

# Accepted Manuscript

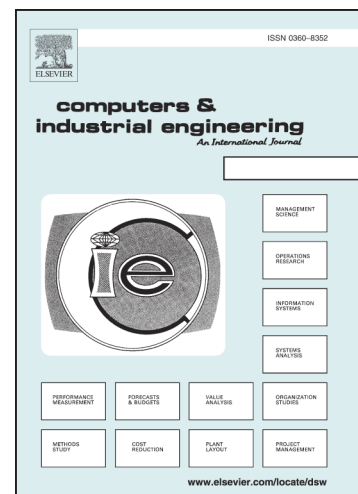
A Hybrid Bayesian Network Model for Predicting Delays in Train Operations

Javad Lessan, Liping Fu, Chao Wen

PII: S0360-8352(18)30102-5  
DOI: <https://doi.org/10.1016/j.cie.2018.03.017>  
Reference: CAIE 5121

To appear in: *Computers & Industrial Engineering*

Accepted Date: 9 March 2018



Please cite this article as: Lessan, J., Fu, L., Wen, C., A Hybrid Bayesian Network Model for Predicting Delays in Train Operations, *Computers & Industrial Engineering* (2018), doi: <https://doi.org/10.1016/j.cie.2018.03.017>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

The final publication is available at Elsevier via <https://doi.org/10.1016/j.jedc.2018.11.005>. © 2018  
This manuscript version is made available under the CC-BY-NC-ND 4.0 license  
<http://creativecommons.org/licenses/by-nc-nd/4.0/>

# A Hybrid Bayesian Network Model for Predicting Delays in Train Operations

Javad Lessan<sup>a,c</sup>, Liping Fu<sup>a,b</sup>, Chao Wen<sup>b,c,\*</sup>

<sup>a</sup> Department of Civil and Environmental Engineering, University of Waterloo, Waterloo, N2L 3G1, Canada

<sup>b</sup> School of Transportation & Logistics, Southwest Jiaotong University, Chengdu Sichuan 610031, China

<sup>c</sup> Railway Research Center, University of Waterloo, Waterloo, N2L 3G1, Canada

Email Addresses:

Javad Lessan: [jlessan@uwaterloo.ca](mailto:jlessan@uwaterloo.ca)

Liping Fu: [lfu@uwaterloo.ca](mailto:lfu@uwaterloo.ca)

Chao Wen: [wenchao@swjtu.cn](mailto:wenchao@swjtu.cn)

\*Corresponding Author:

Email address: [wenchao@swjtu.cn](mailto:wenchao@swjtu.cn) (Chao Wen)

## A Hybrid Bayesian Network Model for Predicting Delays in Train Operations

---

### Abstract

We present a Bayesian network-(BN) based train delay prediction model to tackle the complexity and dependency nature of train operations. Three different BN schemes, namely, heuristic hill-climbing, primitive linear and hybrid structure, are investigated using real-world train operation data from a high-speed railway line. We first use historical data to rationalize the dependency graph of the developed structures. Each BN structure is then trained with the gold standard  $k$ -fold cross validation approach to avoid over-fitting and evaluate its performance against the others. Overall, the validation results indicate that a BN-based model can be an efficient tool for capturing superposition and interaction effects of train delays. However, a well-designed hybrid BN structure, developed based on domain knowledge and judgments of expertise and local authorities, can outperform the other models. We present a performance comparison of the predictions obtained from the hybrid BN structure against the real-world benchmark data. The results show that the proposed model on overage can achieve over 80% accuracy in predictions within a 60-minute horizon, yielding low prediction errors regarding mean absolute error (MAE), mean error (ME) and root mean square error (RMSE) measures.

*Keywords:* High-speed rail; Train operation; Punctuality; Bayesian networks; Delay prediction; Performance evaluation.

---

## 1. Introduction

A railway system comprises several subsystems, such as network infrastructure, rolling-stock, control and communication, and various operational rules and policies with the goal of providing reliable train services to transport passengers or goods. However, many uncertainties may arise from these subsystems that can disturb the planned activities and operations, resulting in unexpected delays (Wen et al., 2017). As a service complaint, train delays impose a huge cost on passengers and operators, contributing to the inefficiency of train operations (Van Oort, 2011). In the United Kingdom, for instance, 14 million train-minute delays were recorded during 2006-2007 on the British national rail network that cost over £1 billion in terms of lost time to the passengers (Office, 2008). Consequently, reducing delays is of great importance to train operators and desirable to passengers (Marković et al., 2015). Specifically, the validity of all levels of railway operations planning, such as creating feasible and realizable timetables, predicting real-time traffic, predicting conflicts, and providing reliable passenger information, depends highly on the accurate estimation of train process times that are subject to delay incidents (Kecman & Goverde, 2015b; Kecman et al., 2015b; Kecman & Goverde, 2015a). Therefore, delays should be predicted and compensated in time, otherwise there may be a disruption or domino effect of the propagated delays (Zhang et al., 2018). While part of the delay factors influencing train process times is predictable and controllable, most of them are not only uncontrollable but also unpredictable, adding to the challenges of managing railway operations.

In real-world train operations, delay prediction relies heavily on the experience and intuition of a local dispatcher rather than a network-wide computational instrument (Martin, 2016). Given the complex structure of a railway network and interdependent train operations between a large set of origins and destinations, a local dispatcher's estimation of delays and the subsequent decisions are strongly dependent on the state of traffic and network and limited to a local geographical area. In large and dense network areas, however, the

domain knowledge and expertise of local dispatchers must be supported by an advanced computational tool that can account for the interdependencies of train operations and interrelated delay factors. Creation of such an advanced tool has been hindered by two fundamental limitations. Firstly, methodologically, there has been a lack of models capable of simultaneously examining multiple components of delay incidents intertwined with stochastic operations and interaction effects. Secondly, technologically, there has been a need for collection and incorporation of massive train operation data. Recently, the integration of graph and probability theories led to the introduction of Bayesian networks (BNs) that enabled practitioners to overtake these limitations. Specifically, BNs methodology is a representational tool meant to capture complex structures and “organize one’s knowledge about a particular situation into a coherent whole” (Darwiche, 2009). At the same time, it allows for incorporation of massive historical data in identifying the contingencies between multiple events and updating the state of different variables given real-time data. These features, convoluting different factors and fusing massive data, have given BNs an advantage over other artificial intelligence techniques.

