A Bayesian network model to predict the effects of interruptions on train operations

Article in Transportation Research Part C Emerging Technologies · May 2020

DOI: 10.1016/j.trc.2020.02.021

7 authors, including:

Ping Huang
Southwest Jiaotong University

C. Wen
Southwest Jiaotong University

Qiyuan Peng
Southwest Jiaotong University

Liping Fu
University of Waterloo

Transportation Research Part C: Emerging Technologies. 2020 May 1; 114:338-58.

A Bayesian network model to predict the effects of interruptions on train operations

Ping Huang [a,b], Javad Lessan [c], Chao Wen [a,b,**], Qiyuan Peng [a], Liping Fu [a,c], Li Li [a], and Xinyue Xu [b]

[a] National Engineering Laboratory of Integrated Transportation Big Data Application Technology, Southwest Jiaotong University, Chengdu, 610031, China;

[b] State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China, 100044;

[c] High-speed Railway Research Centre, University of Waterloo Waterloo, N2L3G1, Canada.

First author: huangping129@my.swjtu.edu.cn

****Corresponding author: wenchao@swjtu.cn**

Abstract

Based on the Bayesian network (BN) paradigm, we propose a hybrid model to predict the three main consequences of disruptions and disturbances during train operations, namely, the primary delay (L), the number of affected trains (N), and the total delay times (T). To obtain an effective BN structure, we first analyze the dependencies of the involved factors on each station and among adjacent stations, given domain knowledge and expertise about operational characteristics. We then put forward four candidate BN structures, integrating expert knowledge, the interdependencies learned from real-world data, and real-time prediction and operational requirements. Next, we train the candidate structures based on a 5-fold cross-validation method, using the operational data from Wuhan-Guangzhou (W-G) and Xiamen-Shenzhen (X-S) high-speed railway (HSR) lines in China. The best performing structure is nominated to predict the consequences of disruptions and disturbances in the two HSR lines. Comparisons results show that the proposed model outperforms three other commonly used predictive models, reaching an average prediction accuracy of 96.6%, 74.8%, and 91.0% on the W-G HSR line, and 94.8%, 91.1%, and 87.9% on the X-S HSR line for variables L, N, and T, respectively.

Keywords: Train operation, Disturbances and Disruptions, Real-time Prediction, Bayesian networks

1. Introduction

A railway system is a complex mode of transportation, which comprises several subsystems operating under various operational rules and is subject to many interrelated and unforeseeable operation-interrupting events (Lessan et al., 2018). For instance, any number of human- or equipment-related faults, failures, or anomalies can cause disturbances and disruptions in train operations, leading to inconveniences for passengers and inefficiency of the rail system (Binder et al., 2017). Disturbances are considered relatively small perturbations of the railway system that can be handled by modifying the timetable without modifying the duties for rolling stock and crew. On the other hand, disruptions are relatively large incidents, requiring both the timetable and the duties for rolling stock and crew to be modified (Binder et al., 2017; Cacchiani et al., 2014). According to statistics from a Dutch railway network, infrastructure-related disruptions occur approximately 22 times per day and each disruption, on average, can last 1.7 h (Jespersen-Groth et al., 2009). The Austrian Federal Railway reported financial losses of more than €100 million every year because of flooding (Kellermann et al., 2016). These figures signify the importance of reducing the consequences of unexpected interruptions if they cannot be avoided. To this end, we need sophisticated tools capable of predicting the effects of disruptions and disturbances in train operations to support timetable re-scheduling, train dispatching, and assist operators in making timely decisions (Lessan et al., 2018). However, train operations behave stochastically largely due to unforeseeable events, and present a highly dynamical and interdependent behavior due to the sharing of limited resources, such as the interlocking equipment and dispatching operations (Corman and Meng, 2015). These reasons bring about nontrivial challenges in estimating the effects of disruptions and disturbances, and thus necessitate a well-designed tool to provide train dispatchers with a real-time, accurate estimation of the effects of unexpected disruptions and disturbances (Corman and Meng, 2015). In this regard, predictive models based on graph and network theory, such as a Bayesian network, turned out to be good choice to model causal relationships of the elements of a complex rail system (Lessan et al., 2018), having presented potentials in many other similar situations (Nielsen and Jensen, 2009). With a well-designed BN structure, one can express the interactive and causal relationships of the involved variables and fuse multiple sources of information through a graphical structure (Lessan et al., 2018). These powerful characteristics motivated us to take advantage of BNs to model the complexity, interdependency, and dynamical behaviors involved in train operations, aiming to predict impacts of disruptions and disturbances on train operations.

The main contribution of this paper is to propose a real-time BN model to predict both the spatial and temporal propagation of interruptions on train operations. Based on analysis of the involved factors, operational interdependencies, and evaluation of their influence, we focus on three main issues. In other words, once a disruption or disturbance occurs, we aim to find out: 1) how the primary delay propagates from the occurrence point to subsequent stations, 2) how many trains will be affected at each station, 3) how the delay time fluctuates at each station given trains dispatching and overtaking operations. The answer to these issues can provide useful guides to track both the temporal and spatial propagations of delays in train operations. These measures, which are also supplemental to any delay propagation predictive model, can aid in effective real-time dispatching decisions in train operations.

The rest of this paper is structured as follows. In Section 2, we describe the issues of real-time delay prediction, identify three factors to measure the effects of disruptions and disturbances, and review the related work on disruption management. In Section 3, we briefly introduce the theoretical foundation of the BN model, and investigate the dependency of the selected variables based on expert knowledge. In Section 4, we describe the operational data used in this study, present the predictive model, and train the BN model based on the cross-validation method. In Section 5, we analyze the goodness-of-fit, predicting residuals of the BN model. Then, we evaluate the proposed model using different evaluation metrics to compare the performance of our BN model against other widely used predictive models. Finally, in Section 6, we present the conclusions of our research.

2. Problem description and related work

2.1. Problem description

Figure 1 shows a fragment of a time-space diagram of trains operating on one of the HSR networks in China, with the horizontal axis representing the timeline and vertical axis representing the sequence of stations.

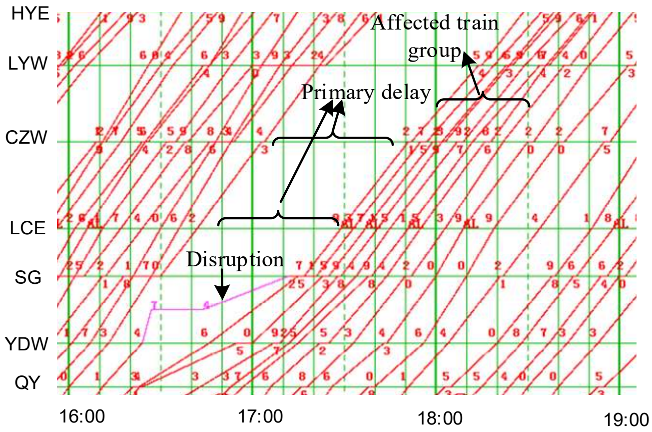


Fig. 1. The effects of disruption shown in a time-space diagram (horizontal axis is time, vertical axis is space, and the numbers represent the rest of the actual time divided by 10 min).

The figure shows that an event occurred in the YDW-SG section and resulted a primary delay to the affected train and subsequent (secondary) delays to a group of other trains. Let us consider a delay of a total of R involved trains, numbered as 1, 2,…, R according to their operation order, at SG station. The arrival delays of these R trains to station SG are t_1, t_2,…, t_R, respectively. Then, following subsequent dispatching operations, such as delay recovery, propagation, and interference of trains, the delay of each train, the number of delayed trains, and the total delay time of affected trains can increase or decrease, as these features are all random variables in the train operations. In this paper, the following factors are introduced to measure the effects of interruptions on train operations:

- N, the number of delayed trains at each station, $N=\{N_1, N_2, ..., N_M\}$ and $N_m=R$, where R is the number of delayed trains at station m, and M is the number of stations the delayed train groups pass through.
- T, the total delay time of trains at station, $T=\{T_1, T_2, ..., T_M\}$, where $T_m = t_{m,1}+ t_{m,2}+ \ldots+ t_{m,R}$ and $t_{m,r}$ is the delay time of train r at station m.
- L, the primary delay, where $L=\{L_1, L_2, ..., L_M\}$ and $L_m = t_{m,1}$. Generally, L_1 is the delay directly caused by interruptions (the delay that first occurs), and L_2, L_3,…, L_M are the propagation consequences of L_1 when it cannot be recovered by rescheduling train movements (Corman et al., 2014; Meester and Muns, 2007). To consider the train overtaking, we assume that if the train with the primary delay is overtaken, L_m is always the delay time of the first delayed train at station m in the actual timetable.

In addition, as the primary delay is critical to N and T, we also consider it as one of the metrics of disruption effects. A primary delay is caused by external sources of interruptions (e.g., rolling stock failures, track failures, power outages) and cannot be recovered by rescheduling train movements (Corman et al., 2014). In other words, it is called a primary delay if the delay is caused by interruptions, and it is called the secondary/knock-on delay if the delay is due to other trains (Meester and Muns, 2007). Therefore, the primary delay in Fig. 1 is defined as L.

2.2. Literature review

Railway disruptions can be caused by external factors, such as bad weather condition, or by factors inside the system, such as rolling stock failures, human-related faults, or temporary speed limitations, and signal failures (Higgins et al., 1995; Wen et al., 2017). Traditionally, the timetable rescheduling problem under disruptions tends to be regarded as an optimization problem (Binder et al., 2017; Cacchiani et al., 2014). In this regard, mathematical optimization tools, such as integer programming (Veelenturf et al., 2015) and mixed-integer linear programming (Zhan et al., 2016) are usually used to meet different objectives such as maximizing the passenger satisfaction, minimizing the operational costs and the train delays (Binder et al., 2017). Also, simulation methods including simulation algorithms and systems are used to manage disruptions, such as the simulation model to analyze the effects of delayed trains (Burdett and Kozan, 2014), and to analyze the congestion delays in railway networks (Murali et al., 2010). In a similar spirit, one can find other simulation systems for evaluating system safety, train speed, and the accuracy of train dwell location (Yasunobu et al., 1993), the RailSys (Wiklund, 2003), and OpenTrack (Nash and Huerlimann, 2004) systems.

Recently, data-driven methods turned out to be an effective tool in improving real-time timetable rescheduling (Wen et al., 2019). Such methods include statistical models, graph- or network-based models, and machine learning models to analyze the characteristics of train delays or predict train delays. Statistical models such as probability density models and regression models are used to understand train operations in terms of their distributional characteristics. Indeed, probability distribution models are one of the widely used methods in railway system. Due to the unexpected disturbances, complex systems such as rail system seem to show heavy-tailed distributional forms (Lessan et al., 2018). Common right-skewed probability distribution models, such as exponential distribution (Briggs and Beck, 2007; Harris, 2006; Huisman and Boucherie, 2001), Weibull distribution (Goverde et al., 2013; Goverde et al., 2001), and log-normal distribution (Huang et al., 2019; Yang et al., 2019) are three of the widely used probability density models in railway systems. Cerreto et al. (2016) investigate the distributions of the actual running times of trains in sections. Lessan et al. (2018) develop a statistical method to estimate the arrival delay distribution, according to the previous departure delay and running time distributions. Another widely used statistical approach for train operation data is regression-based models. Gorman (2009) estimate a linear regression model to predict the total running time of freight trains in the US. Murali et al. (2010) use simulation method to mimic the train operations, and then develop a linear regression model incorporating train operating parameters to estimate train delays. Van der Meer et al. (2009) focus on statistical analysis of train running times among stations to improve the prediction accuracy of delay propagation in railway systems. The analyses show a strong correlation between arrival delays and dwell times, whereas the correlation between running times and departure delays is much weaker. Meester and Muns (2007) use a phase-type distribution model to evaluate the secondary delay distribution according to the primary delay distribution. Kecman and Goverde (2015b) compare the predictive accuracy of the least-trimmed squares (LTS) robust linear regression model and decision trees model in train running time and dwelling time prediction, and show that the local LTS model outperforms in terms of computational time and prediction accuracy. Li et al. (2016) combine linear regression and k-nearest neighbor models to predict the dwelling time of trains, and show that the accuracy is more than 85.8% during peak hours and 80.1% during off-peak hours. Wen et al. (2017) and Huang et al. (2019) use data fitting techniques to study the impacts of disruptions on train operations. Wen et al. (2017) use a regression model—inverse model, to fit the frequency of the affected number of trains while treating the affected number of trains as the independent factor, and their frequencies as the dependent variable. Huang et al. (2019) use probability density models to fit the probability distribution of the affected number of trains, after clustering the disruptions into sub-clusters, according to their influence on train operations. Although, these two studies focus on the influence of disruption on train operations, their contributions are limited, as the former one only focuses on the number of affected trains, and the latter one uses density models which only produces the probability distribution of the number of affected trains and total delay times rather than their predicting their numbers exactly.

