Transportation Research Part C 68 (2016) 350–368

Contents lists available at ScienceDirect

Transportation Research Part C

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

Modeling railway disruption lengths with Copula Bayesian Networks

Aurelius A. Zilko [a,*], Dorota Kurowicka [a], Rob M.P. Goverde [b]

[a] Delft Institute of Applied Mathematics, Delft University of Technology, The Netherlands
[b] Department of Transport and Planning, Delft University of Technology, The Netherlands

ARTICLE INFO

Article history:
Received 6 October 2015
Received in revised form 22 April 2016
Accepted 25 April 2016

Keywords:
Railway disruption
Prediction
Dependence model

ABSTRACT

Decreasing the uncertainty in the lengths of railway disruptions is a major help to disruption management. To assist the Dutch Operational Control Center Rail (OCCR) during disruptions, we propose the Copula Bayesian Network method to construct a disruption length prediction model. Computational efficiency and fast inference features make the method attractive for the OCCR's real-time decision making environment. The method considers the factors influencing the length of a disruption and models the dependence between them to produce a prediction. As an illustration, a model for track circuit (TC) disruptions in the Dutch railway network is presented in this paper. Factors influencing the TC disruption length are considered and a disruption length model is constructed. We show that the resulting model's prediction power is sound and discuss its real-life use and challenges to be tackled in practice.

1. Introduction

It is unavoidable that the operation of a railway system encounters unexpected incidents which disturb or disrupt the railway timetable. Depending on the length of the incident, different measures need to be taken to handle the situation during the down-time. Shorter incidents, usually referred to as *disturbances*, may require only timetable adjustment while longer incidents, usually referred to as *disruptions*, may additionally require rolling stock and crew adjustment. In this paper, we do not distinguish the difference between disturbance and disruption so the term *disruption* is used when referring to these unexpected incidents, regardless of the length.

The topic of disruption management has been a growing research area in the railway operations research. A vast number of different algorithms and models have been proposed for the recovery from the disrupted situation. Cacchiani et al. (2014) provides an overview of these proposed mathematical algorithms and models where some of the presented works mention the uncertainty nature of the disruption length. Given the information about disruption length, the algorithms search an optimal solution to recover from the disrupted situation in the form of timetable rescheduling, rolling stock rescheduling, or crew rescheduling.

However, in reality, disruption length is very uncertain and it is difficult to tell exactly how long a disruption will last. In the Netherlands, this uncertainty creates a big problem that the Operational Control Center Rail (OCCR) in Utrecht faces when a disruption occurs. During this period, train traffic is hindered and timetable is not followed anymore. This is also illustrated as a bathtub model: the normal traffic intensity goes down due to a disruption, stays at a lower level following

* Corresponding author.
E-mail address: A.A.Zilko@tudelft.nl (A.A. Zilko).

http://dx.doi.org/10.1016/j.trc.2016.04.018

some contingency plan, and recovers to the normal situation after the disruption has ended (Ghaemi and Goverde, 2015). The OCCR's main duty is to handle the situation during this period with the goal of recovering the traffic to normal as soon as possible. The disruption length is crucial piece of information that the OCCR needs. With this information, they make decisions regarding the actions they need to take in order to achieve the goal.

In practice, a series of updated disruption length predictions are made as more information about the disruption is gathered. When a disruption occurs, firstly a rough prediction based on history is made. This is called the "P1" prediction and is taken to be the average of a given disruption length in the past. In the meantime, the mechanics are informed about the disruption and are tasked to repair the problem. From now on, the OCCR is in close communication with the mechanics. After arriving at the site, the mechanics have 15 min to diagnose the problem after which they are required to make a prediction, based on their own judgment, regarding the repair time. This mechanics' prediction is called the "P2" prediction. The mechanics are allowed to update the prediction later on and this updated prediction is called the "P2a" prediction. Finally, when a final prediction can be made, the mechanics are required to update the OCCR with the so-called "P3" prediction. Upon completion of the work, the mechanics inform the OCCR that the problem is solved and the disrupted train traffic at the troubled section can be resumed. Then, they are required to record the information about the disruption with an administrative form that will be stored in a data base called the SAP data base. Unfortunately, this procedure does not include recording the mechanics' own predictions. Moreover, information about the cause of failures is also not required to be in a structured format in the database. The data that is used in this paper comes from this SAP data base.

Thus in current practice, the uncertainty in disruption length is handled by means of a series of predictions based on the mechanics' expertise and judgment. One other way to tackle the uncertain disruption length problem is by representing the disruption length with a probability distribution. Having such distribution allows us to generate random samples of disruption length. This approach is relatively new in railway operation but has been used in several earlier studies in highway traffic engineering. For instance, Golob et al. (1986), Giuliano (1989) and Sullivan (1997) use the lognormal distribution and Nam and Mannering (2000) uses the Weibull distribution. In the railway operation field, Meng and Zhou (2011) models the disruption length in a single track rail line in China with the Normal distribution, while Schranil and Weidmann (2013) models the railway disruption length in Switzerland with the exponential distribution.

In this paper, disruption length is the center of attention. The uncertainty of disruption length is going to be modeled with a probability distribution from historical data. Moreover, several influencing factors of disruption length are considered. The goal is to construct a dependence model between the disruption length and these influencing factors. When a disruption occurs, the model is conditionalized on the realization of the influencing factors resulting in a conditionalized disruption length distribution. This conditionalized distribution represents the disruption length specialized to a specific situation. Then, a disruption length prediction is made from this distribution.

To do this, a proposal made by Zilko et al. (2014) is followed where the dependence model is constructed using the Copula Bayesian Network. The Bayesian Network (BN) technique has been used in transportation research field for several different studies. For instance, Gregoriades and Mouskos (2013) quantify accident risk in road traffic in Cyprus and Chen et al. (2015) constructs a dependence model of travelers' preference for toll road utilization in Texas with BNs. In the railway field, Oukhellou et al. (2008) use the technique to perform broken rail diagnosis.

BN modeling consists of two parts: the graphical structure that represents (conditional) independence in the model and the conditional probabilities between the variables to specify the rest of the relationships. There are different types of BNs depending on how the conditional probabilities are modeled. When all the variables are discrete, as in the case of the three studies mentioned above, the conditional probabilities are modeled using conditional probability tables (CPT). In the case of continuous variables in Gregoriades and Mouskos (2013), the variables are discretized to obtain a fully discrete model. When all the variables are continuous and Normally distributed, the Normal Bayesian Network (which uses the multivariate Normal distribution) can be used.

In this paper, the conditional probabilities between the variables are represented with copula (more on copula in Section 4). Copula is a very useful model for the dependence between continuous variables as this allows the separation of marginal distributions and the dependence. We are interested in the use of copula because our variable of interest, disruption length, is continuous. The use of copula is not completely foreign in transportation research. Srinivas et al. (2006) use several different copula families to model the dependence between vehicle axle weights. Wan and Kornhauser (1997) construct a copula-based model to predict the travel time which is used in a routing decision making problem. Ng and Lo (2013) model the air quality conformity in a transportation networks with copulas.

The disruption length model that we build will assist the OCCR by updating the uncertainty of disruption length every time new information about the disruption is available, in a similar manner with how the "P1", "P2", and "P3" predictions help the OCCR. Moreover, because the output of the model is a probability distribution function, this also gives the OCCR full control regarding which value they want to take as a prediction. Do they want to be more conservative by choosing a value in the upper quantile of the distribution? Or do they want to be more optimistic by choosing a value in the lower quantile of the distribution?

A Copula Bayesian Network model needs to be constructed for each disruption type. To help the model construction, a user-friendly and computationally-efficient software called `UNINET` which implements the algorithm of the Copula Bayesian Network will be used. This software was developed at Delft University of Technology and is available at www.lighttwist.net/wp/uninet. As an illustration in this paper, we construct a disruption length model for disruptions caused by train detection problems in the Netherlands, specifically the GRS track circuit (TC) failures.

First, to construct the model, it is necessary to divide the disruption length into two mutually exclusive definitions of time: the *latency time* and the *repair time*. The latency time is the length of time the mechanics need to get to the disrupted site. The repair time is the length of time they need to repair the problem.

A partial model on the latency time has been constructed in Zilko et al. (2015). Zilko and Kurowicka (in press) considers a more complex copula model to represent the mixed discrete–continuous data for the latency time model using another graphical structure called "vines". It is shown that even though the more complex model recovers the data better, the simpler Copula Bayesian Networks model still performs as a statistically sound model for the latency time prediction. For these reasons, in this paper we construct the full disruption length model with a Copula Bayesian Network.

Therefore, the contributions of this paper are threefold: (1) A full disruption model involving both the latency and repair time for TC failures is completed. An issue about the non-informativeness of the cause of failures in the data set is also tackled; (2) The model is validated with a test set and it is shown that it predicts the disruption length time well; and (3) It is shown how the model can be used in a real-life application and how it can help the OCCR during disruptions. How the model plays a role in a planned collaboration with a parallel research on efficient and effective train dispatching during disruption time is also described.

The rest of the paper is organized as follows. Section 2 provides a brief overview of the track circuit system. It is followed by Section 3 where data analysis based on the characteristics of the TC system is performed to find out the necessary influencing factors. The factors are gathered together and the model is constructed in Section 4 where discussion about the algorithm is also provided. We also validate the model in this section. Section 5 discusses the model use in practice via an example and its practical challenges. The paper is summarized with Section 6 where the conclusions and future works are presented.

2. The track circuit

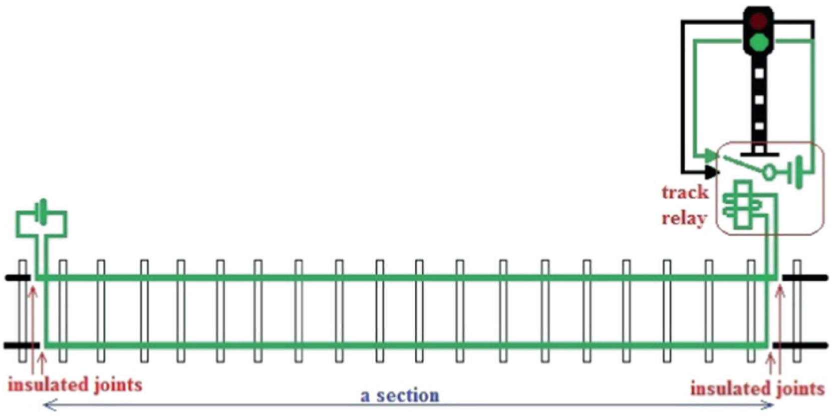
A TC is a train detection device which is part of the train safety system. Its task is to detect whether a block or section of railway tracks is occupied or not.

