A review of train delay prediction approaches

Thomas Spanninger [a],[*], Alessio Trivella [b], Beda Büchel [a], Francesco Corman [a]

[a] *Institute for Transport Planning and Systems, ETH Zurich, Stefano-Franscini-Platz 5, 8093, Zürich, Switzerland*
[b] *Industrial Engineering and Business Information Systems, University of Twente, 7500 AE, Enschede, the Netherlands*

ABSTRACT

Railway operations are vulnerable to delays. Accurate predictions of train arrival and departure delays improve the passenger service quality and are essential for real-time railway traffic management to minimise their further spreading. This review provides a synoptic overview and discussion covering the breadth of diverse approaches to predict train delays. We first categorise research contributions based on their underlying modelling paradigm (data-driven and event-driven) and their mathematical model. We then distinguish between very short to long-term predictions and classify different input data sources that have been considered in the literature. We further discuss advantages and disadvantages of producing deterministic versus stochastic predictions, the applicability of different approaches during disruptions and their interpretability. By comparing the results of the included contributions, we can indicate that the prediction error generally increases when broadening the prediction horizon. We find that data-driven approaches might have the edge on event-driven approaches in terms of prediction accuracy, whereas event-driven approaches that explicitly model the dynamics and dependencies of railway traffic have their strength in providing interpretable predictions, and are more robust concerning disruption scenarios. The growing availability of railway operations data is expected to increase the appeal of big-data and machine learning methods.

1. Introduction

Railway operations are typically based on schedules, defined as "a routine vehicle movement between two or more points at pre-scheduled times, transporting passengers or cargo" (Banerjee et al., 2019). One of the most important service level criterion for railway companies is to operate according to the timetable with minimal deviation – namely to operate punctually (Nabian et al., 2019). The high variability in operational process times and unexpected disruptive events lead to *train delays,* i.e., deviations of the realised arrival and departure times from the planned timetable (Zilko et al., 2016). These delays are inconvenient for passengers and railway operators.

Railway networks constitute a heavily intertwined system as numerous train vehicles share a limited infrastructure of tracks. A single train running behind its schedule is likely to hinder other trains using the same tracks leading to cascading consequences. Due to this high degree of inter-dependency in railway operations, initial delays often propagate to other trains in the railway network (Yuan and Hansen, 2007).

Predicting train delays is an important service quality for passengers reducing the uncertainty about their arrival/departure times (Corman and Kecman, 2018). Additionally, train delay predictions are crucial for anticipative traffic control in real-time. Traffic

---

controllers try to minimise the propagation of delays in railway networks and nowadays base their decisions mainly on experience and real-time information about delays (Hansen et al., 2010). Therefore, the systematic prediction of train delay propagation to be used for dispatching decisions can increase the quality of traffic control significantly, and prevent delays from spreading across the network. Predicting train delays is a soft operational action requiring only marginal investments compared to other management actions to improve the service quality, such as using extra vehicles, stand-by drivers or infrastructural investments.

From a structural point of view, railway services can be divided into passenger and freight services. In terms of spatial coverage as well as technical and operation complexity, we further distinguish between urban railways (metro & tramway), traditional railways, and high-speed railways. Although the precise problem of predicting the future of a train delay may vary based on the nature of the railway service, we can generally formalise this problem by defining a function f, mapping a space of input information Ω to the potential delay realisations Φ:

$$f : \Omega \to \Phi.$$

Assume that train i is planned to be at location x at time t_p. Then, the delay D_t of train i being at location x, computed at a given point in time $t \leq t_p$, is predicted with:

$$D_t\left(i, x, t_p\right) = f(\Omega_t),$$

where the input information Ω_t is conditioned on the current time t.

During the last decades, a variety of approaches has been developed in the context of train delay prediction. This diversity is due to different considered types of railway systems and precise problem definitions. Specifically, the prediction horizons (i.e., $t_p - t$; ranging from very short-term predictions for a few minutes ahead to predictions of delays several hours in advance), the availability and choice of input information Ω_t, the definition of the realisation space Φ, and also the functional form of mapping $f(.)$ may all vary. At a higher level, approaches differ by being data-driven or event-driven, which we refer to as their *modelling paradigm* hereafter and discuss later.

To navigate and understand this multitude of train delay prediction approaches, we classify and discuss the contributions in the literature based on multiple criteria: the underlying modelling paradigm (data-driven vs event-driven), the specific mathematical model, the input data used, the type of prediction (deterministic vs stochastic) and its horizon. We further discuss advantages and disadvantages of applying different prediction techniques for small delays or during a disruption. Our goal is to provide some insights on which prediction model and data to use in a specific situation, and to identify and discuss research gaps and recent trends, i.e, the impact of increasing availability of railway operations data on delay prediction. Summarising, this review contributes to the field of train delay predictive analytics by:

1. Providing a synoptic overview and meaningful categorisations of different approaches;
2. Discussing the used data sources, types of predictions, and prediction horizons;
3. Determining research trends and identifying topics of growing interest.

We believe this review on train delay prediction approaches is timely due to the widely increasing number of scientific publications on the topic. From a practical standpoint, predictive analytics matches well the strong interest from industry to digitalise and optimise railway operations using quantitative methods.

The remainder of the paper is structured as follows. Section 2 explains the methodology behind our review, discusses the publications in scope qualitatively, and introduces the modelling paradigms of event-driven and data-driven approaches. In Section 3, we discuss the different input data sources and their role. In Section 4, we highlight advantages and disadvantages of applying the methods identified in specific contexts. We summarise our findings and future research needs in Section 5.

2. Methodology and modelling paradigm

The methodology of our review is based on Wee and Banister (2016). We started our literature search by looking for combinations of keywords "train", "railway", "delay", "forecasting" and "prediction" in the titles, keywords and abstracts of railway related publications in the well-known scientific search engines of Google Scholar, Scopus and Web of Science and consequently screened their references. We restricted the inclusion of contributions to journal publications and conference proceedings written in the English language. According to the categorisation made by Ghofrani et al. (2018) of descriptive, predictive and prescriptive analytics, we focus only on *predictive* approaches for train delays, finally obtaining a set of 71 publications up to September 2021.

Two recent review papers have some commonalities with the field of train delay prediction but focus on different problems. Specifically, Banerjee et al. (2019) surveyed the prediction of passenger demand for scheduled transport, while Bešinović (2020) considered resilience in railway transport systems. Also worth mentioning are the surveys of Karlaftis and Vlahogianni (2011) on statistical methods and artificial neural networks in transportation research and of Ghofrani et al. (2018) on big data analytics in railway transportation systems. For the prescriptive analytics side, i.e., real-time railway rescheduling models, we refer to the review of Cacchiani et al. (2014).

2.1. Distributions of contributions over time and per scientific journal

Among the 71 reviewed publications, almost 70% are journal publications. The number of yearly contributions has increased

sharply in the recent few years, as shown in Fig. 1. For the sake of visualisation, the label "before 2000" collects five publications from the years 1988, 1990, 1994, 1996, and 1998.

Table 1 reports the scientific journals in which most contributions have appeared, also divided in publication periods. Most published works are found in well-established transportation-related outlets like *Transport Research Part C*, *Transport Research Part B* and *Transportation Science*. The majority of journal contributions has been published in the last six years.

2.2. Classification of prediction models

We classify the approaches according to the formulation of the prediction function $f(.)$ introduced in Section 1. The most used modelling classes to predict train delays are graph models and machine learning methods. Additionally, some works have applied queuing models, equation system models and time-series analysis investigations.

Specifically, we distinguish train delay prediction approaches based on their inherent *modelling paradigm*. We classify approaches that explicitly capture dependencies of train-events (departure, arrival and pass-through) as *event-driven*. Their main intention is to explicitly model railway operation dynamics (procedures and restrictions). For example, the arrival of a train at the second station along its journey, logically, can only take place after the train has departed from the initial station. In terms of infrastructure restrictions, only one train at a time can occupy a track section or stop at a platform (headway constraints). This modelling paradigm is naturally associated with multi-step predictions, since predicting near-future train-events directly affects predictions for delays occurring later. Event-driven approaches with a train-event dependency structure are primarily based on either a graph model or an equation system.

We label instead approaches as *data-driven* if they are not based on an explicitly modelled train-event dependency structure, which extends along prediction horizon. Hence data-driven approaches typically generate one-step predictions also for train-events at the end of the prediction horizon (e.g., at the tenth-next stop), without explicitly modelling the dynamics of traffic flow between time t and time t_p but directly predicting the delay at the desired station or point in time. Most data-driven models are, in fact, machine learning models.

To illustrate the differences in these modelling paradigms, assume that train i has just left its initial station x_1 at time t with the observed departure delay D_t. Now (at time t), we aim to predict train i's arrival delay at its nth station $D_t(i, x_n, t_n)$, planned for time t_n. An event-driven approach would apply function $f(\Omega_t)$ to predict the delay of train i at the second station x_2 ($D_t(i, x_2, t_2)$), then use $D_t(i, x_2, t_2)$ in addition to other input to predict the delay at the third station x_3, et cetera. This iterative process can be seen as an nth functional power or composite prediction function:

$$D_t(i, x_n, t_n) = f(f(f(\ldots f(\Omega_t)))).$$

Note that the parameters of the mapping function $f(.)$ can also vary for time and location. In contrast, a data-driven approach would directly predict the delay $D_t(i, x_n, t_n)$ by a single application of $f(\Omega_t)$. Therefore, the event-driven prediction process can be seen as a chain of prediction steps for the delay at subsequent stations: $x_1 \rightarrow x_2 \rightarrow \ldots x_{n-1} \rightarrow x_n$, whereas data-driven predictions are typically generated directly without using the intermediate predictions: $x_1 \rightarrow x_i$ for $i = 1, \ldots, n$. In our literature study, we found 29 contributions classified as event-driven and 42 as data-driven. In some cases, this classification is not very sharp due to hybrid modelling elements. We describe more in detail the most important mathematical models within these categories in Section 4. Before, we discuss next the different input data sources.
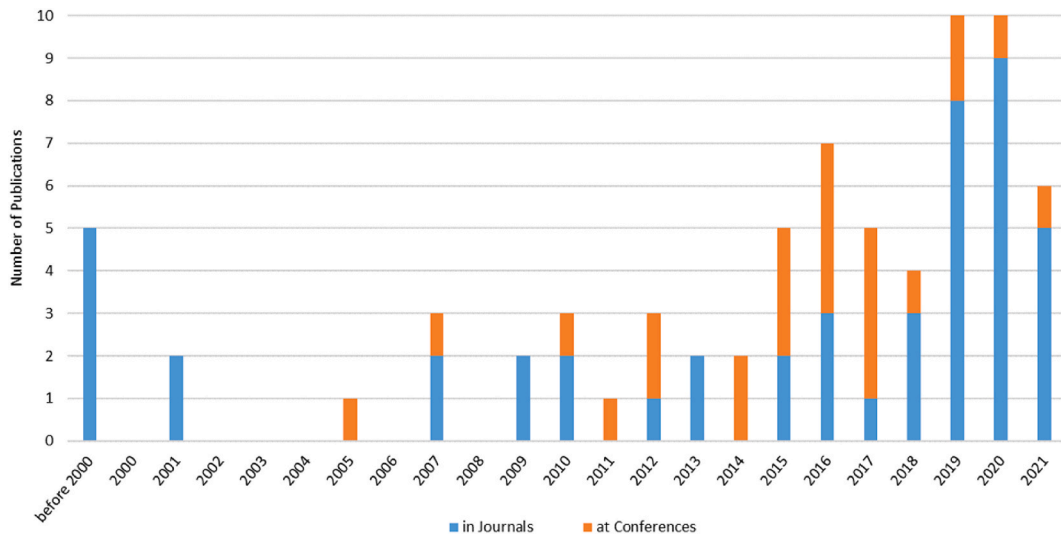

Fig. 1. Number of publications in scope of this review per year in journals and at conferences.