As train operations are composed of successive processes, such as arrival, passage, and departure events, event- and network-based models that can describe the train events are widely used in railway system. In the event-based prediction problems, the graph- and network-based model are widely used to improve timetable rescheduling by predicting train arrival times, running times, dwelling times, or train delays (Büker and Seybold, 2012; Berger et al., 2011; Corman et al., 2014; Goverde, 2010; Kecman and Goverde, 2015a; Milinković et al., 2013). These models are developed by graphically representing the train events, using the nodes of the graph where the relationships of train events are usually expressed by joint probability density functions. Recently, BNs are applied for predicting delay and improving rescheduling operations. In this regard, two different approaches are used in delay prediction. The first one takes the arrival and departure events of each train at each station and treats them as separate nodes of a BN architecture (Lessan et al., 2019). The second one takes the arrival and departure events of two consecutive trains as the nodes of a BN (Corman and Kecman, 2018). Markov model is a widely used graph-based model that considers train operations as discrete processes (Barta et al., 2012; Gaurav and Srivastava, 2018; Kecman et al., 2015; Şahin, 2017). In addition, to estimate the length of disruptions, Zilko et al. (2016) develop a Copula BN model to simultaneously predict the latency time and the repair time of the track circuit failures based on the failure records data. The latency time is the length of time the mechanics need to get to the disrupted site, whereas the repair time is the length of time to repair the problem. Based on the predicted disruption length in Zilko et al. (2016), Ghaemi et al. (2018) use mathematical optimization method to investigate the impacts of disruptions in timetable rescheduling, such as the turning of the trains and the assignment of passenger.

Other alternative approaches are machine learning and data mining models, including supervised learning and unsupervised learning approaches. Supervised learning model includes artificial neural networks (Haahr et al., 2019; Oneto et al., 2017, 2018; Yaghini et al., 2013), support vector machine (Barbour et al., 2018; Marković et al., 2015), deep neural networks (Huang et al., 2020b), and decision trees (Nabian et al., 2019; Nair et al., 2019) to predict train delays, and support vector machine to position the trains (Chen et al., 2015), and to improve the running time prediction ability during disruptions (Huang et al., 2020a). Unsupervised learning includes association analysis (Wallander and Mäkitalo, 2012) and clustering models (Cerreto et al., 2018), to uncover the knowledge and patterns embedded in the train operation data.

A summary of the data-driven methods for timetable rescheduling is shown in Table 1. Our review suggests that there are a handful number of train delay predictive models; however, few of them focus on predicting the effects of disruptions on train operation. To the best of our knowledge, most of the existing works treat disturbance and distribution effects separately. Disturbances are relatively small perturbations on trains operation, which can be handled by adjusting the timetable without altering the duties for rolling stock and crew. On the other hand, disruptions are relatively large incidents, necessitating adjustment of both the timetable and the duties for rolling stock and crew (Binder et al., 2017; Cacchiani et al., 2014). In this study, focusing on the main operational issues, we take into account both disturbances and disruptions effects and model the consequences of the spatial and temporal propagations of delays due to interruptions on train operations. Our biggest contribution is not just in proposing a model that integrates domain knowledge with BN theories to capture the stochastic features of train operations, but in predicting the temporal and spatial effects of propagation of delays, including the primary delay, the number of affected trains, and the total delay times of the affected trains due to both disruptions and disturbances.

Table 1. A summary of the data-driven methods for timetable rescheduling.

Work	Method	Objective
Briggs and Beck (2007)	Probability density	Delay distribution
Goverde et al. (2013)	Probability density	Delay distribution
Goverde et al. (2001)	Probability density	Delay distribution
(Harris, 2006)	Probability density	Delay distribution
Huisman and Boucherie (2001)	Probability density	Delay distribution
Yang et al. (2019)	Probability density	Delay distribution
Lessan et al. (2018)	Probability density	Delay distribution
Gorman (2009)	Regression	Delay estimation
Murali et al. (2010)	Regression	Delay estimation
Wen et al. (2017)	Regression	Delay estimation
Li et al. (2016)	Regression	Delay estimation
Kecman and Goverde (2015b)	Regression	Running time and dwelling time prediction
Meester and Muns (2007)	Phase-type distribution	Secondary delay distribution
Huang et al. (2019)	Probability density	The number of affected trains, and total delay times
Cerreto et al. (2016)	Statistics	Running time distribution
Van der Meer et al. (2009)	Statistics	Correlation analysis
Lessan et al. (2019)	Bayesian network	Delay prediction
Corman and Kecman (2018)	Bayesian network	Delay prediction
Şahin (2017)	Markov chain	Delay prediction
Barta et al. (2012)	Markov chain	Delay prediction
Gaurav and Srivastava (2018)	Markov chain	Delay prediction
Kecman et al. (2015)	Markov chain	Delay prediction
Berger et al. (2011)	Graph model	Delay prediction
Büker and Seybold (2012)	Graph model	Delay prediction
Hansen et al. (2010)	Graph model	Delay prediction
Goverde (2010)	Graph model	Delay prediction
Kecman and Goverde (2015a)	Graph model	Delay prediction
Milinković et al. (2013)	Petri Net	Delay prediction
Corman et al. (2014)	Graph model	Timetable robustness
Zilko et al. (2016)	Bayesian network	Disruption length

Ghaemi et al. (2018)	Bayesian network and mathematical optimization	Train turning, and passenger assignment
Yaghini et al. (2013)	Neural network	Delay prediction
Oneto et al. (2017)	Neural network	Delay prediction
Oneto et al. (2018)	Neural network	Delay prediction
Haahr et al. (2019)	Neural network	Delay prediction
Barbour et al. (2018)	Support vector machine	Delay prediction
Marković et al. (2015)	Support vector machine	Delay prediction
Nabian et al. (2019)	Decision tree	Delay prediction
Nair et al. (2019)	Decision tree	Delay prediction
Huang et al. (2020b)	Deep learning	Delay prediction
Huang et al (2020a)	Support vector machine and Kalman filter	Train running time prediction
Chen et al. (2015)	Support vector machine	Train locating
Wallander and Mäkitalo (2012)	Association analysis	Delay pattern recognition
Cerreto et al. (2018)	Clustering model	Delay pattern recognition
This work	***Bayesian network***	***Primary delay, the number of affected trains, and total delay times***

3. Method

3.1. Bayesian networks

A Bayesian network, $BN = (G, P)$, is a combination of a directed acyclic graph (DAG) and a set of conditional probabilities (Nielsen and Jensen, 2009). $G = (O, A)$ is a directed acyclic graph, where O represents nodes, corresponding with stochastic variables, and A represents arcs connecting the nodes. P is a conditional probability table for discrete variables and a conditional probability distribution function for continuous variables. Each node in the DAG represents a random variable, which can be directly observed, or hidden variables, while the directed edge from O_i to O_j represents the conditional dependence between random variables. O_i is called the parent node, and O_j is called the child node. Each element in the conditional probability table or conditional probability distribution function corresponds to the unique node in the DAG and stores the joint conditional probability of this node for all its direct precursor nodes.

Let us consider n random variables $x_1, x_2, \ldots, x_n$ in a BN. If we let x_i be a node in the graph, and non x_i be any set of nodes that are non-descendant of x_i, and pa x_i be a set of the immediate parents of x_i, then non x_i is conditionally independent of x_i, and the conditional relationship of these variables can be written as:

$$P\left(x_i \mid \mathrm{non}\, x_i, \mathrm{pa}\, x_n\right) = P\left(x_i \mid \mathrm{pa}\, x_i\right) \tag{1}$$

In BNs, the nodes without parents are parameterized as *prior* probability, whereas other nodes are parameterized by P(node | pa node). "pa node" represents the parent node in the network. The joint conditional probability distribution of a combination of arbitrary nodes in the network can be written as:

$$P\left(x_1, x_2, \ldots, x_n\right) = \prod_{i=1}^{n} P\left(x_i \mid \mathrm{pa}\, x_i\right) \tag{2}$$

Constructing a BN to make a prediction requires three steps, namely, variables selecting, structure learning, and parameter tuning. Variables are usually selected according to domain expert knowledge, whereas both structure and parameter can be learned by domain expert knowledge or learned from data (Nielsen and Jensen, 2009). Hybrid BNs can also be obtained by integrating domain knowledge and knowledge from observed data (Lessan et al., 2019). For example, the causality of some variables is not clear and data should not distinguish between network structures that represent the same assertions of conditional independence (Heckerman et al., 1995).

3.2. Variable dependency investigation

The first step to construct a BN model is to determine the causal relationship of the system's variables. Establishing a BN network to predict L, N, and T on a railway line requires the following: 1) the causal relationships of L, N, and T at the same station, and 2) the causal relationships of L, N, and T between adjacent stations.

For the causal relationships at the same station the primary delays will directly affect the number of delayed trains and total delay times at this station. Therefore, at a station m, we have $P(N_m|L_m)$ and $P(T_m|L_m)$, which means L_m is the parent node of both N_m and T_m. In addition, we assume that $P(T_m|N_m)$, given that the total delay time is $T_m = t_1 + t_2 + \ldots + t_R$, which means that higher the number of involved trains, larger the total delay times will be. Therefore, the causal relationship of the variables at the same station, determined according to knowledge, has three arcs as seen in Fig. 2(a).

When we determine the model structure between stations, due to the requirement of dispatching operations, not only the causal relationship of variables should be considered, but also the temporal order of the variables. This means that we cannot use variables that are known later in time to be the parent nodes of variables that are known earlier. If such arcs are included in the BN model, even though lower prediction errors are achieved, the BN model is unpractical for real-time prediction. Therefore, the variables obtained at the latter station cannot be the parent nodes of the variable obtained at the former station, which means, $P(L_{m-1}|L_m)$, $P(N_{m-1}|L_m)$, $P(T_{m-1}|L_m)$, $P(L_{m-1}|N_m)$, $P(N_{m-1}|N_m)$, $P(T_{m-1}|N_m)$, $P(L_{m-1}|T_m)$, $P(N_{m-1}|T_m)$, and $P(T_{m-1}|T_m)$, should all be excluded from the BN. In addition, the nodes N_{m-1} and T_{m-1}, although obtained from the former station, are likely to be known later than L_m, if the effects of disruptions are severe, resulting in large N_{m-1} and T_{m-1}. This situation can be seen in Fig. 1, where, for example, when the time span between the last and first trains in the delayed train group at station SG is longer than the running times of trains from SG to LCE, L_{LCE} is known earlier than N_{SG} and T_{SG}. Therefore, we exclude the arcs from N_{m-1} and T_{m-1} to L_m in the BN model, namely $P(L_m|N_{m-1})$, and $P(L_m|T_{m-1})$.

In this study, we build the arcs of the BN model between two adjacent stations according to domain knowledge that states the delay of a train at one station is directly influenced by its delay at the last station (a Markov property). This is one of the main premises assumed in existing studies using graph- or network-based techniques to model train operations (Büker and Seybold, 2012; Berger et al., 2011; Corman and Kecman, 2018; Goverde, 2010; Hansen et al., 2010; Şahin, 2017). With the domain knowledge, the causal relationships of variables between adjacent stations are $P(L_m|L_{m-1})$, $P(N_m|N_{m-1})$, and $P(T_m|T_{m-1})$, which means that each variable is only directly affected by itself at the last station. Therefore, the BN structure shown in Fig. 2(b) can be obtained according to expert knowledge.