A TC has an electric current source at one end of the section and a detection device at the other. The current flows along the section through the rails. Sections are separated with joint insulators which keep the current flowing within their corresponding section. A track relay acts as the detection device. The track relay is in a pick up position when the section is clear and drops when the section is occupied. In this case, the axles of the train produce a short circuit between the two rails; hence the detection device does not receive any current and the section is detected as occupied (Pachl, 2004). Fig. 1 illustrates how the track circuit system works.

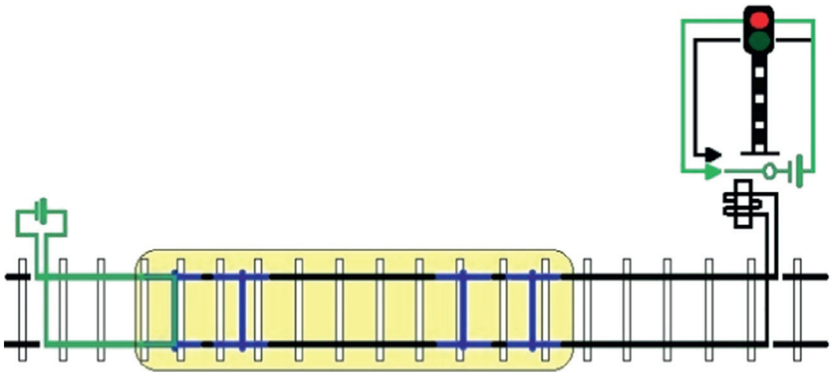
A TC system consists of many different electrical components. Failure of any of these components leads to an unjustified occupation of the corresponding section, and thus to a disrupted train traffic. However, sometimes an external reason also leads to the same problem. The following are the main components of a TC and some external factors which can cause problems (Visser and Steenkamp, 1981).

1. **The joint insulator**. A joint insulator separates the current in two consecutive sections of railway with an insulator made from nylon plates and linings (Fig. 2a). Two causes of joint insulator failure are considered:
   (a) Coins. Sometimes conductive agent, most frequently coins, is deliberately put on the joint insulator hence allowing electric current of one section to flow to the next via this agent. To solve the problem, the mechanics only need to remove the conductive material from the joint insulator.
   (b) Splinter/grinding chips and insulator problem. Other foreign conductive object, such as the splinter/grinding chips, may also fall accidentally on the joint insulator. To solve the problem, the mechanics need to clean the joint insulator from these pollutants. While doing so, sometimes the mechanics also need to replace the nylon plates and linings to renew the insulating function of the insulator.
2. **The relay cabinet**. A relay cabinet contains a few electric components, such as: B2 Vane relay, CR/VTB relay, capacitor, transformer, fuse, and wiring. Failure of any of these components requires the mechanics to replace the defect component with a new one (Fig. 2b).
3. **Arrestor**. An arrestor protects the TC from high voltage and creates a safe current path when the catenary[1] is disrupted and falls on the track. A faulty arrestor is replaced with a new one.
4. **Cable**. The problem with cable arises when the cable itself is broken, its insulation is damaged, or when the cable is stolen due to its valuable copper material. Either way, the faulty cable is replaced.
5. **The track-side electrical junction box**. The track-side electrical junction box connects the heavier track-side cable from the track circuit and the long and thinner cable to the relay cabinet. When it fails, it is replaced by a new one.
6. **The impedance bond**. The impedance bond separates the track circuit's alternating current from the train traction's direct current. When there is problem with an impedance bond, the mechanics need to replace it with a new one.
7. **External reasons**. Some exteral reasons which quite commonly occur in a TC problem:

[1] An overhead wire used to transmit electrical energy to trains.



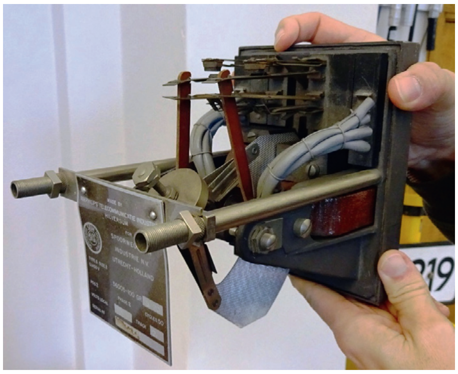
(a) An unoccupied block or section.



(b) An occupied block or section.

Fig. 1. How the track circuit system works.

(a) An insulated joint at Utrecht Centraal Station.



(b) A B2 Vane Relay.

Fig. 2. Two components of the TC.

(a) *Wrong operation by mechanics*. Sometimes, a mechanic makes an accidental mistake that leads to an unjustified section occupation. There is no one specific action that needs to be performed to solve the problem because it depends on what mistake that has been made.

(b) *Adjustment problem*. Sometimes none of the components fail but the TC setting is not correct, most of the time due to high temperature or too low ballast resistance. In this case, the mechanics need to correct the setting by readjusting it.
(c) *Short circuit*. This is the problem when an object causes a short-circuit between the rail and the soil. The type of action needed depends on the damage degree resulted from the short-circuit.

3. Data analysis

The same data set as in Zilko et al. (2015) is used in this paper. The data contains the historical TC problems in the Dutch railway network between 1 January 2011 and 30 June 2013. Considering only incidents which require urgent actions, we are left with 1920 sample points. The latency and repair time are affected by different factors that will be discussed in this section.

3.1. Factors influencing the latency time

The factors influencing the latency time can be categorized into three groups: time, location, and the weather. In the analysis, an extra factor which does not belong to any of these groups also surfaces, the presence of an overlapping disruption.

3.1.1. Time

Three variables are initially thought to represent time: (1) whether or not the disruption occurs during the mechanics' contractual working hours (weekday between 7 AM and 4 PM), (2) whether or not the disruption occurs during the rush hour (weekday between 7 and 10 AM and between 4 and 7 PM), and (3) whether or not the disruption occurs during the weekend. However, information about weekend is already contained in the first two variables. For instance, when an incident occurs during the weekend, it is outside the mechanics' working hours and not during rush hour. Moreover, the characteristic of weekend is similar with the characteristics of non-working hours and non-rush hour time. Therefore, only the first two variables are considered.

The first factor is important because different operations are performed depending whether the incident occurs during the contractual working hours or not. During the working hours, the mechanics leave from their working station to the disruption site. Otherwise, they are not in their station even though they are available on call. In this case, they leave from wherever they are to the disruption site. It turns out that the latency time outside working hours is longer than during working hours. On average, the latency time is around 4.5 min longer when an incident occurs not during the working hours. Fig. 3 shows the empirical distribution of latency time during and outside working hours where difference is visible between the two distribution. Performing the two-sample Kolmogorov–Smirnov (KS) test to the two latency time distributions in Fig. 3 yields a p-value of 9.1080e−07. Similar conclusion is also drawn when the two-sample Cramér-von Mises (CvM) test (Anderson, 1962) is used where the p-value is very close to zero ($\sim$0). This indicates that the observed difference between the two latency time distributions cannot be explained by the sample noise on significance level of 0.05.

To get to the disrupted site, the mechanics travel by cars. During this time, it is more likely to encounter traffic jams which might prolong the latency time. The influence of rush hour to the latency time is smaller than of working hours, where the latency time during rush hour is slightly longer than not during rush hour. The p-value of the two-sample KS test and CvM test in this case are 0.0402 and 0.0332, respectively. On average, the latency time is around 0.5 min longer during rush hour.

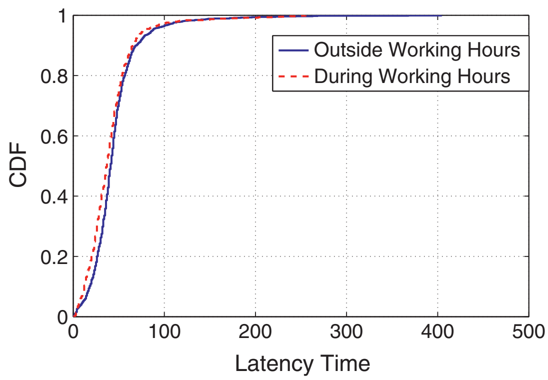


Fig. 3. The empirical distribution of latency time during and outside working hours.

3.1.2. Location

Location needs to be described by some representative properties which affect the latency time length. To help determining which properties need to be examined, first a map of disruptions with the 5% longest latency time is plotted and is shown in Fig. 4. This map corresponds to the disruptions with latency time longer than 90 min.

The map shows that many of the long latency time occurred in the Randstad area in the west of the Netherlands. Being the busiest area in the country, there are more mechanics' working stations situated in the Randstad. Moreover, the road and train traffic density are also higher. The road traffic density has been covered by the variable Rush Hour above. Another characteristic that is under consideration is the level crossing. In the Randstad, there is less level crossing on the railway track than in other parts of the country.

Based on the above observation, Zilko et al. (2014) start the investigation by examining two continuous properties of location: (1) distance to the nearest mechanics' working station and (2) the train traffic density. The first variable is intuitive because the farther the disruption is, the longer it takes for the mechanics to reach it. The second variable represents how busy the location is. For instance, a location with denser traffic indicates more proximity to bigger cities where better infrastructure (e.g. availability of roads or easy access) is available. However, in a follow-up analysis, train traffic density turns out to not affect the latency time time significantly. Moreover, the variable still does not affect the latency time significantly in combination with other variables as well. As a result, this variable is dropped from the model. A third variable which represents the location's accessibility is introduced: (3) distance to the nearest level crossing. To access the disruption site, the mechanics need to park their cars (usually at the nearest level crossing) and walk to the site. For this reason, this variable becomes interesting. Moreover, another property of location that is a binary discrete variable, (4) contract type, is considered as well. Why this variable is seen as a property of location is going to be explained shortly.

In the data set, the disruption site is indicated as between two operational points along the tracks where the latitude and longitude coordinates are known. The disruption site's coordinate is estimated by taking the average of the coordinates of these two operational points. The towns or cities where the mechanics' stations are located are also known. Their exact locations in the towns or cities are approximated by the location of the main railway stations in the corresponding towns or cities. In the Netherlands, the repair work is outsourced to four major contractors companies where each company is responsible for its own region. This information is important to be considered because, for instance, a disruption in the region of contractor *A* is handled by mechanics from contractor *A* even if a working station of contractor *B* is physically closer.

