System-based Vulnerability Measures for Railway Systems

Zhu, Yongqiu; Goverde, Rob

Publication date
2017

Document Version
Proof

Published in
7th International Conference on Railway Operations Modelling and Analysis: RailLille2017

In Proceedings of the 7th International Conference on Railway Operations Modelling and Analysis: RailLille2017
https://raillille2017.sciencesconf.org/

System-based Vulnerability Measures for Railway Systems

Yongqiu Zhu[a,1], Rob M.P. Goverde[a,2]
[a]Department of Transport and Planning, Delft University of Technology
P.O. Box 5048, 2628 CN Delft, The Netherlands
[1]E-mail: y.zhu-5@tudelft.nl, Phone: +31 (0)152784914
[2]E-mail: r.m.p.goverde@tudelft.nl, Phone: +31 (0)152783178

Abstract
The issue of disruptions in railway systems attracts a growing attention due to its severity to the society. Mitigation strategies are proposed from different perspectives to reduce the system vulnerability resulting by disruptions. However, due to financial limitations, it is not realistic to realize all strategies in the real world. Thus, it is necessary to establish a model based on which the system vulnerability can be measured and the mitigation strategy can be evaluated. In this paper, such a model is constructed by applying a Monte Carlo simulation based on the historical disruption data of railways in the Netherlands. Based on this model, we first construct a baseline scenario that is capable of reflecting the current disruption practice and calculate the top ten vulnerable tracks with the indicator of yearly disruption duration from this scenario. Next, we propose mitigation strategies towards each of the top ten vulnerable tracks of different magnitudes, and construct the mitigation scenarios according to the strategies. Finally, comparison between the baseline scenario and each of the mitigation scenarios is performed with the purpose of strategy evaluation.

Keywords
vulnerability, disruption, railway system, Monte Carlo simulation

1 Motivation

Disruptions are inevitable in the daily operation of railways, due to a variety of events such as rolling stock breakdown, switch failure, signal failure, etc. Consequently, on one hand, it results in unwanted monetary costs of either the infrastructure managers (IM) or the railway undertakings (RU) to restore the disrupted infrastructure capacity or service capacity to the normal status. On the other hand, it increases the time costs of passengers who are affected directly or indirectly, and further brings economic loss to the whole society since the society depends on the railways for people's daily mobility to some degree. Obviously, disruption is not only an issue within the railway system, but also a social-economic problem that attracts a growing attention.

According to previous studies, the way to deal with disruptions in the rail system can be divided into two categories based on the timing of specific actions. One is the ex-ante strategy that aims at preventing the occurrences of disruptions, especially the disruptions that resulted from infrastructure failures. For example, the research done by either Johansson et al. (2011) or Deng et al. (2015) provides guidance on which infrastructure should be strengthened with preventive measures. The other is the ex-post strategy that aims at mitigating the negative impacts when disruptions occur, like speeding the response to disruptions by enhancing the cooperation between organizations. Relevant research on the analysis of information flows or involved activities during the disruption management process can be found in Schipper et al. (2015) and Golightly et al. (2013), respectively.

No matter which aforementioned strategy, they share the same purpose of reducing the vulnerability of rail systems. However, due to financial limitations, it is usually impossible to implement all strategies when several candidate strategies are proposed. In such a case, the preferred strategy will be the one that leads to the most reduction on vulnerability, assuming that the investment cost of each candidate strategy is equal. Therefore, the amount of reduced vulnerability is a key indicator for evaluating the strategies. Thus, it is necessary to measure vulnerability beforehand to assist decision makers with strategy selection.

Reggiani et al. (2015) and Mattsson and Jenelius (2015) gave an overview of recent studies on vulnerability measures for transport systems, along with a comparison between vulnerability and resilience. In general, *resilience* is seen as the ability of transport systems to return to a normal status after disruptions have occurred, while *vulnerability* is seen as the risk that disruptions can affect transport systems (Caschili et al., 2015). In subway systems, Adjetey-Bahun et al. (2016) measured resilience with quantified passenger delay and passenger load as indicators. Instead, D'Lima and Medda (2015) measured resilience as the rapidity of systems to return to passenger flows to normal. It can be seen that resilience is more likely the quality of a system responding to disruptions. As Rose (2009) declared, improving resilience can be a solution to vulnerability reduction.

