating after a yellow signal. For example, the peak in the middle panel of Fig. 3 corresponds to 4 trains triggering a yellow signal and re-accelerating (this specific energy profile is also examined more in detail later in Section 4).

3.4. Power peak reduction

Three measures to mitigate the impact of yellow signal propagation are studied: (i) exploiting regenerative braking, (ii) managing regenerative energy by means of an electric energy storage, and (iii) adopting waiting policies for trains that have triggered a yellow signal. We describe these measures next, starting with regenerative braking.

Recall from Section 3.2 that the traction force is computed as in (4). When a train brakes, $a_t < 0$, which may result in $f_t < 0$. Without regenerative braking, a negative traction force does not contribute to energy use. Regenerative braking allows to convert a portion $\phi \in [0, 1]$ of kinetic energy to electric energy. Mathematically, this means computing

$$r_t^n = \max\{0, -\phi f_t\} \cdot (s_t - s_{t-\Delta t}). \tag{6}$$

To elaborate, Eq. (6) considers the maximum between 0 and $-\phi f_t$. This implies that when $f_t < 0$ (i.e., during braking), then $-\phi f_t > 0$ and the max function selects $-\phi f_t$, resulting in a positive amount $r_t^n > 0$. Instead, when $f_t > 0$, (6) produces $r_t^n = 0$. (Notice that this is the opposite behavior compared to (5), where $e_t^n = 0$ if and only if $f_t < 0$.) Thus, the energy recuperated by train n is r_t^n, and a positive energy $r_t^n > 0$ can only be used by other trains $n' \neq n$ at the current time step t. The total regenerative energy produced by all trains is $r_t = \sum_{n=1}^{N} r_t^n$.

Suppose now that a track-based electric energy storage is available to collect energy from regenerative braking and re-distribute it to the vehicles in subsequent time steps $t' > t$. Managing a storage asset requires defining an operating policy to charge/discharge it. Although storage operating policies are usually defined as a function of electricity prices to maximize profit, the goal here is different: to smooth power peaks. We present the storage operating policy in Algorithm 2. In this algorithm, $e_0 := (\gamma_1 + \gamma_2 v_{\text{CRUISE}} + \gamma_3 v_{\text{CRUISE}}^2) \cdot v_{\text{CRUISE}} \Delta t$ denotes the baseline energy consumption (kWh) of a train running at constant cruising speed during a time interval. Moreover, the energy storage is characterized by the following properties: an initial energy level L_0 (kWh), a maximum capacity L^{MAX} (kWh), maximum per-period charging and discharging rates L^{CHG} and L^{DIS} (kWh/Δt), and charging and discharging efficiencies $\eta^{\text{CHG}}, \eta^{\text{DIS}} \in (0, 1]$.

Algorithm 2: Storage operation

Input: Storage parameters (L_0, L^{MAX}, L^{CHG}, L^{DIS}, η^{CHG}, η^{DIS}); Baseline consumption e_0; Policy parameter $\mu \geqslant 0$

for $t = 1, \dots, T$ **do**
- Compute energy consumption e_t^n for $n = 1, \dots, N$ from (5)
- Computed energy recuperated r_t^n for $n = 1, \dots, N$ from (6), and their sum $r_t = \sum_{n=1}^{N} r_t^n$
- Charge storage $L_t \leftarrow \min\{L_{t-1} + \eta^{\text{CHG}} r_t,\ L_{t-1} + L^{\text{CHG}},\ L^{\text{MAX}}\}$
- Initialize storage discharge of current period $D_t \leftarrow 0$
- **for** $n = 2, \dots, N$ **do**
  - **if** $(L_t > 0) \wedge (e_t^n > \mu e_0) \wedge (D_t < L^{\text{DIS}})$ **then**
    - Compute storage discharge for train n as $d_t^n = \min\{L_t,\ L^{\text{DIS}},\ (e_t^n - \mu e_0)/\eta^{\text{DIS}}\}$
    - Discharge storage $L_t \leftarrow L_t - d_t^n$
    - Update energy consumption $e_t^{n,*} \leftarrow e_t^{n,*} - \eta^{\text{DIS}} d_t^n$
    - Update storage discharge of current period $D_t \leftarrow D_t + d_t^n$

Output: New consumption $e_t^{n,*}$ for $t \in \mathcal{T}, n = 1, \dots, N$

At a high level, Algorithm 2 defines a heuristic operating policy that establishes when and by how much to charge and discharge the storage, and to what train supply a certain amount of power. This algorithm requires the storage specification parameters already listed above, a tunable policy parameter μ, and the decision variables to charge the storage (tracked by the storage level L_t) and discharge it to supply power to each train n at time t (denoted by d_t^n).

The intuition behind Algorithm 2 is that at least some trains (followers) will consume more energy than e_0 during a power peak. Thus, the available regenerative energy, collected by the storage and tracked by L_t, compensates for the excess of energy with respect to a quantity μe_0. Regenerative energy not used in the current period charges the storage. The parameter μ encodes how conservative the policy is. Setting high values of μ results in the storage providing energy only during high peaks and increases the chance of not using some regenerative energy before the end of the simulation horizon T. Vice versa, with low μ values the storage more easily supplies electricity, at the risk that this

Table 1
Parameters of trains, stochastic processes, and peak reduction strategies.