Distance to the nearest mechanics' working station is approximated by calculating the straight line distance between the coordinates of the approximated disruption site and the coordinates of the nearest responsible contractors' working station. The rank correlation between the distance to the nearest working station and the latency time is 0.1380 with 95% confidence bound of $(0.0925, 0.1829)$. Zero is not in the confidence bound hence this indicates small positive dependence between the two variables.

The rank correlation between the latency time and the distance to the nearest level crossing is calculated to be 0.0842 with 95% condifence bound of $(0.0383, 0.1297)$. Again, zero is not in the confidence bound hence this indicates small positive dependence between the two variables.

Currently in the Netherlands, there are two types of contract between ProRail, the organization responsible for the railway infrastructure, and the contractors. The old OPC contract (*Output-procescontracten* in Dutch, translated as the Output-based contract) is based on the amount of work the contractors perform while the new PGO contract (*Prestatiegericht Onderhoud* in Dutch, translated as the Performance-based maintenance) introduces a penalty if the work takes too long. The latency time with an OPC is longer than with a PGO contract, with on average 3.3 min longer. Performing the two-sample KS

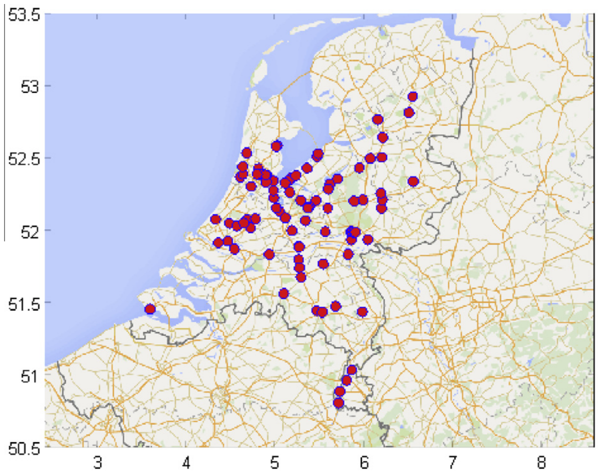


Fig. 4. The map of TC disruptions with latency time longer than 90 min.

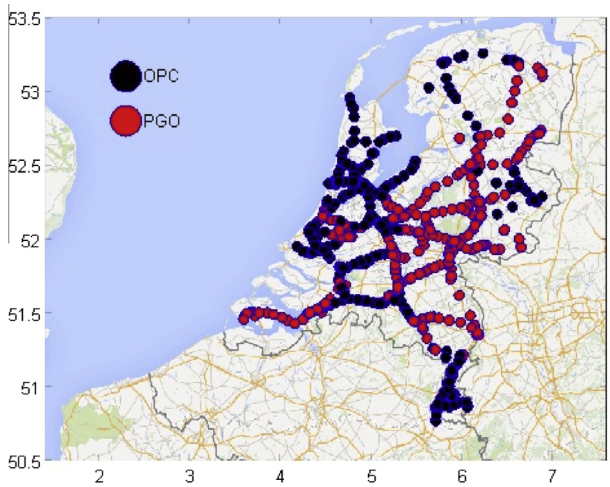


Fig. 5. The spread of the OPC and PGO contract in the Netherlands.

test and CvM test to the latency time of disruptions with OPC and PGO contract yields p-values of 0.0007 and $\sim$0, respectively.

Fig. 5 shows why contract type represents location. It is clear that certain regions in the Netherlands are under the OPC contract while the other regions are under the PGO contract. Interestingly, the Randstad is still with the old OPC contract but this explains why more disruptions with long latency time are observed in this region in Fig. 4. Moreover, this variable also explains the observed long latency time in the province of Limburg in southeast Netherlands.

3.1.3. Weather and overlapping disruptions

As mentioned in Section 2, TC is sensitive to high temperature. In our model, temperature is represented by a binary variable "Warm" with 0 corresponding the temperature below 25 °C and 1 otherwise. The temperature threshold of 25 °C is chosen because this is the most optimal threshold for our data where the strongest effect of the variable Warm on the latency time is observed. Higher threshold leads to lower number of observable warm samples,[2] hence less reliable data analysis, while lower threshold is meaningless from the variable definition's point of view.

This variable, however, does not appear to have direct influence on the latency time. The p-values of the two-sample KS test and CvM test between the latency time when Warm is zero and Warm is one are, respectively, 0.0938 and 0.0727. This indicates that the observed difference between the two latency time distributions can be explained by the sample noise on significance level of 0.05.

However, Warm affects another variable, the presence of an overlapping incident. Two incidents in the data are considered overlapping if their disruption time are overlapping, they are of similar technical problem, and handled by the same contractor. When Warm is zero, the proportion of incidents with an overlapping other incident is 4.15% while when Warm is one, the proportion is higher at 16.07%. This is understandable because high temperature may trigger some TCs to fail (almost) simulatenously. With a limited number of mechanics, some of these incidents can only be handled after the other have been taken care of.

Whether or not there is an overlapping incident affects the latency time. On average, the latency time is around 14.5 min longer when an overlapping incident exists. For this case, the two-sample KS test and CvM test yield p-values of 0.0025 and 0.0056. Fig. 6 presents the latency time distribution when an overlapping disruption exists or not.

3.2. Factors influencing the repair time

3.2.1. Contract type

Contract type is also an influencing factor of repair time. The repair time is significantly longer when the contract type is OPC. On average, the repair time of an incident with OPC contract type is 20.5 min longer than an incident with PGO contract type. Performing the two-sample KS test and CvM test to the two repair time distributions based on contract types yields p-values of 9.9971e−14 and $\sim$0, respectively. Fig. 7 presents the empirical distribution of repair time given the two contract types.

[2] With the threshold of 25 °C, there are only 112 occurrences in the data where the temperature is warm.

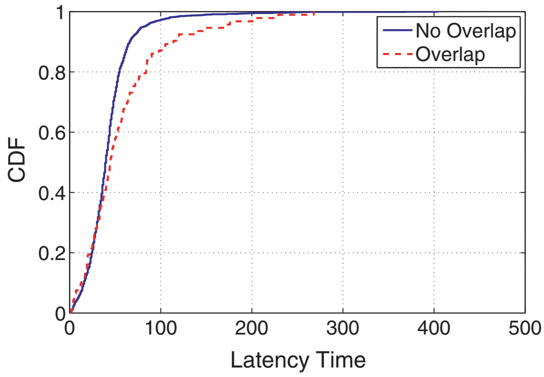


Fig. 6. The empirical distribution of latency time with respect to the presence of an overlapping incident.

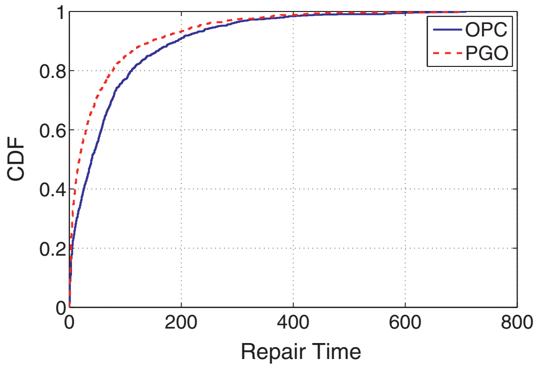


Fig. 7. The empirical distribution of repair time between TC with OPC and PGO contracts.

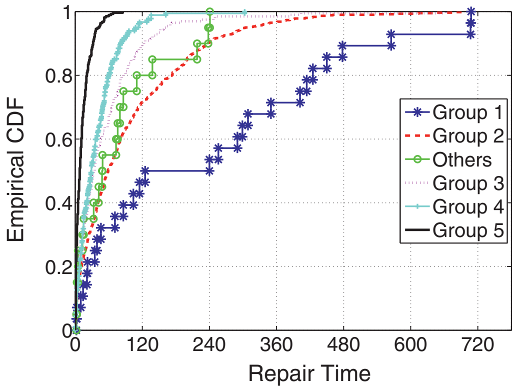
3.2.2. The cause

The type of components failure and the external causes of problems discussed in Section 2 also influence the repair time length. Different failure types lead to different action that needs to be taken, thus different repair time. Based on the characteristics of the required action, the causes are grouped as follows:

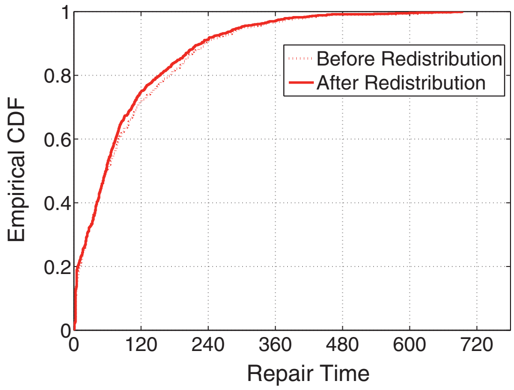
1. Group 1: impedance bond failure.
2. Group 2: the relay cabinet failure, cable problem, track-side electrical junction box problem, and arrestor problem.
3. Group 3: external reasons.
4. Group 4: splinter/grinding chips and insulator problem.
5. Group 5: coins.
6. Others.

The first five groups are ordered based on the length of repair time. Group 1 and 2 require the mechanics to perform replacement job to solve the problem. Impedance bond is separated from the others in Group 2 because usually the mechanics do not carry the necessary tools with them to perform the replacement job needed with an impedance bond failure. For other causes which require replacement, most of the time the mechanics already have the needed tools with them. Consequently, the repair time of impedance bond failure tends to be longer than of the causes in Group 2. The causes in Group 3 mainly require the mechanics to adjust the involved components. Group 4 and Group 5 contain problems with the joint insulator. Two groups are created because the two problems require slightly different type of action. Problems in Group 4 require the mechanics to clean the joint insulator and, sometimes, renew the nylon plates and linings of the insulator. Due to the small size of the joint insulator, the work does not take a lot of time compared to the replacement action for Group 2. The coins problem in Group 5 solely requires the action of coins removal from the troubled joint insulator.