The purpose of this paper is to establish a simulation model based on which a specific mitigation strategy can be evaluated on its reduction of vulnerability. Thus, a resilience measure is included in the evaluation part with indicators of vulnerability differences between the real system and the system simulated by performing a mitigation strategy. Therefore, we focus on vulnerability measures in this paper.

The remainder of this paper is organized as follows. In section 2, we review the literature on vulnerability measure in rail systems. In section 3, we perform independence tests for all random variables that are included in the simulation model, in order to decide whether the distribution of each random variable is marginal or conditional. This is for ensuring that the baseline scenario can reflect the disruption practice well. In section 4, we introduce how the simulation model is established by the Monte Carlo method. In section 5, we implement the vulnerability measure on the baseline scenario and the scenarios on which the mitigation strategies are performed, along with the comparisons between scenarios referring to vulnerability reduction. Conclusions and potential directions for future research can be found in section 6.

2 Background

Literature on vulnerability measures can be differentiated according to whether it is at a topological level or at a system level. Below, we give a review of the relevant studies in rail systems.

The topological vulnerability measure is rooted in graph theory. The real transport system is represented by an abstract network of which the nodes and arcs may have different interpretations depending on the issues of interest. For example, Angeloudis and Fisk (2006) described a subway system with stations as nodes and connections between stations as arcs, in order to study the vulnerability of subway systems with the indicator of degree distribution. Instead, Deng et al. (2015) divided the subway system into thirty-one functional modules, and then established a subway physical network (SPN) of which each node represents a specific module and each arc represents the interdependency between modules. Their intention is to find the most vulnerable module according to the reduced network efficiency by removing each node from the SPN respectively. The network efficiency is represented by the average of reciprocals of all node pair distances.

Obviously, due to the focus on indicators such as degree centrality, betweenness centrality, etc., topological vulnerability measures have a low requirement for input data. This makes it possible to perform a vulnerability comparison between transport systems (Angeloudis and Fisk, 2006; Derrible and Kennedy, 2010; Kurant and Thiran, 2006). However, the advantage of limited data needs, in turn, leads to the limitation that topological vulnerability measures only give a general insight on the structural weakness of transport systems (Mattsson and Jenelius, 2015). As a result, system-based vulnerability measures become more appealing, since they are useful for proposing and assessing mitigation strategies towards a specific system by considering disruption consequences such as the durations of disruptions, the number of affected trains, the number of affected passengers, etc.

Aiming at the number of trains that cannot reach their planned destinations, Johansson et al. (2011) performed an empirical analysis on the southern parts of the Swedish railway system from three perspectives of vulnerability, i.e. global vulnerability analysis, critical component analysis and geographical vulnerability analysis. Likewise, considering disruption impacts on trains, Hong et al. (2015) established a simulation model to estimate the flood-induced vulnerability of each link in the Chinese railway system. On the contrary, Cats and Jenelius (2014) looked at the welfare loss of passengers, and established an interaction model between transport supply and demand to analyse the dynamic passenger betweenness centrality of the Stockholm subway system. In like manner, with increased passenger travel time and unsatisfied passenger demand in mind, Rodríguez-Núñez and García-Palomares (2014) calculated the criticality of each link and the vulnerability of each station for the Madrid subway system based on a trip assignment model.