Table 1

Journals ranked by number of publications on train delay predictions.

Journal	Number of journal publications		
	until 2015	2016–2021	All
Transportation Research Part C	1	5	6
Transportation Research Part B	4	0	4
Transportation Science	4	0	4
Transportation Research Part E	2	1	3
Journal of Rail Transport Planning and Management	1	1	2
Journal of Advanced Transportation	1	1	2
International Journal of Rail Transportation	0	2	2
Transportation Research Record	0	2	2
Recent Advances in Computer Science and Communications	0	2	2
Other journals with one publication	4	16	20
TOTAL	**17**	**30**	**47**

3. Input data

We generalise existing classifications focusing on train delays (e.g., Lee et al. (2016a, 2016b) considered the four categories: station-related, train-related, operation-related and timetable-related delays) by categorising the sources of inputs Ω_t into four categories that capture the entire range of input information for train delay prediction approaches:

- Historic train movements (HTM);
- Infrastructure information (INI);
- Operational information (OPI);
- External factors and weather (EFW).

HTM data covers observations of train-specific scheduled/realised departure and arrival times. The spatial resolution and temporal
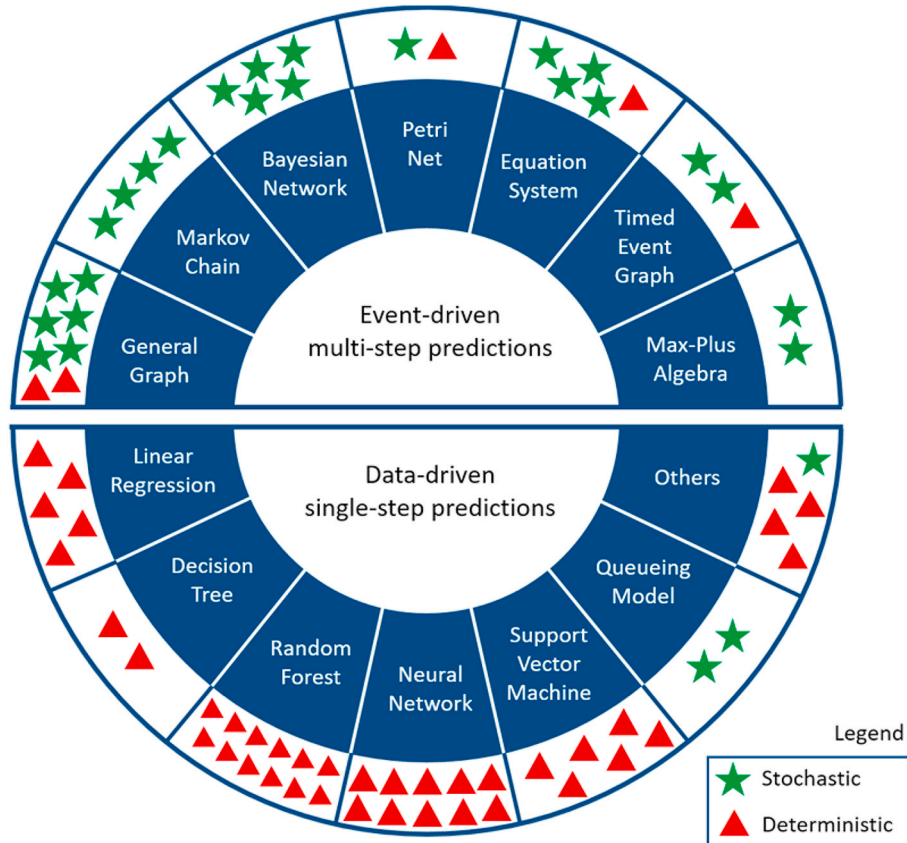
Fig. 2. Classification of train delay prediction approaches.

granularity of HTM data depend on the railway network and on the railway operator's data acquisition and management processes. It may also include pass-through observations at important points of operation (e.g., switches) and basic information about the rolling stock. Typically, railway operators use train describer data to build up an HTM database. Train describer systems keep track of train positions based on train numbers and messages received from elements of the signalling and interlocking systems (sections, switches and signals) (Kecman and Goverde, 2012). Goverde and Hansen (2000), Goverde and Meng (2011) and Kecman and Goverde (2012) discuss how to process train describer records towards input datasets for train delay predictions. In many countries, operators provide open access to HTM data. The most recent (last) available HTM data is particularly relevant for real-time delay predictions, as it includes the knowledge of the actual delay at the latest arrival, departure or pass-through of every running train in the system. Nowadays, more and more railway companies are collecting and publishing these real-time operational data (Ghofrani et al., 2018). The recent increase in data availability definitely boosts the application of state-of-the-art data science and machine learning techniques to analyse train movements and predict train delays.

Information about the typology of a station, section length, number of tracks and switches in a section, overtaking possibilities, and average super-elevation, gradient, and radius of a section are all examples of INI. Data-driven approaches often use location-specific INI as explanatory variables for the prediction. Event-driven approaches sometimes make use of INI to build a dependency structure for train-events.

OPI covers the timetable, crew schedule, buffer times, planned train connections, waiting policies and other potential operational restrictions that are used when operating the railway system, including constraints from the signalling system. In particular, some approaches use section-specific catch-up potential or buffer times, minimal headway times, planned connections or waiting time policies as an explicit source of information (e.g., Berger et al., 2011b; Gao et al., 2020; Goverde, 2010). A few approaches consider the schedules of rolling stock and crew as input data to predict train delays (e.g., Barbour et al., 2018; Li et al., 2020).

Local information about weather indicators like rainfall, temperature, humidity, wind speed, snowfall and visibility are EFW data that have been used in the literature as input for train delay prediction (e.g., Arshad and Ahmed, 2019; Hauck and Kliewer, 2020; Huang et al., 2020b; Jiang et al., 2019b; Sara et al., 2020; Zhou et al., 2020). For instance, Wang and Zhang (2019) combined weather indicators into a single input variable. Another example of EFW information is the presence of a big sporting event or other types of events taking place.

4. Models

We now review the mathematical models applied in the literature and discuss the most important contributions. Fig. 2 provides an overview of the models classified according to their underlying modelling paradigm (event- or data-driven) introduced in Section 2. The markers in the outer circle visualise the number of contributions per model category and are divided by deterministic and stochastic output. We call approaches that provide probability distributions for train delays *stochastic*, whereas *deterministic* approaches result in single-value predictions. It is evident from Fig. 2 that most event-driven approaches produce stochastic predictions (the upper half of the outer circle is predominantly green), whereas data-driven approaches typically result in deterministic predictions (the bottom half of the outer circle is mostly red). This is mainly because event-driven approaches aim to explicitly model process time variabilities and consequently produce potential (probabilistic) predictions of the future dynamics of delay evolution. Additionally, we can observe, that some models are more popular than others, with Neural Networks and Random Forest models being the most widely used on the data-driven side.

4.1. Event-driven approaches

The core idea of event-driven approaches is to explicitly capture and model the dependencies of train arrival, departure and pass-through events in the prediction function $f(.)$. In this way, a chain of consecutive train-events or a network of dependent train-events is constructed for the time horizon of the prediction. For example a late departure of a train at the station 1 influences the arrival delay at station 2 and all other stops downstream of its journey, and might also result in a delay of the following arriving train at station 1. The majority of event-driven contributions rely on a graph model but equation systems models have also been used to define a dependency structure. In the following paragraphs, we outline contributions proposing event-driven models.

4.1.1. General graph models

Graph models generally consist of a collection of nodes and edges. In the context of railway analytics, graphs describe train-events and dependencies in-between them. Nodes represent a state of delay at a train-event (a value or, more in general, a probability function) and directed edges model the impact from one node to the consequent node. Many works on predicting train delays are based on graphs. They represent explicitly a dependency of factors by the nodes and arcs in the graph, which matches the dependencies of train events. They aim to quantify the importance of these dependencies for the process of train delay propagation.

Carey and Kwieciński (1994) analytically model train dependencies explicitly and evaluate the effect of headway constraints. Hallowell and Harker (1996) predict passenger train delays on a partially double-track line, including departure time uncertainty. Yuan and Hansen (2007) propose an analytical stochastic convolution model to estimate delay propagation in stations considering scheduled transfer connections and train conflicts.

Goverde (2010) argues for the first time that train delay prediction problems display the underlying structure of a timed event graph (TEG). Berger et al. (2011b) use waiting policies, driving time profiles, and catch-up potentials explicitly to build a stochastic graph model for train delay prediction in large networks. Büker and Seybold (2012) propose a probability distribution convolution

model for large railway networks, assuming exponentially distributed process times. Keyhani et al. (2012) and Lemnian et al. (2014) present stochastic graph approaches to predict the expected reliability of a scheduled train connection. Kecman and Goverde (2015b) use a microscopic TEG model to predict train events with dynamic edge weights.

4.1.2. Markov Chains

An important class of event-driven approaches to predict train delays uses Markov Chain (MC) models. These models explicitly consider the dependencies of successive train-events for individual trains (i.e., along the path of a train) in the prediction function $f(.)$ as transition of delay states. Effects from one train to another (network effects) are only captured implicitly within those dependencies. Therefore, MC models are especially useful to predict the evolution of delay along a train's journey without expecting a lot of interference with other trains (e.g., connections to other trains, headways on a single-track line).