There are also a few TC problems that are caused by other problems that do not belong to any of the aforementioned causes. These incidents occur very infrequently with a total of 20 incidents in the data. Two examples of TC incidents in this group: a problem where the track and relay cabinet are flooded with water (the mechanics need to clean and pump the water out) and the loss of detection due to sand.



(a) The empirical repair time distribution of all six groups before redistribution



(b) The empirical repair time distribution of Group 2 before and after redistribution

Fig. 8. The empirical distribution of repair time.

The characteristic of the repair time length of the groups is also reflected in the data. Fig. 8(a) shows the empirical distribution of the six groups. Clearly, the group Others is located between Group 2 and Group 3 in term of the length. Moreover, the difference in repair time between different groups is evident.

In our data set, the cause of approximately 30% of the TC incidents is unknown due to unclear or too brief remarks. To tackle this situation, one option would be to discard these "unknowns" from the data set but this would result in the loss of information they carry for the other variables. Another option would be to redistribute them randomly into the six groups based on the group's proportion. However, this approach neglects the repair time length information that we want to associate with the groups.

We propose to perform a redistribution approach which takes into account the dependence between repair time, cause, and contract type. This approach is also known as the "Bayesian Classifier" and is a popular classification technique which has been implemented in many studies, e.g. Marchant and Onyango (2002), Bender et al. (2004), and Wang et al. (2007). With this technique, the probability of each unknown sample X_j to belong to group i given its observed repair time $R = r_j$ and contract type $T = t_j$ is calculated. Using the Bayes' theorem, this probability is proportional to the probability of repair time $R = r_j$ and contract type $T = t_j$ given it belongs to group i, called the *likelihood*, multiplied with the probability of group i, called the *prior*. Dividing the multiplication of likelihood and prior with the probability of the observed repair time $R = r_j$ and contract type $T = t_j$ gives the desired probability. Then, the sample is assigned into the group with the highest probability. Because repair time is continuous, the "probability" of $R = r_j$ is approximated by discretizing the repair time to a range of 5% of r_j, i.e. $\mathbb{P}(0.975r_j \leqslant R \leqslant 1.025r_j)$. This leads to the following formula:

$$\mathbb{P}(Group = i|0.975r_j \leqslant R \leqslant 1.025r_j, T = t_j) = \frac{\mathbb{P}(0.975r_j \leqslant R \leqslant 1.025r_j, T = t_j|Group = i)\mathbb{P}(Group = i)}{\mathbb{P}(0.975r_j \leqslant R \leqslant 1.025r_j, T = t_j)}. \quad (1)$$

Note that there is always a group i where $\mathbb{P}(0.975r_j \leqslant R \leqslant 1.025r_j, T = t_j|Group = i) > 0$ for each X_j. With this technique, an unknown sample with short repair time is more likely to be redistributed into a group with shorter repair time, and vice versa.

Table 1 presents some information about the repair time of each group before and after redistribution of unknowns. Groups with high proportion (Group 2 and Group 5) receive the most assigned unknowns while groups with very low proportion (Group 1 and Others) receive no assigned unknowns. Fig. 8(b) shows the repair time distribution of Group 2, the group that receives the most redistributed unknowns, before (dashed) and after (solid) redistribution. The redistribution appears to slightly shift the repair time distribution to the left.

Table 1
The number of samples, proportion, mean, and standard deviation of repair time of each cause group before and after redistribution. The information after redistribution is presented in brackets.

Cause group	Number of samples	Proportion (%)	Mean	Standard deviation
Group 1	28 (28)	2.14 (1.46)	235.21 (235.21)	218.47 (218.47)
Group 2	602 (944)	46.06 (49.17)	96.67 (89.65)	109.11 (104.12)
Others	20 (20)	1.53 (1.04)	73.90 (73.90)	78.50 (78.50)
Group 3	198 (269)	15.15 (14.01)	55.08 (45.55)	74.43 (68.32)
Group 4	182 (215)	13.93 (11.20)	38.15 (36.83)	38.13 (35.73)
Group 5	277 (444)	21.19 (23.13)	12.64 (10.74)	13.94 (12.37)
All	1307 (1920)	100 (100)	67.03 (61.27)	97.01 (91.76)

A TC disruption length model is constructed with the Copula Bayesian Network method using the finding from the data analysis presented in this section. An overview of the method is provided in the next section and the necessary steps taken in the model construction are explained there as well.

4. The Copula Bayesian Network and the disruption length model

4.1. The Bayesian Network and its structure

Zilko et al. (2014) proposes the use of Copula Bayesian Network as the dependence model between the variables discussed in Section 3 to make inference about the disruption length given the realizations of the influencing variables. As its name suggests, a Bayesian Network (BN) is used to graphically represent the dependence between the variables. A BN is a directed acyclic graph consisting of nodes and arcs, representing the variables and flow of influence between the variables, respectively.

Fig. 9 presents the BN structure of the TC disruption length model in this paper.[3] The ten nodes in the structure correspond to the ten variables in the model and the arcs represent the flow of influence between the variables. The absence of an arc between two nodes indicates (conditional) independence between the variables the two nodes represent. For instance, the variable cause and latency time are conditionally independent given the variables contract, distance to the nearest level crossing (from hereon, this variable's name is shortened as 'level crossing'), and working hours. This means that once information about the contract, level crossing, and working hours is known, the variable cause has no effect on the latency time. This is reasonable because while cause does not have direct influence on latency time, a certain cause occurs more likely in certain areas at certain time and this influences the latency time. Therefore, once information about location (represented by contract and level crossing) and time (represented by working hours) is available, the information about cause becomes irrelevant to the latency time length.

Interdependencies between the influencing factors are also depicted in the structure above. Analyzing the data reveals that an arc is needed between the variable contract and working hours. A possible explanation for this is the incidents that are caused by coins which occur mostly in the evening (thus outside the mechanics' working hours) and not in the Randstad (which has OPC contract). There is influence between cause and contract, level crossing, and working hours for the same reason. Most disruptions caused by coins are observed outside the Randstad as Fig. 10 shows. This means that when the contract type is PGO, hence outside the Randstad, the chance that the disruption is caused by coins is higher. Conversely, when a disruption is caused by coins, it is more likely to occur outside of the Randstad, hence the PGO contract. Similarly, the closer the distance to the nearest level crossing is, the higher chance that the disruption is caused by coins due to easier accessibility. The converse also holds. This goes along with the railway property in the Randstad where less level crossing is present.

The structure in Fig. 9 is obtained based on the definition of the variables and the analysis of the data. Another way to determine the structure is by learning from data. A number of algorithms have been developed for this purpose when the variables at hand are all discrete. One of the algorithms is the hill-climbing greedy search in the space of all possible BN structure. It is a score-based algorithm which assigns a score to each possible BN structure based on the data and the structure that maximizes the score is chosen. The score of a structure $\mathcal{G}$ is defined as the probability of the structure given the data $\mathcal{D}$, i.e. $\mathbb{P}(\mathcal{G}|\mathcal{D})$ Margaritis (2003). Because our model consists of six discrete and four continuous variables, we need to discretize the continuous variables to be able to use the algorithm. Because of the limited number of samples, the continuous variables are discretized into four discrete states. Performing the hill climbing search results in a BN structure as presented in Fig. 11. The structure is obtained by executing the algorithm using the package `bnlearn` in `R` Scutari (2010).

Resemblance between the structure in Figs. 9 and 11 can be observed even with a few flipped and missing arcs. Margaritis (2003) studies the performance of the algorithm where it is not always able to fully recover the directionality of the arcs. Moreover, the result in Fig. 11 is obtained by discretizing the continuous variables. The missing arcs and reversed direction may be artefacts of the discretization.

Two models with both structures are constructed in the next subsection to see which structure is more suitable for our data.

4.2. The Copula Bayesian Network and copula

The TC Disruption Length BN is quantified with the Copula Bayesian Network. The method is introduced in Kurowicka and Cooke (2005) and extended in Hanea et al. (2006) and Hanea et al. (2010) to construct a dependence model between continuous variables. This method implements the use of *copula* as the backbone model of dependence. A copula is the n-dimensional joint distribution in the unit hypercube of n uniform random variables. The theorem of Sklar (1959) serves as the basis of the copula application. He states that any cumulative distribution function $(X_1, \ldots, X_n)$, denoted as $F_{1,\ldots,n}$, can be rewritten in terms of the corresponding copula C as

$$F_{1,\ldots,n}(x_1, \ldots, x_n) = C(F_1(x_1), \ldots, F_n(x_n)) \tag{2}$$

[3] All BN figures in this paper are created with `UNINET`.

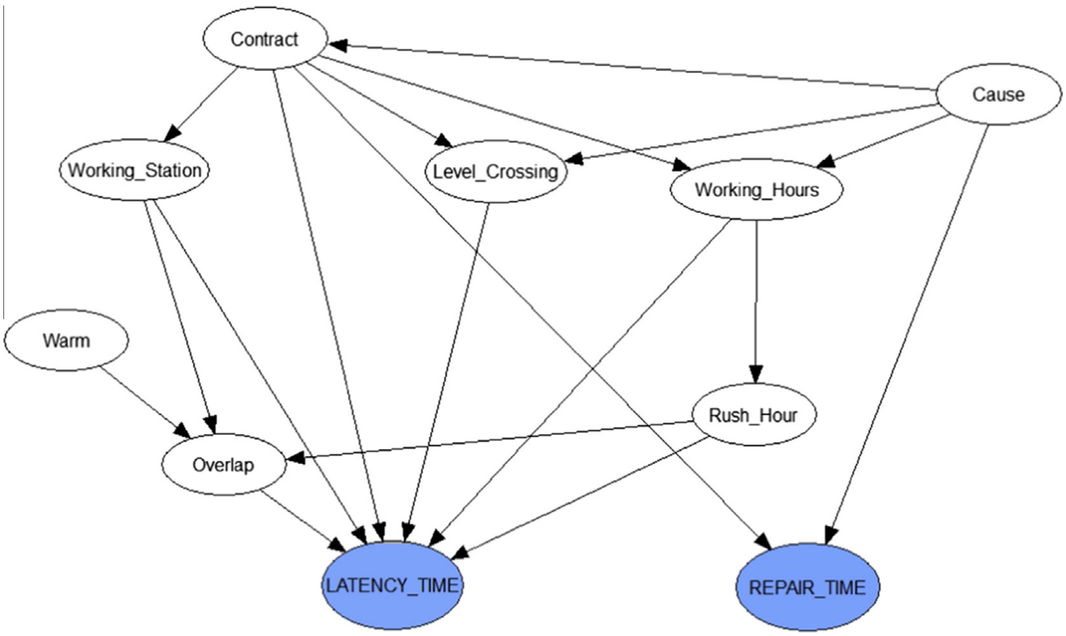


Fig. 9. The track circuit BN.

