A literature review of machine learning models for train delay prediction

D. White, N. Al-Moubayed

November 20, 2019

Abstract

Context: train delay prediction (TDP)

Train delays impose a huge cost on both train operators and passengers. A preliminary analysis of historical delay attribution data released by Network Rail (NR) shows that 35 million minutes of delay were experienced by passengers in the 2018 - 2019 financial year. The prediction of train delays allows the rescheduling or re-routing of crews and rolling-stock, the reduction in the the amount of delay, and improvement of information available to passengers, and subsequent better decision-making.

Objective: the goal of this work is to synthesise available research results to inform evidence-based selection of machine learning (ML) models for TPD.

Method: relevant studies about ML techniques were gathered via a systematic literature review. Supporting contextual studies were also gathered.

Results: 19 studies were selected. 13 distinct models are used: Bayesian networks, support vector regression, random forests, neural networks, fuzzy Petri nets, extreme learning machines, various forms of regression () The data used, and results obtained, are compared.

Conclusion: to do.

Contents

1 Research questions

We are interested in answering the following research questions:

- RQ1: What ML models are commonly used for TDP?
- RQ2: What exogenous data is used to improve the performance of those models? / big data analysis
- RQ3: What are the research challenges in TDP?
- RQ4: No idea

2 Introduction

2.1 Delays

A *delay* may be defined as "the difference between the minimum, or unopposed, travel time, and the actual travel time or [as] the difference between the scheduled running time and actual running time" [6].

A *delay* is a "positive deviation between the realized time and scheduled times of [an] activity" [3].

They are reduced by the provision of *slack* in railway timetables [3]. Slack is the amount of time a train can delayed without the delay propagating.

Mention relationship between slack, capacity, and so on and so forth.

Although precise terminology differs, the literature agrees that there are two principal classes of delay [16]: primary and secondary.

A *primary* or *exogenous* delay is "caused by external stochastic disturbances" [17]. The causes of primary delays are varied and numerous [1], [14], [22], [3], and many different classifications exist. For the purposes of this dissertation, the following classification is proposed:

- *Weather*: severe heat, flooding, landslips, leaves, snow, and ice
- *Passenger*: prolonged alighting and boarding times, large flows
- *Maintenance*: construction work, repair work
- *Other*: accidents, vandalism, trespassing, fatalities, strikes, holidays

A large component of this dissertation is the effect of the inclusion of exogenous data on the predictive capability of various machine learning techniques: this classification broadly follows the sets of data that will be considered.

The relative importance of these factors is poorly studied. A limited study by [16] found that the punctuality of trains is correlated with the number of passengers, occupancy ratio, departure punctuality, and operational priority rules.

The number of passengers affects the *dwell time*, the time "devoted to the loading and unloading processes of the train" [26]. For this dissertation, this refers to the alighting and boarding of passengers. Dwell time is a key parameter of system performance, service reliability and quality [21]. Passenger volume is considered a key factor influencing both dwell time [26] and punctuality [16].

A *secondary* (*knock-on*, *consecutive*) delay is "generated by operations conflicts" [3], i.e. primary delays. Secondary delays often affect both the route on which the primary delay occurred and any connecting routes; delays 'cascade' as "trains, drivers, and crews aren't in the right place at the right time to run other services" [23], or o trains are held according to waiting policies between trains [1].

Although further classifications of secondary delay exist [5], the current level of detail will suffice for this dissertation.

Secondary delays cannot be exactly forecast [1], [14] because they are influenced by multiple interacting factors: the severity of the primary delay, the timetable of the train, the infrastructure, and even the behaviour of the driver, who may drive faster than planned, or reduce dwell time at stations, in order to make up time.

The goal of this work is to synthesise available research results to inform evidence-based selection of ML techniques.

2.2 Timescales

There are several different timescales at which delays can be predicted, such as short-term (predicted using real-time operating data) and long-time (3 days to a week in advance), as in [28]; tactical (in which models are applied to both timetabling and resource planning), and operational (in which models are used for the real-time prediction of train delays), as in [13]. This dissertation is concerned with the short-term / operational level, henceforth referred to as the *real-time* level.

Models for real-time traffic have so far focused on overcoming the combinatorial complexity of train rescheduling, rolling stock and crew scheduling, and delay management [11]. Real-time train delay prediction (RTTDP) models are *online*, i.e. updated as data on train movements becomes periodically available. Many different models have been proposed; they will be discussed later.