Barta et al. (2012) propose an MC model to predict the evolution of delays for freight trains. Kecman et al. (2015) model the uncertainty in the evolution of train delays as a stochastic process, showing that including real-time information about the most recent delay observations increases prediction accuracy. Sahin (2017) uses an MC model to predict train delay states (based on a 1-min discretisation) at departure/arrival events, analysing the time allowances in the timetable. Gaurav and Srivastava (2018) present an MC and regression approach to predict train delays in the Indian railway network. Within their test-case setting, they conclude that most delay evolutions along a train's journey follow a 1-order Markovian Process, i.e., the latest delay information for a train is sufficient to predict future delays of that train whilst earlier observations are not needed. To the best of our knowledge, this is the only study to specifically investigate the order of MC for train delay evolution, although other contributions using MC models to predict train delays also assume first-order models.

4.1.3. Bayesian Networks

Bayesian Network (BN) approaches can be seen as a generalisation of MC approaches since they also include the explicit modelling of inter-train-event dependencies (network effects) and potential external influencing factors. This results in more complex structures of the prediction function $f(.)$ that still provide an interpretable framework to understand the spreading of predicted train delays throughout the railway network (also from train to train).

Zilko et al. (2016) propose the Copula BN method to predict the length of a railway disruption. Corman and Kecman (2018) are the first to apply a BN to directly predict train delays, assuming normally distributed changes of delay from node to node. They analyse how new information about the latest observed train delays changes the predicted delay probabilities throughout the railway network. Lessan et al. (2019) also use a BN approach and investigate methods for dependency structure learning. Huang et al. (2020a) propose a BN to predict the primary delay, the number of affected trains and the total delay time during disruptions.

4.1.4. Petri Nets

Apart from MC and BN approaches, directed bipartite Petri Net (PN) models are also models that represent interconnection of processes by means of a graph structure, and that have also been applied to predict train delays. Petri Nets are a specific type of bipartite graph, with two sets of nodes: events and processes. Events are only connected to processes and vice versa. The interconnection of those elements can describe well the dependency structure of operations. Petri Nets also model a current state of the system over time, by means of tokens (possibly coloured) which describe the state of the system along a few possible characteristics. In this sense, they can extend pure graphical models to include aspects of simulation models, a study of a-priori dynamics of the systems (for instance, reachability) and inclusion of if-then rules representing business rules.

Milinković et al. (2013) apply a PN model to predict train delays. They implement a fuzzy logic to model train delay propagation under incomplete information about train delay states. Zhuang et al. (2016) develop a PN approach with fuzzification of time intervals in a train timetable to predict delays, detect potential conflicts and support traffic management.

4.1.5. Equation systems

As graph models, equation system models (EQS) describe the effects of delay propagation for a given prediction horizon by characterising the dependencies of train-events in railway networks. Their close relation to graph models is highlighted by Goverde (2010) translating their TEG approach into a system of linear equations to predict train delays.

Chen and Harker (1990) predict train running times by explicitly considering departure times. By solving a nonlinear EQS, Higgins and Kozan (1998) quantify expected train delays that also account for knock-on delay propagation at scheduled train connections. Huisman and Boucherie (2001) analyse the role of heterogeneity of train speeds towards delay.

Goverde (2010) introduces the Max-plus algebra methodology based on recursive linear equations to represent a TEG and consequently predict the propagation of train delays assuming that the structure of the network dependencies repeats in every period, but the exogenous variation, i.e., the delay, happens once at the beginning of the study. Ma et al. (2016) further develop the Max-plus algebra theory to predict train delay propagation in a Chinese railway system.

Recently, Zhang et al. (2020) formulate a linear system of ordinary differential equations, solved with control theory techniques, to predict the mitigation of train delays to neighbouring stops of delayed trains.

4.1.6. Synthesis of event-driven approaches

Event-driven approaches range from theoretical models of equation systems to more visualisable graphical models. Typically, process time variability is modelled by assuming distribution functions for process times (or delays at events) and fitting the parameters of the assumed distributions according to historic observations under the same (or similar) circumstances. Different

probability distributions have been used to model train delays including the Exponential, Weibull, Normal, Log-normal, Erlang and Chi-squared functions (Shi et al., 2021) as well as phase-type distributions (Meester and Muns, 2007). Among those, exponentially distributed processes are the most popular as they are the easiest to convolute while performing equally well in terms of prediction accuracy as the others.

The complexity of the graphical models has increased substantially throughout the last decades due to the explicit capturing of more and more dependencies. Especially, the inclusion of inter-train dependencies in the recent applications of BN models as a natural extension to MC models sets a milestone in graphical event-driven approaches. On the theoretical side of event-driven approaches, it is necessary to point out the Max-plus algebra approaches, as they rely on a specific formulation of equation systems that has been especially introduced for modelling train delay propagation.

Table 2 summarises the contributions of event-driven train delay prediction approaches. The table includes the specific model category, the three-character ISO country code of the railway network used in the numerical experiments, the input data sources used for prediction (see Section 3), and an indicator of the prediction output being deterministic (D) vs stochastic/probabilistic (S), which is discussed more in detail in Section 5.

4.2. Data-driven approaches

We classify train delay prediction approaches as data-driven, if they do not explicitly model and exploit a train-event dependency structure for the prediction horizon. As stated earlier, most contributions in this class fall into the field of machine learning, i.e., algorithms that "improve automatically through experience" (Jordan and Mitchell, 2015), meaning in our case that the performance of the prediction function $f(.)$ increases when fed with historical data.

The vast majority of contributions in this category rely on supervised learning methods because historical observations for both input variables and realised delay (i.e., the true value of f) are used to learn the prediction function f. In particular, HTM observations are commonly considered as a source of experience to train a model. Furthermore, influencing factors for train delays (infrastructure, operational information, external factors) may also be used as a source of experience to improve predictions. Machine learning techniques used in the discussed papers include Linear Regression, Decision Trees, Random Forests, Neural Networks, and Support Vector Machines. In addition, we classify time-series analysis and queuing models as data-driven approaches, too.

4.2.1. Linear regression

The simplest form of regression technique is Linear Regression (LR), which models a train delay (i.e., output variable) as a linear combination of different input features (i.e., explanatory variables).

Gorman (2009) presents an application of this method to US freight rail data, predicting train delays by splitting the total running time of a train as the sum of the so-called free running time and an additional time caused by congestion. Murali et al. (2010) use LR to

Table 2
Event-driven train delay prediction approaches.

Reference	Model category	D/S	Country	HTM	OPI	INI	EFW
Huang et al. (2020a)	BN	S	CHN	x	x		
Zhang et al. (2020)	EQS	D	N.A.	x	x		
Lessan et al. (2019)	BN	S	CHN	x	x		
Corman and Kecman (2018)	BN	S	SWE	x	x		
Gaurav and Srivastava (2018)	MC	S	IND	x			
Sahin (2017)	MC	S	TUR	x	x		
Zhuang et al. (2016)	PN	S	CHN	x			
Davydov et al. (2016)	EQS	S	RUS	x			
Martin (2016)	BN	S	N.A.	x	x		
Zilko et al. (2016)	BN	S	NLD	x		x	x
Ma et al. (2016)	Max-plus	D	CHN	x	x		
Kecman et al. (2015)	MC	S	NLD	x	x		
Kecman and Goverde (2015b)	TEG	D	NLD	x			
Lemnian et al. (2014)	General graph	D	DEU	x			
Milinković et al. (2013)	PN	D	SRB	x		x	
Keyhani et al. (2012)	General graph	S	DEU	x	x		
Büker and Seybold (2012)	General graph	S	DEU	x	x		
Barta et al. (2012)	MC	S	CHE	x			
Berger et al. (2011b)	TEG	S	DEU	x	x		
Hansen et al. (2010)	TEG	D	NLD	x	x		
Goverde (2010)	Max-plus	D	NLD		x		
D`Ariano and Pranzo (2009)	General graph	D	NLD	x	x		
Yuan and Hansen (2007)	General graph	S	NLD	x	x	x	
Yuan (2007)	General graph	S	NLD	x			
Huisman and Boucherie (2001)	EQS	S	NLD		x		
Higgins and Kozan (1998)	EQS	S	AUS	x	x	x	
Hallowell and Harker (1996)	General graph	S	USA		x	x	
Carey and Kwieciński (1994)	General graph	S	GBR	x			
Chen and Harker (1990)	EQS	S	USA	x	x	x	

estimate train delays over track segments considering actual delays in the railway network and infrastructure information. Wang and Work (2015) apply an LR model after train trips have started showing that incorporating real-time delay information significantly increases the prediction quality. Hauck and Kliewer (2020) use an LR model to predict train delays in the German railway network and show that their LR model delivers similar results as their NN implementation on the same test-case scenario.

4.2.2. Decision trees and Random Forests

Decision Trees (DT) are popular supervised learning models due to their simplicity and interpretability. In the context of train delay predictions, their goal is to find rules based on decisive information criteria in Ω_t (e.g., current delay > 3min, section length > 5 km) using historical observations to determine delay predictions. Random Forests (RF) are ensembles of decision trees based on bootstrapping techniques, in which multiple trees are used to generate individual predictions and the most-likely prediction resulting from these trees is then selected as output. This makes RF generally more accurate than DT and more robust to modelling errors.

Kecman and Goverde (2015a) compare the accuracy of several models, and conclude for their test-cases, that the RF model outperforms the LR and DT models to predict train running and dwelling times. Wang and Zhang (2019) analyse the effect of different weather types and headways using RF. Nair et al. (2019) present a large-scale application of RF, while Nabian et al. (2019) discuss a bi-level RF approach for short-term passenger train delay prediction. Shi et al. (2020) apply a gradient-boosting DT to delay prediction. Finally, Gao et al. (2020) use a two-stage RF model to increase the prediction accuracy of train recovery times.

4.2.3. Neural networks

Artificial Neural Networks (NN) are inspired by biological neural networks and consist of units (neurons) and connections to transfer information (signals) between them. Their structure is usually based on multiple layers of neurons, and the propagation of signals occur by neurons processing data, applying a non-linear function, and producing an output that is weighted and sent to the following layer. This way, input data (explanatory variables) is propagated through the network until the final output layer is reached. We refer to the methodology section in Oneto et al. (2018) for more technical details on NN and their use for train delay prediction.

The first application of NN to train delays is by Malavasi and Ricci (2001), predicting the total delay and the number of delayed trains in a railway network. Yaghini et al. (2013) predict passenger train delays in the Iranian railway network, analysing which definition of input variables and NN design yields the best performance.

An extensive analysis of the application of NN models in the form of shallow and deep extreme learning machines to predict train delays is provided by Oneto et al. (2018). They suggest that using NN to exploit historic train movement data allows constructing high performance and robust prediction models for train delays. Wen et al. (2019) show that their application of a long short-term memory NN prediction model outperforms an RF approach on a test case based on the Dutch railway network. Huang et al. (2020b) apply different types of NN models to predict train delays by including state, location and trajectory-based input data.