Name	Value	Unit	Name	Value	Unit	Name	Value	Unit
m	500	t	d_{MIN}	3.0	km	w	1	–
ρ	1.06	–	v_{CRUISE}	35	m/s	ϕ	0.7	–
γ_1	5.8	kN	v_{MAX}	40	m/s	η^{CHG}	1	–
γ_2	0.072	kN s/m	v_{APPROACH}	20	m/s	η^{DIS}	1	–
γ_3	0.013	kN (s/m)^2	a_{DET}	−0.55	m/s	L^{CHG}	inf	kWh/Δt
α_n	$2 \cdot 10^{-5}$	–	θ_1	1.05	–	L^{DIS}	inf	kWh/Δt
β_n	0.02	–	θ_2	1	–	L^{MAX}	inf	kWh
σ_n	0.05	–	λ_1	2	–			
d_0	3.2	km	λ_2	1	–			

electricity does not go towards shaving the peak points. This storage operating model captures the essential features of a battery storage including capacity, efficiencies, and charging and discharging rates. Those are considered as input parameters in the mode and we are not interested, for instance, in optimizing the size of a storage system (see De La Torre et al., 2014 for a related study). Onboard (i.e., moving) storage devices may also be considered, but in this case a policy to allocate regenerative energy across multiple onboard storage units would need to be defined as well. Thus, Algorithm 2 would need to be tweaked to work in this situation.

We eventually consider another potential peak reduction measure that does not rely on assumptions on the electric rail system technology but on *fixed waiting time* rules for train control. More precisely, after a vehicle decelerates and reaches the given approach speed v_{APPROACH}, it must wait for a predetermined amount of time (δ seconds) before being allowed to re-accelerate. The intuition behind this measure is that, by introducing waiting times, the re-acceleration phases of different trains after a yellow signal is triggered are better spread over time, thus limiting the generation of high power peaks. Note that due to the safety system, more downstream trains may have to reach speeds lower than v_{APPROACH}, or even stop, to respect the headway constraints. In this case, the fixed waiting time applies to the lowest speed reached during the supervised trajectory phase.

4. Numerical study and insights

The numerical experiments and insights are presented in this section, starting in Section 4.1 with the computational setting. Section 4.2 analyzes a yellow signal event and how it can propagate to follower trains. Section 4.3 examines the performance of the system under different stochastic models. Section 4.4 studies the effect of the three peak reduction strategies. Finally, Section 4.5 highlights the key trade-offs between traffic regularity and energy efficiency objectives.

4.1. Parameters and computational setup

The simulation parameters employed are summarized in Table 1 and largely follow Corman et al. (2021) for the stochastic process models and Wang et al. (2020) for the train parameters (e.g., resistance parameters), also adopted by Trivella et al. (2021). We consider a system of $N = 6$ consecutive trains, which may occur in high capacity corridors. These trains are initially spaced at a regular headway d_0 one from another, and assumed to be identical, i.e., subject to the same dynamics, which is common for homogeneous urban and intercity rail traffic (Abbas-Turki et al., 2011).

To estimate the KPIs described in Section 3.2, we simulated 5000 trajectories of this system for a time horizon $T = 2000$ s discretized with $\Delta t = 1$ s. In this setting, the total computation time for the simulation was roughly 40 and 72 s for the OU and DMR model, respectively, when using Matlab R2021b on a laptop with a processor i7-10610U and 16 GB RAM.

Concerning the deterministic model, we consider two specifications, named DET_0 and DET_+, at varying initial conditions. DET_0 is such that $\bar{v}_n = 35$ m/s for all trains, whereas DET_+ uses different speeds $\bar{v}_1 = 35$ m/s and $\bar{v}_n = 36$ m/s for $n = 2, \dots, 6$ (see the related discussion in Section 3.1). For these models, the computational time to estimate the KPIs is negligible as one trajectory is sufficient to fully characterize the system.

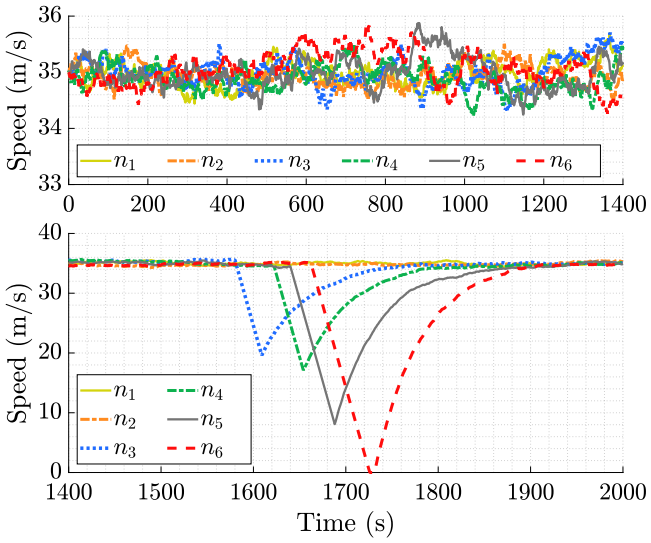

Fig. 4. Time-speed profiles of six trains in one trajectory.

Regarding the storage, since the aim is to assess the value of the extra flexibility brought by this technology rather than characterizing it, we assume it has perfect efficiency, enough capacity to store the energy produced by braking, and that injection and withdrawal rates are non-binding. We do however analyze parameter μ in Algorithm 2 as it defines the storage operating policy.

The results and insights are discussed next. Recall that Sections 4.2 and 4.3 do not make use of the peak reduction strategies, whose effect is investigated later in Sections 4.4 and 4.5.

4.2. Analysis of a trigger event

In this section, the OU energy profile displayed in the middle panel of Fig. 3 is used to better illustrate the propagation of an individual yellow signal event. The peak in this profile is due to four trains accelerating in a short time frame as shown in the time-speed profiles in Fig. 4. Note that the scale of the y-axis is varied in the top and bottom panels to clearly distinguish the stochastic speed variations (mostly occurring within a band of ±0.5 m/s) from the yellow signal effect.