In this paper, we present three different BN designing architectures, namely, a heuristic, a naive, and a hybrid method, to represent the relationship and superposition of interdependent variables identified in the delay chain of trains. Using information obtained from historical data, we rationalize the contingency graph of the proposed BN structures. Next, we apply the gold standard  $k$ -fold cross-validation method to train and evaluate the proposed BNs. The hybrid BN structure, having a higher performance compared to the other models, is then tested against real-world benchmark data under different performance measures. To the best of our knowledge, this is the first hybrid BN-based delay prediction model introduced into the relevant prediction literature. The main idea behind the hybrid structure introduced here is to distinguish between the delay due to the most recent performed operation and the delay propagated from previous operations. The proposed ideas were made possible through examining the similarities and differences between the naive and heuristic structures supported

by domain knowledge and expertise of local authorities. Our results can be generalized to similar problems in other networks in order to better support train dispatching and delay management decisions.

65 The remainder of this paper is structured as follows. The next section presents a brief overview of the related literature and summarizes our contributions. Section (3) provides the methodological framework and formal description of the terms and the concepts used in this study. Section (4) describes the historical data and our general assumptions, and continues with training  
70 and validating results of the candidate BNs. Section (5) focuses on evaluating the performance of the hybrid BN model discussed under different performance measures. Finally, conclusions and future research directions are presented in Section (6).

## 2. Literature Review

75 Train timetables are traditionally scheduled using train motion equations with the input of the estimated running and dwelling times at individual stations and sections. To minimize the probability of schedule deviation in actual operations, the parameters of these equations are usually tuned or optimized based on historical train data (Kecman & Goverde, 2015b). However, these  
80 techniques are not adaptive, often failing to address the time-varying nature of train operation settings. For example, each new operational configuration would require re-optimizing the timetables, which is computationally extensive. Some of these drawbacks could be overcome by applying data-driven approaches and statistical models to estimate the process times based on various contribut-  
85 ing factors (Kecman & Goverde, 2015b). The underlying problem is related to the delay prediction practice that has received considerable attention due to its vital importance to train operations management and passenger information provision (Meester & Muns, 2007).

A number of prediction models have been developed in the literature, which  
90 can be classified by their scope, model types and solution methods (Marković

et al., 2015). Traditional methods such as regression models have been introduced to predict delays. However, these methods require frequent updates of train positions and rich data. Micro- and macro-level simulation tools have been applied to simulate delays at different level of details. The simulation  
95 models, developed based on fixed distributions, require frequent updates from train positions and real-time train data (Kecman et al., 2015b). The update requirements are mostly due to time-varying operational conditions and the interaction between different subsystems (stations, sections and trains) under the effects of infrastructure and operational rules. Yuan (2006) and Yuan et al.  
100 (2002) presented a delay prediction model that deals with the stochastic behavior, dependency of train delays and delay propagation to assess stability and punctuality of a published timetable against primary delays. An artificial neural network model was proposed to predict the delay of passenger trains in Iranian Railways (Yaghini et al., 2013). The accuracy level of the proposed model was  
105 found to be superior to other statistical models such as decision tree and multinomial logistic regression methods. Peters et al. (2005) developed an intelligent real-time delay prediction model that predicts the delay of the upstream or downstream trains based on the delays currently incurred in the network. The prediction accuracy of the proposed model was compared against a rule-based  
110 system with a set of predefined rules in a deterministic manner. However, these models are not flexible enough to incorporate the domain knowledge of experts and local dispatchers as well as the operational characteristics.

A generic statistical model for estimating the running and dwelling times was proposed by Kecman & Goverde (2015b). Three global predictive models:  
115 robust linear regression, regression trees, and random forests are presented based on advanced statistical learning techniques. Moreover, based on the robust linear regression and some refinements, they calibrated local models for each particular train line, station or block section. The presented models were evaluated using an aggregated set of historical data on the level of block sections.  
120 In another effort, the real-time prediction of train delays was used to detect instabilities in the timetable and retrieve a feasible train schedule (Marković et al.,

2015). Kecman (2014) also proposed a real-time delay prediction model based on historical arrival and departure data. Event graphs were used in Hansen et al. (2010), to forecast running and arrival times. A stochastic model for delay propagation in large transportation networks was proposed by Berger et al. 125 (2011), to process massive streams of real-time data. In the same way, the statistical models are not adaptive enough to incorporate the domain knowledge of local dispatchers and the networks' characteristics.

A model for real-time prediction of train delays using Bayesian reasoning can be found in Kecman et al. (2015a). They used two months of historical traffic 130 realization data from the Swedish infrastructure manager in a simulated real-time environment. The computational results indicated that the predictions are reliable for up to 30-minute horizons. Their main assumption, however, is that the train orders and routes within the prediction horizon are known, which 135 is often not the case in the real-world. A Bayesian model for predicting the propagation of delays can be found in Kecman et al. (2015b), which uses real-time events based on their specific order. Martin (2016) proposed a prototype rail advisory system that applies a series of predictive reasoning and machine learning models, to predict the effects of various disruptions. Also train movement data, collected from the infrastructure track occupation records, sensors 140 in rolling-stock, or mobile GPS devices, were used by Flier et al. (2009) to find robust train paths. Marković et al. (2015) presented a comparison between the performance of support vector regression and neural networks for analyzing passenger train arrival delays and the influence of infrastructure on arrival delays. 145 Using numerous test instances they show that support vector regression outperforms other models in predicting arrival delays. However, to date, identifying which BN architectures are most valid/reliable for predicting train delays for each particular network structure has not been well studied.

Clearly, there is still a need for better predictive models that account for 150 massive real-world train operation data, domain knowledge and expertise of local authorities. In this paper, for the first time we propose a hybrid BN-based predictive model for predicting arrival and departure delays, built upon test-



ing different BN architectures, wealth of train operation records, and domain-specific knowledge. The proposed model is easy to interpret and generalize  
 155 while at the same time computationally efficient. The work presented in this paper contributes to the literature with a new delay prediction model obtained based on heuristic and naive prediction models, train movement data, and domain knowledge. The application of the proposed BN model in real-life train operations can contribute to the development of a more robust train operation  
 160 management and information system for improved dispatching and user services.

### 3. Methodological Framework and Preliminaries

#### 3.1. Bayesian Networks

Let us consider  $n$  random variables  $X_1, X_2, \dots, X_n$ , and a directed acyclic graph (DAG) where each node  $j$  ( $1 \leq j \leq n$ ) of the graph is associated to the  
 165 variable  $X_j$ . Then the graph is a BN, representing the dependencies of  $X_1, X_2, \dots, X_n$ , if:

$$P(X_1, X_2, \dots, X_n) = \prod_{j=1}^n P(X_j | \text{parents}(X_j)), \quad (1)$$

where  $\text{parents}(X_j)$  denotes the parent nodes of  $X_j$ . The parent nodes are the set of all nodes  $X_i$  each of which is directly linked to node  $j$  with an arc in the graph in BN, i.e.,  $i \rightarrow j$ . The distribution  $P(X_j | \text{parents}(X_j))$  is viewed as  
 170 a local distribution function, which can be expressed by a probabilistic classification or regression function (Nielsen & Jensen, 2009).