We can observe that NN have mostly been applied to predict the delay of passenger trains with a recent focus on real-time utilisations. They currently provide the most accurate prediction results (Bao et al., 2021), although they do neither need nor provide insights into the dynamics of the railway system.

4.2.4. Support Vector Machines

Support Vector Machines (SVM) are popular supervised learning methods used for classification. SVM determine a hyperplane in a high- or infinite-dimensional space separating two classes of observations, such that distance or margin between this hyperplane and the nearest training data from the two classes is maximal. A well-known SVM variant is used for regression.

Marković et al. (2015) compared the results of NN and SVM to predict train delays in the Serbian railway system, highlighting that purely data-driven NN models have a considerable risk of overfitting, i.e., of calibrating the model too closely to the training data, which leads to poor predicting performance on new data. Lee et al. (2016a, 2016b) develop an SVM approach that considers sequences of diagnostic events (abnormal or particular situations) to increase prediction performance. Huang et al. (2020c) use a combination of SVM and Kalman filter to predict train running times during railway disruptions. They show that a reduction of prediction error is possible with their approach, especially for trains that are already running with a delay.

4.2.5. Others

Greenberg et al. (1988) propose a queueing model (QM) to predict the average delay of trains on a single-track line. Meester and Muns (2007) show that phase-type distributions can calculate knock-on train delays given probability distributions of primary delays in a global mathematical queueing model. An online model to recognise delay and predict its propagation to the next station is proposed by Hansen et al. (2010). Pongnumkul et al. (2014) use time-series analysis (TSA) to predict train arrival delays in the Thai railway network based on autoregressive and moving-average processes. Dekker et al. (2019) provide the only unsupervised machine learning contribution in scope we are aware of, by predicting the transition between train delay states in the Dutch railway network using a clustering algorithm.

4.2.6. Synthesis data-driven approaches

As we have noted, most data-driven approaches are associated with machine learning and regression models. There, the application of non-linear regression models marks an important progress, since also non-trivial non-linear dependencies (and interdependencies) can be captured within these models. Hence, the first application of an NN model by Malavasi and Ricci (2001) to the domain of train delay prediction has to be highlighted. Also, the first application of an SVM model by Marković et al. (2015) to overcome the problem of overfitting in data-driven approaches can be considered as a seminal work. Furthermore, the application of NN models saw a huge

increase in complexity by the enhancements of Oneto et al. (2018) adding multiple layers of neurons in shallow and deep extreme learning machines, including weather as external factor and its applicability in real-time. Generally speaking, machine learning approaches are considered of being capable of processing big datasets with complex formats and non-trivial structures efficiently. Table 3 summarises the literature on data-driven train delay prediction, including the same information reported in Table 2 previously discussed (i.e., model category, deterministic or stochastic outputs, reference country in the case study, and input data sources).

Finally, we considered all papers from Tables 2 and 3 based on their year of publication and mathematical model class. We observed the following:

1. Event-driven approaches have experienced a shift from general graph models (with contributions ranging from 1994 to 2015) to more specific MC, PN and BN models (all related publications are more recent than 2012).
2. There has been a substantial boost in the number of publications dealing with data-driven approaches in the last few years. In fact, only eight papers were published before 2015, while 33 up from 2015.

5. Discussion

5.1. Event-driven versus data-driven methods

A modelling paradigm gap separates event-driven and data-driven approaches. Event-driven prediction approaches aim to describe a train-event dependency structure that holds for a given prediction horizon and exploit this structure to generate predictions of delay developments (i.e., knock-on delays in terms of propagation of delays across trains). Such predictions are constructed incrementally according to the dependencies for all train-events within the horizon. Therefore, event-driven approaches are generally more easily

Table 3
Data-driven train delay prediction approaches.

Reference	Model category	D/S	Country	HTM	OPI	INI	EFW
Bao et al. (2021)	NN	D	CHN	x	x	x	
Shi et al. (2021)	RF	D	CHN	x	x	x	
Chen et al. (2021)	LR, DT, SVM	D	CHN	x			x
Arshad and Ahmed (2021)	RF	D	IND	x			x
Wang et al. (2021)	SVM	D	CHN	x		x	
Pradhan et al. (2021)	LR	D	IND	x			x
Shi et al. (2020)	DT	D	CHN	x	x		
Huang et al. (2020b)	NN	D	CHN	x	x	x	x
Huang et al. (2020c)	SVM	D	CHN	x	x		
Zhou et al. (2020)	NN	D	CHN	x			
Sara et al. (2020)	SVM	D	FRA	x			x
Hauck and Kliewer (2020)	LR	D	DEU	x			x
Gao et al. (2020)	RF	D	CHN	x	x	x	
Li et al. (2020)	RF	D	NLD	x	x	x	x
Dekker et al. (2019)	Clustering	D	NLD	x			
Mou et al. (2019)	NN	D	NLD	x			
Wen et al. (2019)	NN	D	NLD	x			
Wang and Zhang (2019)	DT	D	CHN	x	x		x
Nair et al. (2019)	RF	D	DEU	x	x	x	x
Nabian et al. (2019)	RF	D	NLD	x	x		x
Arshad and Ahmed (2019)	RF	D	IND	x	x		x
Jiang et al. (2019a)	RF	D	CHN	x		x	
Jiang et al. (2019b)	RF	D	SWE	x		x	x
Oneto et al. (2018)	NN	D	ITA	x	x		x
Barbour et al. (2018)	SVM	D	USA	x	x	x	x
Shih et al. (2017)	QR	S	USA	x			
Wen et al. (2017)	RF	D	CHN	x	x		
Davydov et al. (2017)	Others	S	RUS	x	x		
Liu et al. (2017)	NN	D	CHN	x			
Lee et al. (2016a, 2016b)	SVM	D	NLD	x			
Kosolsombat and Limprasert (2016)	RF	D	THA	x			
Marković et al. (2015)	SVM	D	SRB	x		x	
Wang and Work (2015)	LR	D	USA	x			
Kecman and Goverde (2015a)	RF	D	NLD	x		x	
Pongnumkul et al. (2014)	TSA	D	THA	x			
Yaghini et al. (2013)	NN	D	IRN	x			
Murali et al. (2010)	LR	D	USA	x	x	x	
Gorman (2009)	LR	D	USA	x	x		
Meester and Muns (2007)	QM	S	NLD		x		
Peters et al. (2005)	NN	D	DEU	x			
Malavasi and Ricci (2001)	NN	D	ITA	x	x	x	
Greenberg et al. (1988)	QM	S	USA		x	x	

interpretable, i.e., delay predictions for train-events at the end of the horizon can be tracked and explained by their preceding dependency. However, event-driven approaches usually do not capture complex interrelations of historic train operation data themselves. Instead, HTM data is used to calibrate model parameters aiming to describe the a-priori assumed dependencies. Regardless of the prediction method being event- or data-driven, it is important that it accounts for knock-on delays or there would the risk that the method suffers from a structural deficit. Event-driven approaches usually capture these effects explicitly with the dependency structure, while data-driven implicitly by training on the HTM data.

Data-driven approaches on the other side do not explicitly model the dynamics of railway operations (i.e., internal mechanisms of the system) during the prediction horizon, and hence are usually less interpretable. Consequently, this also means that data-driven models typically require less domain knowledge and physical behaviour and instead rely more on the HTM data source. One key strength of data-driven approaches is that they are capable of handling multi-dimensional noisy data and can take into account all kinds of (hidden) relationships of variable dependencies as well as external and human influences, that might not be immediately evident at first sight. The downside, however, is that tracing the reasons for delay predictions (i.e., the debugging of the prediction process, and the detection of any incorrect input parameter value or setting) can be very hard. In fact, the most sophisticated data-driven approaches such as NN and SVM are often referred to as black-box models (Jiang et al., 2019a, 2019b).

Event-driven approaches typically generate stochastic predictions (i.e., probability distributions) for future delays, as far as they model explicitly process time variability and uncertainty in the system. Differently, most data-driven approaches produce single-value deterministic predictions. This is also evident from Tables 2 and 3, where we classify the type of prediction in all reviewed papers. Single-value predictions are straightforward to read and provide an easily understandable outlook into the future of train delay development. However, they do not convey any information about their (un)certainty and ignore (or drastically simplify) the random occurrence of disturbances in the prediction horizon. Predicted probability distributions carry substantially more information on the potential (and random) future propagation of delays in the network (e.g., their variance).

In terms of passenger information, discrete values as the expected time of arrival or a 95% probability of getting a connection are most useful, while providing full probability distributions/density functions would be more time consuming and harder to understand. Nevertheless, stochastic predictions of future delays might be especially meaningful for risk-based traffic management. Furthermore, probability distributions for delays can always be reduced to a single-value prediction (e.g., the mean or quantiles). If the influence of the most important factors affecting train delays is precisely known, deterministic predictions should be the preferred choice of model as they generally produce higher prediction accuracy (Bao et al., 2021).

In the end we can say that both data-driven and event-driven approaches are suited to predict train delays in real time, even though constructing/assuming an understandable and useful dependency structure (i.e., as large as needed and as small as possible) is a challenging part of event-driven approaches. We believe, however, that the definition of a dependency structure might give event-driven approaches an advantage to predict uncommon delays, such as large disruptions, and their propagation throughout the railway network. For example, Huang et al. (2020c) overcome a shortcoming in their data-driven SVM approach to predict rare train delays for which limited experience is available. In general, the quality of train delay predictions is difficult to compare across published results due to many reasons that we detail in Subsection 5.3. However, we anticipate here that data-driven methods have proven to deliver state-of-the-art results in terms of prediction accuracy, albeit such results are harder or not possible to interpret as previously discussed.

In conclusion, our analysis reveals complementary strengths of event- and data-driven approaches. While event-driven methods are more interpretable and suitable in cases of disruptions and uncommon train delays, state-of-the-art data-driven methods usually deliver the highest accuracy and can more easily be applied in real-time.

5.2. Data usage and limitations

As discussed in Section 3, HTM, INI, OPI and EFW information can all be used as input data to predict train delays. Digitalisation and increasing data availability push towards the usage of more and more data for both data- and event-driven approaches. Especially the increase of the quality in terms of granularity and precision of HTM data allows increasing the robustness and accuracy of train delay prediction models.

Of course, both data-driven and event-driven approaches rely on diverse data-sources to generate train delay predictions. Their different underlying modelling paradigm, however, leads to a conceptually different usage of HTM, INI and OPI data to some extent. While data-driven models mostly use indicators of actual delays, infrastructure and operations as direct regressors for delay predictions, event-driven approaches rely specifically on infrastructure and operational data to build a dependency network of train-events for the prediction horizon first, and then also use diverse data sources to calibrate their model parameters.