For scheduled train services, a trade-off exists between efficiently utilising the capacity of a railway network and improving the reliability and punctuality of train operations.

2.3 Metrics

Punctuality is "a feature consisting in that a predefined vehicle arrives, departs, or passes at a predefined point at a predefined time" [25]. This definition has the interesting effect that trains that arrive *early* cannot be considered punctual. However, the use of punctuality hides a lot of information [27]; reliability and variability are better metrics [16].

reliability has several measures [24]:

- The probability that a train arrives x minutes late (punctuality)
- The probability of an early departure
- The mean difference between the expected arrival time and the scheduled arrival time
- The mean delay of an arrival given that one arrives late
- The mean delay of an arrival given that one arrives more than x minutes late

- The standard deviation of arrival times

variability is a "measurement of the uncertainty of trip journey times in transportation" [16]. It relates to the distribution of arrival times for a train [15]: a train that arrives the same amount of minutes behind schedule every day has low variability, but not would be considered punctual.

The Office of Road and Rail (ORR), the UK's railway regulator, uses Public Performance Measure (PPM) to assess punctuality, and Cancellations and Significant Lateness (CaSL) to assess reliability.

A systematic literature review (SLR) is "a means of evaluating and interpreting all available research relevant to a particular research question or topic area or phenomenon of interest" [30]. The specific objectives of this SLR are:

- To identify categories of ML techniques
- To summarise current research solutions for TDP
- To synthesis the current results from ML techniques for TDP
- To identify the research challenges and needs in the area of TDP

3 Overview of literature review method

A full systematic literature review is beyond the scope of this paper. Instead, a recent literature review focusing on big data analytics (BDA) in railway transportation systems (RTS) [9] was used as a basis for the authors' own. The methodology of [9] is therefore thoroughly discussed. The studies identified in [9] are used to select key search terms for further database queries. Additionally, the cited references of those studies were used to find other relevant papers. The studies found in [9], by database search, and from cited references, are then selected for inclusion in this review by title, abstract, and finally by content. Although the search was by no means exhaustive, the authors are satisfied that the papers gathered represent a comprehensive review of the application of ML to TDP.

3.1 Recent applications of big data analytics in railway transportation systems: A survey [9]

Readers are invited to read the paper themselves; only an overview of content relevant to this literature review is presented here. The authors identified the following data-related keywords: "Data analytics, Big data, Data mining, Machine learning, Descriptive analytics, Predictive analytics" and the following RTS-related keywords: "Rail, Railway Engineering, Railway Systems, Railway Operations, Railway Safety, Railway Maintenance". Of particular significance is "Rail Operations", which covers the actual *running* of trains on a RTS, and therefore delays. The authors limited the scope of their search to papers in scientific journals, conferences, and dissertations in English from the last 15 years (i.e. 2003 - 2017). The authors specifically included only papers with quantitative results, disregarding those about qualitative challenges of BDA in RTS or purely mathematical modelling of RTS problems. The authors searched ScienceDirect, Emeralds, Scopus, EBSCO, and IEEE Xplore, and also used cited references of studied papers as a source. 115 papers were found and were classified by a four-layer structure:

1. Area of RTS: Maintenance, Operations, Safety
2. Analytic category: descriptive, predictive, prescriptive
3. BDA model: clustering, numeric prediction, association, statistical analysis, image processing, simulation, classification, semantic analysis, text analysis, optimisation
4. Implementation technique: Bayesian network, SVM, SVR, Decision Tree, ANN, Regression

We are interested in Operations; papers in the other areas are disregarded. Within Operations, the authors discuss the applications of BDA to RTS, data collection and sources in RTS, and finally the studies themselves. Only those that focus on TPD are considered. In total, 19 papers were selected by title for inclusion in this literature review; they provided the foundation for a subsequent database search.

3.2 Database search

The papers selected from [9] were used to identify the following key search terms: "train", "delay" and "prediction". Alas, "train" is a common word in scientific literature, and this confounded initial results somewhat. Appropriate synonyms were identified for each term. Where databases allowed Boolean operators, the following formulations were used:

- (((trains OR train) NOT training) OR rail OR railway OR railways OR railroad OR railroads)
- (delay OR delays OR "event times")
- (prediction OR predicting OR analysis OR analyzing OR estimation OR estimating)

The following databases, based on those used by [10], were searched:

- ACM Digital Library
- IEEE Xplore
- ScienceDirect
- SpringerLink

Where possible, the discipline was restricted to Computer Science. Initially, approximately 3000 studies were identified. Some post-processing was necessary to reduce the number of studies retrieved from IEEE Xplore, in particular, resulting in 69 studies.