Train 3 is the first to trigger a yellow signal and decelerate, which causes train 4 to decelerate too, followed by trains 5 and 6. Interestingly, whilst the first affected train decelerates to $v_{\text{APPROACH}} = 20$ m/s before re-accelerating, the downstream vehicles may have to reach lower speeds, and potentially even stop, before the headway is restored and they can accelerate again. This implies that the traffic regularity is more sensitive to yellow signals when more consecutive trains are running.

To visualize the same effect from another angle, Fig. 5 shows the space lost by the different trains compared to a baseline with constant speed v_{CRUISE}. As before, downstream vehicles are disproportionately affected by a triggered yellow signal. In this example, the span between first and last simulated vehicle increases by roughly 3 km during the yellow signal event.

A heatmap of train accelerations is presented in Fig. 6. The different shades of orange correspond to the background variations due to the stochastic process, whereas deceleration and acceleration phases caused by the triggering of a yellow signal are clearly visible by the

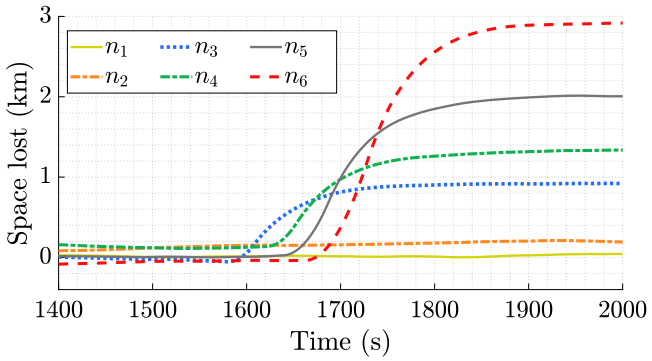

Fig. 5. Space lost with respect to a deterministic speed baseline.

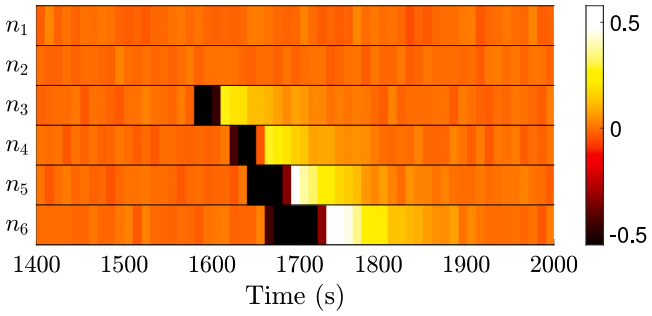

Fig. 6. Synchronization of acceleration and deceleration activities.

Table 2
Traffic regularity KPIs.

KPI	Model	Train (1 is the leader)					
		1	2	3	4	5	6
K1 (% triggers)	OU	0	15.2	29.0	41.7	52.0	60.1
	DMR	0	2.8	4.7	6.0	7.6	8.7
	DET_0	0	0	0	0	0	0
	DET_+	0	100	100	100	100	100
K2 (FTTY (s))	OU	–	1891	1795	1702	1625	1560
	DMR	–	1981	1970	1961	1950	1944
	DET_0	–	>2000	>2000	>2000	>2000	>2000
	DET_+	–	201	228	255	282	309

Table 3
Throughput of the corridor.

KPI	OU	DMR	DET_0	DET_+
K3 (vehicles/hour)	35.4	38.9	39.4	27.0

darker and lighter colors, respectively. By overlaying this heatmap with the trajectories in Fig. 4, notice that the peak point occurs at 1700–1800 s, corresponding to the period in which most trains are indeed accelerating, thus requiring a large joint traction force.

4.3. Analysis of key performance indicators

The aggregate performance of the system is analyzed from a traffic regularity perspective in Tables 2–3 for both stochastic models (OU, DMR) and deterministic models (DET_0, DET_+).

We observe the following important points:

- The OU model incurs many more triggers than the DMR model (see K1), and when this happens, it is earlier for any of the followers as shown by the FTTY indicator K2. These results suggest that an ATO model aware of the location of the traffic ahead leads to a higher level of traffic regularity that enables to better exploit the capacity of the corridor. This is confirmed by an almost 10% higher throughput in DMR compared to OU (see K3).
- The percentage of yellow signals increases substantially when moving from the first follower (train 2) to the last (train 6). This implies that it is common for yellow signals to propagate backwards, that is, a train n decelerating will very likely trigger train $n+1$ to decelerate or brake too. This is a major cause of peaks in energy consumption.
- Deterministic strategies showcase two extreme behaviors: DET_0 is the unrealistic situation of perfect operations, where no yellow signal occurs, whereas in DET_+ all followers triggers a yellow signal due to their higher speed compared to the first leader (although by just 1 m/s).

Table 4
Energy related KPIs.

KPI	OU	DMR	DET_0	DET_+
K4 (kWh)	2876	2814	2800	3312
K5 (kWh/30 s)	64.7	52.0	42.4	83.9
K6 (%)	59.1	9.3	0	100

Next, we now focus on the energy properties of the six-train system and report in Table 4 the related KPIs. Recall that K4, K5, and K6 represent the average total consumption, average maximum consumption, and percentage of energy profiles with a detected peak, respectively (see Section 3.2).

What we observe from the table is the following:

- [K4]: The energy consumption under OU is 2.2% higher than DMR. Indeed trains under DMR less frequently have to decelerate, brake, and accelerate following a yellow signal. This difference translates to a substantial saving of energy and costs with an ATO-based system, especially when considering the consumption of an entire railway network for a year. Compared to the ideal situation represented by DET_0, the total consumption with DMR is only 0.5% higher, while with OU is 2.7% higher. The consumption under DET_+ is very large because this deterministic profile includes a yellow signal with certainty.
- [K5]: OU energy profiles exhibit a highest consumption point that is on average 25% higher than that of DMR profiles, which is a significant difference. Notice that this KPI averages all profiles with and without detected peaks and that an individual peak can exceed 80 kWh/30 s. The two deterministic models display extreme behaviors: DET_0 has a constant, low energy consumption over the horizon, while DET_+ has a peak of about 84 kWh/30s.
- [K6]: Our method detected peaks in about 60% of the OU profiles, while the same number is less than 10% for DMR. Naturally, the estimates of this KPI are quite related to those of K1 for the last follower, underscoring a relation between traffic and energy properties of the system.