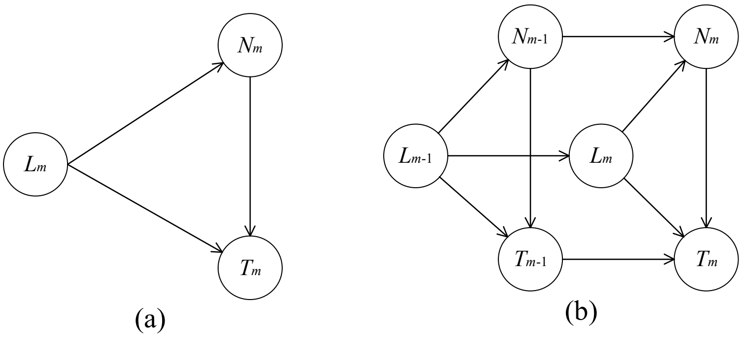


Fig. 2. The BN structure determined according to knowledge: (a) BN structure at one station, (b) BN arcs connecting adjacent stations.

4. Predictive model

4.1. Data description

We calibrate our model using data from two HSR lines in China, namely, the Wuhan-Guangzhou (W-G) HSR and Xiamen-Shenzhen (X-S) HSR. The W-G HSR line has a length of 1,096 km, which is one of the busiest passenger railway lines in China. Trains operating on this line are all equipped with the Chinese train control system class-3 (CTCS-3), which allows a maximum speed of 350 km/h. We consider the movement of trains in the northbound direction on this line, from QY station to HYE station, which encompass 7 stations and 6 sections in total, as shown in Fig 3 (a). The X-S HSR line has a length of 543 km. Trains operating on this line are all equipped with the Chinese train control system class-2 (CTCS-2), which allows a maximum speed of 250 km/h. We consider the movement of trains in the northeast bound direction on this line, from HZS to PN, which includes 7 stations and 6 sections totally, as shown in Fig 3 (b).

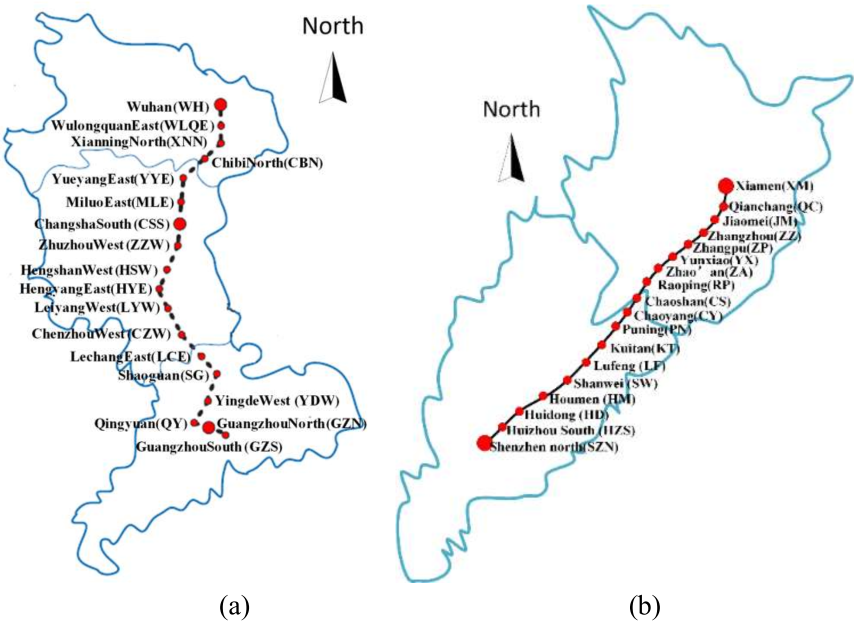


Fig. 3. Map of W-G HSR line (a) and X-S HSR line (b).

The operational data collected from the QY-HYE section of the W-G HSR line includes a total of 57,796 train services, and those collected from HZS-PN on X-S HSR includes a total of 41,186 train services. The data covers operational records from March 24, 2015 to November 10, 2016, comprising scheduled/actual arrival/departure records for each train at each station, train numbers, dates, and information on occupied tracks. All the data are recorded with a time format in minutes owing to the accuracy of the system, as listed in Table 2.

Table 2. Train operation data format in database.

Train	Station	Date	Actual arrival	Actual departure	Scheduled arrival	Scheduled departure	Occupied track*
G280	GZN	2015/3/24	7:02	7:02	7:01	7:01	II
G548	SG	2015/8/27	17:26	17:30	17:26	17:29	4
G1118	HYE	2015/10/2	13:37	13:40	13:38	13:40	8
G6025	HSW	2015/11/28	20:57	20:57	20:58	20:58	I

*The passage tracks at the station are labeled with Roman characters, while the dwelling tracks are labeled with numbers.

Using these operational records, we obtain data about the three variables (*L, N,* and *T*) at each station. Because the frequency of disruptions is very low and longer disruptions are less frequent, we exclude the disruptions whose *L*, *N*, and *T* are too large, to ensure the statistical significance of the extracted data. This data pre-processing or filtering out longer disruptions is a common technique in other BN-based delay prediction studies. For example, both Corman and Kecman (2018) and Lessan et al. (2019) consider the delay horizon of 60 min. Because the frequency of long disruptions is very low and excessively long disruptions are less frequent to happen, we excluded the disruptions whose L, N, and T are too large. By seeing that observations with L in 60 min horizon accounts for over 97% of all observations in the dataset, we deleted the instances with L, N and T larger than 60 min, 20 trains, and 300 min, respectively, which is only at most 3% of our dataset that does not have statistical significance effect on the observed behavior in the extracted data compared to the original data. Table 3 shows three cases included in the data. In addition, the distributional characteristics of each variable are shown in Fig. 4 and Fig. 5, which show that these three variables on the two HSR lines, similar to the train delays distribution in railway systems, all follow long-tail distributions, where large values have low frequencies. Finally, to avoid over-fitting, the data are randomly split to form the training dataset, which contains 70% of the disruptions for model training, and the testing dataset, which contains 30% of the disruptions for model testing and evaluation.

Table 3. Examples of data for modeling for W-G HSR.

Variable	Number	QY	YDW	SG	LCE	CZW	LYW	HYE
	1	9	13	12	11	43	44	43
L(min)	2	50	49	48	48	49	48	48
	3	13	14	14	14	15	15	14
	1	2	1	2	2	13	14	7
N	2	2	2	2	2	2	2	1
	3	8	7	7	8	8	8	8

	1	13	13	17	16	343	345	252
T(min)	2	88	86	85	79	81	81	48
	3	162	164	168	181	194	196	192

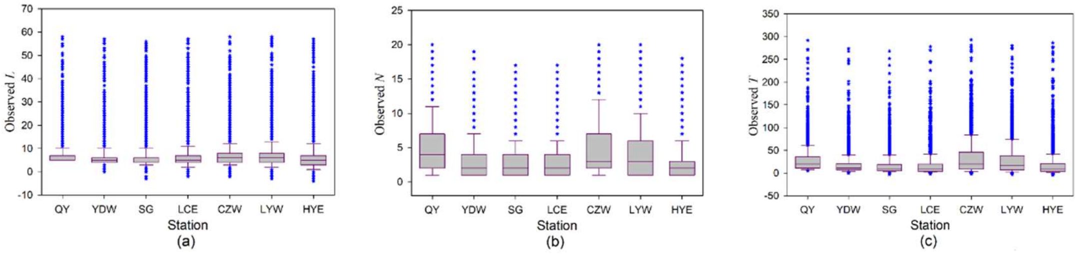


Fig. 4. Distribution characteristics of the variables for the W-G HSR line.

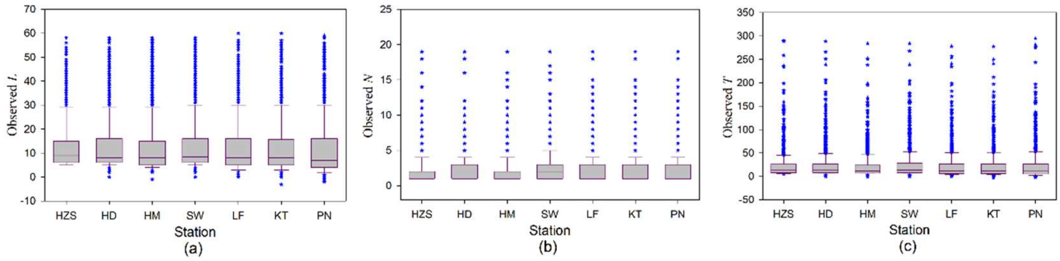


Fig. 5. Distribution characteristics of the variables for the X-S HSR line.

4.2. Experiments for structure determination

The BN structure is either based on domain (expert) knowledge or by learning from data. If the variables' relationships are complicated or hard to discover, the BN structure needs to utilize both of these methods to achieve robust results (Heckerman et al., 1995). As introduced in Section 3.2, in this study, the causal relationships of L, N, and T have been partially determined based on expert knowledge. Therefore, in this section, we perform experiments to investigate the BN structure from data. To test the causal relationships of variables at the same station, we use data from each station of the W-G HSR line to learn the model structure using the max-min hill-climbing (MMHC) algorithm (Tsamardinos et al., 2006). MMHC is an efficient and powerful method for structure learning of BN based on observed data. Beginning with an empty graph, the algorithm first identifies the parents and children set of each variable. Then, it performs a greedy hill-climbing search with some operations such as edge addition, deletion, or direction reversal to achieve the largest performance score (Heckerman et al., 1995). Using our data, the structures learned from data at each station are shown in Fig. 6. This figure indicates that six out of seven structures learned from data are similar to the structure determined by domain knowledge shown in Fig. 2 (a). Only at QY station is $P(N_{QY}|L_{QY})$ missing, which indicates that the structures at station can be determined as shown in Fig. 2(a).

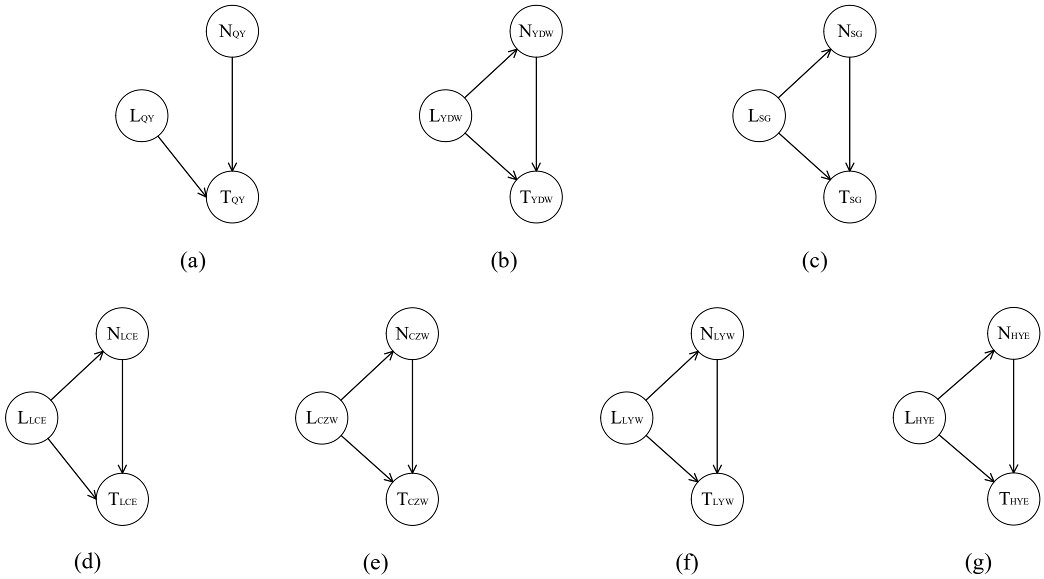


Fig. 6. BN structures learned from data of each station of the W-G HSR.

To construct a BN model of the whole railway line, we conduct some experiments to investigate the possible arcs between stations under the given relationships introduced in Fig. 2, which means that we manipulate the arcs obtained by knowledge in Fig. 2, and exclude the arcs that should not be included in the BN introduced in Section 3.2. After including the arcs from domain knowledge and excluding the arcs that are contrary to the time order, we use the MMHC algorithm to obtain the structure from data from each section on the W-G HSR line. Figure 7 indicates that the structures are different from section to section, which means there is no structure that can be a perfect fit for all sections of a railway line. Given the structures in Fig. 7, we then can classify the arcs between stations into four categories:

Category 1: This category is similar to the structure obtained by domain knowledge, as shown in Fig. 2(b). See Fig. 7 (e) as an example.