3.3 Study selection

Study selection was a three-stage process:

1. Initial selection by title
2. Selection by abstract
3. Exclusion by content

Of the 88 total studies found, 5 were duplicates. Studies were further selected by abstract. If the authors were content from the abstract a paper was relevant, it was selected; otherwise, the paper was read to determine suitability. In total, 19 studies were identified for inclusion in this literature review. Of those, 4 were discovered through other means - either from a prior, less structured, search, or from cited references.

4 Overview of studies

We identified 19 studies in the literature that focus on the application of ML to TPD. An preliminary analysis shows that all work occurred in the past decade (i.e. during, or after, 2010), with most studies (47%) published in 2017 and 2019.

Year	#
2010	1
2011	1
2012	1
2014	1
2015	2
2016	2
2017	4
2018	3
2019	5
Total	19

13 distinct ML techniques were identified. 29 were applied in total, as several papers compared and contrasted different techniques, or variations of the same technique (as in [12][17][14][13]), or even combined multiple models (e.g. the random forests regression of [**wen˙et˙al˙2017**], or emsemble model of [**nair˙et˙al˙2019**]; in these cases, each distinct 'usage' is counted separately.

The most popular techniques are random forests (20%) and extreme learning machines (17%)., although these figures are skewed by the inclusion of four papers CITE HERE which are closely related.

Some studies use techniques that may be more accurately classified as 'statistics' rather than ML (i.e. simple variants of regression, as in [20][29]) or which are closer to *algorithmic* models than ML ([**hansen˙goverde˙van˙der˙meer˙2010**]); however, as these lines is blurred, and for completeness' sake, all were selected for inclusion in this review.

ML model	Acronym	#
Bayesian network	BN	4
Kernel method	KN	1
Extreme learning machine	ELM	5
Random forest	RF	6
Fuzzy Petri net	FPN	1
Adaptive neural fuzzy inference system	ANFIS	1
Gradient-boosted regression trees	GBRT	1
k-nearest neighbour	k-NN	1
Artificial neural network	ANN	2
Support vector regression	SVR	2
Kernel regression	KR	2
Markov		2
Decision tree	DT	1
Total		29

5 ML models

In this section, each of the distinct ML models identified previously is discussed.

5.1 Bayesian networks

We shouldn't have specific sections for each. We're trying to group by model, remember. A Bayesian network (BN) is a "probabilistic graphical model that uses Bayesian inference for probability computations" [**towards˙data˙science˙BN˙intro**]. Each directed edge models a conditional independence, allowing "the incorporation of massive historical data" [12].

The study compared heuristic hill-climbing, primitive linear, and hybrid structure BNs.

There is no common dataset for TDP, unlike other ML areas such as computer vision. That said, data trends can be observed in the papers gathered. Several use the TNV-Extract tool developed by Goverde ? and thus use data from the Netherlands. Several use HSR data from China. Four - those use Italian rail data. The fields of each dataset are explored later on.

5.1.1 A hybrid Bayesian network model for predicting delays in train operations [12]

Objectives

introduced the first hybrid BN-based to the area of TDP.

The hybrid heuristic BN, built using naive and heuristic structures and refined by domain knowledge and expert judgements. Achieved 80% accuracy over a 60-minute prediction horizon.

Advantages: it is simple, and so computational efficient. The authors note that results could be improved by including the 'spatiotemporal' properties of each section, which we have taken to mean the speed at which trains can run.