There are two parts of a BN which must be determined. The first part is the structure of the graph, which can be created either randomly by data learning using heuristic methods, or can be designed on the basis of domain-specific  
 175 knowledge. Hybrid BN structures can also be developed, for example, through domain-specific knowledge and refining or merging heuristic/random structures based on observed dynamics and dependences between variables. This is mostly done by refining a given structure through adding or removing variables or

connections if desired, to achieve a superior BN structure in terms of a pertinent  
 180 explanatory measure. The second part of designing a BN is the solution of the  
 network parameters or the state of a node given the state(s) of its parent node(s)  
 (Kecman et al., 2015a). The states, which are defined by either conditional  
 probabilities or regression models, can be inferred or regressed directly from  
 observations through either a diagnostic or causal inference method (Nielsen &  
 185 Jensen, 2009).

### 3.2. Dependency Representation

In this paper, we use the *event-activity* modeling approach to formulate de-  
 lay dependencies. It models the running and dwelling operations defined in  
 a timetable as alternating activities in a network structure which is called an  
 190 *event-activity* graph. It is also a convenient way of describing delay propagation  
 that represents a train run by an interconnected sequence of events and activ-  
 ities. The main train activities are running, dwelling and waiting operations,  
 each of which needs a minimum amount of time to be accomplished. Each event,  
 such as departure, arrival or passage at any track section, represents the begin-  
 195 ning or the end of a process. The events are either arrival or departure types  
 that can occur simultaneously (in case of more than one train). The events  
 are connected by the corresponding running and dwelling activities. This is also  
 similar to the logic applied in other studies to represent the relationship between  
 train delays (Dollevoet et al., 2014; Kecman et al., 2015b; Martin, 2016).

200 Consider, for example, a typical train passes through  $S$  number of stations  
 according to its published timetable, which is scheduled to arrive at time  $t_s^A$  at  
 the station  $s \in S$ , and scheduled to depart at time  $t_s^D$  from the same station.  
 However, due to various disturbances during its operation, the train can deviate  
 from the scheduled operations and have the realizations of actual arrival time  
 205  $\hat{t}_s^A$  and actual departure time  $\hat{t}_s^D$ . Figure (1) depicts two successive stations  
 $s$  and  $s'$ , where the parameters inside the parentheses display the scheduled  
 and actual time of the events, respectively. In railway operations terminology,  
 delays are commonly called arrival and departure delay. The (positive) differ-



Figure 1: The general scheme of train movements at two successive stations  $s$  and  $s' = s + 1$ . Up: nominal and actual arrivals ( $A$ ), Down: nominal and actual departures ( $D$ ).

ence between actual and scheduled times,  $(\hat{t}_s^A - t_s^A)^+ = \max(\hat{t}_s^A - t_s^A, 0)$  and  
 210  $(\hat{t}_s^D - t_s^D)^+ = \max(\hat{t}_s^D - t_s^D, 0)$ , are defined as arrival and departure delays at  
 station  $s$ , respectively. Similarly, the actual dwell time of a train at station  $s$  is  
 defined as the difference between its departure and arrival times  $(\hat{t}_s^D - \hat{t}_s^A)$ , and  
 the actual running time is defined, for two consecutive stations  $s$  and  $s'$ , as the  
 time taken to traverse the section connecting the two stations  $(\hat{t}_{s'}^A - \hat{t}_s^D)$ . Like-  
 215 wise,  $(t_s^D - t_s^A)$ , and  $(t_{s'}^A - t_s^D)$  are the scheduled dwell time and the scheduled  
 running time defined in the published timetable. In practice, some additional  
 time - buffer time - is added to improve schedule reliability and reduce the effect  
 of disturbances and recovering operations from unexpected delays during trains'  
 running and dwelling activities.

220 To model the delay chain of trains, we represent each arrival/departure event  
 of a train run as a node of a BN. In this way, each arc of the BN represents cor-  
 responding train activity. The number of events corresponds to the number of  
 scheduled arrival and departure events. All components are connected together  
 through successive events and activities based on their relationship, such that  
 225 the model represents the logical connection among the different events and ac-  
 tivities. The principle idea is to predict the process time of each train's running  
 and dwelling activity depending on the state of its parent nodes that reflect  
 the actual position (or delay) of the immediate upstream events. Without loss  
 of generality, we assume that the parent nodes play the role of explanatory  
 230 variables to each of their child nodes.

## 4. Training and Validation of BN Models

### 4.1. Description of the Data Set

The data used in this study come from train operation records on Wuhan-Guangzhou (WH-GZ) high-speed rail (HSR) line in China. The WH-GZ HSR connects Wuhan in Hubei province to Guangzhou in Guangdong province with a 1096km double-track HSR line including 18 stations, as shown in Figure (2). A total of 15 stations and 14 sections, from GZS to CBN, are managed by the Guangzhou Railway Bureau and the rest is managed by the Wuhan Railway Bureau. The train movement data were extracted from the Guangzhou Bureau database for the period from Feb. 2015 to Nov. 2015, which includes 378,510 arrival and departure events between the stations on the specified line, excluding early arrivals and departures. We use 75% of the collected observations for training and comparing the candidate BN structures. The remaining 25% of the observations are withheld to test the superior prediction model and evaluate its prediction performance.

As reported by China Railway Corporation, the operation punctuality of its HSR lines on average is about 85% because of delays during train operations (China-Railway-Corporation, 2016). On one hand, the departure delay is due to the late arrivals or due to disturbances in train operation at stations. On the other hand, the arrival delay is due to departure delay in the previous station or due to a disturbance during traversal time in track sections. Therefore it is important to focus on both the departure and arrival delays. Figures (3) and (4) reveal that arrival and departure delays follow the same distribution, and there exists a linear relationship (or a chain) with a high correlation between the arrival and departure delays at stations. Indeed, the correlation of the arrival and departure delays at different stations is found to be at least 94%. We use these findings to characterize and calibrate delay dependencies in the proposed BN structures with different complexity level.