In addition to different perspectives on disruption impacts, differences still exist in the adopted approaches to analyse system-based vulnerability. According to Murray et al. (2008), the approaches of network vulnerability analysis can be unified to *scenario-specific*, *simulation, strategy-specific* and *mathematical modelling*. *Scenario-specific* analysis is used in constructing the disruptive scenarios perceived to be important. For example, Johansson et al. (2011) constructed seven scenarios within each disabling a specific functional system (e.g. traction power, telecommunication, signal, etc.). As to *simulation*, this is used in constructing scenarios of diverse disruptive magnitudes. For example, each aforementioned scenario in Johansson et al. (2011) was further extended to ten sub-scenarios with a disruptive magnitude from “0” to “100%”. In such a case, a range of possible vulnerability can be obtained. *Strategy-specific* analysis is used in constructing the scenarios of which specific strategies are performed on. For example, Cats and Jenelius (2014) constructed four scenarios differentiated by the extent of availability of *Real-time Information Provision* to passengers. Finally, *mathematical modelling* is used in constructing all scenarios with the purpose of seeking the most vulnerable scenario. For example, in order to identify the most vulnerable link, Rodríguez-Núñez and García-Palomares (2014) constructed a sequence of scenarios where in each a link is disrupted. Likewise, Cats et al. (2016) identified the top twenty vulnerable links by a full-scan of all links.

In light of the aforementioned reviews, we can find that system-based vulnerability analysis is more suitable for our case, since we intend to evaluate mitigation strategies afterwards. As yet, we did not find any research that provided a system-based vulnerability measure on railway systems at a national level by taking all types of disruptions into account. However, it is of vital importance to recognize the characteristics of the whole system rather than part of it, because local patterns cannot fully represent the global pattern. Thus, in this paper, we established a simulation model to provide a system-based vulnerability measure for the entire railway system of the Netherlands.

First, in terms of historical disruption data, a baseline scenario is constructed based on which we can measure system-based vulnerability and perform full-scan analysis among all disrupted tracks to find out the top ten vulnerable tracks in the real world. Second, we propose several mitigation strategies from different magnitudes towards the top ten vulnerable tracks respectively and constructed new scenarios based on these strategies. For each new scenario, the system-based vulnerability was measured and compared with that of the baseline scenario, with the purpose of strategy evaluation. The approaches of *scenario-specific*, *simulation*, *strategy-specific* and *mathematical modelling* are all adopted in this paper, which are helpful for us to gain wide insights into the entire system.

3 Data analysis

In the Netherlands, a website called *Rijden de Treinen* (in Dutch) serves passengers with real-time disruption information such as the start time of disruptions, the expected durations of disruptions, the affecting tracks, the causes, etc. In addition to the real-time data, all historical disruption records are also available, which additionally contain the information of actual durations of disruptions. Netherlands Railways (NS), the main railway undertaking in the Netherlands, is responsible for the data support to this website. In this paper, the information of disruptions during 2012 and 2015 were derived from *Rijden de Treinen*.

With the purpose of measuring vulnerability, a Monte Carlo simulation model is established, which is capable of simulating daily disruption scenarios based on which observing temporal changes would be possible. For constructing daily disruption scenarios, five random variables are used, which are *daily counts*, *cause*, *track*, *start* and *duration*. *Daily counts* represent the number of disruptions occurring in one day. *Cause* refers to the type of cause that led to each disruption, such as *rolling stock breakdown*, *damaged catenary*, *signal failure*, etc. *Track*, *start* and *duration* represent the affecting track, start hour and the duration of disruptions respectively.

In the simulation model, the values of random variables are produced in accordance with given probabilities. For the baseline scenario, all probabilities are generated from historical disruption data. For the scenarios on which mitigation strategies are performed, the probabilities of the random variables that are influenced by strategies are modified, while other probabilities are the same as the ones we generate for the baseline scenario. Note that, as we intend to perform empirical vulnerability analysis, we generate empirical probabilities for all random variables. The empirical probability can also be called relative frequency (Mood et al., 1974). For example, for a specific hour, its empirical probability is the ratio of the number of disruptions that have started at this hour versus the number of disruptions that have started at any hour. Clearly, the empirical probabilities of outcomes of *start* sum to 1.

A simple way to generate an empirical probability is to assume that all random variables are statistically independent, thus probabilities would be marginal and can be directly generated from the data set. For example, the disruption probability of a specific track is the relative disruption frequency of the track in historical data. However, such an assumption may be untrue, if the random variables are significantly correlated or time dependent. For example, suppose the disrupted tracks vary with the disruption causes, then the probability distribution of *track* will depend on the values of *cause*. This means that the disruption probability of a specific track is the relative disruption frequency of the track in the historical data of corresponding disruption cause. Therefore, to make the simulation performance similar to the disruption practice to the greatest extent, we need to generate the correct probability (marginal or conditional) for each random variable. Whether a probability is marginal or conditional depends on whether the corresponding random variable is independent of any other factors. As such, it is necessary to perform an independence test for each random variable, which is the contribution of this section.