Category 2: In this category, L at the former station influences N and/or T at the latter station, i.e., $P(N_m|L_{m-1})$ and/or $P(T_m|L_{m-1})$. See Fig. 7 (a) and (f) as examples.

Category 3: In this category, N at the former station influences T at the latter station and/or T at the former station influences N at the latter station, i.e., $P(T_m|\ N_{m-1})$, and/or $P(N_m|\ T_{m-1})$, respectively. See Fig. 7 (c) and (d) as examples.

Category 4: In this category, both L at the former station influences N and/or T at latter station and N at former station influences T at the latter station and/or T at the former station influences N at the latter station. See Fig. 7 (b) as an example.

Therefore, based on the classified categories, we propose four different structures as candidates for the variable relationships between stations, each corresponding to a category, which are:

Structure A: this structure (Fig. 2(b)) is originally obtained by exerting domain knowledge, which is similar to the Category 1.

Structure B: this structure (Fig. 8(a)) includes the arcs in Fig. 2(b) and category 2.

Structure C: this structure (Fig. 8(b)) includes the arcs in Fig. 2(b) and category 3.

Structure D: this structure (Fig. 8(c)) includes the arcs in Fig. 2(b) and category 4.

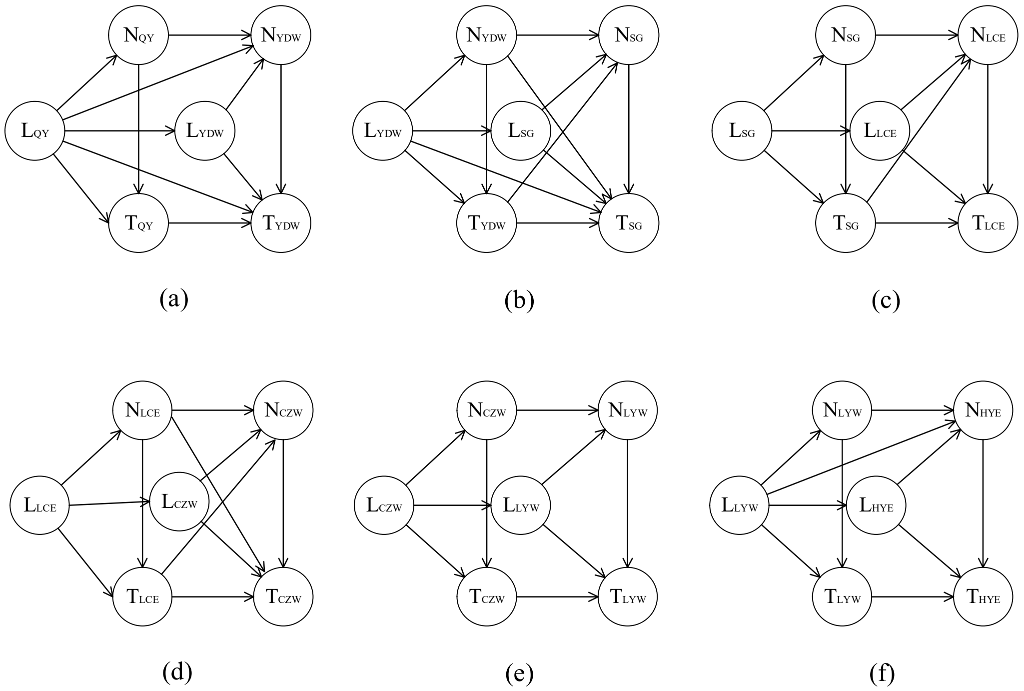


Fig. 7. BN structures for adjacent stations learned from data on the W-G HSR line.

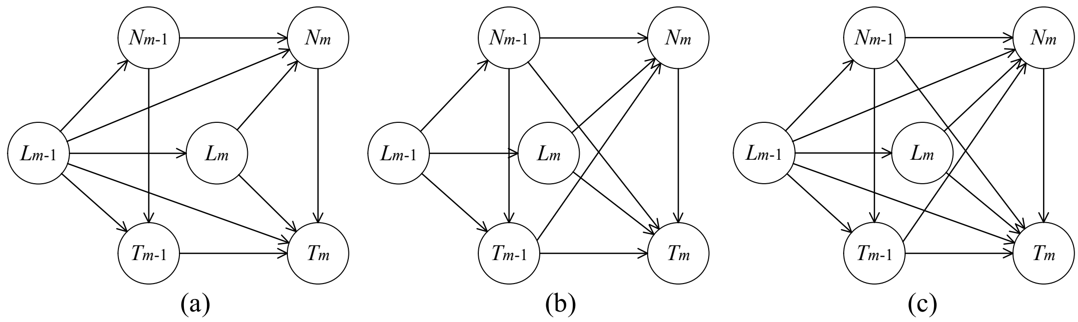


Fig. 8. Candidate BN structures for the HSR lines.

4.3. Parameter tuning

Before a BN structure can be used for prediction, it needs careful parameter tuning to obtain a model that can satisfactorily capture the properties of the incorporated variables/nodes (Nielsen and Jensen, 2009). Parameter tuning involves the estimation of the probability distribution of parameters according to a known dataset. In this study, the nodes L, N, and T are considered as continuous variables, and a Gaussian distribution is set as the prior distribution of each node. Then, the maximum likelihood method is used to estimate the parameter probability distribution, which would then be used to estimate the posterior probability distribution of the nodes.

Given an acyclic graph, G, over n variables, we know from Section 3.1 that the joint distribution of the n variables is given by Eq. (2). Therefore, the parameters we need to learn from the training data are the conditional distribution $P(x_i \mid \mathrm{pa}\, x_i)$. To simplify variable expression, we use $P(x_i \mid \mathrm{pa}\, x_i) = \theta_{x_i|\mathrm{pa}\, x_i}$ to specify the parameters. Given the training data with d complete observations, $D = \left\{\left(x_1^t, x_2^t, ..., x_n^t\right), t = 1, 2, ..., d\right\}$, the loss function (log-likelihood) of the maximum likelihood method can be written as (Neapolitan, 2004):

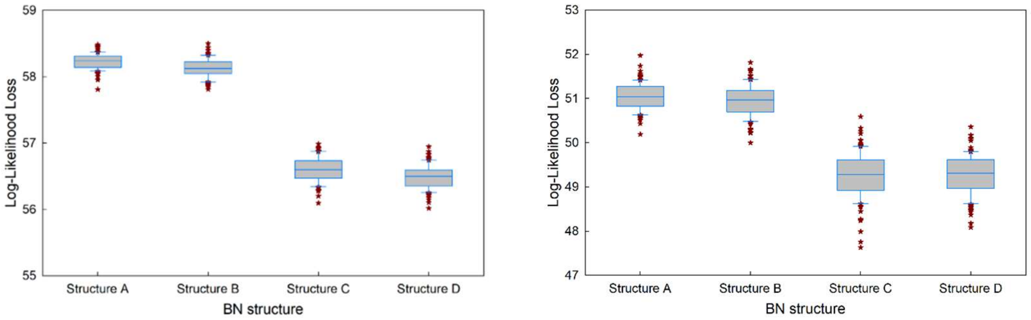
$$
\begin{aligned}
l(D,\theta,\mathrm{G}) &= \log P(D \mid \theta) \\
&= \sum_{t=1}^{d} \log P\left(x_1^t,\dots,x_n^t\right) \\
&= \sum_{t=1}^{d} \sum_{i=1}^{d} \log \theta_{x_i^t \mid \mathrm{pa}\, x_i^t} \\
&= \sum_{t=1}^{d} \sum_{x_i,\mathrm{pa}\, x_i} d\left(x_i, \mathrm{pa}\, x_i\right) \log \theta_{x_i^t \mid \mathrm{pa}\, x_i^t}
\end{aligned}, \tag{3}
$$

where d is the number of observed instances in the training dataset and (x_i, pa x_i) is a particular setting of the variable and its parents. According to the maximum likelihood method, the optimal estimated parameter $\hat{\theta}_{x_i \mid \mathrm{pa}\, x_i}$ that maximizes the loss function has a simple closed form expression (Neapolitan, 2004):

$$
\hat{\theta}_{x_i \mid \mathrm{pa}\, x_i} = \frac{d\left(x_i, \mathrm{pa}\, x_i\right)}{\sum_{x_i'=1}^{r_i} d\left(x_i', \mathrm{pa}\, x_i\right)}. \tag{4}
$$

The simplicity of the parameter tuning using the maximum likelihood method is that $\theta_{x_i \mid \mathrm{pa}\, x_i}$ can be chosen freely for each set of parents, pa x_i, when the BN structure is known and all the variables are fully observable (Neapolitan, 2004). We first train the candidate BN structures defined in Section 4.2 using data from the W-G HSR line based on 5-fold cross-validation. The log-likelihood loss is chosen as the performance function of the BN network. We perform 100 runs of predictions for each structure, and the loss function distribution of each candidate BN structure is shown in Fig. 9 (a). This figure shows that the losses of Structure A and B are evidently higher than those of Structure C and D, which indicates that the effects of $P(T_m| N_{m-1})$ and $P(N_m| T_{m-1})$ on the goodness-of-fit of model are stronger than $P(N_m|L_{m-1})$ and $P(T_m|L_{m-1})$ in the model. Structure D has the smallest errors, but it is not significant compared to Structure C. We thus compared their efficiency in the training process, where Structure C cost 1.89 s for 100 runs, and Structure D cost 2.05 s for 100 runs (on an Intel core i5 8250U processor). As the extra time cost of Structure D is extremely low compared to Structure C, we thus choose Structure D to be our predictive model, which has the smallest loss function among the proposed structures. In addition, because the outperforming BN structure is selected using data from the W-G HSR line, here, to test the generalizability of the proposed BN model, we train the candidate models using data from another HSR line, namely the X-S HSR line. Figure 9 (b) shows that Structure D is also the best structure for the X-S HSR, because its loss distribution is more centralized and its average loss for 100 runs is slightly lower; 49.296 compared to 49.257 for Structures C and D, respectively. The training of Structure C cost 1.16 s, while that of Structure D cost 1.22 s for 100 runs.

Figure 10 depicts the BN structure we choose for predicting the effects of disruption on the HSRs. In the predictive model, L at station m is only directly influenced by itself at station m-1, namely $P(L_m| L_{m-1})$. N at station m is directly influenced by four variables, including L at station m, and L, N, and T at station m-1, namely, $P(N_m| L_m, L_{m-1}, N_{m-1}, T_{m-1})$. T at station m is directly influenced by five variables, including L and N at station m, and L, N, and T at station m-1, namely, $P(T_m| L_m, N_m, L_{m-1}, N_{m-1}, T_{m-1})$.



(a) (b)

Fig. 9. Values of loss function of the candidate structures trained using data from W-G (a) and X-S (b) HSRs.

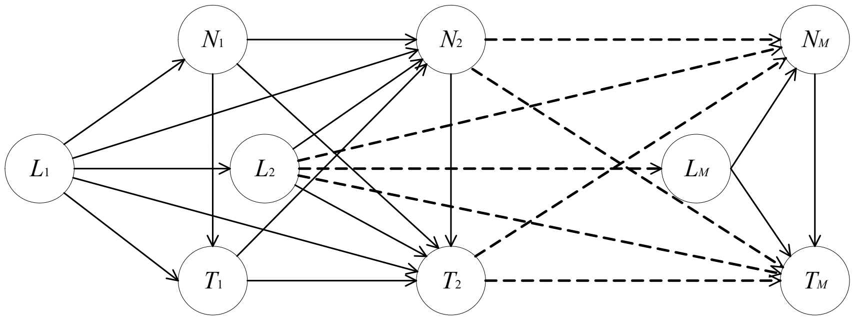


Fig. 10. The causal relationship of the variables *L*, *N*, and *T* for HSR disruptions.

4.4. Train overtaking

Train overtaking makes two trains' states highly dependent, as the trains are competing to use limited resources (e.g., block sections, and stations). For example, if two trains both are delayed, the overtaking can increase (decrease) the delay of one train while that of the other can decrease (increase). On the other hand, if one train is delayed and the other train is on schedule, the delay may propagate from the delayed one to the other, which can lead to the change in the number of delayed trains and total delay times. Train overtaking is thus another potential factor that can affect the values of variable *L*, *N*, and *T*.