The MAE prediction error was around 30s; the RMSE for both predicted arrival and departure delays was less than 2 minutes

However. Predictions from the hybrid model matched observations only 56% of the time. The authors attribute this to primarily to the discrete prediction space, and so employed discretisation to convert continuous variables into bins.

explore three different Bayesian network schemes: heuristic hill-climbing, primitive linear, and hybrid. Hybrid, incorporating domain knowledge and judgements of local experts, was found to outperform other models, with an accuracy of over 80% in predictions within a 60-minute horizon. The authors define a railway system as several interconnected subsystems: infrastructure, rolling stock, control and communication, and various operational rules and policies. It was found that arrival and departure delays follow the same distribution, with a linear relationship (chain) with a high correlation between arrival and departure delays at the same station (at least 94

5.1.2 Stochastic prediction of train delays in real-time using Bayesian networks [corman˙kecman˙2018]

present a stochastic model for predicting the propagation of train delays based on Bayesian networks (BNs). BNs allow the updating of probability distributions and reduce the uncertainty of future train delays in real-time as more data continuously comes available from the monitoring system. This authors extend this approach by modelling the interdependence between trains that share the same infrastructure or have a scheduled passenger train. The model is tested on historical train realisation data from a bus corridor in Sweden

5.2 Support vector regression

Support vector regression (SVR) uses support vector machines (SVMs) as

A SVM maximises the margin between two or more classes to find the optimal *hyperplane*, the seperator between two classes. In SVR, the hyperlaine is used to

A kernel

SVRs can be used for continuous values; SVMs are for classification.

6 Ensemble methods

Ensemble methods use multiple methods to generate forecast. The rationale behind such a framework is simple: gathering forecasts from a diverse set of models reduces bias and error rates. [**nair˙et˙al˙2019**] uses a purely data-driven ensemble method; [**oneto˙et˙al˙2018**] combine a RF and a experience-based model.

7 Analytical models

A brief overview of analytical models is provided for context, with emphasis on the shortcomings of such models, and how data-driven approaches can ameliorate these shortcomings. In short: analytical models perform worse, but are easy to explain, understand, and interpret. Analytical models cannot capture the complexity of such models.

An *analytical model* is "primarily quantitative or computational in nature and represents the system in terms of a set of mathematical equations that specific parametric relationships and their associated parameter values as a function of time, space, and/or other system parameters" [8]. Current state-of-the-art TDPS use analytical models [17].

Simplistic early models, such as [7] made overly restrictive assumptions about railway operations by, for example, forbidding overtakes, assuming that departure times are uniformly distributed, and that the speed of each train is unique and constant.

Subsequent work in this area has largely relaxed these assumptions, by including factors such as overtakes, different speeds, priority systems, and uncertainties associated with train departure time [19], [4]. More complex models have also emerged, incorporating stochastic approximation [2] and the impact of dispatching strategies on train delays and passenger waiting time [18]. Although the state-of-the-art advances constantly, a good example of an recent *in-use* system is [1], which is currently used in the German rail network.

8 Datasets

There is no common dataset for TDP, unlike other ML areas such as computer vision. That said, data trends can be observed in the papers gathered. Several use the TNV-Extract tool developed by Goverde ? and thus use data from the Netherlands. Several use HSR data from China. Four - those use Italian rail data. The fields of each dataset are explored later on.

9 Fields

10 Exogenous data

It is widely accepted amongst ML practitioners that the greater the quantity of information available for the creation of a model, the greater the performance of that model will be. Features can either be *engineered* or exogenous data can be incorporated. This is the realm of 'big data', which involves "multiple datasets and a complicated structure" [**Ghofrani˙et˙al˙2018**]. This section is broken down by the classification defined in the introduction.

Data is exogenous if it is independent of other input data but the output data depends on it. The scope for inclusion is essentially limitless: any source of data which may affect railway operations is a viable candidate. In the studies selected for this review, there are two main sources: infrastructure [**markovic˙et˙al˙2015**] [**milinkovic˙et˙al˙2013**] via *expert opinion*, and weather [**oneto˙et˙al˙2017**] [**oneto˙et˙al˙2018**], [**oneto˙et˙al˙2019**].

Special mention must go to [**nair˙et˙al˙2019**], which used network traffic states, such as likely stretch conflicts and current headways, weather, event information, work zone information, inferred occupation conflicts, train connections, and rolling stock rotations.

10.1 Weather

Weather is a common cause of primary delay. The classification defined earlier included severe heat, flooding, landslips, leaves, snow, and ice. It is expected that weather-induced delays are seasonal. Severe heat is likely to cause delays in summer; leaves in autumn; and snow and ice in winter. [**brazil˙2017**] found that most weather-caused delays occurred in the last third of the year, with a peak in November.

Weather was first included in a TDP model, to the best of the authors' knowledge, in [**oneto˙et˙al˙2016**].