The main implication from these results is that an ATO system (represented here by the DMR process) has the potential to not only reduce the overall energy requirements but also to prevent the occurrence of critical peaks in consumption. What is discussed here is also evident from Fig. 7, showing the first 50 energy profiles for OU (top panel) and DMR (middle panel). The figure also shows the deterministic profile DET_+ (bottom panel) and the fixed baseline consumption DET_0.

While a considerable number of OU profiles include at least one peak (32/50), only a few DMR profiles do (3/50). In the considered setting, the peaks often reach double the regular energy consumption level (i.e., with no train affected by a yellow signal), and mostly appear in the second part of the horizon as the initial state of the system fulfills the minimum safety distance. The usefulness of peak reduction strategies is investigated next.

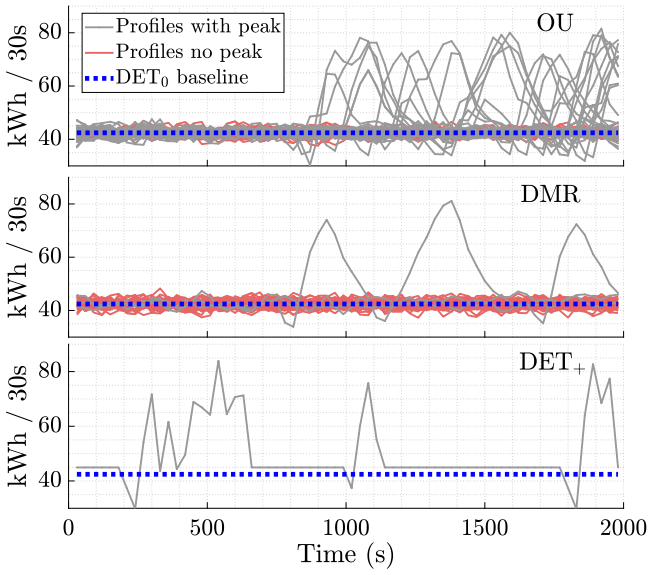

Fig. 7. Simulated energy profiles for model OU, DMR, and DET_+.

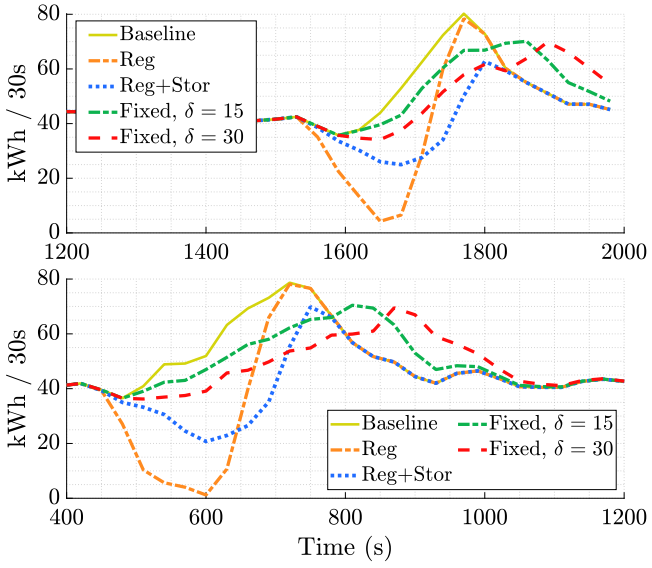

Fig. 8. Energy profiles under different peak reduction strategies.

4.4. Analysis of peak reduction strategies

The focus hereafter is on the OU process, as this is the stochastic model that could benefit most from fewer peaks and smaller peak size. Fig. 8 illustrates the effect of peak reduction strategies on the two energy profiles already shown in the middle and bottom panels of Fig. 3. Specifically, the fixed waiting time rules are considered at varying δ (labeled "Fixed"), and the use of regenerative braking, with storage ("Reg+Stor") and without it ("Reg"), assuming $\phi = 0.7$ (different values have been tested too but the key findings do not change). When managing the storage, $\mu = 1$ is used (see Algorithm 2) while different values of this parameters are analyzed later in this section.

Each strategy has a different effect on the original profile (continuous line). Regenerative energy alone reduces the overall consumption but does not smooth the peak in its highest point. When storage is available, however, the resulting peak is much smaller. The effect of fixed waiting strategies is different as the peak is both reduced in size and spread over time, while it is unclear if the total consumption decreases in these cases.

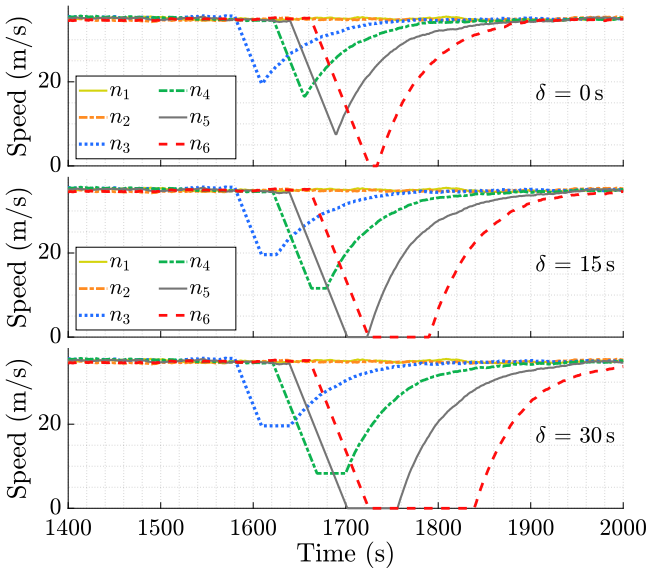

Fig. 9. Speed trajectories during a trigger event at varying waiting time δ.