The idea of using HTM data to predict train delays is to find recurrent patterns of delay and delay propagation that is similar to the situation at the current prediction time, and use this experience to draw conclusions on future delays, to identify appropriate countermeasures, and finally overcome sources as well as spreading of delays. The underlying assumption therefore is that delays develop/propagate in a typical way – so by comparing the current situation (current delays, used rolling stock, weather, etc.) with observed HTM data, a similar situation of the past should provide an estimation for the future development of delays. To this end, the timetable in place when making a prediction should be the same as the one used to record the HTM data. Atomic elements in the sense of repetitive (periodic) schedules of operations can range from an hourly basis to daily basis.

However, in modern and complex railway systems, every day of operations is unique, especially when the infrastructure is used for both passenger and freight services, and for maintenance services. In fact, freight trains are much less regular in volume, and actual path followed, with a mix of planned abstract train paths, and last minute insertion of additional train paths (Cacchiani et al., 2010).

When looking at full days, a study reports how for a year of 365 days, 314 patterns have been identified, when considering intervals of entire days (Gestrelius et al., 2015). When considering smaller intervals, a higher amount of the same patterns repeated is expected. In this context, Oneto et al. (2018) argue that timetable changes restrict the usage of HTM data significantly. To overcome the problem of different timetables during the training/testing period and prediction period, it is very important that the approach can abstract from these schedule changes and focus on invariant traffic dynamic properties. However, this limitation is the key reason why many approaches can use no more than a few months of HTM data to train their models. Among 40 contributions specifying the length of their training HTM dataset, we find that more than half (21) report using six or less months of observations (13 of which three months or less) while in only six papers a model is trained on more than two years of data.

In terms of EFW data, weather variables including temperature, wind speed and rainfall have been used in the literature. However, only about 20% of studies consider EFW data. It seems reasonable to expect that "normal" weather conditions do not affect railway operations massively, whereas severe weather conditions might have a huge impact (e.g., heavy snowfall). If there is enough experience in terms of HTM data in combination with data on weather indicators, train delay prediction approaches can make use of actual weather conditions or even weather forecasts.

The necessity to use different input data sources is closely related to the domain of the railway service and to the model to solve the precise prediction problem. While infrastructure information is extremely important for expected-arrival-time predictions some days in advance (typically for freight services), the most recent delay information is essential to predict passenger train delays in real-time.

In the future, digitalisation may lead to additional and more abstract data sources that can support train delay predictions. For instance, Ghofrani et al. (2018) name smart ticket sales as a potential data source to explain abnormal demand fluctuations in rail services. Artificial intelligence will increasingly explain the impact on train punctuality of more explanatory variables (see, e.g., Roßler et al., 2021). It will be a key task of scientific analyses to quantify the relevance of novel upcoming data sources and consequently from railway operating companies to provide them eventually. Learning from historical observations will remain nonetheless critical as the amount and quality of information increases and as this data source inherits the results of all potential influencing factors of train delays.

Especially data-driven approaches face the well-known problem in machine learning of overfitting. This problem arises when a model is trained "too close" to a given set of training data; consequently delivering poor predictions when applied out-of-sample, i.e., to test data not used for training and new to the model. For example, Marković et al. (2015) show that Neural Networks for train delay prediction tend to easily run into overfitting using HTM data, whereas Support Vector Machines appear more robust in this sense. Wen et al. (2019) argue that Random Forests as a generalisation of Decision Trees are more likely to avoid overfitting.

Another important distinction in the context of input data usage stems from the methodological handling of human based decisions of traffic control within the prediction horizon. Clearly, predicting train delays, in the long run, requires making assumptions on how railway traffic is handled. For this reason, train delay prediction approaches have also been divided into optimisation-based, simulation-based and analytic-parametric methods (Gorman, 2009). Although we do not focus on this alternative classification, we mention these approaches in the following to briefly discuss if and how they account for traffic control decisions, which affect predictions.

Optimisation-based approaches aim to find optimal routes for trains for a required destination and/or an optimal sequence or train order to pass a bottleneck of the infrastructure to minimise total delay and simultaneously predict delay development of trains. Simulation-based methods on the other side describe future developments of delays using richer railway dynamics and data while applying simple heuristic decision rules for managing railway traffic as done by human dispatchers. Thus, simulation-based methods need detailed information on the infrastructure, physical behaviour of the rolling stock and dynamics of the railway systems. Without reservation, they can be seen as detailed representations of the dynamic of a railway system as soon as the modelled rules match reality. To achieve that, simulation-based methods require some huge effort to be configured and maintained and are not easily transferable to other railway systems. RailSys (Radtke and Hauptmann, 2004), OpenTrack (Nash and Huerlimann, 2004), LUKS (Janecek et al., 2010) and OnTime (Büker and Seybold, 2012) are known examples of simulation software for train delay propagation. Additionally, a recent work by Liebchen et al. (2021) points out that simulation tools are still too simple to provide a clearly matching picture of the dynamics withing a railway system.

Analytic-parametric approaches in comparison make more general assumptions, for instance, on the underlying distributions of process times and influences of external factors to delays. Therefore, they require less data compared to simulation-based methods and are easier to maintain and transfer. Nevertheless, they can only provide predictions within their framework of assumptions and may fail for unseen and therefore unexpected situations where such assumptions do not hold.

5.3. Prediction horizon and quality

Prediction approaches for train delays range from a very short-term prediction horizon (i.e., $s - t$) of a couple of minutes ahead to predictions made several days in advance. With an increasing prediction horizon, more and more rescheduling actions could be taken by human dispatchers (e.g., retiming and rerouting), affecting the railway traffic flow and hence making the prediction task more complex. We could assign 58 publications to the following six classes of prediction horizons (note that the first four classes are defined based on time, the last two classes on space):

- very short-term (up to 30 min): 3 contributions
- short-term (up to 2 h): 7 contributions
- medium-term (up to 4 h): 11 contributions
- long-term (more than 4 h): 16 contributions

- next station: 3 contributions
- train journey (more than one station ahead): 18 contributions

Note that different structures of railway services have different spectra of acceptable delays or even delay definitions. For freight trains, arrivals within 30 min of the planned arrival time are mostly considered to be punctual, whereas for passenger services punctuality is usually defined in the range of 3–10 min delayed arrivals after the scheduled arrival time. We illustrate in Fig. 3 the ranges of prediction horizons in minutes, for those 18 contributions that explicitly mention the investigated prediction horizon in their respective case studies. The figure highlights the diversity of train delay prediction horizons covered in the literature.

To evaluate the accuracy of train delay predictions, the prediction error (i.e., the difference between the predicted delay of a train at a certain train-event and the train's realised delay at this event) is analysed. Once the realised delay of the predicted event is measured, it is straightforward to calculate this error for single-value predictions. There is instead no unique definition of prediction error for probability distributions, but most papers calculate it using single-value properties of the distribution such as its mean or median. Often, the mean absolute error (MAE), the mean squared error (MSE) and the root mean squared error (RMSE) are used to evaluate the prediction accuracy.

Answering the question of which train delay prediction approach delivers the most accurate results is challenging due to several reasons. First, the prediction horizons differ strongly throughout the literature. Second, comparing deterministic and stochastic predictions requires making assumptions on how the stochastic predictions are evaluated. Third, the evaluation measures used for prediction errors are diverse. Fourth, the quality of predictions heavily relies on the considered dataset as well as the region and structure of the railway system considered in the case study (Tables 2 and 3 include 16 different countries where the case studies have been conducted). Despite these challenges, we try to compare the results in the literature, focusing on the contributions that report an analysis or enough details of the prediction quality of their models.

Berger et al. (2011a, 2011b) report an increasing error for a prediction horizon of 30–240 min in a German case study. Specifically, the MAE increases in the first 150 min of the horizon and then remains rather flat until the end of a 240-min horizon. Kecman and Goverde (2015a) show in a Dutch case study that the MAE for a 0 to 20-min prediction horizon increases up to about 40 s and then stays at this level for the rest of a 120-min horizon. Oneto et al. (2018) differentiate the prediction horizon in terms of stations ahead and show that for stations 1 to 5 their evaluated MAE increases from one-and-a-half to two-and-a-half minutes, using dynamic predictions based on NN. Nair et al. (2019) compute RMSE statistics for different prediction horizon intervals (0–10, 10–30, 30–90, 90–180 and 180+ minutes). They find that the RMSE increases from 110 s for the first period to 400–500 for the last. Corman and Kecman (2018) show that their BN approach provides stochastic predictions where the MAE (based on distribution mean) increases from 30 to 100 s in a prediction horizon of 0–60 min.

Based on these results, we provide some insights on the relation between the accuracy of train delay predictions and the length of the horizon. First, there exists a small but non-negligible prediction error already for very short-term train-events happening a couple of minutes in the future. This could result from uncertainty about dwelling time, rather independent from the prediction horizon, leading to a baseline error arising for the very short-term, too. Second, the prediction error increases with larger prediction horizons, which is intuitive as the amount of uncertainty due to small and big disturbances in the railway system increases over time. However, the prediction error does not increase anymore (or even slightly decrease in one study) after a certain point far enough in the future (e.
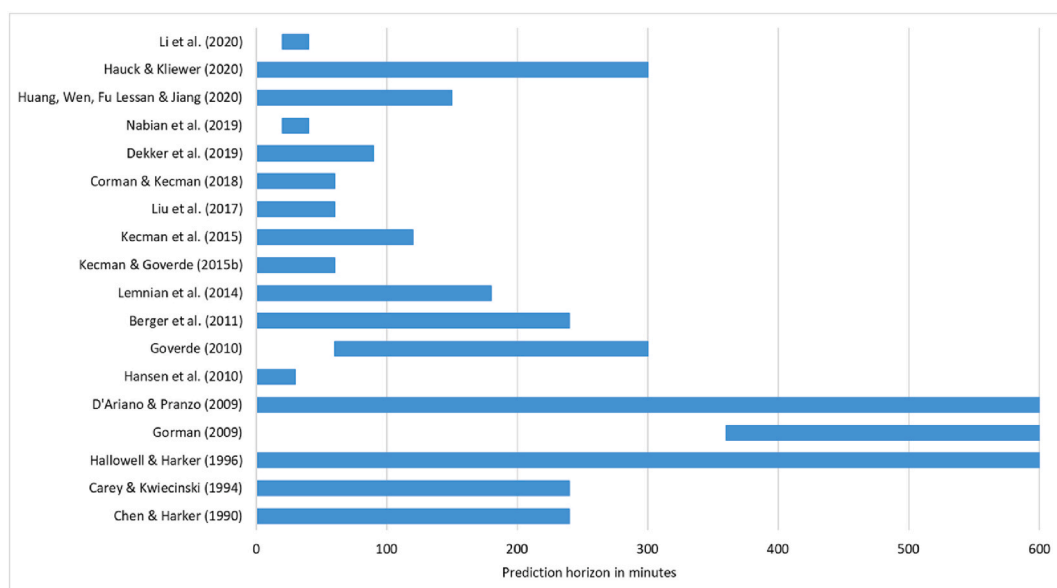

Fig. 3. Contributions classified by prediction horizon.

g., 150 min in the case of Berger et al. (2011a, 2011b)). These findings suggest a functional form of increasing, saturating prediction error for an increasing prediction horizon, identifying the medium-term accuracy as a challenging aspect.