Subsequent studies have established the impact of severe weather on train delays "BRAZIL201769, title = "Weather and rail delays: Analysis of metropolitan rail in Dublin". Papers largely agree that, dependent on climate, weather delays most trains during the last third of the year, with November a particular culprit, likely due to the sustained impact of leaf-fall.

[**wang˙et˙al˙2019**] observed that in locations less prepared for specific severe weather - such as snowy weather in southern cities - delays were greater. They found that in severe weather trains delays are determined by mainly the type of bad weather, but in ordinary weather they are determined mainly by historical delay time and the delay frequency of trains.

Fields tend to be largely consistent: there are only so many weather variables of note. [**wang˙et˙al˙2019**] used lowest temperature, highest temperature, weather category (e.g. "overcast", "light rain"), Beaufort scale (wind speed) and air quality index (seemingly unique to China). Data was collected from 344 cities along the route in question. However, the timeframe used was between 1st January and 31st March. As weather-related delays are seasonal, this reduces the validity of conclusion relating to the importance of weather.

[**oneto˙et˙al˙2016**] note that weather conditions can additionally influence passenger flow and consequently dwell times, which have already been described as a key influence on delays.

[**oneto˙et˙al˙2016**] CITE ALL HERE use temperature, relative humidity, wind direction, wind speed, rain level, pressure and solar radiation. [**nair˙et˙al˙2019**] use weather data from 92 weather observatories, including snow conditions, visibility, and temperature. They found that weather has only a small impact of delays; an analysis of delay-attribution showed that less than 3% of delays were directly attributed to weather.

The proposed dataset for this dissertation uses similar fields. For interoperability with forecasts, the comprehensive data provided by the Met Office has been mapped to a more simplistic set of fields: wind gust, relative humidity, visiblity, wind direction, wind speed, temperature, weather type (category), and precipitation probability.

[**wang˙et˙al˙2018**] collected weather data from 344 cities along the route in question, Beijing to Guangzhou. It is worth nothing that the two are approximately 2200km, and so delays are of a magnitude not frequently found

[**oneto˙et˙al˙2016**] found that the inclusion of weather data improved the accuracy of their RF model by approximately 10%, with the caveat that the further ahead in the future the forecast is (and thus the less accurate, the smaller this increase was.

10.2 Infrastructure

Infrastructure naturally matters to trains. Extant work has been surprisingly lacking incorporating infrastructure characteristics (single or double track, station layouts, interlocking, cant, speed limit), and so on.

[**milinkovic˙et˙al˙2013**] groups infrastructure opinions. Collected opinions from traffic dispatchers, operators, and experts famiilar with the functioning of the system. Was used more broadly to define input variables, and the primary causes of delay (not the causes of primary delay). Defined three input parameters: the train category, timetable influence, and the distance travelled by the train. Timetable influence was used as a catch-all of sorts; the study is vague on specifics. It included the influence of infrastructure parameters, timetable characteristics, operation time, the type of locomotive, local conditions, technological solutions, principles for safety and signalling, and weather conditions. This is for the FPN!

For the ANFIS, which used real-life data (go into detail here), an 'infrastructure influence', which included the percentage of restricted speed sections, the number of junctions, and the number of stations). Included section length, section plans, restricted speed, and track routes.

The authors note that the average track occupancy of a section can indicate possible bottlenecks of a system. This is close to the *tactical* level briefly discussed earlier: the use of data to make decisions on improving infrastructure.

The dataset for the proposed study includes infrastructure characteristics used by ? to actually plan train delays.

Used the Delphi method.

It seems inherently obvious is should have a huge effect on the propagation of delays.

[**markovic˙et˙al˙2015**] explores an expert opinion in much better defined terms. The influence of multiple factors along a rail line (single-tracking, reduced speeds, characteristics of block and interlocking systems, number of stations, stops, loops, road-rail level crossings, and junctions) is aggregated into one variable with a value determined by the expert opinions of five dispatchers. Estimates obtained via the Delphi method: experts evaluate a route over multiple rounds until a consensus is reached. Strong correlation found between expert opinions and train delays.

The condition of the Serbian railways is considerably worse than that of many other countries explored. Characterised by: recently renewed lines (enabling maximum speed), lines with sections with TSRs, single and double-track lines, many junctions and railroad crossing, lines split int sections with different signalling and safety equipment.

Evaluated each route on a scale of 1 - 10. 1 denotes a route with the highest number of infrastructural factors that could cause unplanned delays.