3.1 Choice of potential influence factors for each random variable

The first step is to find the factors that could influence the random variables. In Table 1, we list the candidate explanatory variables (influence factors) for each response variable (random variable).

Table 1: The potential influence factors of each random variable

Response variable	**Candidate explanatory variables**				
daily counts	*month*	*day*			
cause	*month*	*day*			
track	*month*	*day*	*cause*		
start	*month*	*day*	*cause*		
duration	*month*	*day*	*cause*	*track*	*start*

Temporal factors *month* and *day* are chosen as potential influence factors for all random variables. Here, *day* represents the day of a week, which has seven possible outcomes. For *daily counts*, it is not hard to imagine that in particular months with extreme weather condition, disruptions occur more frequently. Besides, we can imagine that disruptions happen more often in weekdays compared with those in weekends, since the infrastructure that supports dense traffic on weekdays could encounter more issues. For *cause*, it is natural to think that some disruption causes like *snowfall* only take place at specific months, while some disruption causes like *rolling stock breakdown* mostly happen at weekdays. For *track*, we also take *cause* as one of the candidate explanatory variables. Normally, only a part of the track is affected by a particular disruption cause. For example, *damaged railway bridge* only happens at the tracks where there is a railway bridge. For *start*, we also take *cause* as one of the candidate explanatory variables. We can imagine that some disruption causes are more likely to start at some particular hours, like *copper theft* usually takes place at midnight. Moreover, according to Zilko et al. (2016), disruption durations are composed of repair time and latency time. The types of disruption cause can affect the repair time, while the locations of workstations and working hours can affect the latency time. As such, for *duration*, in addition to temporal factors, we also take *cause*, *track* and *start* into account.

3.2 Independence tests for numerical response variables

The response variables, *daily counts* and *duration*, are both numerical variables. Their corresponding explanatory variables are all nominal variables. Hence, the method of Analysis of Variance (ANOVA) was firstly employed. However, we found that some outliers exist in the data sample of either *daily counts* or *duration*, which led to a violation of the assumption of ANOVA about a normal distribution of the residuals. Therefore, we finally chose the Kruskal-Wallis Test (KWT), which is the nonparametric version of one-way ANOVA without requiring normality.

Instead of testing numeric values, KWT focuses on the numeric indices of ordered data. It ranks all data across all groups and then computes the median rank for each group. The null hypothesis is that the median ranks of all groups are equal. If the null hypothesis is accepted under the desired significance level, we can make the conclusion that all groups come from the same distribution. This means there is no effect originating from the groups. However, if at least one of the median ranks is significantly different from the others, the null hypothesis will be rejected and we can conclude that the groups' effects are active. Given the circumstances, we further test whether the effects are different between groups. If not, we classify the groups with the same effects into one cluster and keep the groups with different effects separated. On one hand, the intention is to reduce the number of outcomes of explanatory variables, because too many outcomes will split the data sample into many parts of each with very small size. Since we aim to generate empirical probabilities, the issue of insufficient data has to be avoided. On the other hand, clustering groups that have the same effect can help us to recognize the characteristics of disruption practice better.

In this paper, clustering is realized by optimal *k*-means clustering through dynamic programming (Wang and Song, 2011). Some modifications to this method are made for our case. The modified model can be formulated as follows:

$$\text{maximize} \quad \sum_{i=1}^{k} p_{\mathrm{v}}\left(c_i\right)/k \tag{1}$$

$$\text{subject to} \quad p_{\mathrm{v}}\left(c_i\right) > l, \qquad \forall\ i \in K \tag{2}$$

$$p_{\mathrm{v}}\left(c_i, c_j\right) < u, \quad \forall\ i, j \in K, i \neq j \tag{3}$$

$$p_{\mathrm{v}}\left(c_i\right) = 1, \qquad \text{if } \left|c_i\right| = 1 \tag{4}$$

$$c_i = \left\{g_s, \cdots, g_e\right\}, \qquad 1 \le s \le e \le n \tag{5}$$

$$1 \le k \le n, \tag{6}$$

with $K = \{1, \cdots, k\}$ that refers to the indexes of clusters.