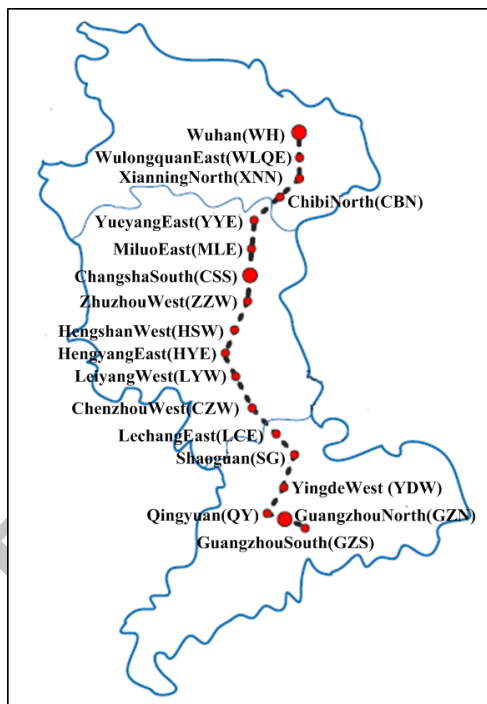
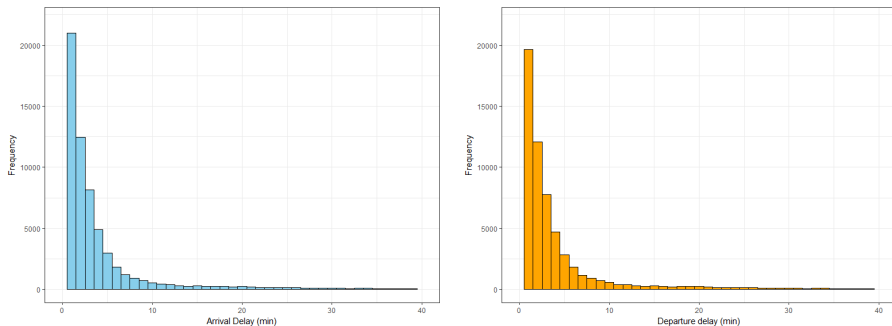


Figure 2: Map of Wuhan-Guangzhou HSR.



(a) Histogram of arrival delay observations (b) Histogram of departure delay observations

Figure 3: Frequency distribution of arrival and departure delay at all stations (bin: 1 min).

#### 4.2. Training and Validation

260 In this study, we trained and examined three different BN structures in R-project using the “bnlearn” package. The estimation of parameters is carried out using the maximum likelihood estimation (MLE) method to build the relationship between parent and child variables (Nagarajan et al., 2013).

265 The first network structure, denoted as HC, is developed using the hill-climbing method to learn heuristically the network structure from the empirical data. This approach starts from a DAG with  $n$  number of nodes and no arcs but being added one-by-one sequentially. More specifically, a network score is used to measure the additional explanatory power of the model when a new arc is added, or a current arc is reversed or removed between any pairs of nodes. In  
 270 each step the action (addition, removal or reversion) that achieves the highest score is picked. The iterative procedure continues until a higher value of the score measure cannot be obtained (Buntine, 1996). Figure (5) displays the BN structure obtained with the hill-climbing method.

275 Our second architecture is a primitive linear structure, denoted as PL, with  $n$  nodes, in which each event is connected to its immediate upstream event in the order they appear in the timetable. With the origin and destination nodes GZN and CBN, respectively, in this structure, events occur in a fixed sequence  $j \rightarrow j + 1$ , where  $j = 1, 2, \dots, n - 1$ , and  $n$  is the total number of events in a train

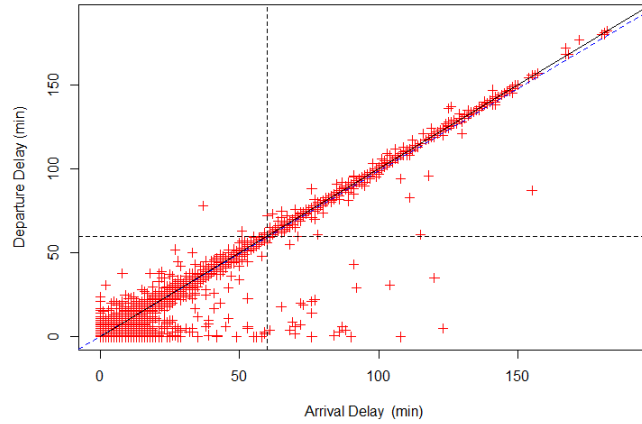


Figure 4: Scatter plot of all arrival vis-à-vis departure delay instances.

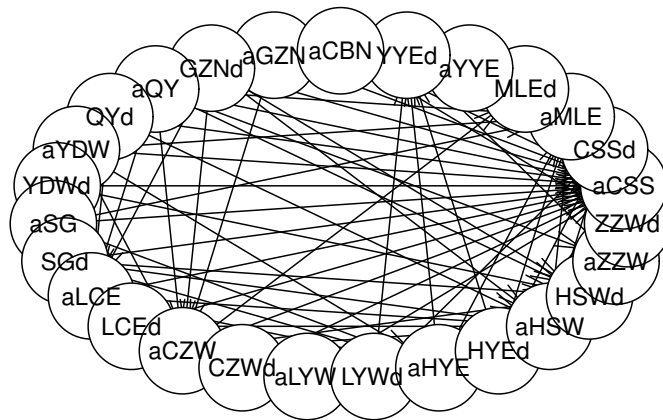


Figure 5: BN structure obtained by the hill-climbing heuristic.

run. In other words, every local node  $j$  is conditioned only on the most recent  
 280 node  $j - 1$ , where  $j - 1 \rightarrow j$  is the newly performed operation. In probabilistic  
 reasoning, this means that the arrival (departure) time distribution of a train  
 at each station is a function of the departure (arrival) time distribution of the  
 most recent performed operation.

The architecture of the third proposed BN model, named hybrid BN model  
 285 and denoted as HB, is based on the structure obtained by the heuristic hill-  
 climbing method and the structure of the primitive linear, which is then refined  
 using domain knowledge and the expert judgments about the sequence of sta-  
 tions and the relationships between consecutive train operations. More specifi-  
 cally, we investigated the network structure obtained from the heuristic method  
 290 and found out that each arrival (departure) node at the network is connected di-  
 rectly to its immediate upstream departure (arrival) event i.e.,  $j \rightarrow j + 1$ . This is  
 similar to the sequence of events in the primitive linear model as well. However  
 in the heuristic BN structure (HB), most of the arrival (departure) events were  
 connected to their second previous arrival (departure) event, i.e.,  $j - 1 \rightarrow j + 1$ .  
 295 This was in line with our preliminary knowledge that a train when coming to  
 (or departing from) a station, it would inhere a propagated arrival (departure)  
 delay from the last visited station. In addition, it can possibly incur a newly  
 formed delay during its most recent operation. As depicted in Figure (5), most  
 of the arrival (departure) events were connected to their second previous arrival  
 300 (departure) event. Moreover, the heuristic structure has some nodes with more  
 than two parents; however, we could not find any generalizable logic behind  
 this. Overall, the main idea underlying the hybrid structure is to distinguish  
 between the delay propagated from upstream operations and the delay due to  
 most recent (newly) performed operation in the prediction process. We note  
 305 that this can be generalized to the similar problems in the other networks.