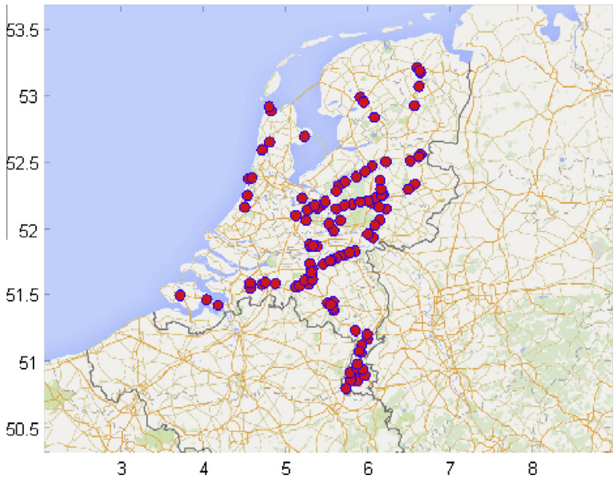


Fig. 10. The location of the observed TC disruptions caused by coins.

where $F_i(X_i)$ denotes the marginal distribution of the i-th variable.

There are many different copula families. One that is of interest in this paper is the multivariate Normal, or Gaussian, copula C_Σ. It is defined as

$$C_\Sigma(u_1, \ldots, u_n) = \Phi_\Sigma(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_n)) \tag{3}$$

where Φ^{-1} denotes the inverse cumulative distribution of a univariate standard normal distribution and Φ_Σ denotes the cumulative joint distribution of a multivariate normal distribution with zero mean and correlation matrix Σ. With the Copula Bayesian Network, the parameter Σ corresponds to the arcs in the BN structure. We are interested in this copula because it allows conditionalization to be computed rapidly, a very useful feature in the real-time decision making environment of the OCCR.

Copula has been proven to be an attractive model for the dependence between continuous variables. In our case however, some of the influencing variables are discrete. This results in a mixed discrete–continuous model that needs to be

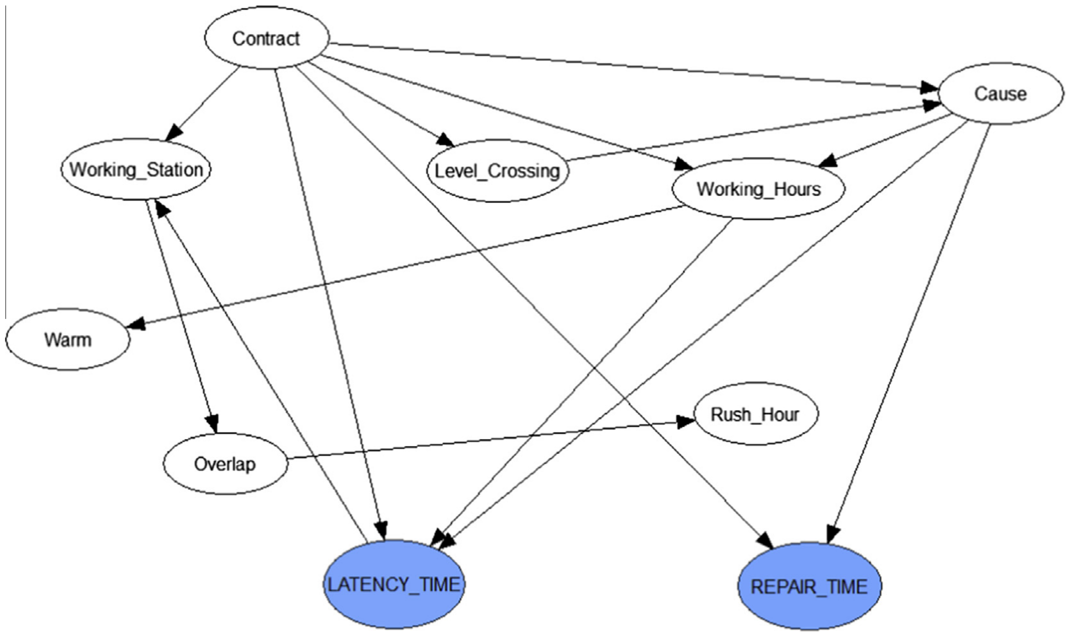


Fig. 11. The learned track circuit BN structure from data.

constructed. Unfortunately, the use of copula in discrete model is known to be troublesome. For this reason, Zilko and Kurowicka (in press) investigate the matter further for the latency time model. A more complex copula model construction (with the so-called copula-vine approach) is also considered to perfectly represent the discrete part of the model by adding more parameters. It turns out that even though the Copula Bayesian Network model with Normal copula does not perfectly capture the discrete part of the model, its performance for the latency time prediction is still statistically sound. The copula-vine approach increases the likelihood of the model but the gain in the latency time prediction is not significant. For this reason in this paper, the TC BN model is constructed with the Copula Bayesian Network method with the Normal copula.

For the mixed discrete–continuous TC Disruption Length model, the parameter of the Normal copula Σ is calculated using the maximum-likelihood approach. At the moment, this functionality is not available at UNINET so one needs to manually compute Σ in another software and then inputs the result into UNINET.

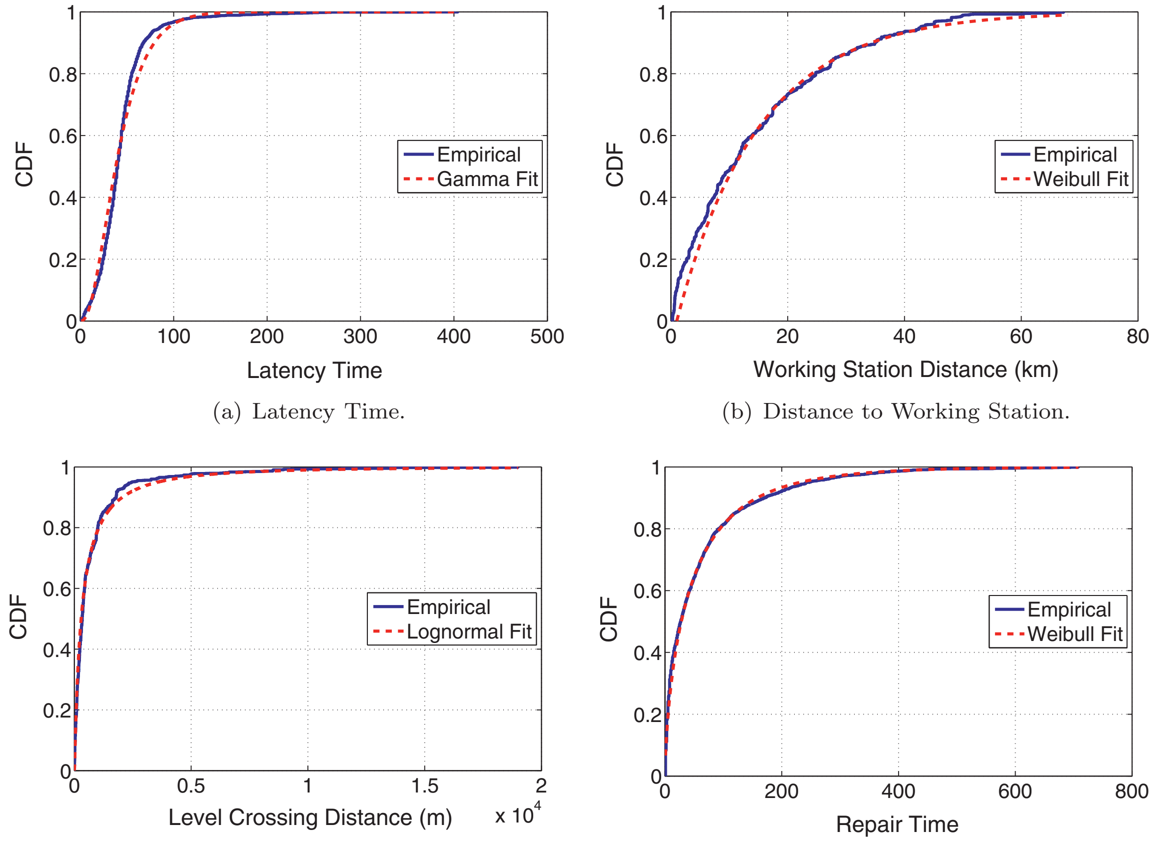
4.3. Fitting parametric distributions to the continuous variables

Two ingredients are needed in the theorem of Sklar's in Eq. (2), the copula C and the marginal distribution $F_i(X_i)$. For the copula, we have chosen to work with the Normal copula C_Σ. The marginal distributions can be represented by the empirical distributions from data. However for the continuous variables, there is an option to possibly represent them by known parametric distributions. One advantage of fitting a parametric distribution to a continuous variable is the ability to conditionalize the variable more extreme than what is observed in the data.

Four parametric distributions are considered for fitting: exponential, lognormal, Gamma, and Weibull distribution. The use of these four distributions is not new in the railway operation research. The exponential distribution is used to estimate the disruption length in Schranil and Weidmann (2013). The lognormal distribution is under consideration because, by definition, the four continuous variables are non-negative. In a slightly different setting, Yuan (2006) considers the lognormal, Gamma, and Weibull distribution to model several different kinds of train delay. Later on, Corman et al. (2011) and Jensen et al. (2015) use the Weibull distribution to model the arrival, departure, and dwell delay based on the work of Yuan (2006).

For each continuous variable and each parametric distribution, the distribution's parameter(s) is (are) calculated using the maximum likelihood approach. Then, goodness of fit test can be performed with the KS test. This results in the latency time to be best-fitted with gamma distribution, distance to the nearest level crossing with the lognormal distribution, and both distance to the nearest working station and repair time with the Weibull distribution. Fig. 12 shows the empirical distributions (solid blue lines) and the best fitted parametric distributions (red dashed lines) of the four continuous variables. However, the KS test rejects the hypothesis that any of the variables can be represented with the parametric distributions with p-values way lower than 0.05. The same conclusion is drawn when the Anderson–Darling (AD) test is performed to each case (Anderson and Darling, 1952).