In this study, the BN model is built on a Markov property assumption, which is a common assumption in railway train operation research (Büker and Seybold, 2012; Berger et al., 2011; Corman and Kecman, 2018; Goverde, 2010; Hansen et al., 2010; Şahin, 2017). According to the Markov property, we can assume a train's state at a station depends on its state at the most recent station. This means that train overtaking only influences the values of *L*, *N*, and *T* at the station where overtaking happens, which then can indirectly affect the train states at the following stations through delay propagation. In other words, even though train overtaking can change the values of *L*, *N*, and *T* at the following stations, these influences are indirect consequences of delay propagation. Therefore, we can add complementary variables pertinent to the overtaking events at the overtaking stations to include the effects of overtaking operations in BN model.

To include the effect of overtaking, we consider two different situations that may have an impact on *L*, *N*, and *T*. The first one is whether the train with the primary delay overtakes other trains, or other trains overtake it. The other situation is the number of overtaking events in the affected train groups (including the one due to primary delay). As aforementioned, the BN model in this study aims to meet real-time prediction requirements. Therefore, we consider overtaking in the scheduled timetables, for which we define the following factors:

- *O*, if the train with the primary delay overtakes other trains or if other trains overtake it in the scheduled timetable. $O=\{O_1,O_2,...,O_M\}$, where O_m is a binary variable denoted by 1 if the primary delay train has an overtaking event with other trains at station *m*, and 0 otherwise.
- *V*, the number of overtaking events in the affected train group at each station. $V=\{V_1,V_2,...,V_M\}$, where V_m is a discrete variable denoted by the number of overtaking of the delayed trains at station *m*.

To consider the influences of these two factors, we add them as complementary variables to the nominated BN structure as shown in Fig. 10. According to the definitions of variables *O* and *V*, variable *O* directly influences the value of *L*, whereas variable *V* influences the value of *N* and *T*. This means that variable *O* is the parent node of *L* at each station, i.e., $P(L_m|O_m)$, and *V* is the parent node of both *N* and *T* at each station, i.e., $P(N_m|V_m)$ and $P(T_m|V_m)$, as shown in Fig. 11.

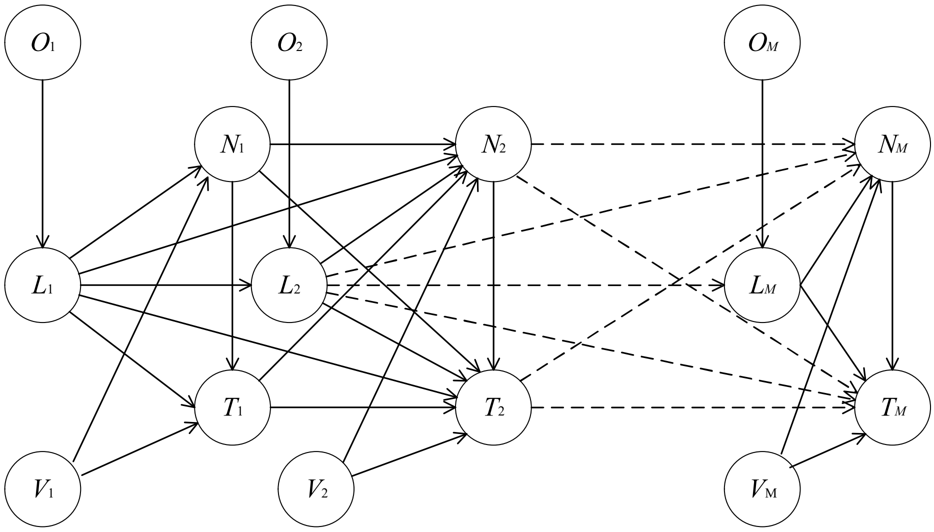
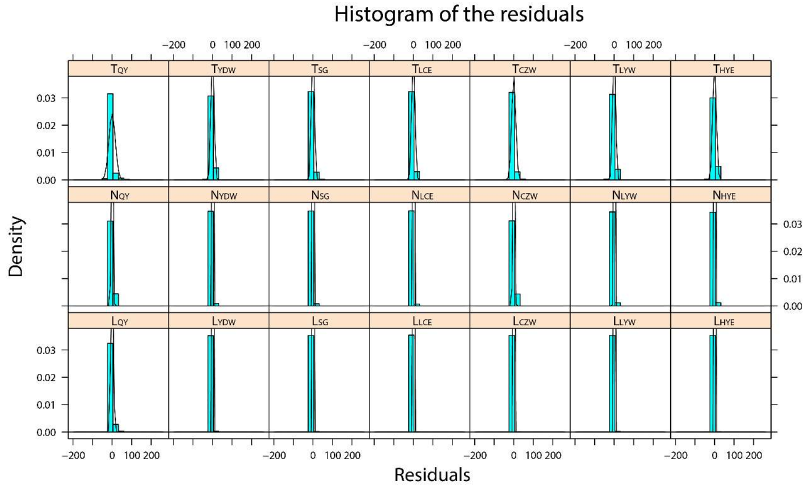


Fig. 11. The causal relationship of variable *L*, *N*, and *T* for HSR disruptions considering train overtaking.

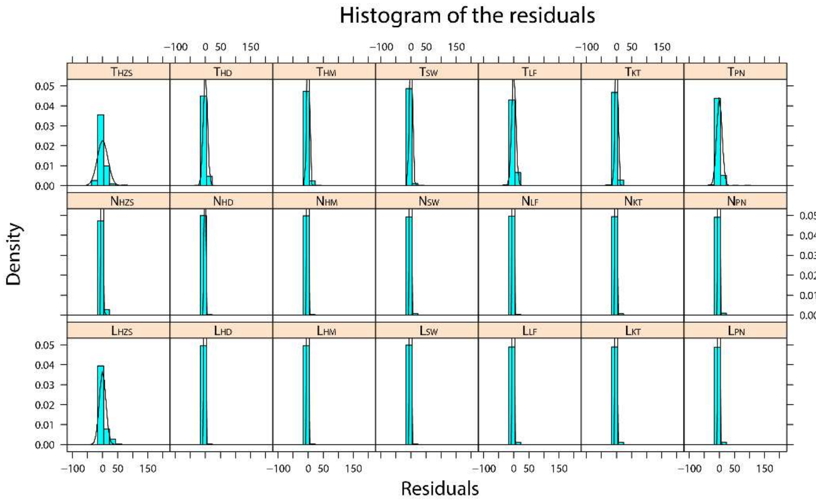
5. Results and discussion

5.1. Predictive result analysis

First, to evaluate the performance of the model in predicting disruption effects, we examine the goodness-of-fit of the nominated BN model by investigating the residuals of each node using training data from the W-G and X-S HSR lines, respectively. Figures 12 (a) and (b) show that the residuals of predictions at each node. In these figures, each sub area represents the residual distribution of a variable at a station; the horizontal axis represents residuals and the vertical axis represents the probability density of residuals; which variable and which station the residuals are from is indicated on the top of each sub plot; the unit of variable *L* and *T* is min, and the variable *N* has no unit. We can see in almost all of the sub plots that the residuals are around zero; however, as QY and HZS are the first stations in our dataset of the W-G and X-S HSR lines, the fitting residuals at these two stations are larger than others which is expected as they have no parent nodes. Overall, we can see that the proposed BN model can achieve a satisfactory goodness-of-fit.



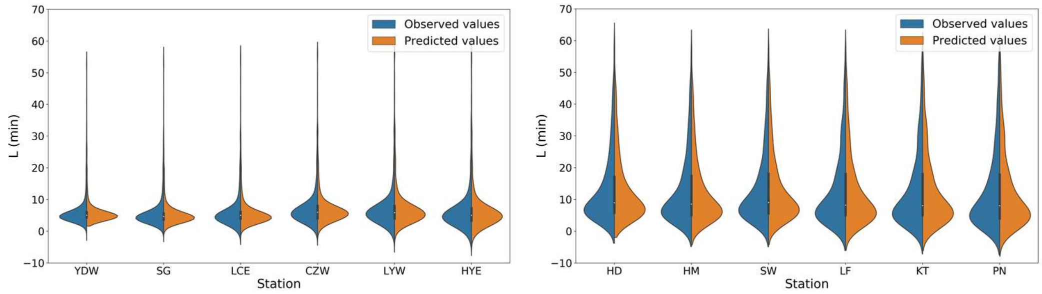
(a)



(b)

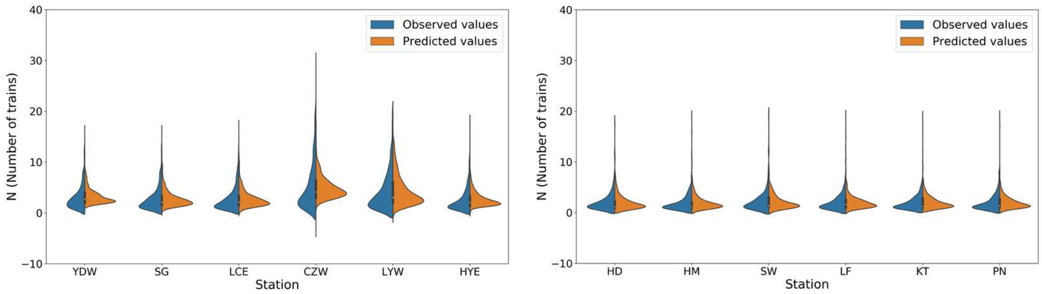
Fig. 12. Fitted residuals of each node of the BN model of (a) W-G and (b) X-S HSRs

Next, we investigate the prediction results of the BN model for the testing dataset. The comparisons of the predicted values against the observed values of the variables *L, N*, and *T* on the W-G and X-S HSRs are shown in Fig. 13 to Fig. 15, where the left part of the violin plot indicates the observed values, and the right part of the violin plot represents the predicted values. Because the stations QY and HZS have no parent stations in our dataset, we just analyze the predictive results from YDW to HYE stations on the W-G HSR and from HD to PN stations on the X-S HSR. These plots show that the predicted *L*, *N*, and *T* at each station of both lines have a very close distribution with the observed values, suggesting reliability of predictions from the proposed structure. We also investigate the predicted residuals of each node of the BN model. Figure 16, which has a similar layout to that of Fig. 12, shows that the predicted residuals of each of the variables of two HSR lines are also centralized around zero. This validates normality and zero-mean of the predicted residuals of regression models.



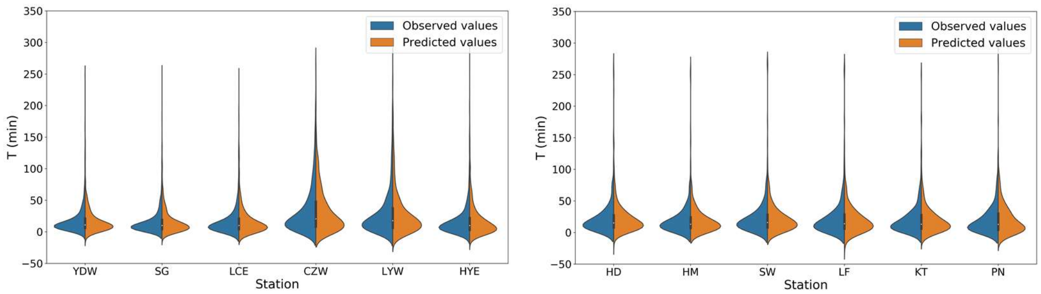
(a) (b)

Fig. 13. Observed *L* VS predicted *L* of W-G (a) and X-S (b) HSRs.



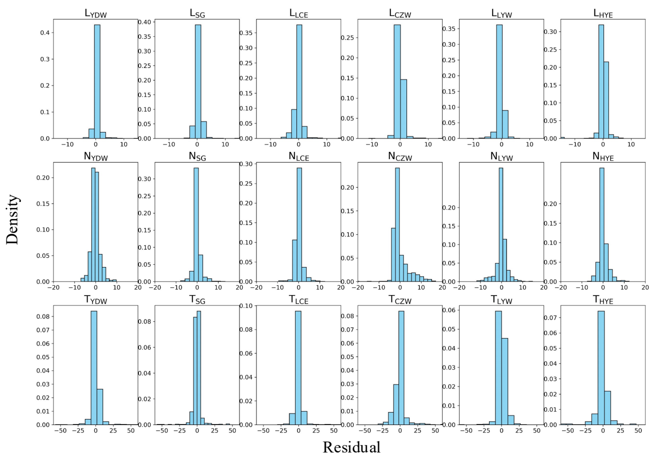
(a) (b)

Fig. 14. Observed *N* vs. predicted *N* of W-G (a) and X-S (b) HSRs.