Figure (6) depicts the graphical representation of the hybrid BN structure  
 with the origin and destination nodes GZN and CBN, respectively. In this struc-  
 ture, events occur in a fixed sequence  $j \rightarrow k$  and  $j \rightarrow k + 1$  and  $k \rightarrow k + 1$ , where  
 $k = j + 1, j = 1, 2, \dots, n - 1$ , and  $n$  is the total number of events in a train run. In



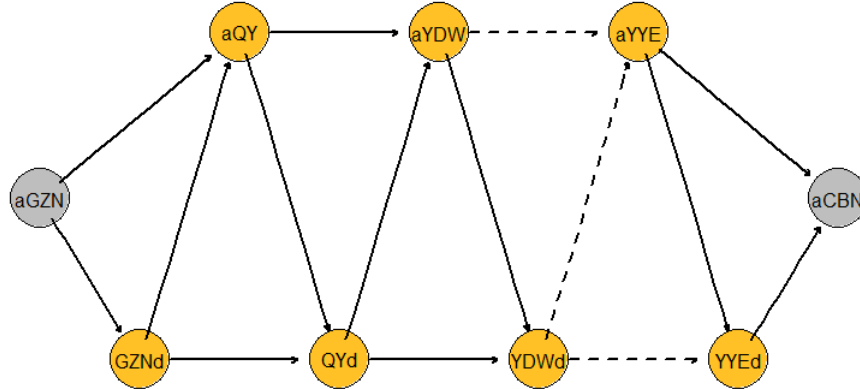


Figure 6: (Color online) Schematic representation of the hybrid BN architecture; station-level arrival and departure events are connected through running and dwelling activities along with the propagated delay links; for example “aQY” and “QYd”, respectively, stand for arrival (at) and departure (from) station “QY”.

310 other words, except for the first and last nodes, every local node is conditioned on the two most recent nodes, where  $j \rightarrow k$  is the newly performed operation resulting in event  $k$ , and  $j \rightarrow k+1$  accounts for the propagated delay from previous operations. In probabilistic reasoning, this means that the arrival (departure) time distribution of a train at each station is a function of two components: the process time distribution of all previous operations, and the predicted process time of the most recently performed operation. More explicitly, the former part casts the superposition (propagation) of delays and the latter term considers the chance of having a newly formed delay. In this representation, each node has only a small number of parameters to estimate, compared to the complex structure obtained by the heuristic hill-climbing method.

315  
320

To estimate the parameters while avoiding over fitting, the proposed BN structures were trained and evaluated using the  $k$ -fold cross-validation method. We compared the predictive performance of the proposed BN models using the gold standard cross-validation method with ten runs of 10-fold cross-validation.

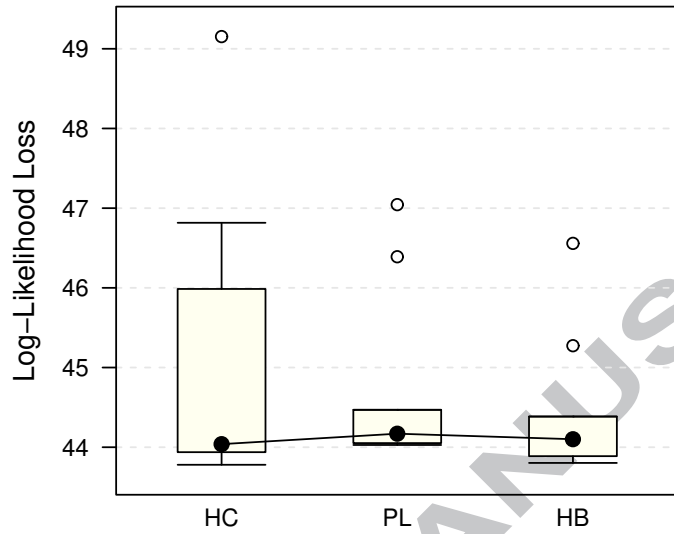


Figure 7: Ten runs of 10-fold cross validation of HC, PL and Hybrid BN structures; median points are connected.

325 The results are presented in Figure (7), where the Log-likelihood loss value  
 (average prediction error) of each BN structure is provided in box plots. The  
 value of the loss function for each BN structure shows that the PL structure is  
 not as good as the HB structure. Especially, the whiskers of the hybrid structure  
 are lower than the respective whiskers of the PL, as well as the outliers, meaning  
 330 that hybrid BN structure outperforms PL structure in terms of robustness, and  
 that HC is the worst of the three. Indeed, the results show that the Log-  
 likelihood loss of hybrid structure is about 44 with a standard deviation of 1.23.  
 Therefore, we selected the hybrid structure as the prediction model to evaluate  
 against the real-world benchmark data.

## 335 5. Performance Evaluation

To investigate how the predictions from the proposed model match with  
 the real-world benchmark data, we performed the following series of analyses.  
 Firstly, a comparison of the real and predicted arrival and departure delay

distributions for each station is provided in Figures (8) and (9). Moreover, in Figure (10) the scatter plots of the observed vs. the predicted values for arrival and departure delays are presented. From these results, we can see that the predictions match well with observations, both for arrival and departure events. Especially, in the interquartile range, the whiskers and the right tail match closely between these figures for each station. Moreover, as can be seen in Figure (10), the majority of predictions are close to the depicted diagonal lines for both arrival and departure events. This indicates that the predictions are satisfactorily close to the observed one.