Therefore, empirical distributions are going to be used for the continuous variables in the model.



(a) Latency Time. (b) Distance to Working Station. (c) Distance to Level Crossing. (d) Repair Time.

Fig. 12. Fitting parametric distributions to the continuous variables.

4.4. Model construction

Zilko et al. (2015) conclude that the Normal copula is able to reasonably represent the latency time model. In our case, two more variables are added to the model: cause and repair time. First, it will be checked whether the Normal copula models these two variables well. This is done by fitting bivariate Normal copula between cause and repair time using maximum likelihood. The goodness of fit is performed by sampling from the Normal copula. For each cause, the repair time distribution from the data and the Normal copula are compared with the two-sample KS and CvM Test. If the two repair time distributions are similar for all causes, the Normal copula is said to model the dependence well.

The parameter of the bivariate Normal copula is found to be −0.5853. The Normal copula is not rejected for Group 1, Group 2, Group 3, Group 4, and Others. For Group 5, the Normal copula is rejected with *p*-values of 0.0178 and 0.0149 for the KS test and CvM test, respectively. For this group, the Normal copula models the short repair time well but is conservative for the longer one where it overestimates the repair time length.

The parameters of two multivariate Normal copulas, each corresponding to the structures in Figs. 9 and 11, are calculated using the maximum likelihood technique. The computation is computationally heavy where with an Intel(R) Core i5-3470 3.2 GHz processor and 8 GB RAM, the parameter estimation takes approximately 24 h with the software MATLAB for each structure. However, this parameter estimation needs to be performed only once.

To measure which structure models the data better, the Akaike Information Criterion (AIC) introduced in Akaike (1974) is used. It is defined as

$$AIC = 2k - 2\ln(L) \tag{4}$$

where k represents the number of parameters and L represents the likelihood. The structure with the lower AIC models the data better. For our data set, the AIC score for the structure in Fig. 9 is 34498.4 while for the structure in Fig. 11 is 34718.8. Therefore, we proceed with the BN structure as in Fig. 9.

Algorithms for testing the goodness of fit of a copula to a set of data have been developed for purely discrete or purely continuous models. Consequently for our mixed discrete–continuous model, goodness of fit test of the multivariate Normal copula is performed on two groups: the discrete variables and the continuous variables. For the discrete variables, we observe the difference between the joint discrete distribution from the data and from the Normal copula. Five of the six discrete variables are binary and one has six states so the joint discrete distribution has 192 cells. Measuring the difference in the sense of Kullback–Leibler divergence (for more details about this, see e.g. Cover and Thomas (2006)), the distance is found to be 0.0718. This corresponds to the p-value of 2.2119e−06 implying that the difference is too high to be explained by the statistical fluctuation in the data. This means there is some misfit at the discrete part of the model. The largest problem is found to be in the cell corresponding to the situation of PGO contract, outside the mechanics' working hours, under 25 °C temperature, outside of rush hour, no overlap, and caused by coins.

For the continuous variables, a goodness of fit test as in Breymann et al. (2003) which is based on the Rosenblatt's probability integral transform (Rosenblatt, 1952) is used. The test is as follows: Let $\mathbf{U} = (\widetilde{U}_1, \widetilde{U}_2, \widetilde{U}_3, \widetilde{U}_4)$ be random vector with uniform margins. Then, define U_1, U_2, U_3, U_4 as:

$$U_1 = \widetilde{U}_1, \quad U_2 = C_{\Sigma_{[1,2]}}(\widetilde{U}_2|U_1), \quad U_3 = C_{\Sigma_{[1,2,3]}}(\widetilde{U}_3|U_1, U_2), \quad U_4 = C_{\Sigma}(\widetilde{U}_4|U_1, U_2, U_3)$$

where $\Sigma_{[1,2]} = \rho_{12}$ is an element in the correlation matrix Σ representing the dependence between the first and the second variable and $\Sigma_{[1,2,3]}$ corresponds to a sub-matrix of Σ representing the dependence between the first, second, and third variable. If the copula is correct, U_i should be uniform on $(0, 1)$ for each i and are independent. This means that the test-statistics

$$V = \left(\Phi^{-1}(U_1)\right)^2 + \left(\Phi^{-1}(U_2)\right)^2 + \left(\Phi^{-1}(U_3)\right)^2 + \left(\Phi^{-1}(U_4)\right)^2 \tag{5}$$

is Chi-square distributed with 4 degrees of freedom.

The vector $\mathbf{U}$ is obtained through transforming the four continuous variables to uniform using their empirical distributions. By the KS-Test, the hypothesis that each of U_1, U_2, U_3, and U_4 is uniform is not rejected with the smallest p-value being 0.6682. Moreover, the null hypothesis that V comes from a Chi-square distribution with 4 degree of freedom is also not rejected with p-value of 0.5278. Therefore, the test does not reject the null-hypothesis that the Normal copula fits the data for the continuous variables.

The purpose of the model is to produce disruption length prediction given the information of all the other influencing factors. From this point of view, the observed misfit at the discrete part of the model does not mean the performance of the model will be bad for the prediction. This will be checked later in Section 5 where we perform validation with the training set and a test.

5. Model use and validation

5.1. Model use

This section discusses how the model is planned to be used in real-time practice following the disruption handling procedure in the Netherlands. As an illustration, a TC incident on 1 July 2014 in the village of Sloterdijk near Amsterdam is chosen. It was caused by loose cable due to broken cable shoes (Cause Group 2). The nearest mechanics' working station was in Amsterdam with approximated distance of 4.492 km with the nearest level crossing was 3.884 km away. It occurred during the afternoon rush hour time (Rush Hour = 1) and outside the mechanics' working hour (Working Hours = 0). The temperature at the time was 18.8 °C (Warm = 0) and there was no overlapping incident (Overlap = 0). The contract type was the OPC contract (Contract = 0). The actual latency time and repair time were 51 and 101 min, respectively.

When a TC disruption occurs, before any further information about the incident comes in, we have the unconditional model which covers all historical TC disruptions in the data base. Fig. 13 shows the unconditional TC BN model at hand. This is the same BN as presented in Fig. 9 where now the marginal distributions of the variables are presented along with their corresponding means and standard deviations at the bottom of the nodes. An eleventh node, disruption length, is added to the BN. This node is a *functional* node that is defined as the sum of latency and repair time.

At this point, we can already see a decision making problem that one faces in providing the prediction: which value of the disruption length distribution needs to be taken as the prediction? An option would be to take its mean which in this case is 104 min. This is essentially what is done in current practice as the "P1" disruption length prediction. Because the distribution is left-skewed, the mean is larger than its 50% quantile (the median). In fact, the mean of 104 min corresponds to the 67% quantile of the distribution. Another option would be to choose the median as the prediction, which in this case takes value of 72 min. However, for our example, both choices underestimate the actual disruption length of 51 + 101 = 152 min which corresponds to the 80% quantile of the disruption length distribution.

With the available information about the influencing factors of latency time (at this point, the cause is not known yet), the BN is updated. Conditionalizing the model on the information yields a conditioned BN that is presented in Fig. 14. It is visible that the model adjusts itself to the situation at hand. The disruption length is predicted to be longer with mean 130 min and median 92 min. Moreover, it also predicts the latency time to be longer than average. The mean of latency time increases from 43.2 to 53.2 min while the median from 40 to 45 min. Notice that the distribution of cause also changes. Because the contract type is OPC with the nearest level crossing to be very far away (a common property of the Randstad), the probability that the disruption is caused by coins decreases from 23% to 16.5%. As a result, the predicted repair time's mean increases by 15.8 min (the median's increase is 11 min).

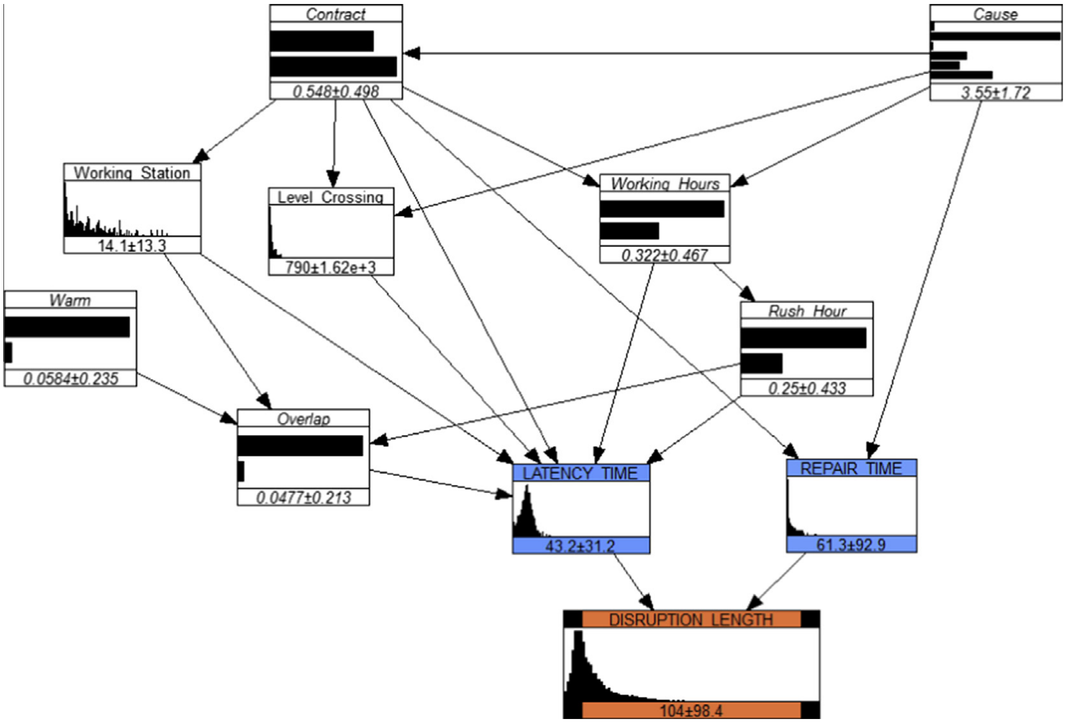


Fig. 13. The unconditional TC BN.