Table 5
KPIs of the system under different peak reduction strategies.

Technology	KPI	Fixed waiting time δ (s)						
		0	10	20	30	40	50	60
–	K3	35.4	34.0	33.0	31.7	30.6	30.1	29.3
	K4	2876	2877	2870	2864	2856	2845	2837
	K5	64.7	63.6	62.2	61.6	60.6	59.7	59.2
Reg	K3	35.4	34.0	33.0	31.7	30.6	30.1	29.3
	K4	2781	2773	2768	2761	2757	2753	2748
	K5	64.4	63.3	62.0	61.4	60.5	59.6	59.2
Reg+Stor ($\mu = 1$)	K3	35.4	34.0	33.0	31.7	30.6	30.1	29.3
	K4	2795	2788	2780	2775	2769	2765	2763
	K5	59.3	59.5	59.1	58.6	58.1	57.7	57.4

To better understand the aforementioned changes, Fig. 9 illustrates the speed trajectories of the six trains associated with the energy profile reported the top panel of Fig. 8 (which was the same example also considered in Section 4.2), at varying δ. Recall that the use of regenerative braking, with and without storage, does not alter the speed trajectories, whereas fixed waiting rules affect the train dynamics as well. This is why the speed trajectories under regenerative energy are not displayed: they coincide with the baseline trajectories in the top panel of Fig. 9 ($\delta = 0$). From this figure, it is evident that the higher is the waiting time δ, the more the trains are spread apart from each other after a yellow signal, and with that the acceleration activities. Moreover, the more downstream trains may have to wait longer than δ before the headway is restored and they can re-accelerate. Thus, although fixed waiting rules can help in shaving power peaks, they may also slow the traffic down and affect the throughput of the system, as examined next.

Table 5 reports a subset of the system KPIs under regenerative braking, fixed waiting rules, and a combination of them. The main findings from these results are the following:

- [K3]: The throughput decreases, approximately linearly, when the fixed waiting time δ increases. In particular, this KPI goes from 35.4 to 29.3 vehicles/hour when δ varies from 0 to 60 s, which is a 17% reduction.
- [K4]: The total energy consumption is reduced only marginally by variations of δ. However, exploiting regenerative braking brings savings of 3.3% on average.

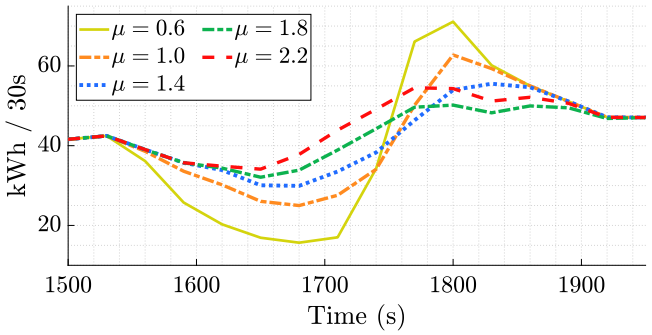

Fig. 10. Energy profile with regenerative braking and storage at varying μ.

Table 6
Policy performance at varying storage threshold multiplier μ.

KPI	0.6	0.8	1.0	1.2	1.4	1.6	1.8	2.0	2.2
K4	2784	2790	2796	2799	2800	2802	2806	2813	2819
K5	62.5	61.2	59.3	57.1	55.2	53	52.1	52.7	53.4
K6	52.2	51.1	49.4	45.5	39.9	31.6	36.3	46.4	51.4

- [K5]: The highest consumption point decreases by 8.5% on average when δ increases from 0 to 60 s. When restricting to energy profiles with detected peaks, this improvement exceeds 10%, which is substantial. Regenerative braking alone does not help much in shaving the peak, while it is very effective when it is coupled with an electricity storage, with an improvement of 8.5%.

Beside δ, another important tunable parameters is μ, which defines the storage operating policy. Thus, we examine how the KPIs change when using regenerative braking with storage at varying μ. Fig. 10 shows an energy profile under values of μ from 0.6 to 2.2. Increasing μ is beneficial to smooth the peak and, for instance, the highest point in this profile decreases from 71 to 50 kWh when μ goes from 0.6 to 1.8. However, increasing μ from 1.8 to 2.2 results in a higher peak of 54.5 kWh, and a higher energy consumption too as it can be seen from the area under the profiles. In fact, μ represents a threshold for the storage to be discharged and setting too high values of it may result in very conservative policies.

Table 6 reports estimates of the energy KPIs for operating strategies at varying μ (notice that the traffic KPIs K1–K3 do not change with μ; hence, they are not included in this table). The energy consumption (K4) indeed increases with μ but only marginally. However, the percentage of trajectories with a peak (K6) varies from 31.6% to 51.4% when μ is increased from 1.6 to 2.2, which is a substantial increase indicating that the behavior of K6 is not monotone in μ. The same holds for K5 (average peak), which shows a small increment after $\mu = 2$. Overall, the value $\mu = 1.6$ achieves the best trade-off for the considered experimental setup.

4.5. Trade-offs between objectives

The findings in Sections 4.3–4.4 underscored several conflicts that exist among different objectives, especially between the traffic-related and the energy-related KPIs. Some of the trade-offs that arise under different peak reduction strategies are investigated more systematically in the following.