The study only considered a limited number of routes: those passing through Rakovica station. Models for larger areas cannot rely on human assessment of infrastructure conditions. 39 lines were evaluated. Specifically: number of stations/stops/junctions/loops/crossings, percentage of single track, percentage with restricted speed, length with restricted speed, block section, track clear section (station distance, braking distance, automatic block system, centralised traffic control, axle counters).

[**nair˙et˙al˙2019**] take exactly this approach. The authors reconstruct the network and estimate capacity directly from passing messages. Furthermore, the method used generated train-class specific networks. The inferred method is employed for various downstream tasks: inferring train paths, conflict status estimation, typical travel time estimation. "Passing" messages are sorted by date, time, and train. If there are sufficient observations, the control point and track stretch is recorded as an edge. The frequency of transitions from each outgoing edge, the mean and standard deviation of travel times are also recorded for each edge. A feasibility matrix for each outgoing edge is recorded at each vertex, which records pairwise edge feasible flows at each section by identifying movements by two trains in a short time window; this is used to identify potential conflicts between trains when there are deviations from the schedule. Reconstructed networks around several major hubs were inspected by hand and found to be accurate.

Station attributes used included the designated platform, station attributes, historical mean delay at tracks, platforms, actual platform, track allocation, and track / platform change status.

10.3 Maintenance

Only one paper was found that incorporated the "maintenance" class: [**nair˙et˙al˙2019**]. The authors used work zone information, indicating location, duration, and the likely impact of different train categories.

10.4 Other

No papers were found to incorporate accidents, vandalism, trespassing, or fatalities, or strikes. Holidays are, however, included in [**nair˙et˙al˙2019**].

References

[1] Annabell Berger et al. "Stochastic Delay Prediction in Large Train Networks". In: *11th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems*. Ed. by Alberto Caprara and Spyros Kontogiannis. Vol. 20. OpenAccess Series in Informatics (OASIcs). Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2011, pp. 100–111. ISBN: 978-3-939897-33-0. DOI: 10.4230/OASIcs.ATMOS.2011.100. URL: http://drops.dagstuhl.de/opus/volltexte/2011/3270.

[2] Malachy Carey and Andrzej Kwieciński. "Stochastic approximation to the effects of headways on knock-on delays of trains". In: *Transportation Research Part B: Methodological* 28.4 (1994), pp. 251–267. DOI: 10.1016/0191-2615(94)90001-9.

[3] Fabrizio Cerreto et al. "Causal Analysis of Railway Running Delays". In: *11th World Congress on Railway Research (WCRR 2016)*. 2016.

[4] Bintong Chen and Patrick T. Harker. "Two Moments Estimation of the Delay on Single-Track Rail Lines with Scheduled Traffic". In: *Transportation Science* 24.4 (1990), pp. 261–275. DOI: 10.1287/trsc.24.4.261.

[5] Winnie Daamen, Rob M. P. Goverde, and Ingo A. Hansen. "Non-Discriminatory Automatic Registration of Knock-On Train Delays". In: *Networks and Spatial Economics* 9.1 (2008), pp. 47–61. DOI: 10.1007/s11067-008-9087-2.

[6] Mark Dingler et al. "Determining the causes of train delay". In: *AREMA Annual Conference Proceedings*. 2010.

[7] Ove Frank. "Two-Way Traffic on a Single Line of Railway". In: *Operations Research* 14.5 (1966), pp. 801–811. DOI: 10.1287/opre.14.5.801.

[8] Sanford Friedenthal, Alan Moore, and Rick Steiner. *A Practical Guide to SysML*. 2nd ed. Morgan Kaufmann Publishers, 2012.

[9] Faeze Ghofrani et al. "Recent applications of big data analytics in railway transportation systems: A survey". In: *Transportation Research Part C: Emerging Technologies* 90 (2018), pp. 226–246. DOI: 10.1016/j.trc.2018.03.010.

[10] Sarah Heckman and Laurie Williams. "A systematic literature review of actionable alert identification techniques for automated static code analysis". In: *Information and Software Technology* 53.4 (2011), pp. 363–387. DOI: 10.1016/j.infsof.2010.12.007.

[11] Pavle Kecman, Francesco Corman, and Lingyun Meng. "Train delay evolution as a stochastic process". In: *Proceedings of the 6th International Conference on Railway Operations Modelling and Analysis: RailTokyo2015*. 2015. URL: http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-119746.