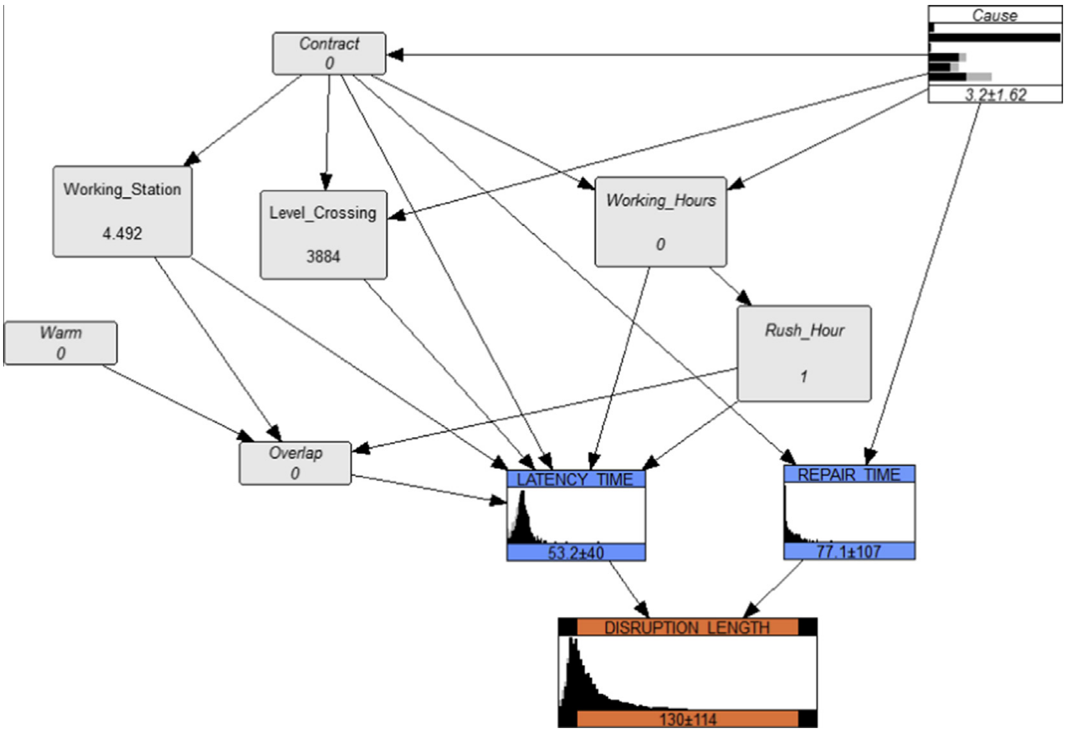


Fig. 14. The TC BN conditioned on the latency time's influencing factors.

The mechanics arrive at the site 51 min after the disruption starts so the actual latency time is 51 min. After investigation, the problem with cable (Group 2) is found. The model is further conditioned on the new information and the BN is updated

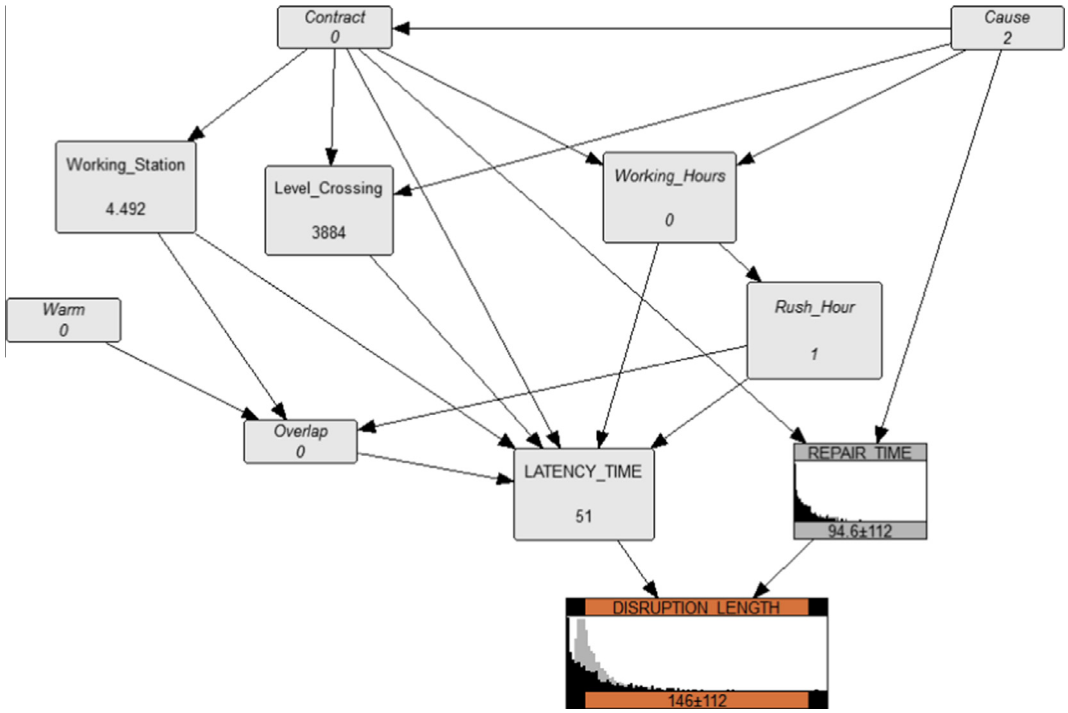


Fig. 15. The TC BN conditioned on all factors.

Table 2
The performance of different prediction choices compared to the truth for the illustration.

Prediction choices	Latency time	Repair time	Disruption length
"P1" prediction	43.2	61.3	104
Prediction with mean	53.2	94.6	146
Prediction with median	45	57	108
Truth	51	101	152

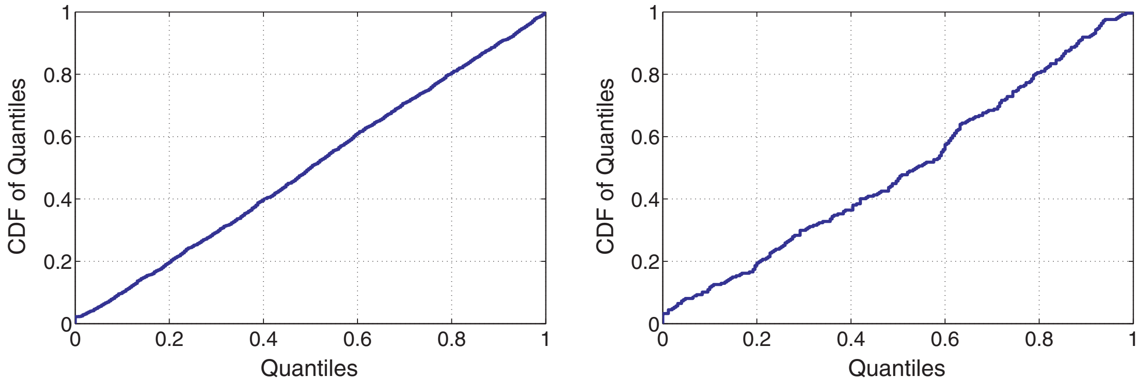
as presented in Fig. 15. The model, again, adjusts itself to the situation. The disruption length prediction is updated to 146 min in the mean and 108 min in the median. The mean of repair time is also updated to 94.6 min and the median to 57 min.

Table 2 presents the performance of three different prediction choices in comparison to the truth. For our example, the prediction with mean performs better than the prediction with median or the "P1" prediction.

The example shows how our model adjusts with the situation. The model processes the influencing factors and updates the disruption length accordingly. As more information is processed, the model learns that this particular incident is a longer one hence it returns longer disruption length. Moreover, the computation of the conditional probability distribution of the disruption length is performed quickly. Generating 1000 samples from the conditional distribution for each run of the model with an Intel(R) Core i5-3470 3.2 GHz processor and 8 GB RAM computer takes on average 6.6773 s with the software MATLAB.[4]

In practice, conditionalization on the variable Cause can only be performed after the mechanics diagnose the problem and find the cause. The time the mechanics need to diagnose the problem is called the diagnosis time. However, the diagnosis time is not available in the data set and is actually contained in the definition of "repair time". Unfortunately, the data does not provide any information to separate the diagnosis time from the actual repair time. For this reason, one should be aware that when the TC BN model is used, the repair time prediction also contains the diagnosis time.

[4] Computation with the software UNINET in the same computer takes less than one second.



(a) With the training set. (b) With the test set.

Fig. 16. The distribution of the quantiles of repair time.

5.2. Model validation

The output of the Copula Bayesian Network model is the conditional distribution of latency and repair time given observation of the influencing factors. Since these two variables are our variables of interest, model validation is performed by measuring the accuracy of the conditional distributions with data. For each incident in the data, the quantile of this conditional distribution corresponding to the actual observed time is computed. The model represents the data well if the collection of these quantiles forms a uniform distribution on $(0, 1)$.

When a TC disruption occurs, the model yields the conditional latency time distribution by conditioning on the latency time's influencing factors. Zilko et al. (2015) has validated the model's conditional latency time distribution where, indeed, the quantiles of the realization forms a uniform distribution on $(0, 1)$. This indicates that the latency time is well-modeled.

Next, after the mechanics arrive at the site and the cause of failure is known, the model is further conditioned on the actual latency time and cause. This will result in the conditional repair time and disruption length distributions as the outputs of the model. At this stage, because the latency time is already "known", the conditional distribution of disruption length is just the summation of the conditional repair time distribution and a constant (the actual latency time). Therefore, it is sufficient to perform model validation on the conditional repair time distribution only.

Fig. 16(a) shows the resulted quantiles distribution of the repair time for the training set. Overall, the distribution of the quantiles appears to be uniform. Performing the one-sample KS-Test against the uniform distribution on $(0, 1)$ yields a p-value of 0.2863. Quesenberry and Miller (1977) suggests the use of the Watson test (introduced in Watson (1961)) as a goodness-of-fit test statistics for a uniform distribution on the unit interval. Performing this test, we also do not reject the null-hypothesis that the distribution of the quantiles is uniform as the test-statistics (0.0574) is below the critical value (0.187) for the significance level of 5% (Pearson and Hartley, 1972). This means that the repair time and disruption length are well-modeled.

The model also needs to be validated against a test set, a separate set of data that is not used in the model construction. For this, TC disruptions data collected from 1 May 2014 up to 31 October 2014 is used. Within these six months period, 339 urgent TC incidents were recorded. Just like in the training set, the cause of approximately 30% of the TC problems in the test set is also unknown. For validation, however, the unknown samples are discarded from the test set and, thus, the test is performed only on the 247 samples with known causes. The latency time model's prediction power for the test set has been investigated in Zilko and Kurowicka (in press) where it is discovered that the quantiles forms a uniform distribution on $(0, 1)$.