Updating delay predictions is important to reduce the amount of uncertainty in future railway dynamics, which is valuable in real-time traffic control. However, relying on medium- and long-term delay propagation predictions should be done with caution, since real-time information can help reduce uncertainty in such horizons only up to a certain point. For passengers, this translates to delay information for short trips being usually rather reliable, while this may not be the case for longer trips of e.g., 1 h, hence planning based on some personal buffer time may be advised.

5.4. Prediction purposes and requirements, for different processes and stakeholders

A prediction for railway traffic can be of interest to the many stakeholders of the system in different ways. We here systematically review the most relevant stakeholders and identify possible interest in prediction outcomes; highlighting how the availability of information about the predicted state could possible improved their actions, and discussing the possible implications of a wrong prediction.

5.4.1. Customer-passengers
The prediction horizon lasts from a few minutes up to a few hours. Predictions lead to a psychologic reduction of discomfort due to known delay against unknown delay; in case of connections, the predicted probability of catching the connecting train creates awareness (see Keyhani et al. (2012). Effects can be changes in route choice resulting in a potential replanning of transport chains (e.g., transfers, different trains). Even more, replanning actions could have effects beyond the transport chain (e.g., shift of location or time for performing activities, dropping activities) (Leng and Corman, 2020). In the long term, delay predictions affect passengers' mode choices and can potentially lead to inclusions of buffer times. The impact of wrong predictions are additional travel times, regrets due to suboptimal choice and discomfort.

5.4.2. Customer-freight
The prediction horizon lasts from a few minutes up to a few days for the purposes of planning of downstream processes and planning of resources required to further handle the goods. In the long term, delay predictions of freight trains can help to optimise the planning of buffer times, defining of economic storage quantity, sizing of batches of transport and might affect the mode choice. The impact of wrong freight train delay predictions are penalties for delayed delivery, opportunity costs for loss of customers and additional resource costs.

5.4.3. Infrastructure manager, traffic control
The prediction horizon lasts from a few minutes up to a few hours. Traffic control managers can use delay predictions for train retiming, reordering, rerouting, short turning and even cancellation (Berger et al., 2011a, 2011b). In the long term, delay predictions for infrastructure managers reveal needs for extra resources in case of contingencies (for more infrastructure capacity available to run the scheduled traffic: more tracks; more controlled operations; less traffic running). The impacts of wrong predictions are cascading effects of the wrong control actions, need for further control action, costs and penalties due to delays, or less traffic running.

5.4.4. Railway operator, crew and rolling stock
The prediction horizon ranges from less than an hour up to a day. Railway operators use delay predictions for shift retiming, swapping tasks within same resource, swapping tasks with other resources, cancellation, insertion and also decisions to use resources overtime. For very long train runs (e.g., long distance passenger traffic, night trains and freight traffic), the prediction horizon might need to cover the total circulation time of the train between its two ends, which can go up to a few days. In the long term, delay predictions can point out needs for additional standby crews and rolling stock resources. The impact of wrong predictions are extra costs for overtime usage of resources, costs for additional standby resources and penalties for not being able to deliver train services as planned.

5.4.5. Infrastructure manager, planning of internal special resources (maintenance, possession)
The prediction horizon ranges from less than an hour, up to multiple days. Here, predictions only have a limited impact to the current traffic situation. Nevertheless, the impact of wrong predictions are needs for unplanned emergency operations rather than the usage of planned intervals, and penalties due to reduced productivity of the railway resources.

Infrastructure manager, planning of services, scheduling of extra trains:

The prediction horizon ranges from a few days, up to a few months. The impact to current traffic situation is limited, apart from temporal and geographical availability of resources. The impacts of wrong prediction are extra resource costs (crew, preparation of vehicles that are not necessary in the end) and losses of revenue for not monetizing on available resources.

6. Conclusion and further research directions

Train delay prediction is an active and growing field of research with a variety of approaches developed in the literature, including methods based on event graphs and machine learning. A natural question that arises is which method to apply in which context. Answering this question is challenging as the applicability and accuracy of prediction methods largely depend on the precise purpose

and use-case. We have nevertheless critically examined a large number of contributions on this topic, distinguishing between event-driven and data-driven approaches and identifying several aspects relevant for research.

State-of-the-art results in terms of prediction accuracy are achieved by data-driven machine learning methods, which include the widely applied Neural Networks and Random Forests. Stating this, we want to point out at the same time that data-driven approaches inherit some risk in terms of overfitting. Nevertheless, increasing data availability and accessibility (e.g., HTM data) enhances model calibration and out-of-sample performance, making those methods increasingly popular. Due to limited historical observations on rare events, forecasting delays during disruptions is harder with such techniques. Moreover, the lack of interpretability of the most advanced data-driven approaches makes their adoption by practitioners slower than in academia.

Event-driven approaches provide easier-to-interpret predictions for train delays, as they model the evolution and propagation of train delays throughout the railway network. Therefore, event-driven approaches can be more flexibly adapted to handle disruptions as well as different temporal and spatial domains. Graph models, Markov Chains, equation systems and Bayesian Networks are the main event-driven approaches for train delay prediction. Bayesian Networks are interesting as they combine the explicit modelling of a train-event dependency structure for the prediction horizon with the inclusion of multiple data sources.

Following this discussion, we identify that hybrid methods can combine the strengths of data-driven and event-driven approaches. For instance, an SVM has been recently combined with a Kalman filter to overcome the prediction inability of the SVM alone to produce useful forecasts in unexpected disruptions scenarios (Huang et al., 2020c). Interestingly, Ulak et al. (2020) point out that Bayesian Networks are to some extent already at the intersection between machine learning and techniques that explicitly exploit a dependency network. This combination of network modelling and big-data usage might become even more relevant in future conditions with increasing data availability and even busier railway systems.

Regarding the explicit inclusion of more data sources, the robustness and resilience of railway timetables have been studied in detail, but we see some potential in explicitly including more properties of the timetable such as spatially and temporally diverse scheduled safety margins for running times (i.e., recovery times, buffer times), plus specific inclusions of prescriptive business rules, operation policies. Having sufficient data, with enough variety and including all the relevant variables, those could be possibly learned from data; but a model based approach can naturally express them. Explicitly including more timetable and operational variables could enhance prediction accuracy.

A fair comparison of the quality of the reviewed prediction approaches is hardly possible due to profound differences in application settings and evaluation methods. Stochastic prediction models that deliver output probability distributions are especially difficult to evaluate and compare with deterministic predictions. Overall, our analysis reveals that the prediction error increases with the time horizon, at a decreasing rate, until the error eventually stabilises. This is to some extent due to the decreasing predictability of human decisions to manage the railway traffic flow.

Given our review and discussion, we propose seven directions of future research outlined in the following.

- (i) Including novel data sources in prediction models such as social media postings, which can be used to infer travel demand as well as behavioural patterns of passengers on platforms, or additional weather indicators.
- (ii) Exploiting the availability of increasingly granular observations of train movement data. While observations have been available only for stations a couple of decades ago, railway companies are now able to measure train passing times on an operation point level (e.g., signals, switches, etc.) and even equipping trains with GPS sensors that have the ability to provide a continuous stream of delay information. This development offers the chance to update delay predictions with increasing frequency whenever new information becomes available.
- (iii) Reconciling data sources according to their influence and importance for specific train delay predictions problems. This could be achieved by splitting delay predictions according to their mostly influencing factors and circumstances in order to capture the essential drivers for train delays.
- (iv) Better specifying the scope of the problem. What is actually important to be forecasted and what can be neglected? Moreover, do predictions for passengers and railway traffic controllers have the same requirements? In the upcoming years, we expect more and more contributions to focus on predicting train delays under very specific circumstances and settings.
- (v) Quantifying the uncertainty embedded in a delay prediction (e.g., by means of reasonable confidence intervals) for more informed decision making. To this end, making use of increasingly granular delay observations can help to better understand the trade-off between the length of the prediction horizon and the reliability of the prediction.
- (vi) Taking advantage of the recursive dynamics of predicting train delays and optimizing traffic flow. Very often, train delay predictions are not accurate, because wrong assumptions of future traffic-control decisions have been made. Internalising those human control decisions in the prediction process and the vice-versa usage of predictions to control the traffic flow could reduce the uncertainty and consequently increase the prediction reliability and traffic performance at once.
- (vii) Capturing only necessary and useful inter-train dependencies and their effects to train delay propagation. In other words: modelling every "arc" in a graph where a dependency exists, but avoiding blown-up network structures with dependencies that can be numerically estimated, but do not correspond to physically representable relations. Bayesian Networks already capture effects to train delay predictions but rely on a standard procedure as their underlying network structure either relies heavily on expert knowledge or is compiled by machine learning algorithms. Clearly, statistical analyses on the importance of potential inter-train dependencies can help to construct a more reasonable underlying network structure to predict train delays.

Declaration of competing interest

No potential conflict of interest.

Acknowledgements

References

Arshad, M., Ahmed, M., 2019. Train delay estimation in Indian railways by including weather factors through machine learning techniques. Recent Adv. Comput. Sci. Commun. 12 https://doi.org/10.2174/2666255813666190912095739, 1–00.

Arshad, M., Ahmed, M., 2021. Train delay estimation in Indian railways by including weather factors through machine learning techniques. Recent Adv. Comput. Sci. Commun. 14 (4), 1300–1307.

Banerjee, N., Morton, A., Akartunalı, K., 2019. Passenger demand forecasting in scheduled transportation. Eur. J. Oper. Res. 286 (3), 797–810. https://doi.org/10.1016/j.ejor.2019.10.032.

Bao, X., Li, Y., Li, J., Shi, R., Ding, X., 2021. Prediction of train arrival delay using hybrid ELM-PSO approach. J. Adv. Transport. 2021 (7763126) https://doi.org/10.1155/2021/7763126.

Barbour, W., Martinez, J.C.M., Kuppa, S., Work, D.B., 2018. Prediction of arrival times of freight traffic on US railroads using support vector regression. Transport. Res. C Emerg. Technol. 93, 211–227. https://doi.org/10.1016/j.trc.2018.05.019.