Fig. 11 illustrates 4 trade-offs under 4 strategies. Recall that "Reg", "Stor", and "Fixed" in the name denote, respectively, the use of regenerative energy, storage, and fixed waiting rules. Strategies with "Fixed" are studied for varying $\delta \in [0, 60]$, while "Reg+ Stor" for varying $\mu \in [0.6, 2.2]$. "Fixed+ Reg+ Stor" uses $\mu = 1.6$. To help identifying the best solutions, in each subplot, the improving direction of the trade-off is indicated by a star-shaped marker. In other words, the best solutions are those that are as close as possible to the marked corner.

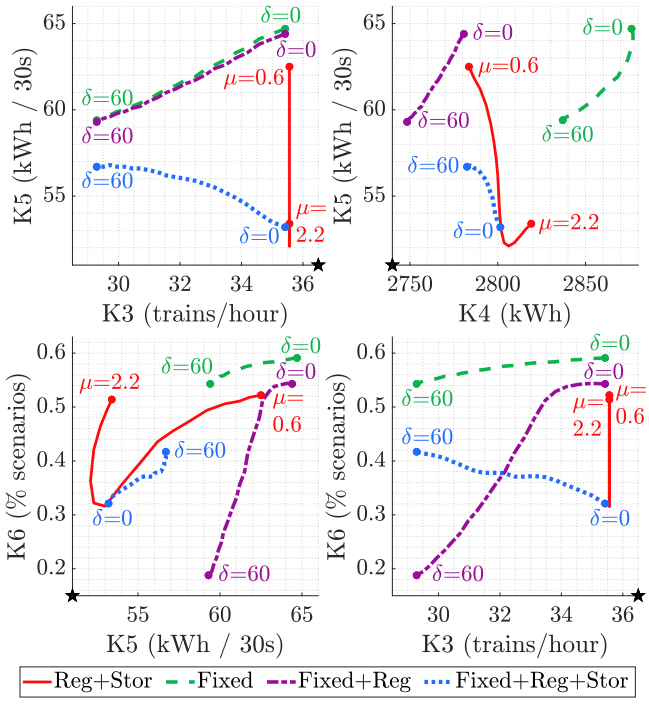

Fig. 11. Trade-off between KPIs under different peak reduction measures.

Drawing conclusions about these results is non-trivial as no strategy dominates the others in managing all KPIs. Fixed waiting rules behave quite differently depending on whether regenerative braking and storage are used. In general, coupling regenerative braking and a well tuned storage operating policy, i.e., with $\mu = 1.6$, seems to work well in most cases. However, deviating from this value of μ always worsens at least one KPI. Similarly, the effect of varying δ in the other strategies is mixed and never improves all KPIs jointly. The main takeaway from this analysis is that practical approaches to reduce power peaks in busy railway corridors should be designed carefully and tested across multiple dimensions, limiting the risk that improving one KPI negatively affects others.

5. Conclusion

This work is the first to analyze a dynamic railway traffic system involving a platoon of leader–follower trains. By simulating different stochastic processes, we observed that the propagation of yellow signals in such a system and the consequent simultaneous acceleration of trains may generate significant peaks in energy use, which is a main concern in the modern railway industry due to high energy prices and environmental concerns. We thus proposed and analyzed different peak reduction strategies that rely on technological assumptions (regenerative braking, energy storage) and/or train control (fixed waiting rules).

In an extensive numerical study, the performance of the dynamic system is examined in terms of traffic regularity and energy consumption, and insights are provided on the effectiveness of different peak reduction strategies, potentially combined. For example, a key finding is that using regenerative energy alone (i.e., based on an instantaneous matching of acceleration and deceleration) is significantly less effective than combining it with energy storage devices for managing and redistributing energy flows over time. The storage operating policy itself has a significant impact on the extent of power peak shaving and hence needs to be designed and tested carefully.

This paper also investigated the benefits of an ATO controller for both smoothing peaks and improving traffic regularity compared to a

human driver. Conceptually, the superior performance of the former model stems from the sharing of information (in particular, location data) across vehicles, and the exploitation of this data to improve control under stochastic disturbances that occur in real operations. Thus, this study contributes to assessing the potential benefits of ATO systems, which is much needed in practice. Finally, this paper identified and evaluated key trade-offs and conflicts that arise between objectives, suggesting that energy-related and traffic-related KPIs must be considered jointly when designing or analyzing advanced railway systems.

Future research directions include calibrating the parameters of the stochastic processes based on real data. At the time of writing, a comprehensive dataset to allow for such calibration is not available to the authors, and these parameters are defined based on the railway literature and simulation experiments. Moreover, additional sensitivity analysis may be conducted on other parameters than the ones studied here, such as the initial headway between trains, examining how this relates to the common target of achieving a 80% of capacity utilization (see UIC Code, Norm 406).

Another avenue would be extending the stochastic simulation framework developed in this paper from a corridor to an entire railway network, which is relevant to better describe network effects that go beyond the interaction of rail vehicles based on leader–follower dynamics.

Finally, the paper employs relatively simple custom algorithms for power peak reduction. Future research could target developing alternative peak reduction strategies based on more advanced optimization frameworks, for instance, to operate the energy storage or to coordinately control vehicles when a yellow signal has been triggered. As shown in Section 4.5, finding a good peak reduction strategy entails solving an optimization problem with many conflicting objectives given by the different KPIs. Moreover, calculating objective values is expensive as it requires simulating thousands of trajectories from a stochastic systems. For these reasons, promising approaches include hybrid heuristics and metaheuristics like evolutionary multi-objective optimization methods (Coello, 2006). Since the problem also includes inherent randomness and robustness characteristics, adaptive algorithms and reinforcement learning methods could be developed too, which have proven successful to manage both energy storage (Powell & Meisel, 2015) and train rescheduling (Jusup et al., 2021). Although embedding such techniques in the proposed simulation framework may be computationally challenging, a significant body of literature across diverse domains (e.g., energy and transport) has shown that advanced optimization algorithms have the potential to enhance the solution approach for challenging decision problems (see, e.g., Corman & Meng, 2014 and Frangopoulos, 2018.)