Fig. 16(b) shows the resulted quantiles distribution of the repair time with the test set. Overall the distribution of the quantiles appears very close to the uniform distribution. Moreover, performing the KS-Test against uniform distribution yields a p-value of 0.3676. Moreover, performing the Watson test also does not reject the null-hypothesis that the distribution of quantiles is uniform on $(0, 1)$ as the test statistics (0.1170) is below the 5% critical value (0.187). This means that the TC Copula Bayesian Network model performs well for the 247 TC disruptions in the test set.

We also compare the performance of our model with the "P1" prediction that is used in practice. The "P1" prediction is the mean disruption length of the unconditional model. Taking the mean as our TC model's prediction, we calculate each model's Root Mean Squared Error (RMSE) with the training and test set. With the training set, the RMSE of the "P1" prediction is 97.0795 while our model's is 87.2795. With the test set, the RMSE of the "P1" prediction is 113.9478 while our model's is 100.1521. With lower RMSEs, it indicates that the TC BN model predicts the actual disruption length better than the "P1" prediction.

6. Conclusions

To help the OCCR handle the uncertainty of disruption length, we propose the Copula Bayesian Network method to predict the disruption length based on the available information during disruption. As an example, a model has been constructed for disruptions caused by TC failure in the Netherlands. Influencing factors of the latency and repair time length have been considered and added to the model. It is also shown that despite the misfit with data at the discrete part of the model, the model's prediction power for disruption length is still sound. Moreover, even though the model's parameter estimation takes a lot of time with the maximum likelihood technique, the model use in real-life application is fast. Models for disruptions caused by other factors can be constructed by following the same procedure.

A challenge has been mentioned in Section 5 regarding the choice of prediction value. The output of the Copula Bayesian Network model is a conditional distribution from which a value from the distribution is chosen as the prediction. Statistically, the mean is commonly chosen because it minimizes the prediction's mean square error (MSE). However, this may not be the most desired choice in practice. The purpose of the model is to support the OCCR in their decision-making process to handle the disrupted train traffic by providing a disruption length prediction. With this information, certain decisions about the train traffic are made with the goal of returning to normal train operation as soon as possible. Thus, another choice of prediction may benefit this goal better.

To tackle the challenge, collaboration is planned with an ongoing parallel project dealing with efficient and effective dispatching. A study on when to be optimistic or conservative during different stages of the disruption is going to be performed. Being very optimistic means a higher probability that the actual disruption will take longer than the prediction. On the other hand, being very conservative means that the operation during disruption assumes a bad scenario while the actual situation may actually not be that bad.

The model constructed in this paper is based on the available data and is shown to work well for this data set. In principal, the model can be improved if and when better data is available. This data should contain more precise information about the mechanics' operation during disruption, for instance their actual departure point, their actual route, and all of their actual actions which they perform, to name a few. However, this potentially leads to a complex study on its own as one reaches the field of human behavioral modeling which needs to be quantified, somehow. For the moment, this is out of our research scope.

Acknowledgement

This work is funded by ExploRail, a partnership programme of the Dutch Technology Foundation STW and ProRail, project No. 12257: "Smart information and decision support for railway operation control centres (SmartOCCR)". The authors would like to acknowledge Dirk Kes, André Duinmeijer, and Erwin van Wonderen from ProRail for the supervision and access to the data set.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Autom. Control 19 (6), 716–723.
Anderson, T.W., 1962. On the distribution of the two sample cramer-von mises criterion. Ann. Math. Stat. 33 (3), 1148–1159.
Anderson, T.W., Darling, D.A., 1952. Asymptotic theory of certain goodness-of-fit criteria based on stochastic processes. Ann. Math. Stat. 23, 193–212.
Bender, A., Mussa, H.Y., Glen, R.C., Reiling, S., 2004. Molecular similarity searching using atom environments, information-based feature selection, and a naïve bayesian classifier. J. Chem. Inform. Comput. Sci. 44 (1), 170–178.
Breymann, W., Dias, A., Embrechts, P., 2003. Dependence structures for multivariate high-frequency data in finance. Quant. Finan. 3, 1–14.
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. Transport. Res. B 63, 15–37.
Chen, C., Zhang, G., Wang, H., Yang, J., Jin, P.J., Walton, C.M., 2015. Bayesian network-based formulation and analysis for toll road utilization supported by traffic information provision. Transport. Res. C 60, 339–359.
Corman, F., D'Ariano, A., Pranzo, M., Hansen, I.A., 2011. Effectiveness of dynamic reordering and rerouting of trains in a complicated and densely occupied station area. Transport. Planning Technol. 34 (4), 341–362.
Cover, T.M., Thomas, J.A., 2006. Elements of Information Theory. Wiley.
Ghaemi, N., Goverde, R.M.P., 2015. Review of railway disruption management practice and literature. In: Proc. 6th International Conference on Railway Operations Modelling and Analysis (RailTokyo2015, Tokyo, Japan, March 2015).
Giuliano, G., 1989. Incident characteristics, frequency, and duration on a high volume urban freeway. Transport. Res. A 23 (5), 387–396.
Golob, T.F., Recker, W.W., Leonard, J.D., 1986. An analysis of the severity and incident duration of truck-involved freeway accidents. Accident Anal. Prevent. 19 (4), 375–395.
Gregoriades, A., Mouskos, K.C., 2013. Black spots identification through a bayesian networks quantification of accident risk index. Transport. Res. C 28, 28–43.
Hanea, A., Kurowicka, D., Cooke, R., 2006. Hybrid method for quantifying and analyzing bayesian belief nets. Quality Reliab. Eng. Int. 22, 709–729.
Hanea, A., Kurowicka, D., Cooke, R., Ababei, D., 2010. Mining and visualizing ordinal data with non-parametric continuous bbns. Comput. Stat. Data Anal. 54 (3), 668–687.
Jensen, L.W., Landex, A., Nielsen, O.A., 2015. Assessment of stochastic capacity consuption in railway networks. In: Proc. 6th International Conference on Railway Operations Modelling and Analysis (RailTokyo2015, Tokyo, Japan, March 2015).
Kurowicka, D., Cooke, R., 2005. Distribution-free continuous bayesian belief nets. Modern Stat. Math. Methods Reliab., 309–323
Marchant, J.A., Onyango, C.M., 2002. Comparison of a bayesian classifier with a multilayer feed-forward neural network using the example of plant/weed/soil discrimination. Comput. Electron. Agric. 39, 3–22.
Margaritis, D., 2003. Learning bayesian network model structure from data Ph.D. thesis. School of Computer Science, Carnegie-Mellon University, Pittsburgh, PA.
Meng, L., Zhou, X., 2011. Robust single-track train dispatching model under a dynamic and stochastic environment: a scenario-based rolling horizon solution approach. Transport. Res. B 45 (7), 1080–1102.
Nam, D., Mannering, F., 2000. An exploratory hazard-based analysis of highway incident duration. Transport. Res. A 34 (2), 85–102.
Ng, M.W., Lo, H.K., 2013. Regional air quality conformity in transportation networks with stochastic dependencies: a theoretical copula-based model. Netw. Spatial Econ. 13 (4), 373–397.

Oukhellou, L., Cme, E., Bouillaut, L., Aknin, P., 2008. Combined use of sensor data and structural knowledge processed by bayesian network: application to a railway diagnosis aid scheme. Transport. Res. C 16 (6), 755–767.
Pachl, J., 2004. Railway Operation and Control. VTD Rail Publishing.
Pearson, E.S., Hartley, H.O., 1972. Biometrika Tables for Statisticians, vol. II. Cambridge University Press.
Quesenberry, C.P., Miller Jr., F.L., 1977. Power studies of some tests for uniformity. J. Stat. Comput. Simul. 5 (3), 161–191.
Rosenblatt, M., 1952. Remarks on a multivariate transformation. Ann. Math. Stat. 23, 470–472.
Schranil, S., Weidmann, U.A., 2013. Forecasting the duration of rail operation disturbances. In: Proc. 92nd Annual Meeting of the Transportation Research Board, Washington, D.C.
Scutari, M., 2010. Learning bayesian networks with the bnlearn r packageh. J. Stat. Softw. 35 (3), 1–22.
Sklar, A., 1959. Fonctions de rpartition n dimensions et leurs marges. Publ. Inst. Statist. Univ. Paris 8, 229–231.
Srinivas, S., Menon, D., Meher Prasad, A., 2006. Multivariate simulation and multimodal dependence modeling of vehicle axle weights with copulas. J. Transport. Eng. 132 (12), 944–955.
Sullivan, E.E., 1997. New model for predicting freeway incidents and incident delays. J. Transport. Eng. 123 (4), 267–275.
Visser, A.J., Steenkamp, H., 1981. Spoorstroomlopen. Nederlandse Spoorwegen, (NS). August.
Wan, K., Kornhauser, A.L., 1997. J. Transport. Eng. 123 (4), 267–275.
Wang, Q., Garrity, G.M., Tiedje, J.M., Cole1, J.R., 2007. Naïve bayesian classifier for rapid assignment of rrna sequences into the new bacterial taxonomy. Appl. Environ. Microbiol. 73 (2), 5261–5267.
Watson, G.S., 1961. Goodness-of-fit tests on a circle. Biometrika 48, 109–114.
Yuan, J., 2006. Stochastic modelling of train delays and delay propagation in statistics Ph.D. thesis. Delft University of Technology.
Zilko, A.A., Hanea, A.M., Kurowicka, D., Goverde, R.M.P., 2014. Non-parametric bayesian network to forecast railway disruption lengths. In: Proc. 2nd International Conference on Railway Technology: Research, Development and Maintenance (Railways 2014, Ajaccio, France, April 2014).
Zilko, A.A., Kurowicka, D., 2016. Copula in a multivariate mixed discrete-continuous model. J. Comput. Stat. Data Anal., in press.
Zilko, A.A., Kurowicka, D., Hanea, A.M., Goverde, R.M.P., 2015. The copula bayesian network with mixed discrete and continuous nodes to forecast railway disruption lengths. In: Proc. 6th International Conference on Railway Operations Modelling and Analysis (RailTokyo2015, Tokyo, Japan, March 2015).