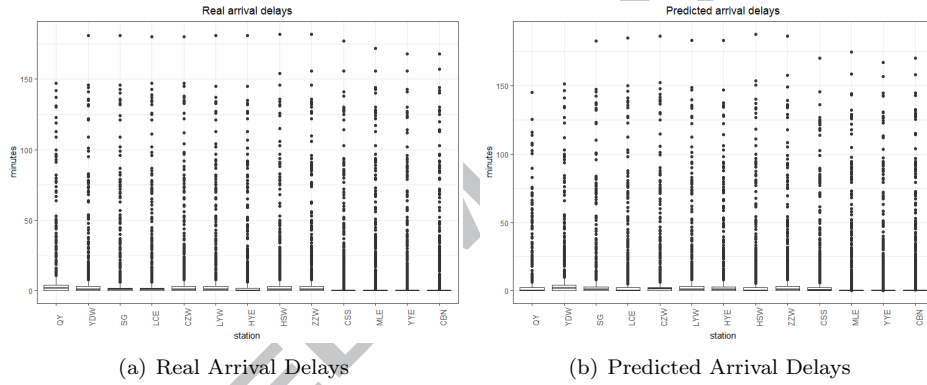


Figure 8: Comparison of predicted and observed arrival delay distribution for different stations.

Secondly, prediction errors are measured using three criteria, namely, the MAE, ME and RMSE as defined in Equations (2)-(4). These errors are calculated for predicted arrival and departure delays at each station  $s \in S$ .

$$MAE_s = \frac{\sum_{i=1}^n |p_i^s - o_i^s|}{n}, \quad (2)$$

$$RMSE_s = \frac{\sqrt{\sum_{i=1}^n (p_i^s - o_i^s)^2}}{n}, \quad (3)$$

$$ME_s = \frac{\sum_{i=1}^n p_i^s - o_i^s}{n}. \quad (4)$$

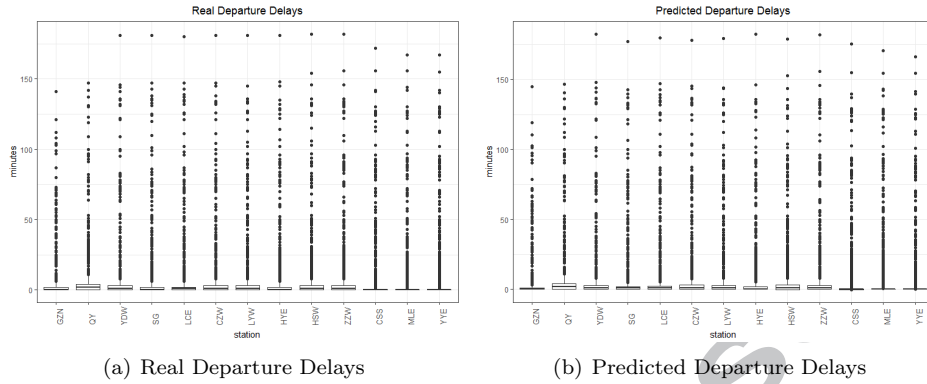


Figure 9: Comparison of predicted and observed departure delay distribution for different stations.

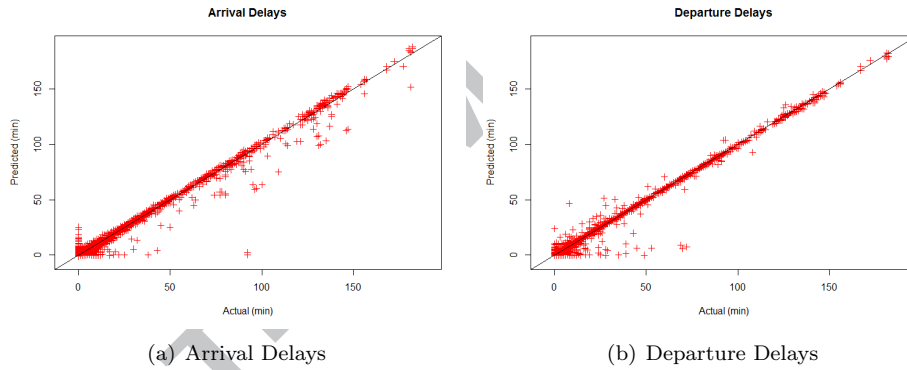
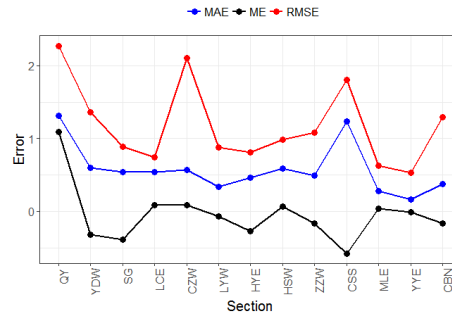


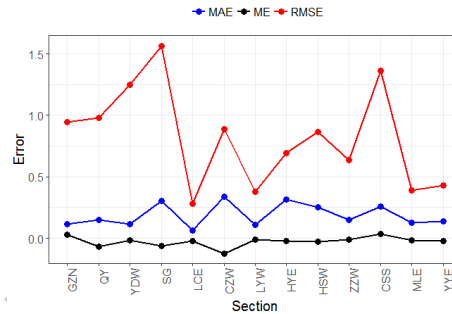
Figure 10: Scatter plots of actual vis-à-vis predicted arrival and departure delays.

where,  $p_i^s$ , and  $o_i^s$  are, respectively the predicted and observed delay values for  $i^{th}$  arrival or departure events at station  $s$ , and  $n$  is the total number of observations. These measures quantify the average deviation of the predictions from the observed values. The closer these measures are to zero, the better is the model's performance level. All three measures for the predicted arrival and departure delays for all of the stations are presented in Figure (11). We can see that the prediction errors are very low. For example, the mean absolute prediction error for all predicted events is around 30 seconds, while the maximum prediction absolute error is less than 90 seconds. Moreover, RMSE for both

360 predicted arrival and departure delays are less than two minutes, which are relatively larger than the MAE suggesting the existence of a few outlier prediction errors. These findings support the prediction power of the proposed model.



(a) Arrival Delays



(b) Departure Delays

Figure 11: (Color online) Magnitude of prediction errors in terms of  $MAE$ ,  $ME$  and  $RMSE$  for different stations.

In Section (4), we found that the average loss over ten runs of 10-fold cross validation is about 44%. In other words, the predictions from HB matches in 365 56% of the times with observations, which is not satisfactorily high for our purposes. This is, mostly, due to the fact that the prediction space is discrete. To overcome this problem, we employed discretization to transfer the continuous variables to bins (intervals) for prediction and inference purposes. By discretization, we can use quantitative and qualitative factors to measure the 370 model's predictability. We considered bins with three minutes width as the

prediction intervals. Our choice of three minutes intervals was obtained from domain knowledge that late arrivals of less than 90 seconds are not considered as delays. We note that as the width of intervals increases the predictions accuracy will increase, since each prediction will have a higher probability of falling  
 375 within the corresponding interval. However, this causes our predictive model to resemble a randomized prediction scheme. Next, each actual and estimated value of delay events, respectively, is assigned to the bin it lies within. The predicted bin is then compared with the true one obtained from the actual observation, using confusion (error) matrix measures for binary classification as  
 380 represented in Table (1).