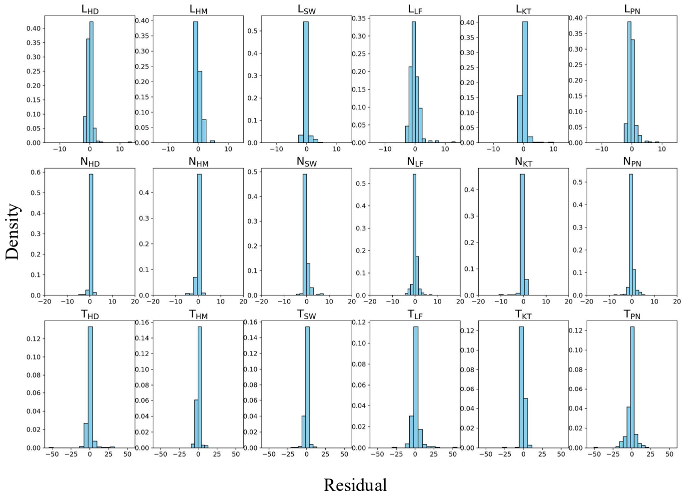


(a) (b)

Fig. 15. Observed *T* vs. predicted *T* of W-G (a) and X-S (b) HSRs.



(a)



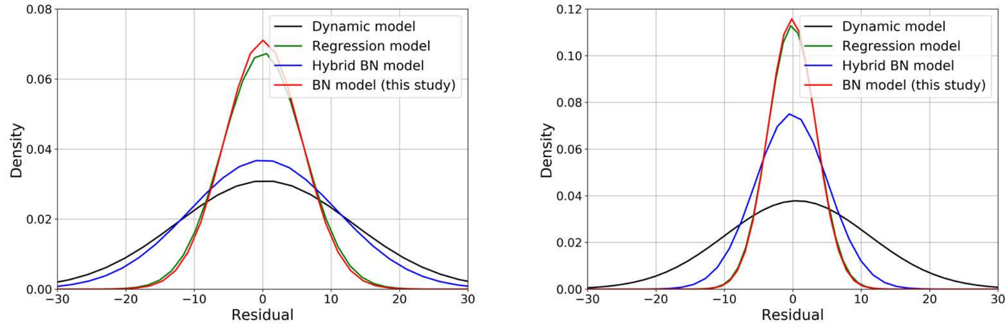
(b)

Fig. 16. Predicted residuals distribution of each variable of the W-G (a) and X-S (b) HSRs.

5.2. Comparison against different benchmarks

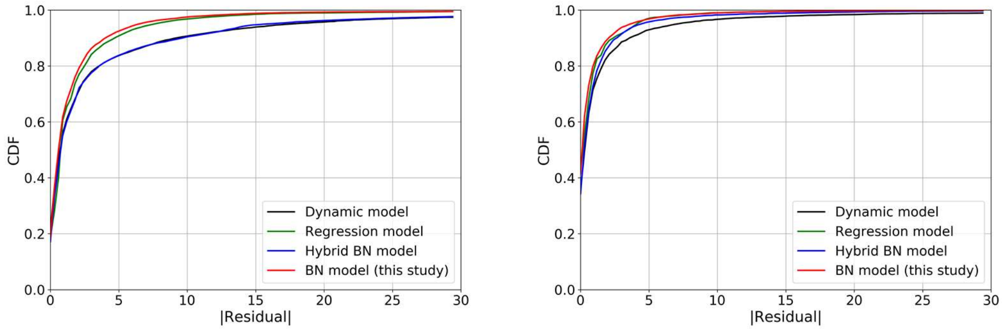
To check the performance of the proposed BN model against other predictive models, we choose three common and recently proposed models for real-time train dispatching as benchmark models. The first one is the non-stationary Markov chain model (Kecman et al., 2015), called the “dynamic model” in this study. In the dynamic model, the train delay is treated as a dynamic process, where it is updated over the time and space simultaneously, and using a dynamical transition probability matrix, this model tackles the dynamic process of train delays. This model can be seen as a representative of many other models that are based on the assumption of Markov property, such as Refs. (Barta et al., 2012; Gaurav and Srivastava, 2018; Goverde, 2010; Şahin, 2017). The second benchmark model is a BN-based model proposed by Lessan et al. (2019), called the “hybrid BN model”, where the authors construct a BN model considering both domain knowledge and the structure learned by the heuristic algorithm. In the hybrid model, the train events (arrival and departure) are treated as nodes in the BN model, and each node takes the latest two train events as its parent nodes. The third model is based on a regression technique, so-called the “regression model” in this study, which is widely used in train operation modeling to predict train delays (Murali et al., 2010), train dwelling times (Li et al., 2016), and the probability of the number of delayed trains (Wen et al., 2017). We predict the variables L, N, and T at each station by taking their parents nodes as the input of the regression model.

We first investigate the residual distribution of each model for all the variables. The predictive residual distributions of L, N, and T of each model are showed in Fig. 17. In this figure, the higher probability of the residuals around 0 means the model has lower prediction errors. In other words, the sharper the probability density curve, the better the model. Figure 17 shows that the residuals of the proposed BN model have a shaper distribution for variable L, N, and T on both the W-G and X-S HSR lines. In addition, the cumulative distribution function (CDF) of the residuals is shown in Fig. 18, which also shows that the CDF of the proposed BN model surrounds the CDF curves of other models, suggesting a better predictive performance.



(a) (b)

Fig. 17. The residual distribution of L, N, and T for the (a) W-G and (b) X-S HSR lines.



(a) (b)

Fig. 18. The CDF of L, N, and T for the (a) W-G HSR and (b) X-S HSR lines.

To evaluate the accuracy of the predictions from the proposed model, we use two metrics, namely, the mean absolute error (MAE) and root mean square error (RMSE), as shown below:

$$\text{MAE} = \frac{1}{K}\sum_{i=1}^{K}\left|y_i - \hat{y}_i\right| \tag{5}$$

$$\text{RMSE} = \sqrt{\frac{1}{K}\sum_{i=1}^{K}\left(y_i - \hat{y}_i\right)^2}, \tag{6}$$

where K is the testing sample size, $\hat{y}_i$ is the predicted value, and y_i is the observed value.

With these two metrics on hand, we first investigate the predictive errors of the proposed BN model on each variable at all stations of these two HSR lines against those of the benchmark models. The MAE and RMSE of the predictions of each model are listed in Table 4. Table 4 shows that the predictive performance of the proposed BN model on variable L is slightly better than other models, however, the proposed BN model significantly outperforms other models on N and T, which are two critical factors used to measure the influence of disruptions on train operations. The reason for not seeing outstanding performance compared to that of others is that the primary delay (variable L) is just influenced by overtaking of trains and its state at the most recent station, indicating that all the models have limited input and that all models treat the prediction of this variable in a similar way, which makes the performance of each model quite similar. The proposed BN outperforms the hybrid BN model and the dynamic model on variables N and T, because the hybrid BN model and the dynamic model only use the train operation information from the previous stations, without considering the influence of the factors at the prediction station. The proposed BN model outperforms the regression model on the variables *N* and *T*, because the regression model only captures the influence of the independent factors on the dependent factor, without considering the dependency between the independent factors. However, the conditional probability in the BN model enables the BN model to capture the dependency between any pair of nodes, which is precisely why the BN model is powerful in uncertainty modeling (Nielsen and Jensen, 2009).

Table 4. Predictive errors of *L*, *N*, and *T* of each model on the testing samples.

Line	Model	MAE			RMSE		
		L (min)	*N*	*T* (min)	*L* (min)	*N*	*T* (min)
W-G HSR	BN (this study)	**0.834**	**1.412**	**3.789**	**1.610**	**2.230**	**9.031**
	Hybrid BN model	0.838	1.495	9.509	1.648	2.379	18.514
	Dynamic model	0.916	1.447	10.406	2.057	2.421	22.109
	Regression model	0.977	1.627	4.255	1.766	2.532	9.757
X-S HSR	BN (this study)	0.724	**0.482**	**1.998**	**1.126**	**0.957**	**5.055**
	Hybrid BN	**0.722**	0.521	3.218	1.175	0.944	9.070
	Dynamic model	0.836	0.510	5.527	1.452	1.068	18.199
	Regression model	0.802	0.503	2.059	1.243	0.991	5.604

The highlighted numbers represent the best results.

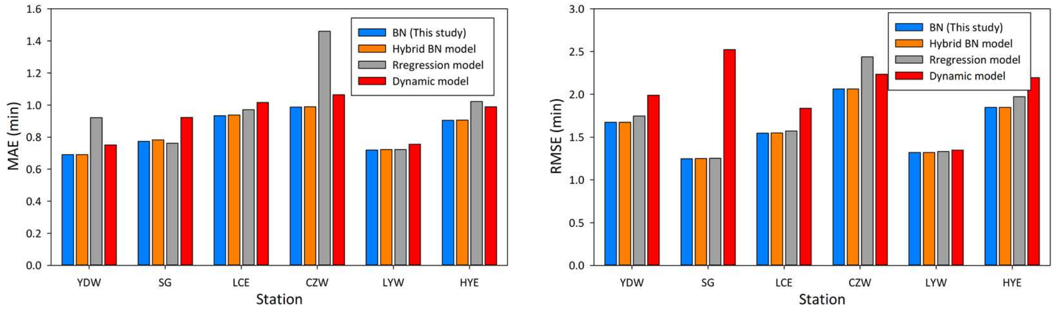
In addition, to analyze the performance results of the proposed BN model on these disruptions, which have severe effects on train operations, we also investigate the predictive results of all the models on disruptions whose N is more than 5, and T is larger than 100 min. The MAE and RMSE of the predictions of each model on disruptions with severe influences are listed in Table 5. Table 5 shows that the proposed BN model also has lower predictive errors on the disruptions with severe influences on both the W-G and X-S HSR lines.

Table. 5 Predictive errors of L, N, and T of each model on samples with severe influence.

Line	Model	MAE			RMSE		
		L (min)	N	T (min)	L (min)	N	T (min)
W-G HSR	BN (this study)	**2.150**	**4.735**	**23.169**	**5.173**	6.255	**38.144**
	Hybrid BN model	2.171	5.047	54.900	5.181	**6.213**	68.571
	Dynamic model	2.470	5.664	71.100	5.803	6.915	88.218
	Regression model	2.436	5.414	25.538	5.416	6.845	39.921
X-S HSR	BN (this study)	0.786	**1.337**	**10.280**	0.932	**2.012**	**16.759**
	Hybrid BN	0.775	1.545	13.026	0.922	2.230	25.665
	Dynamic model	0.974	3.207	110.100	1.179	4.401	138.365
	Regression model	**0.717**	1.420	11.104	**0.898**	1.994	16.870

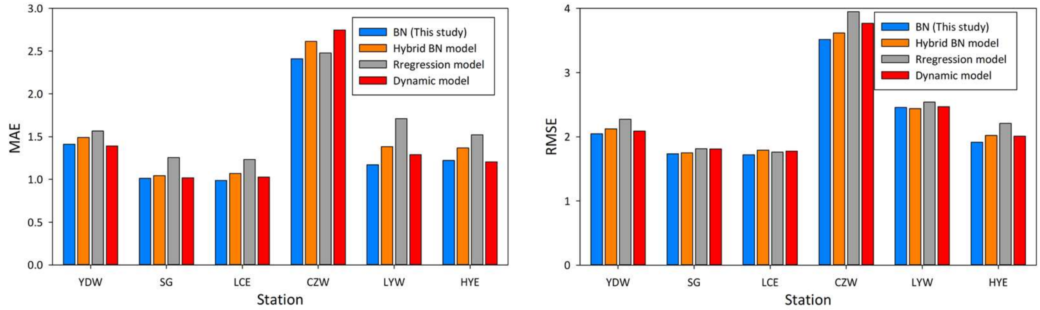
The bold fronts represent the best results.

In the next step, we compare the performance of each model for *L*, *N*, and *T* at each station. The comparative results of each model in term of MAE and RMSE for *L*, *N*, and *T* from the YDW to HYE station on the W-G line and from the HD to PN station on the X-S HSR are shown in Fig. 19 to Fig. 24. These figures show that the performance of the proposed BN model on variable *L* is not lower than the other benchmark models', however, the proposed BN model significantly outperforms other models on N and T at each station on both HSR lines.