CRediT authorship contribution statement

Alessio Trivella: Conceptualization, Methodology, Software, Validation, Writing – original draft, Writing – review & editing, Visualization. **Francesco Corman:** Conceptualization, Methodology, Writing – review & editing.

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.

Data availability

The link to the code is shared in the paper

Acknowledgments

This paper builds on an extended abstract that was accepted for presentation at the 10th Symposium of the European Association for Research in Transportation (hEART 2022; Trivella & Corman, 2022). The second author was supported by the Swiss National Science Foundation under Project 1481210/DADA.

References

Abbas-Turki, A., Zaremba, E., Grunder, O., & El-moudni, A. (2011). Perfect homogeneous rail traffic: A quick efficient genetic algorithm for high frequency train timetabling. In *2011 14th International IEEE conference on intelligent transportation systems* (pp. 1495–1500). IEEE.

Albrecht, T. (2010). Reducing power peaks and energy consumption in rail transit systems by simultaneous train running time control. *WIT Transactions on State-of-the-Art in Science and Engineering*, *39*.

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

Banić, M., Miltenović, A., Pavlović, M., & Ćirić, I. (2019). Intelligent machine vision based railway infrastructure inspection and monitoring using UAV. *Facta Universitatis, Series: Mechanical Engineering*, *17*(3), 357–364.

Bärmann, A., Martin, A., & Schneider, O. (2017). A comparison of performance metrics for balancing the power consumption of trains in a railway network by slight timetable adaptation. *Public Transport*, *9*, 95–113.

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

Brandenburger, N., Naumann, A., & Jipp, M. (2021). Task-induced fatigue when implementing high grades of railway automation. *Cognition, Technology & Work*, *23*, 273–283.

Cacchiani, V., & Toth, P. (2018). Robust train timetabling. In *Handbook of optimization in the railway industry* (pp. 93–115). Springer.

Caprara, A., Fischetti, M., & Toth, P. (2002). Modeling and solving the train timetabling problem. *Operations Research*, *50*(5), 851–861.

Coello, C. C. (2006). Evolutionary multi-objective optimization: a historical view of the field. *IEEE Computational Intelligence Magazine*, *1*(1), 28–36.

Corman, F., & Meng, L. (2014). A review of online dynamic models and algorithms for railway traffic management. *IEEE Transactions on Intelligent Transportation Systems*, *16*(3), 1274–1284.

Corman, F., Trivella, A., & Keyvan-Ekbatani, M. (2021). Stochastic process in railway traffic flow: Models, methods and implications. *Transportation Research Part C (Emerging Technologies)*, *128*, Article 103167.

De La Torre, S., Sánchez-Racero, A. J., Aguado, J. A., Reyes, M., & Martínez, O. (2014). Optimal sizing of energy storage for regenerative braking in electric railway systems. *IEEE Transactions on Power Systems*, *30*(3), 1492–1500.

De Martinis, V., & Corman, F. (2018). Data-driven perspectives for energy efficient operations in railway systems: Current practices and future opportunities. *Transportation Research Part C (Emerging Technologies)*, *95*, 679–697.

De Simone, L., Caputo, E., Cinque, M., Galli, A., Moscato, V., Russo, S., Cesaro, G., Criscuolo, V., & Giannini, G. (2023). LSTM-based failure prediction for railway rolling stock equipment. *Expert Systems with Applications*, Article 119767.

EU (2021). European union - 2030 climate target plan. Accessed 20 March 2023.

Frangopoulos, C. A. (2018). Recent developments and trends in optimization of energy systems. *Energy*, *164*, 1011–1020.

González-Gil, A., Palacin, R., & Batty, P. (2013). Sustainable urban rail systems: Strategies and technologies for optimal management of regenerative braking energy. *Energy Conversion and Management*, *75*, 374–388.

González-Gil, A., Palacin, R., Batty, P., & Powell, J. (2014). A systems approach to reduce urban rail energy consumption. *Energy Conversion and Management*, *80*, 509–524.

Hansen, I., & Pachl, J. (2014). *Railway timetabling and operations: Analysis, modelling, optimisation, simulation, performance, evaluation*. Eurail press, Hamburg.

Ho, T. K., Tsang, C. W., Ip, K. H., & Kwan, K. (2012). Train service timetabling in railway open markets by particle swarm optimisation. *Expert Systems with Applications*, *39*(1), 861–868.

Horizon Europe (2022). Digital & Automated up to Autonomous Train Operations. Accessed 20 March 2023.

Howlett, P. (2000). The optimal control of a train. *Annals of Operations Research*, *98*(1), 65–87.

Jusup, M., Trivella, A., & Corman, F. (2021). A review of real-time railway and metro rescheduling models using learning algorithms. In *30th International joint conference on artificial intelligence (IJCAI-21)*.

Khodaparastan, M., Mohamed, A. A., & Brandauer, W. (2019). Recuperation of regenerative braking energy in electric rail transit systems. *IEEE Transactions on Intelligent Transportation Systems*, *20*(8), 2831–2847.

Liu, X., & Li, K. (2020). Energy storage devices in electrified railway systems: A review. *Transportation Safety and Environment*, *2*(3), 183–201.

Lu, S., Weston, P., Hillmansen, S., Gooi, H. B., & Roberts, C. (2014). Increasing the regenerative braking energy for railway vehicles. *IEEE Transactions on Intelligent Transportation Systems*, *15*(6), 2506–2515.

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

Luo, J., Peng, Q., Wen, C., Wen, W., & Huang, P. (2022). Data-driven decision support for rail traffic control: A predictive approach. *Expert Systems with Applications*, *207*, Article 118050.