[12] Javad Lessan, Liping Fu, and Chao Wen. "A hybrid Bayesian network model for predicting delays in train operations". In: *Computers & Industrial Engineering* 127 (2019), pp. 1214–1222. DOI: 10.1016/j.cie.2018.03.017.

[13] Nikola Marković et al. "Analyzing passenger train arrival delays with support vector regression". In: *Transportation Research Part C: Emerging Technologies* 56 (2015), pp. 251–262. DOI: 10.1016/j.trc.2015.04.004.

[14] Sanjin Milinković et al. "A fuzzy Petri net model to estimate train delays". In: *Simulation Modelling Practice and Theory* 33 (2013), pp. 144–157. DOI: 10.1016/j.simpat.2012.12.005.

[15] Robert B. Noland and John W. Polak. "Travel time variability: A review of theoretical and empirical issues". In: *Transport Reviews* 22.1 (2002), pp. 39–54. DOI: 10.1080/01441640010022456.

[16] Nils O.E. Olsson and Hans Haugland. "Influencing factors on train punctuality—results from some Norwegian studies". In: *Transport Policy* 11.4 (2004), pp. 387–397. DOI: 10.1016/j.tranpol.2004.07.001.

[17] Luca Oneto et al. "Advanced Analytics for Train Delay Prediction Systems by Including Exogenous Weather Data". In: *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)* (2016). DOI: 10.1109/dsaa.2016.57.

[18] Süleyman Özekici and Selim Şengör. "On a Rail Transportation Model with Scheduled Services". In: *Transportation Science* 28.3 (1994), pp. 246–255. DOI: 10.1287/trsc.28.3.246.

[19] E. R. Petersen. "Over-the-Road Transit Time for a Single Track Railway". In: *Transportation Science* 8.1 (1974), pp. 65–74. DOI: 10.1287/trsc.8.1.65.

[20] Suporn Pongnumkul et al. "Improving arrival time prediction of Thailand's passenger trains using historical travel times". In: *2014 11th International Joint Conference on Computer Science and Software Engineering (JCSSE)* (2014). DOI: 10.1109/jcsse.2014.6841886.

[21] Andre Puong. "Dwell Time Model and Analysis for the MBTA Red Line". In: (Mar. 2000).

[22] Network Rail. *Delays explained*. 2019. URL: https://www.networkrail.co.uk/running-the-railway/looking-after-the-railway/delays-explained/ (visited on 11/19/2019).

[23] Network Rail. *Knock-on delays*. 2019. URL: https://www.networkrail.co.uk/running-the-railway/looking-after-the-railway/delays-explained/knock-on-delays/ (visited on 11/19/2019).

[24] P Rietveld, F.R Bruinsma, and D.J van Vuuren. "Coping with unreliability in public transport chains: A case study for Netherlands". In: *Transportation Research Part A: Policy and Practice* 35.6 (2001), pp. 539–559. DOI: 10.1016/s0965-8564(00)00006-9.

[25] Andrzej Rudnicki. "Measures of Regularity and Punctuality in Public Transport Operation". In: *IFAC Proceedings Volumes* 30.8 (1997), pp. 661–666. DOI: 10.1016/s1474-6670(17)43896-1.

[26] Hor Peay San and Mohd Idrus Mohd Masirin. "Train Dwell Time Models for Rail Passenger Service". In: *MATEC Web of Conferences* 47 (2016), p. 03005. DOI: 10.1051/matecconf/20164703005.

[27] R Skagestad. "Kritiske prestasjonsindikatorer i jernbanedrift". PhD thesis. Norwegian University of Science and Technology, 2004.

[28] Pu Wang and Qing-peng Zhang. "Train delay analysis and prediction based on big data fusion". In: *Transportation Safety and Environment* 1.1 (2019), pp. 79–88. DOI: 10.1093/tse/tdy001.

[29] Ren Wang and Daniel B. Work. "Data Driven Approaches for Passenger Train Delay Estimation". In: *2015 IEEE 18th International Conference on Intelligent Transportation Systems* (2015). DOI: 10.1109/itsc.2015.94.

[30] C.C. Williams and J.K. Hollingsworth. "Automatic mining of source code repositories to improve bug finding techniques". In: *IEEE Transactions on Software Engineering* 31.6 (2005), pp. 466–480. DOI: 10.1109/tse.2005.63.