Table 1: Schematic of a confusion matrix with bins for observed and predicted delays.

Bins	Predictions	
Observations	No. of True Positive (TP)	No. of False Negative (FN)
	No. of False Positive (FP)	No. of True Negative (TN)

$$Accuracy = \frac{No. \text{ of } TP \text{ Cases}}{Total \text{ No. of Cases}}, \quad (5)$$

$$Sensitivity = \frac{No. \text{ of } TP \text{ Cases}}{(No. \text{ of } TP \text{ Cases}) + (No. \text{ of } FN \text{ Cases})}, \quad (6)$$

$$Specificity = \frac{No. \text{ of } TN \text{ Cases}}{(No. \text{ of } FP \text{ Cases}) + (No. \text{ of } TN \text{ Cases})}. \quad (7)$$

The confusion matrix provides different performance measures, such as Accuracy, Sensitivity and Specificity for evaluating the predictions (Sokolova & Lapalme, 2009). We restate the definitions of these measures in Equations (5)-  
 385 (7) as they are applied to our work. The results of Accuracy, Sensitivity and Specificity measures are depicted in Figure (12). The Accuracy measure indicates the overall effectiveness of predictions, that is, whether the BN model passes the minimum requirements. We obtained overall accuracy to be over

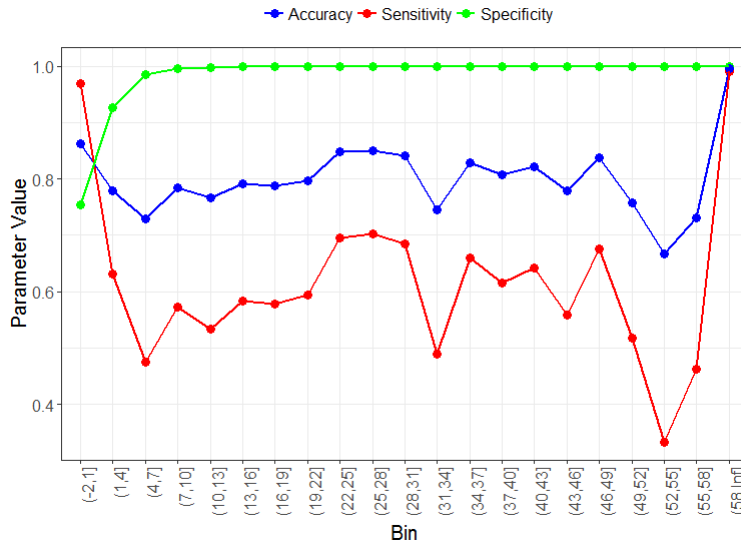


Figure 12: (Color online) Prediction power in terms of Accuracy, Sensitivity and Specificity.

80%, with a no information rate of 58%. The overall accuracy needs to be higher  
 390 than the no-information rate for the model, and this occurs for the proposed  
 model. Sensitivity (True Positive Rate, or Recall) measures the proportion of  
 cases that are correctly identified, which in our case is the percentage of bins  
 that are correctly identified; it is more than 60% on average over all stations.  
 Finally, Specificity (True Negative Rate) represents how effectively the proposed  
 395 BN model avoids the wrong predictions. An alternative to no information rate  
 is the Kappa statistic, which is illustrated through the use of a confusion ma-  
 trix. This statistic shows the overall agreement between the observed accuracy  
 against an expected accuracy (random chance), which generally means it is less  
 misleading than simply using accuracy as a metric. This statistic takes values  
 400 between -1 and 1. An absolute Kappa value of 1 shows a complete agreement  
 while a value of zero shows complete disagreement. Kappa statistics higher than  
 30% are considered acceptable. For the presented BN, we found the Kappa value  
 to be 69%, which confirms that the prediction strength of the proposed model  
 is substantial (Landis & Koch, 1977).

405 A prediction instance is shown in Figure (13). The presented time-space

diagram shows the planned (grey) paths, the predicted train paths (orange), the realized (green) train paths in space and time. As can be seen in the first half of this figure (on the left side), all the predictions are close to the respective actual observed operations until the CZW station, after which the prediction error starts to grow. This is because of the error accumulation which could be addressed fairly easily in real-world operation as predictions could be updated using real-time data (e.g., arrival time at the preceding station, the position of the corresponding train along the track and the adjusted timetable).

We finally turn to the computational aspect of the proposed prediction model. Overall, the computational time used for training and testing of the proposed model did not exceed ten minutes, which is not computationally time intensive.

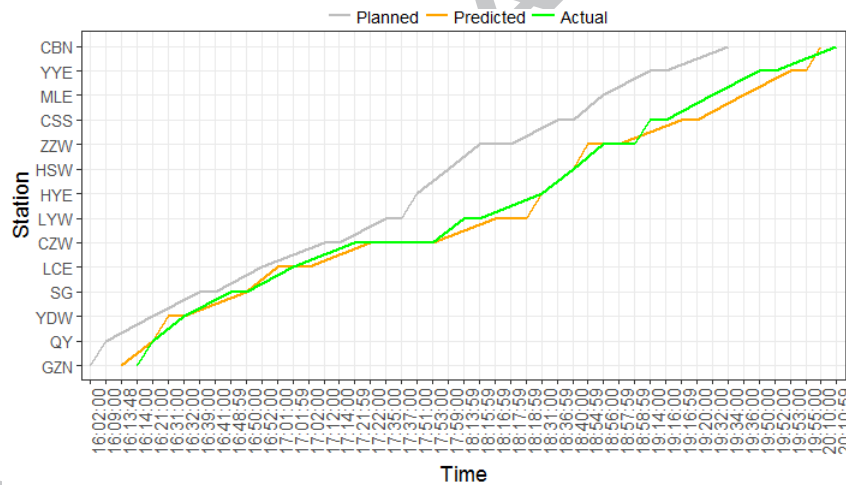


Figure 13: (Color online) Time-distance diagram of planned, predicted and realized train paths.

## 6. Concluding Remarks

The research presented in this paper employs Bayesian reasoning to construct a delay prediction model for train operations. Three different candidate



BN structures were trained and tested using the gold standard  $k$ -fold cross validation method against historical train operation data. The results indicated that a hybrid heuristic BN structure, built upon naive and heuristic structures and refined by domain knowledge and experts judgments, can achieve a higher prediction performance, compared to other structures. Using real-world benchmark data, different performance comparison measures indicated that the hybrid BN structure performs satisfactorily in predicting train delays. Indeed, the proposed model was shown to have significant performance in terms of accuracy, sensitivity and specificity measures. Specifically, the proposed model can achieve over 80% accuracy for a 60-minute prediction horizon.