OWiD (2016). Our world in data - emissions by sector. Accessed 20 March 2023.

Poulus, R., van Kempen, E., & van Meijeren, J. (2018). *Automatic train operations: Driving the future of rail transport*: *Technical report*, TNO.

Powell, W. B., & Meisel, S. (2015). Tutorial on stochastic optimization in energy—Part II: An energy storage illustration. *IEEE Transactions on Power Systems*, *31*(2), 1468–1475.

Qin, Y., Hu, X., He, Z., & Li, S. (2020). Longitudinal emissions evaluation of mixed (cooperative) adaptive cruise control traffic flow and its relationship with stability. *Journal of the Air & Waste Management Association*, *70*(7), 670–686.

Quaglietta, E., Wang, M., & Goverde, R. M. (2020). A multi-state train-following model for the analysis of virtual coupling railway operations. *Journal of Rail Transport Planning & Management*, *15*, Article 100195.

Railenergy (2016). Innovative integrated energy efficiency solutions for railway rolling stock, rail infrastructure and train operation. Accessed 20 March 2023.

Railtech (2021). Netherlands decides about ATO implementation by end of 2025. Accessed 20 March 2023.

Ran, X.-C., Chen, S.-K., Liu, G.-H., & Bai, Y. (2020). Energy-efficient approach combining train speed profile and timetable optimisations for metro operations. *IET Intelligent Transport Systems*, *14*(14), 1967–1977.

Ratniyomchai, T., Hillmansen, S., & Tricoli, P. (2014). Recent developments and applications of energy storage devices in electrified railways. *IET Electrical Systems in Transportation*, *4*(1), 9–20.

Regueiro Sánchez, D. (2021). *Quantification and reduction of power peaks in railway networks: a simulation-based approach* (Master's thesis), ETH Zurich.

SBB (2021). Load management - smart grid at SBB. Accessed 17 May 2023.

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

Sels, P., Cattrysse, D., & Vansteenwegen, P. (2016). Automated platforming & routing of trains in all belgian railway stations. *Expert Systems with Applications*, *62*, 302–316.

Singh, P., Dulebenets, M. A., Pasha, J., Gonzalez, E. D. S., Lau, Y.-Y., & Kampmann, R. (2021). Deployment of autonomous trains in rail transportation: Current trends and existing challenges. *IEEE Access*, *9*, 91427–91461.

Singh, P., Elmi, Z., Meriga, V. K., Pasha, J., & Dulebenets, M. A. (2022). Internet of things for sustainable railway transportation: Past, present, and future. *Cleaner Logistics and Supply Chain*, *4*, Article 100065.

Tang, R., De Donato, L., Besinović, N., Flammini, F., Goverde, R. M., Lin, Z., Liu, R., Tang, T., Vittorini, V., & Wang, Z. (2022). A literature review of artificial intelligence applications in railway systems. *Transportation Research Part C (Emerging Technologies)*, *140*, Article 103679.

Tashmetov, K., Aliev, R., Aliev, M., & Tashmetov, T. (2022). Expert system for diagnosing faults railroad switch of automation and telemechanic systems. In *AIP conference proceedings, Vol. 2432*. AIP Publishing LLC, Article 030083.

Trivella, A., & Corman, F. (2019). Modeling uncertainty dynamics in public transport optimization. In *19th Swiss transport research conference (STRC 2019)*. STRC.

Trivella, A., & Corman, F. (2022). An analysis of power peaks in stochastic models of railway traffic. In *HEART 2022 - 10th symposium of the European association for research in transportation*.

Trivella, A., Wang, P., & Corman, F. (2021). The impact of wind on energy-efficient train control. *EURO Journal on Transportation and Logistics*, *10*, Article 100013.

UIC (2017). International union of railways - energy efficiency and CO2 emissions. Accessed 20 March 2023.

Wang, P., Bešinović, N., Goverde, R. M., & Corman, F. (2022). Improving the utilization of regenerative energy and shaving power peaks by railway timetable adjustment. *IEEE Transactions on Intelligent Transportation Systems*.

Wang, M., Li, H., Gao, J., Huang, Z., Li, B., & Van Arem, B. (2017). String stability of heterogeneous platoons with non-connected automated vehicles. In *2017 IEEE 20th international conference on intelligent transportation systems (ITSC)* (pp. 1–8). IEEE.

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

Wang, G., Xu, T., Tang, T., Yuan, T., & Wang, H. (2017). A Bayesian network model for prediction of weather-related failures in railway turnout systems. *Expert Systems with Applications*, *69*, 247–256.

Yang, X., Chen, A., Ning, B., & Tang, T. (2016). A stochastic model for the integrated optimization on metro timetable and speed profile with uncertain train mass. *Transportation Research, Part B (Methodological)*, *91*, 424–445.

Yang, X., Li, X., Ning, B., & Tang, T. (2015). A survey on energy-efficient train operation for urban rail transit. *IEEE Transactions on Intelligent Transportation Systems*, *17*(1), 2–13.

Ye, H., & Liu, R. (2017). Nonlinear programming methods based on closed-form expressions for optimal train control. *Transportation Research Part C (Emerging Technologies)*, *82*, 102–123.

Yin, J., Tang, T., Yang, L., Gao, Z., & Ran, B. (2016). Energy-efficient metro train rescheduling with uncertain time-variant passenger demands: An approximate dynamic programming approach. *Transportation Research, Part B (Methodological)*, *91*, 178–210.

Zhang, Y., Bai, Y., Hu, J., & Wang, M. (2020). Control design, stability analysis, and traffic flow implications for cooperative adaptive cruise control systems with compensation of communication delay. *Transportation Research Record*, *2674*(8), 638–652.