Barta, J., Rizzoli, A.E., Salani, M., Gambardella, L.M., 2012. Statistical modelling of delays in a rail freight transportation network. In: Proceedings of the 2012 Winter Simulation Conference (WSC), pp. 1–12. https://doi.org/10.1017/CBO9781107415324.004.

Berger, A., Blaar, C., Gebhardt, A., Müller-Hannemann, M., Schnee, M., 2011a. Passenger flow-oriented train disposition. In: Demetrescu, C., Halldórsson, M.M. (Eds.), Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 6942 LNCS. Springer Berlin Heidelberg, pp. 227–238. https://doi.org/10.1007/978-3-642-23719-5_20.

Berger, A., Gebhardt, A., Müller-Hannemann, M., Ostrowski, M., 2011b. Stochastic delay prediction in large train networks. In: 11th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems. https://doi.org/10.4230/OASIcs.ATMOS.2011.100.

Bešinović, N., 2020. Resilience in railway transport systems: a literature review and research agenda. Transport Rev. 40 (4), 457–478. https://doi.org/10.1080/01441647.2020.1728419.

Büker, T., Seybold, B., 2012. Stochastic modelling of delay propagation in large networks. J. Rail Trans. Plann. Manag. 2 (1–2), 34–50. https://doi.org/10.1016/J.JRTPM.2012.10.001.

Cacchiani, V., Caprara, A., Toth, P., 2010. Scheduling extra freight trains on railway networks. Transp. Res. Part B Methodol. 44 (2), 215–231. https://doi.org/10.1016/j.trb.2009.07.007.

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. Part B Methodol. 63, 15–37. https://doi.org/10.1016/j.trb.2014.01.009.

Carey, M., Kwieciński, A., 1994. Stochastic approximation to the effects of headways on knock-on delays of trains. Transport. Res. Part B 28 (4), 251–267. https://doi.org/10.1016/0191-2615(94)90001-9.

Chen, B., Harker, P.T., 1990. Two moments estimation of the delay on single-track rail lines with scheduled traffic. Transport. Sci. 24 (4), 261–275. https://doi.org/10.1287/trsc.24.4.261.

Chen, Z., Wang, Y., Zhou, L., 2021. Predicting weather-induced delays of high-speed rail and aviation in China. Transport Pol. 101, 1–13. https://doi.org/10.1016/j.tranpol.2020.11.008. May 2020.

Corman, F., Kecman, P., 2018. Stochastic prediction of train delays in real-time using Bayesian networks. Transport. Res. C Emerg. Technol. 95, 599–615. https://doi.org/10.1016/j.trc.2018.08.003.

D'Ariano, A., Pranzo, M., 2009. An advanced real-time train dispatching system for minimizing the propagation of delays in a dispatching area under severe disturbances. Network. Spatial Econ. 9 (1), 63–84. https://doi.org/10.1007/s11067-008-9088-1.

Davydov, B., Chebotarev, V., Kablukova, K., 2017. Online train traffic adjustments: probabilistic modeling and estimating. Sci. Tech. Conf. Trans. Syst. Theor. Pract. 55–60 https://doi.org/10.1007/978-3-319-62316-0.

Davydov, B., Kablukova, K., Godyaev, A., 2016. Prediction of the train traffic in a transport corridor. Procedia Eng. 165, 1430–1436. https://doi.org/10.1016/j.proeng.2016.11.875.

Dekker, M.M., Panja, D., Dijkstra, H.A., Dekker, S.C., 2019. Predicting transitions across macroscopic states for railway systems. PLoS One 14 (6), 1–26. https://doi.org/10.1371/journal.pone.0217710.

Gao, B., Ou, D., Dong, D., Wu, Y., 2020. A data-driven two-stage prediction model for train primary-delay recovery time. Int. J. Software Eng. Knowl. Eng. 30 (7), 921–940. https://doi.org/10.1142/S0218194020400124.

Gaurav, R., Srivastava, B., 2018. Estimating train delays in a large rail network using a zero shot Markov model. In: 2018 21st International Conference on Intelligent Transportation Systems (ITSC), pp. 1221–1226. https://doi.org/10.1109/ITSC.2018.8570014. November.

Gestrelius, S., Bohlin, M., Aronsson, M., 2015. On the uniqueness of operation days and delivery commitment generation for train timetables. In: 6th International Conference on Railway Operations Modelling and Analysis (RailTokyo2015), Tokyo, Japan.

Ghofrani, F., He, Q., Goverde, R.M.P., Liu, X., 2018. Recent applications of big data analytics in railway transportation systems: a survey. Transport. Res. C Emerg. Technol. 90, 226–246. https://doi.org/10.1016/j.trc.2018.03.010.

Gorman, M.F., 2009. Statistical estimation of railroad congestion delay. Transport. Res. E Logist. Transport. Rev. 45 (3), 446–456. https://doi.org/10.1016/j.tre.2008.08.004.

Goverde, R.M.P., 2010. A delay propagation algorithm for large-scale railway traffic networks. Transport. Res. C Emerg. Technol. 18 (3), 269–287. https://doi.org/10.1016/j.trc.2010.01.002.

Goverde, R.M.P., Hansen, I.A., 2000. TNV-prepare: analysis of Dutch railway operations based on train detection data. Adv. Trans. 50 (7), 779–788.

Goverde, R.M.P., Meng, L., 2011. Advanced monitoring and management information of railway operations. J. Rail Trans. Plann. Manag. 1 (2), 69–79. https://doi.org/10.1016/j.jrtpm.2012.05.001.

Greenberg, B.S., Leachman, R.C., Wolff, R.W., 1988. Predicting dispatching delays on a low speed, single track railroad. Transport. Sci. 22 (1), 31–38. https://doi.org/10.1287/trsc.22.1.31.

Hallowell, S.F., Harker, P.T., 1996. Predicting on-time line-haul performance in scheduled railroad operations. Transport. Sci. 30 (4), 364–378. https://doi.org/10.1287/trsc.30.4.364.

Hansen, I.A., Goverde, R.M.P., Van Der Meer, D.J., 2010. Online train delay recognition and running time prediction. In: 13th International IEEE Conference on Intelligent Transportation Systems, pp. 1783–1788. https://doi.org/10.1109/ITSC.2010.5625081.

Hauck, F., Kliewer, N., 2020. Data analytics in railway operations: using machine learning to predict train delays. In: Neufeld, J.S., Buscher, U., Lasch, R., Möst, D., Schönberger, J. (Eds.), Operations Research Proceedings 2019. Operations Research Proceedings. Springer, Cham, pp. 741–747. https://doi.org/10.1007/978-3-030-48439-2_90.

Higgins, A., Kozan, E., 1998. Modeling train delays in urban networks. Transport. Sci. 32 (4), 346–357. https://doi.org/10.1007/0-306-48058-1_1.

Huang, P., Lessan, J., Wen, C., Peng, Q., Fu, L., Li, L., Xu, X., 2020a. A Bayesian network model to predict the effects of interruptions on train operations. Transport. Res. C Emerg. Technol. 114, 338–358. https://doi.org/10.1016/j.trc.2020.02.021.

Huang, P., Wen, C., Fu, L., Lessan, J., Jiang, C., 2020b. Modeling train operation as sequences : a study of delay prediction with operation and weather data. Transport. Res. Part E 141, 102022. https://doi.org/10.1016/j.tre.2020.102022.

Huang, P., Wen, C., Fu, L., Peng, Q., Li, Z., 2020c. A hybrid model to improve the train running time prediction ability during high-speed railway disruptions. Saf. Sci. 122, 104510. https://doi.org/10.1016/j.ssci.2019.104510.

Huisman, T., Boucherie, R.J., 2001. Running times on railway sections with heterogeneous train traffic. Transp. Res. Part B Methodol. 35 (3), 271–292. https://doi.org/10.1016/S0191-2615(99)00051-X.

Janecek, D., Weymann, F., Schaer, T., 2010. LUKS. Integrated tool for analysing the performance of railway junctions and lines; LUKS. Integriertes Werkzeug zur Leistungsuntersuchung von Eisenbahnknoten und -strecken. ETR. Eisenbahntechnische Rundschau 59 (1–2).

Jiang, C., Huang, P., Lessan, J., Fu, L., Wen, C., 2019a. Forecasting primary delay recovery of high-speed railway using multiple linear regression, supporting vector machine, artificial neural network, and random forest regression. Can. J. Civ. Eng. 46 (5), 353–363. https://doi.org/10.1139/cjce-2017-0642.

Jiang, S., Persson, C., Akesson, J., 2019b. Punctuality prediction: combined probability approach and random forest modelling with railway delay statistics in Sweden. 2019 IEEE Intell. Trans. Syst. Conf., ITSC 2797–2802. https://doi.org/10.1109/ITSC.2019.8916892, 2019.

Jordan, M.I., Mitchell, T.M., 2015. Machine learning: trends, perspectives, and prospects. Science 349 (6245), 255–260.

Karlaftis, M.G., Vlahogianni, E.I., 2011. Statistical methods versus neural networks in transportation research: differences, similarities and some insights. Transport. Res. C Emerg. Technol. 19 (3), 387–399. https://doi.org/10.1016/j.trc.2010.10.004.

Kecman, P., Goverde, R.M.P., 2012. Process mining of train describer event data and automatic conflict identification. Comput. Railw. XIII: Comput. Syst. Design Oper. Railw. Other Trans. Syst. 127, 227–238. https://doi.org/10.2495/CR120201.

Kecman, P., Corman, F., Meng, L., 2015. Train Delay Evolution as a Stochastic Process. 6th International Conference on Railway Operations Modelling and Analysis - RailTokyo2015. *March, 1–27*.

Kecman, P., Goverde, R.M.P., 2015a. Predictive modelling of running and dwell times in railway traffic. Publ. Trans. 7 (3), 295–319. https://doi.org/10.1007/s12469-015-0106-7.

Kecman, P., Goverde, R.M.P., 2015b. Online data-driven adaptive prediction of train event times. IEEE Trans. Intell. Transport. Syst. 16 (1), 465–474. https://doi.org/10.1109/TITS.2014.2347136.

Keyhani, M.H., Schnee, M., Weihe, K., Zorn, H.P., 2012. Reliability and delay distributions of train connections. In: 12th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems. https://doi.org/10.4230/OASIcs.ATMOS.2012.35.

Kosolsombat, S., Limprasert, W., 2016. Arrival time prediction and train tracking analysis. In: Pacific Rim International Conference on Artificial Intelligence, pp. 170–177. https://doi.org/10.1007/978-3-319-60675-0_15.

Lee, W.H., Yen, L.H., Chou, C.M., 2016a. A delay root cause discovery and timetable adjustment model for enhancing the punctuality of railway services. Transport. Res. C Emerg. Technol. 73, 49–64. https://doi.org/10.1016/j.trc.2016.10.009.