The proposed BN model distinguishes between two different delay elements in the train operations, namely, propagated delay and a possible delay in the current operation. This property is expected to be important to ensure that the model is able to be generalized well from the training data of the specific HSR line to any data from other HSR lines. Moreover, the proposed model has two main advantages: (a) it is simple, which makes it interpretable and computationally efficient, and (b) it incorporates the interrelationships of causal factors and superposition of arrival and departure delay components. These properties allow the proposed model to incorporate various specific variables, such as online traffic condition parameters, causes of delay, and quantitatively compute and capture route conflicts. It is expected that the prediction error could be reduced if the spatiotemporal properties of each track section are also included in the prediction model.

Our future research includes extensions of the proposed model to integrate online traffic data to support delay management and passenger information systems by providing predictions about delays. This research aims to provide a tool that runs online contingency scenarios. To create such a contingency tool, we need decision support tools that can learn over time with new operation data, and then inform the affected passengers and propose optimum dispatching decisions in case of necessary changes due to delays. The HB reported here is central to that aim.

## References

- Berger, A., Gebhardt, A., Müller-Hannemann, M., & Ostrowski, M. (2011).  
455 Stochastic delay prediction in large train networks. In *OASICs-OpenAccess Series in Informatics*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik volume 20.
- Buntine, W. (1996). A guide to the literature on learning probabilistic networks from data. *IEEE Transactions on knowledge and data engineering*, 8, 195–  
460 210.
- China-Railway-Corporation (2016). Punctuality reaches 98.8%. URL: <http://news.gaotie.cn/yunying/2016-07-23/336706.html>.
- Darwiche, A. (2009). *Modeling and reasoning with Bayesian networks*. Cambridge University Press.
- 465 Dollevoet, T., Huisman, D., Kroon, L., Schmidt, M., & Schöbel, A. (2014). Delay management including capacities of stations. *Transportation Science*, 49, 185–203.
- Flier, H., Graffagnino, T., & Nunkesser, M. (2009). Finding robust train paths in dense corridors. In *3rd International Seminar on Railway Operations Modelling and Analysis*.  
470
- Hansen, I. A., Goverde, R. M., & van der Meer, D. J. (2010). Online train delay recognition and running time prediction. In *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on* (pp. 1783–1788). IEEE.
- 475 Kecman, P. (2014). *Models for Predictive Railway Traffic Management*. Ph.D. thesis TU Delft, Delft University of Technology.
- Kecman, P., Corman, F., & Meng, L. (2015a). Train delay evolution as a stochastic process. In *6th International Conference on Railway Operations Modelling and Analysis-RailTokyo2015*.

- 480 Kecman, P., Corman, F., Peterson, A., & Joborn, M. (2015b). Stochastic prediction of train delays in real-time using bayesian networks. In *Conference on Advanced Systems in Public Transport*.
- Kecman, P., & Goverde, R. M. (2015a). Online data-driven adaptive prediction of train event times. *IEEE Transactions on Intelligent Transportation Systems*, *16*, 465–474.
- 485
- Kecman, P., & Goverde, R. M. (2015b). Predictive modelling of running and dwell times in railway traffic. *Public Transport*, *7*, 295–319.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, (pp. 159–174).
- 490 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.
- Martin, L. J. (2016). Predictive reasoning and machine learning for the enhancement of reliability in railway systems. In *International Conference on Reliability, Safety and Security of Railway Systems* (pp. 178–188). Springer.
- 495
- Meester, L. E., & Muns, S. (2007). Stochastic delay propagation in railway networks and phase-type distributions. *Transportation Research Part B: Methodological*, *41*, 218–230.
- Nagarajan, R., Scutari, M., & Lèbre, S. (2013). Bayesian networks in r.
- 500 *Springer*, *122*, 125–127.
- Nielsen, T. D., & Jensen, F. V. (2009). *Bayesian networks and decision graphs*. Springer Science & Business Media.
- Office, G. B. N. A. (2008). *Reducing passenger rail delays by better management of incidents*. HC (Series) (Great Britain. Parliament. House of Commons). Stationery Office. URL: <https://books.google.ca/books?id=t26vGUGL89EC>.
- 505

- Peters, J., Emig, B., Jung, M., & Schmidt, S. (2005). Prediction of delays in public transportation using neural networks. In *Computational Intelligence for Modelling, Control and Automation, 2005 and International Conference on Intelligent Agents, Web Technologies and Internet Commerce, International Conference on* (pp. 92–97). IEEE volume 2.
- 510
- Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45, 427–437.
- 515 Van Oort, N. (2011). *Service Reliability and Urban Public Transport Design Service Reliability*. Ph.D. thesis Ph. D. thesis, TRAIL Research School.
- 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*, (pp. 1–20).
- 520
- Yaghini, M., Khoshraftar, M. M., & Seyedabadi, M. (2013). Railway passenger train delay prediction via neural network model. *Journal of advanced transportation*, 47, 355–368.
- Yuan, J. (2006). *Stochastic modelling of train delays and delay propagation in stations* volume 2006. Eburon Uitgeverij BV.
- 525
- Yuan, J., Goverde, R., & Hansen, I. (2002). Propagation of train delays in stations. *WIT Transactions on The Built Environment*, 61.
- Zhang, H., Li, S., & Yang, L. (2018). Real-time optimal train regulation design for metro lines with energy-saving. *Computers & Industrial Engineering*, .

**Acknowledgement**

This work was supported by the National Nature Science Foundation of China [grant number 61503311], National Key R&D Plan of China [grant number 2017YFB1200701] and NSERC (National Sciences and Engineering Research Council of Canada). We acknowledge the support of the Railways Technology Development Plan of China Railway Corporation [grant number 2016X008-J]. Parts of the work were supported by State Key Lab of Railway Control and Safety Open Topics Fund [grant number RCS2017K008]. We are grateful for the contributions made by our project partners.

ACCEPTED MANUSCRIPT

**A Hybrid Bayesian Network Model for Predicting Delays in Train Operations**

## Research Highlights

- Three different Bayesian network structures are introduced to tackle the complexity and superposition nature of delays in train operations.
- Train operation data is used to learn the BN structures under the golden standard  $k$ -fold cross validation method.
- The outperforming heuristic-based structure is elaborated as a hybrid BN delay prediction model.
- The hybrid BN model is tested against real-world data using different performance measures.