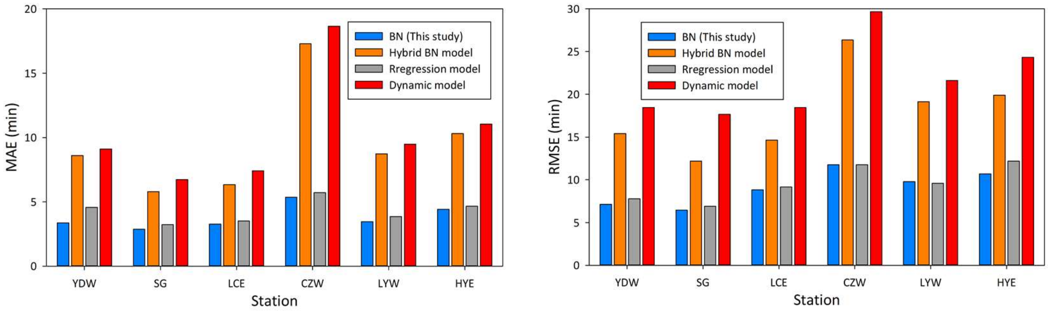
(a) (b)

Fig. 19. Predictive MAE (a) and RMSE (b) of L on the W-G HSR line.



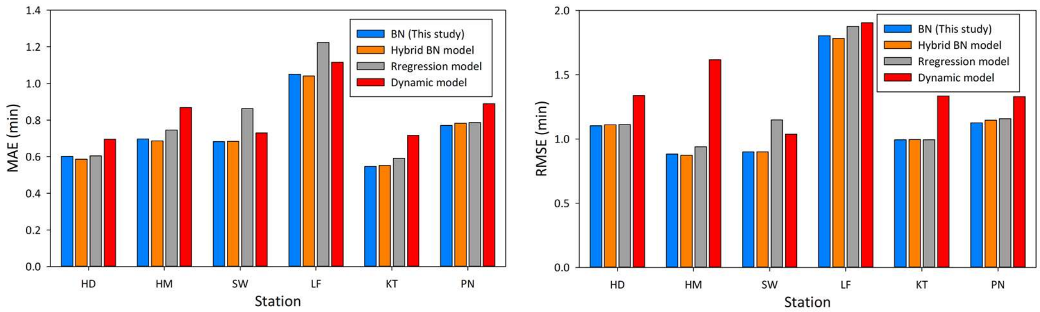
(a) (b)

Fig. 20. Predictive MAE (a) and RMSE (b) of N on the W-G HSR line.



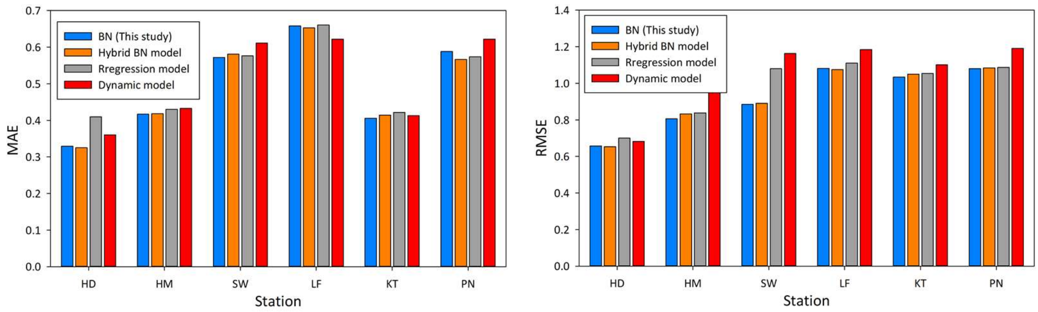
(a) (b)

Fig. 21. Predictive MAE (a) and RMSE (b) of T on the W-G HSR line.



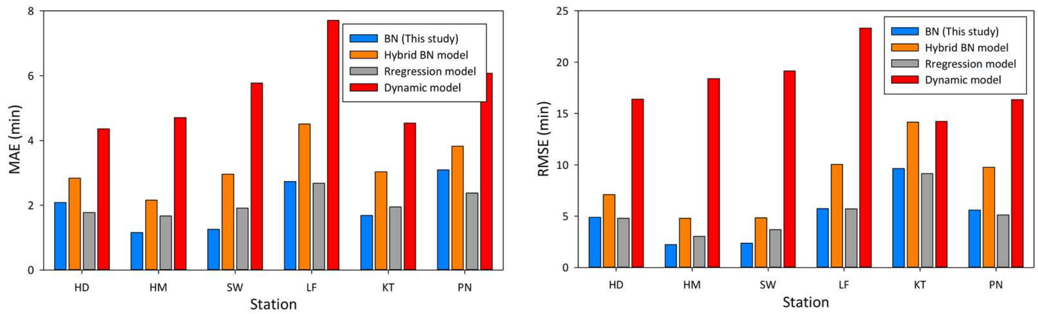
(a) (b)

Fig. 22. Predictive MAE (a) and RMSE (b) of L on the X-S HSR line.



(a) (b)

Fig. 23. Predictive MAE (a) and RMSE (b) of *N* on the X-S HSR line.



(a) (b)

Fig. 24. Predictive MAE (a) and RMSE (b) of *T* on the X-S HSR line.

In real-time dispatching, however, the most useful action for dispatchers is to predict and track the changes of these variables station by station, to make timely decisions. Therefore, accurate prediction of the possible increase or decrease of these variables can improve the efficiency of dispatching work. For example, if the predicted value of any of these variables shows an increase, then the dispatchers would need to re-schedule trains to avoid train conflicts. On the other hand, by seeing a sudden decrease in any of these variables, dispatchers would need to reschedule the delayed trains to recover the delays. Now, we investigate the model performance against these unexpected changes. To this end, when we are evaluating the model performance for variable L at station m, we choose those samples whose value of L changes from station m-1 to m. Considering the horizons of variables and the horizons of changes, for variable L, any delay increase or recovery is treated as unexpected change. For variable N, any increase or decrease was treated as unexpected change, and for variable T, any change exceeding 10 min is treated as unexpected change. Then, we treat disruptions in ordinal terms by dividing the domain of each variable into discrete intervals. The intervals of each variable are defined as follow:

L: [Min, 10], (10, 20], and [20, Max], labeled as short, medium, and long disruptions, respectively.

N: [Min, 5], (5, 10], and [10, Max], labeled as short, medium, and long disruptions, respectively.

T: [Min, 50], (50, 150], and [150, Max], labeled as short, medium, and long disruptions, respectively.

We select one of the most widely used metrics for discrete problems—accuracy, as shown in Eq. (7), to investigate the model performance. Table 6 gives five cases of the data on variable L used to calculate the accuracy of the BN model at the YDW station of the W-G HSR, where it shows that the value of L for these five cases has changes exceeding 1 min from station QY to YDW.

We first calculate the average accuracy of each model for all stations on W-G and X-S HSR lines, respectively, as shown in Table 7. Table 7 shows that the proposed BN model has satisfactory performance in predicting variables with unexpected changes, compared to the benchmark models. In addition, the accuracy of the models at each station for the W-G HSR and X-S HSR is shown in Tables 8 and 9, respectively. Table 8 shows that the accuracy of the proposed BN model for *L*, *N*, and *T* at each station of the W-G HSR is higher than 55%, and the highest can reach up to 98%. Table 9 shows that the accuracy of the proposed model for *L*, *N*, and *T* at each station of the X-S HSR is higher than 75%, with the maximum achieved an accuracy of 100%. Although, some benchmark models have higher accuracy at some stations, however, the proposed model shows high accuracy over the whole line and most of its stations.

$$\text{Accuracy} = \frac{\text{No. of True cases}}{\text{Total cases}} \times 100\% \tag{7}$$

Table 6. An instance of the data for variable change investigation.

Observed LQY (min)	Observed LYDW (min)	Change of L from QY to YDW (min)	Predicted LYDW (min)	Actual label of *L* at YDW	Predicted label of *L* at YDW
9	15	6	9.4	Medium	Short
34	40	6	36.2	Long	Long
13	8	5	12.5	Short	Medium
20	18	2	19.5	Medium	Medium

Table 7. The accuracy of each model for all stations.

Model	W-G HSR			X-S HSR		
	L	*N*	*T*	*L*	*N*	*T*
BN (This study)	0.966	**0.748**	**0.910**	**0.948**	0.911	**0.879**
Hybrid BN model	**0.967**	0.740	0.760	0.938	**0.913**	0.788
Regression model	0.961	0.708	0.900	0.926	0.909	0.856
Dynamic model	**0.967**	0.739	0.739	0.940	0.898	0.710

The highlighted numbers represent the best results.

Table 8. The accuracy of each model at each station for the W-G HSR.

Variable	Model	Station					
		YDW	SG	LCE	CZW	LYW	HYE
L	BN (This study)	**0.981**	**0.966**	0.960	0.962	**0.961**	**0.965**
	Hybrid BN model	0.978	0.965	**0.979**	0.966	0.956	0.956
	Regression model	0.978	0.965	0.959	0.943	0.960	0.958
	Dynamic model	0.981	0.964	**0.979**	**0.968**	0.949	0.958
N	BN (This study)	0.813	**0.792**	**0.824**	0.573	0.639	**0.848**
	Hybrid BN model	0.801	0.782	0.815	**0.590**	**0.649**	0.805
	Regression model	0.765	0.779	0.813	0.500	0.635	0.753
	Dynamic model	**0.827**	0.776	0.815	0.575	0.617	0.822
T	BN (This study)	**0.978**	**0.863**	**0.918**	**0.880**	**0.900**	**0.921**
	Hybrid BN model	0.873	0.807	0.800	0.654	0.608	0.817
	Regression model	0.971	0.862	0.900	0.853	0.907	0.908
	Dynamic model	0.880	0.807	0.782	0.625	0.515	0.825

The highlighted numbers represent the best results.

Transportation Research Part C: Emerging Technologies. 2020 May 1; 114:338-58.

Table 9. The accuracy of each model at each station for the X-S HSR.

Variable	Model	Station					
		HD	HM	SW	LF	KT	PN
L	BN (This study)	**0.931**	**0.941**	**0.985**	**0.950**	**0.935**	**0.945**
	Hybrid BN model	0.923	0.941	0.984	0.899	0.934	0.945
	Regression model	0.931	0.941	0.907	0.899	0.934	0.945
	Dynamic model	0.923	0.941	0.984	0.915	0.934	0.940
N	BN (This study)	**0.938**	0.910	0.874	**0.968**	**0.911**	0.867
	Hybrid BN model	0.937	**0.925**	**0.908**	0.947	0.892	0.866
	Regression model	0.937	0.910	0.873	0.947	0.910	**0.877**
	Dynamic model	0.937	0.865	0.873	0.936	0.910	0.866
T	BN (This study)	**0.750**	**0.850**	**1.000**	0.875	**0.875**	**0.923**
	Hybrid BN model	0.750	0.690	0.913	**0.906**	0.625	0.846
	Regression model	0.750	0.846	0.869	0.875	0.875	0.923
	Dynamic model	0.625	0.692	0.695	0.813	0.625	0.807

The highlighted numbers represent the best results.

6. Conclusions

In this paper, we propose a BN model to predict the effects of disturbance and disruption propagation in time and space dimensions during train operations. To this end, we first select three main factors, namely the primary delays, the number of affected trains, and the total delayed times as the metrics of the disruption effects. We then investigate the performance of four different BN structures, incorporating the combination of domain knowledge and the data-driven method. Next, we run a 5-fold cross-validation technique to evaluate the predictive performance and generalizability of the proposed model using the operational data from two HSR lines in China. In the next step, we examine the performance of the nominated BN model at each station against MAE and RMSE metrics, through comprehensive experiments to show the generalizability of the proposed model to other railway systems and demonstrate the potential application of the model in the future.

Our future work includes the extension of the proposed model to consider more variables, such as disruption cause, timetable structure, infrastructure influence, and passenger flow. This means that the BN models can be based on disruptions at different periods (peak-hour, off-peak-hour, ordinary days, holidays etc.), and disruptions happening at stations with different track and platform characteristics. Further, we will seek to improve the proposed BN model to be capable of predictions on very large size railway networks. This is readily possible, as for each disturbance or disruption, the proposed BN model just needs one prediction at a station, compared to the existing train delay prediction models that need to predict the delays of all trains at each station.