Lee, W., Tax, D.M.J., Duin, R.P.W., 2016b. Beyond condition-monitoring : comparing diagnostic events with word sequence Kernel for train delay prediction. In: European Conference of the Prognostics and Health Management Society, pp. 1–6.

Lemnian, M., Rückert, R., Rechner, S., Blendinger, C., Müller-Hannemann, M., 2014. Timing of train disposition: towards early passenger rerouting in case of delays. In: 14th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems. https://doi.org/10.4230/OASIcs.ATMOS.2014.122.

Leng, N., Corman, F., 2020. The role of information availability to passengers in public transport disruptions: an agent-based simulation approach. Transport. Res. Pol. Pract. 133 (April 2019), 214–236. https://doi.org/10.1016/j.tra.2020.01.007.

Lessan, J., Fu, L., Wen, C., 2019. A hybrid Bayesian network model for predicting delays in train operations. Comput. Ind. Eng. 127, 1214–1222. https://doi.org/10.1016/j.cie.2018.03.017.

Li, Z.C., Wen, C., Hu, R., Xu, C., Huang, P., Jiang, X., 2020. Near-term train delay prediction in the Dutch railways network. Int. J. Real. Ther. 1–20. https://doi.org/10.1080/23248378.2020.1843194, 00(00).

Liebchen, C., Dutsch, S., Jin, S., Tomii, N., Wang, Y., 2021. The ring never relieves. In: 9th International Conference on Railway Operations, Modelling and Analysis (ICROMA 2021).

Liu, Y., Tang, T., Xun, J., 2017. Prediction algorithms for train arrival time in urban rail transit. In: 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), pp. 1–6. https://doi.org/10.1109/ITSC.2017.8317609.

Ma, H., Qin, Y., Han, G., Jia, L., Zhu, T., 2016. Forecast of train delay propagation based on max-plus algebra theory. In: Proceedings of the 2015 Chinese Intelligent Systems Conference, pp. 661–672. https://doi.org/10.1007/978-3-662-48386-2.

Malavasi, G., Ricci, S., 2001. Simulation of stochastic elements in railway systems using self-learning processes. Eur. J. Oper. Res. 131 (2), 262–272. https://doi.org/10.1016/S0377-2217(00)00126-0.

Marković, N., Milinković, S., Tikhonov, K.S., Schonfeld, P., 2015. Analyzing passenger train arrival delays with support vector regression. Transport. Res. C Emerg. Technol. 56, 251–262. https://doi.org/10.1016/j.trc.2015.04.004.

Martin, L.J.W., 2016. Predictive reasoning and machine learning for the enhancement of reliability in railway systems. Int. Conf. Reliab., Saf. Secur. Railw. Syst. 178–188.

Meester, L.E., Muns, S., 2007. Stochastic delay propagation in railway networks and phase-type distributions. Transp. Res. Part B Methodol. 41 (2), 218–230. https://doi.org/10.1016/j.trb.2006.02.007.

Milinković, S., Marković, M., Vesković, S., Ivić, M., Pavlović, N., 2013. A fuzzy Petri net model to estimate train delays. Simulat. Model. Pract. Theor. 33, 144–157. https://doi.org/10.1016/j.simpat.2012.12.005.

Mou, W., Cheng, Z., Wen, C., 2019. Predictive model of train delays in a railway system. In: RailNorrköping 2019. 8th International Conference on Railway Operations Modelling and Analysis (ICROMA), pp. 913–929.

Murali, P., Dessouky, M., Ordóñez, F., Palmer, K., 2010. A delay estimation technique for single and double-track railroads. Transport. Res. E Logist. Transport. Rev. 46 (4), 483–495. https://doi.org/10.1016/j.tre.2009.04.016.

Nabian, M.A., Alemazkoor, N., Meidani, H., 2019. Predicting near-term train schedule performance and delay using Bi-level random Forests. Transport. Res. Rec. 2673 (5), 564–573. https://doi.org/10.1177/0361198119840339.

Nair, R., Hoang, T.L., Laumanns, M., Chen, B., Cogill, R., Szabó, J., Walter, T., 2019. An ensemble prediction model for train delays. Transport. Res. C Emerg. Technol. 104, 196–209. https://doi.org/10.1016/j.trc.2019.04.026.

Nash, A., Huerlimann, D., 2004. Railroad simulation using OpenTrack. WIT Trans. Built Environ. 74, 45–54.

Oneto, L., Fumeo, E., Clerico, G., Canepa, R., Papa, F., Dambra, C., Mazzino, N., Anguita, D., 2018. Train delay prediction systems: a big data analytics perspective. Big Data Res. 11, 54–64. https://doi.org/10.1016/j.bdr.2017.05.002.

Peters, J., Emig, B., Jung, M., Schmidt, S., 2005. Prediction of delays in public transportation using neural networks. In: International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06), 2, pp. 92–97. https://doi.org/10.1109/cimca.2005.1631451.

Pongnumkul, S., Pechprasarn, T., Kunaseth, N., Chaipah, K., 2014. Improving arrival time prediction of Thailand's passenger trains using historical travel times. In: 2014 11th Int. Joint Conf. On Computer Science and Software Engineering (JCSSE), pp. 307–312. https://doi.org/10.1109/JCSSE.2014.6841886.

Pradhan, R., Kumar, A., Kumar, M., Sharma, B., 2021. Simulating and analysing delay in Indian railways. IOP Conf. Ser. Mater. Sci. Eng. 1116 (1), 012127 https://doi.org/10.1088/1757-899x/1116/1/012127.

Radtke, A., Hauptmann, D., 2004. Automated planning of timetables in large railway networks using a microscopic data basis and railway simulation techniques. WIT Trans. Built Environ. 74, 615–625.

Roßler, D., Reisch, J., Hauck, F., Kliewer, N., 2021. Discerning primary and secondary delays in railway networks using explainable AI. Trans. Res. Procedia 52, 171–178. https://doi.org/10.1016/j.trpro.2021.01.018.

Sahin, I., 2017. Markov chain model for delay distribution in train schedules: assessing the effectiveness of time allowances. J. Rail Trans. Plann. Manag. 7 (3), 101–113. https://doi.org/10.1016/j.jrtpm.2017.08.006.

Sara, L., Soumaya, O., Houda, J., Mohamed, A., 2020. Predict France trains delays using visualization and machine learning techniques. Procedia Comput. Sci. 175, 700–705. https://doi.org/10.1016/j.procs.2020.07.103.

Shi, R., Wang, J., Xu, X., Wang, M., Li, J., 2020. Arrival train delays prediction based on gradient boosting regression tress. In: 4th International Conference on Electrical and Information Technologies for Rail Transportation (EITRT) 2019. Lecture Notes in Electrical Engineering, vol. 640, pp. 307–315. https://doi.org/10.1007/978-981-15-2914-6_29.

Shi, R., Xu, X., Li, J., Li, Y., 2021. Prediction and analysis of train arrival delay based on XGBoost and Bayesian optimization. Appl. Soft Comput. 109, 107538. https://doi.org/10.1016/j.asoc.2021.107538.

Shih, M., Dick, C.T., Mussanov, D., Barkan, C.P.L., 2017. A parametric model of the train delay distribution based on traffic conflicts. In: Proceedings of the 7th International Conference on Railway Operations Modelling and Analysis. RailLille2017), pp. 1549–1566.

Ulak, M.B., Yazici, A., Zhang, Y., 2020. Analyzing network-wide patterns of rail transit delays using Bayesian network learning. Transport. Res. C Emerg. Technol. 119, 102749. https://doi.org/10.1016/j.trc.2020.102749.

Wang, P., Zhang, Q., 2019. Train delay analysis and prediction based on big data fusion. Trans. Saf. Environ. 1 (1), 79–88. https://doi.org/10.1093/tse/tdy001.

Wang, R., Work, D.B., 2015. Data driven approaches for passenger train delay estimation. In: 2015 IEEE 18th International Conference on Intelligent Transportation Systems, pp. 535–540. https://doi.org/10.1109/ITSC.2015.94.

Wang, Y., Wen, C., Huang, P., 2021. Predicting the effectiveness of supplement time on delay recoveries: a support vector regression approach. Int. J. Real. Ther. 1–18. https://doi.org/10.1080/23248378.2021.1937355.

Wee, B. Van, Banister, D., 2016. How to write a literature review paper? Transport Rev. 36 (2), 278–288. https://doi.org/10.1080/01441647.2015.1065456.

Wen, C., Lessan, J., Fu, L., Huang, P., Jiang, C., 2017. Data-driven models for predicting delay recovery in high-speed rail. In: 2017 4th International Conference on Transportation Information and Safety. ICTIS 2017 - Proceedings, pp. 144–151. https://doi.org/10.1109/ICTIS.2017.8047758.

Wen, C., Mou, W., Huang, P., Li, Z., 2019. A predictive model of train delays on a railway line. J. Forecast. 39 (February), 470–488. https://doi.org/10.1002/for.2639.

Yaghini, M., Khoshraftar, M.M., Seyedabadi, M., 2013. Railway passenger train delay prediction via neural network model. J. Adv. Transport. 47 (3), 667–678. https://doi.org/10.1002/atr.

Yuan, J., 2007. Dealing with stochastic dependence in the modeling of train delays and delay propagation. Int. Conf. Trans. Eng. 3908–3914. https://doi.org/10.1061/40932(246)641, 2007.

Yuan, J., Hansen, I.A., 2007. Optimizing capacity utilization of stations by estimating knock-on train delays. Transp. Res. Part B Methodol. 41 (2), 202–217. https://doi.org/10.1016/j.trb.2006.02.004.

Zhang, L., Feng, X., Ding, C., Liu, Y., 2020. Mitigating errors of predicted delays of a train at neighbouring stops. IET Intell. Transp. Syst. 14 (8), 873–879. https://doi.org/10.1049/iet-its.2019.0365.

Zhou, P., Chen, L., Dai, X., Li, B., Chai, T., 2020. Intelligent prediction of train delay changes and propagation using RVFLNs with improved transfer learning and ensemble learning. IEEE Trans. Intell. Transport. Syst. 1–13. https://doi.org/10.1109/tits.2020.3002785.

Zhuang, H., Feng, L., Wen, C., Peng, Q., Tang, Q., 2016. High-speed railway train timetable conflict prediction based on fuzzy temporal knowledge reasoning. Engineering 2 (3), 366–373. https://doi.org/10.1016/J.ENG.2016.03.019.

Zilko, A.A., Kurowicka, D., Goverde, R.M.P., 2016. Modeling railway disruption lengths with Copula Bayesian networks. Transport. Res. C Emerg. Technol. 68, 350–368. https://doi.org/10.1016/j.trc.2016.04.018.