Declarations of interest

None.

Transportation Research Part C: Emerging Technologies. 2020 May 1; 114:338-58.

Acknowledgments

This work was supported by the National Nature Science Foundation of China [grant numbers 71871188, U1834209], and the Science & Technology Department of Sichuan Province [grant number 2018JY0567], the Doctoral Innovation Fund Program of Southwest Jiaotong University [grant number D-CX201827], and the China Scholarship Council [grant number 201707000038]. We are grateful for the useful contributions made by our project partners.

Reference

Büker, T., Seybold, B., 2012. Stochastic modelling of delay propagation in large networks. *Journal of Rail Transport Planning & Management* 2(1-2), 34-50.

Barbour, W., Mori, J.C.M., Kuppa, S., Work, D.B., 2018. Prediction of arrival times of freight traffic on us railroads using support vector regression. *Transportation Research Part C: Emerging Technologies* 93, 211-227.

Barta, J., Rizzoli, A.E., Salani, M., Gambardella, L.M., 2012. Statistical modelling of delays in a rail freight transportation network, Proceedings of the 2012 Winter Simulation Conference (WSC). IEEE, pp. 1-12.

Berger, A., Gebhardt, A., Müller-Hannemann, M., Ostrowski, M., 2011. Stochastic delay prediction in large train networks, *11th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

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

Briggs, K., Beck, C., 2007. Modelling train delays with q-exponential functions. *Physica A: Statistical Mechanics and its Applications* 378(2), 498-504.

Burdett, R., Kozan, E., 2014. Determining operations affected by delay in predictive train timetables. *Computers & Operations Research* 41, 150-166.

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. *Transportation Research Part B: Methodological* 63, 15-37.

Cerreto, F., Nielsen, B.F., Nielsen, O.A., Harrod, S.S., 2018. Application of data clustering to railway delay pattern recognition. *Journal of Advanced Transportation* 2018. doi:10.1155/2018/6164534.

Cerreto, F., Nielsen, O.A., Harrod, S., Nielsen, B.F., 2016. Causal analysis of railway running delays, 11th World Congress on Railway Research (WCRR 2016), Milan, Italy.

Chen, D., Wang, L., Li, L., 2015. Position computation models for high-speed train based on support vector machine approach. *Applied Soft Computing* 30, 758-766.

Corman, F., D'Ariano, A., Hansen, I.A., 2014. Evaluating disturbance robustness of railway schedules. *Journal of Intelligent Transportation Systems* 18(1), 106-120.

Corman, F., Kecman, P., 2018. Stochastic prediction of train delays in real-time using Bayesian networks. *Transportation Research Part C: Emerging Technologies* 95, 599-615.

Corman, F., Meng, L., 2015. A review of online dynamic models and algorithms for railway traffic management. *IEEE Transactions on Intelligent Transportation Systems* 16(3), 1274-1284.

Gaurav, R., Srivastava, B., 2018. Estimating Train Delays in a Large Rail Network Using a Zero Shot

Transportation Research Part C: Emerging Technologies. 2020 May 1; 114:338-58.

Markov Model, *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, pp. 1221-1226.

Ghaemi, N., Zilko, A.A., Yan, F., Cats, O., Kurowicka, D., Goverde, R.M., 2018. Impact of railway disruption predictions and rescheduling on passenger delays. *Journal of Rail Transport Planning & Management* 8(2), 103-122.

Gorman, M.F., 2009. Statistical estimation of railroad congestion delay. *Transportation Research Part E: Logistics and Transportation Review* 45(3), 446-456.

Goverde, R.M., 2010. A delay propagation algorithm for large-scale railway traffic networks. *Transportation Research Part C: Emerging Technologies* 18(3), 269-287.

Goverde, R.M., Corman, F., D'Ariano, A., 2013. Railway line capacity consumption of different railway signalling systems under scheduled and disturbed conditions. *Journal of Rail Transport Planning & Management* 3(3), 78-94.

Goverde, R.M., Hansen, I., Hooghiemstra, G., Lopuhaa, H., 2001. Delay distributions in railway stations, *9th World Conference on Transport Research, Seoul, Korea, July 22-27, 2001*. WCTRS.

Haahr, J.T., Hellsten, E.O., van der Hurk, E., 2019. Train Delay Prediction in the Netherlands through Neural Networks.

Hansen, I.A., Goverde, R.M., van der Meer, D.J., 2010. Online train delay recognition and running time prediction, *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on*. IEEE, pp. 1783-1788.

Harris, M., 2006. Analysis and modelling of train delay data. MSc Thesis, University of York, UK.

Heckerman, D., Geiger, D., Chickering, D.M., 1995. Learning Bayesian networks: The combination of knowledge and statistical data. *Machine learning* 20(3), 197-243.

Higgins, A., Kozan, E., Ferreira, L., 1995. Modelling delay risks associated with train schedules. *Transportation Planning and Technology* 19(2), 89-108.

Huang, P., Wen, C., Fu, L., Peng, qiyuan, Li, Z., 2020a. A hybrid model to improve the train running time prediction ability during high-speed railway disruptions. Safety Science doi:10.1016/j.ssci.2019.104510.

Huang, P., Wen, C., Fu, L., Peng, Q., Tang, Y., 2020b. A deep learning approach for multi-attribute data:A study of train delay prediction in railway systems. Information Sciences 516, 234–253.

Huang, P., Wen, C., Peng, Q., Jiang, C., Yang, Y., Fu, Z., 2019. Modeling the Influence of Disturbances in High-Speed Railway Systems. *Journal of Advanced Transportation*. doi:10.1155/2019/8639589.

Huisman, T., Boucherie, R.J., 2001. Running times on railway sections with heterogeneous train traffic. *Transportation Research Part B: Methodological* 35(3), 271-292.

Jespersen-Groth, J., Potthoff, D., Clausen, J., Huisman, D., Kroon, L., Maróti, G., Nielsen, M.N., 2009. Disruption management in passenger railway transportation, *Robust and online large-scale optimization*. Springer, pp. 399-421.

Kecman, P., Corman, F., Meng, L., 2015. Train delay evolution as a stochastic process, *6th International Conference on Railway Operations Modelling and Analysis-RailTokyo2015*.

Kecman, P., Goverde, R.M., 2015a. Online data-driven adaptive prediction of train event times. *IEEE Transactions on Intelligent Transportation Systems* 16(1), 465-474.

Kecman, P., Goverde, R.M., 2015b. Predictive modelling of running and dwell times in railway traffic. *Public Transport* 7(3), 295-319.

Kellermann, P., Schönberger, C., Thieken, A.H., 2016. Large-scale application of the flood damage model RAilway Infrastructure Loss (RAIL). *Natural Hazards and Earth System Sciences* 16(11),

Transportation Research Part C: Emerging Technologies. 2020 May 1; 114:338-58.

2357-2371.

Lessan, J., Fu, L., Wen, C., 2019. A hybrid Bayesian Network Model for Predicting Delays in Train Operations. *Computers & Industrial Engineering* 127, 1214-1222.

Lessan, J., Fu, L., Wen, C., Huang, P., Jiang, C., 2018. Stochastic model of train running time and arrival delay: a case study of Wuhan–Guangzhou high-speed rail. *Transportation Research Record* 2672(10), 215-223.

Li, D., Daamen, W., Goverde, R.M., 2016. Estimation of train dwell time at short stops based on track occupation event data: A study at a Dutch railway station. *Journal of Advanced Transportation* 50(5), 877-896.

Marković, N., Milinković, S., Tikhonov, K.S., Schonfeld, P., 2015. Analyzing passenger train arrival delays with support vector regression. *Transportation Research Part C: Emerging Technologies* 56, 251-262.

Meester, L.E., Muns, S., 2007. Stochastic delay propagation in railway networks and phase-type distributions. *Transportation Research Part B: Methodological* 41(2), 218-230.

Milinković, S., Marković, M., Vesković, S., Ivić, M., Pavlović, N., 2013. A fuzzy Petri net model to estimate train delays. *Simulation Modelling Practice and Theory* 33, 144-157.

Murali, P., Dessouky, M., Ordóñez, F., Palmer, K., 2010. A delay estimation technique for single and double-track railroads. *Transportation Research Part E: Logistics and Transportation Review* 46(4), 483-495.

Nabian, M.A., Alemazkoor, N., Meidani, H., 2019. Predicting Near-Term Train Schedule Performance and Delay Using Bi-Level Random Forests. *Transportation Research Record*, 2673 (5), 564–573.

Nair, R., Hoang, T.L., Laumanns, M., Chen, B., Cogill, R., Szabó, J., Walter, T., 2019. An ensemble prediction model for train delays. *Transportation Research Part C: Emerging Technologies* 104, 196-209.

Nash, A., Huerlimann, D., 2004. Railroad simulation using OpenTrack. *WIT Transactions on The Built Environment* 74.

Neapolitan, R.E., 2004. *Learning bayesian networks*. Pearson Prentice Hall Upper Saddle River, NJ.

Nielsen, T.D., Jensen, F.V., 2009. *Bayesian networks and decision graphs*. Springer Science & Business Media.

Oneto, L.,Fumeo, E., Clerico, G., Canepa, R., Papa, F., Dambra, C., Mazzino, N., Anguita, D., 2017. Dynamic delay predictions for large-scale railway networks: Deep and shallow extreme learning machines tuned via thresholdout. IEEE Transactions on Systems, Man, and Cybernetics: Systems 47 (10), 2754–2767.

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 research* 11, 54-64.

Şahin, İ., 2017. Markov chain model for delay distribution in train schedules: Assessing the effectiveness of time allowances. *Journal of Rail Transport Planning & Management*, 7 (3), 101–113

Tsamardinos, I., Brown, L.E., Aliferis, C.F., 2006. The max-min hill-climbing Bayesian network structure learning algorithm. *Machine learning* 65(1), 31-78.

Van der Meer, D.J., Goverde, R.M., Hansen, I.A., 2009. Prediction of train running times using historical track occupation data, *WCTR*. Delft University of Technology, pp. 1-62.

Veelenturf, L.P., Kidd, M.P., Cacchiani, V., Kroon, L.G., Toth, P., 2015. A railway timetable rescheduling approach for handling large-scale disruptions. *Transportation Science* 50(3), 841-862.

Wallander, J., Mäkitalo, M., 2012. Data mining in rail transport delay chain analysis. *International*

Transportation Research Part C: Emerging Technologies. 2020 May 1; 114:338-58.

Journal of Shipping and Transport Logistics 4(3), 269-285.

Wen, C., Huang, P., Li, Z., Lessan, J., Fu, L., Jiang, C., Xu, X., 2019. Train Dispatching Management With Data-Driven Approaches: A Comprehensive Review and Appraisal. IEEE Access 7, 114547–114571.

Wen, C., Li, Z., Lessan, J., Fu, L., Huang, P., Jiang, C., 2017. Statistical investigation on train primary delay based on real records: evidence from Wuhan–Guangzhou HSR. *International Journal of Rail Transportation* 5(3), 170-189.

Wiklund, M., 2003. Serious breakdowns in the track infrastructure: calculation of effects on rail traffic. *VTI Meddelande*(959).

Yaghini, M., Khoshraftar, M.M., Seyedabadi, M., 2013. Railway passenger train delay prediction via neural network model. *Journal of advanced transportation* 47(3), 355-368.

Yang, Y., Huang, P., Peng, Q., Jie, L., Wen, C., 2019. Statistical delay distribution analysis on high-speed railway trains. *Journal of Modern Transportation*, 27(3), 188-197.

Yasunobu, S., Miyamoto, S., Ihara, H., 1993. Fuzzy control for automatic train operation system, *Readings in Fuzzy Sets for Intelligent Systems*. Elsevier, pp. 348-354.

Zhan, S., Kroon, L.G., Zhao, J., Peng, Q., 2016. A rolling horizon approach to the high speed train rescheduling problem in case of a partial segment blockage. *Transportation Research Part E: Logistics and Transportation Review* 95, 32-61.

Zilko, A.A., Kurowicka, D., Goverde, R.M., 2016. Modeling railway disruption lengths with Copula Bayesian Networks. *Transportation Research Part C: Emerging Technologies* 68, 350-368.