Suppose an explanatory variable has n outcomes (groups) and we intend to assign them into k clusters. As shown in equation (1), the objective is to maximize the mean of within-cluster p-values (i.e. $p_v\left(c_i\right)$) that are computed by implementing KWT on the within-cluster samples c_i. To ensure that the groups within one cluster come from the same distribution, we set a lower bound l (at least 5%) for $p_v\left(c_i\right)$ (equation (2)). Meanwhile, we set an upper bound u (at most 5%) for the p-value computed by implementing KWT on two clusters while labelling the data within the same cluster as one group (equation (3)). The intention is to ensure that the groups in different clusters come from different distributions at a significance level. In addition, we assign the value of 1 for the clusters with single group, since applying KWT on one group sample is meaningless (equation (4)). Moreover, the number of groups contained in each cluster c_i cannot be less than one and be larger than n (equation (5)). Likewise, the desired number of clusters should at least be one, but cannot exceeds n (equation (6)).

As to the value of *k*, we try from 2 until some k^0 under which we can gain a feasible clustering solution. Such a feasible solution is the optimal one under the given *k*, but may not be the globally optimal one through all possible *k*. The solution could be improved if we continue to try larger *k* after k^0, and the globally optimal solution will be the one k^* after which no feasible solution can be gained.

3.3 Independence tests for nominal response variables

The response variables *cause, track* and *start*, are all nominal variables. Their corresponding candidate explanatory variables are also nominal variables. Besides, *cause, track* and *start* all have more than two outcomes. In such a case, we decide to use Multinomial Logistic Regression (MLR) to analyse whether the candidate explanatory variables have an effect on the response variables.

Suppose a response variable R has m outcomes and an explanatory variable E has n outcomes, and we choose the last outcome of R as the reference. Based on MLR, we can compute the relative probability of being one outcome of R versus being in the reference outcome, using a linear combination of the predictors which are the n outcomes of E, i.e. E_j,

$$\mu_{i,m} = \ln\left(\frac{\pi_i}{\pi_m}\right) = \alpha_i + \sum_{j=1}^{n-1} \beta_{i,j} E_j, \forall\ i \in \{1,\cdots,m-1\}. \tag{7}$$

In (7), π_i represents the probability of the outcome i of R under the condition E. Note that, E_j is a binary variable with the value of either 1 or 0, as the explanatory variables here are nominal variables. For testing whether the explanatory variable E has an effect, the hypothesis that E does not affect the response variable R can be written as (Long, 1997):

$$H_0 : \beta_{i,j} = 0, \forall i \in I, j \in J, \tag{8}$$

with $I = \{1,\cdots,m-1\}, J = \{1,\cdots,n-1\}$. For testing whether E significantly affects $\mu_{i,m}$, the hypothesis that the outcome i of R and the outcome m of R are indistinguishable can be written as (Anderson, 1984):

$$H_0 : \beta_{i,j} = 0, \forall j \in J. \tag{9}$$

If the hypothesis represented by equation (9) is rejected and

$$\exists\ a,b \in J, \beta_{i,a} \neq 0,\ \ \beta_{i,b} \neq 0 \text{ and } \beta_{i,a} \neq \beta_{i,b}, \tag{10}$$

then we can conclude that the outcome a of E and the outcome b of E have different effects on R. Next, the set S_i is constructed, which includes such a and b. All elements of $\bigcup_{i\in I} S_i$ have distinct effects on R. All elements of $J \setminus \bigcup_{i\in I} S_i$ have the same effect on R.

3.4 Results

The p-values of performing KWT on *daily counts* grouped by *month* and *day* respectively are all equal to zero. The p-values of performing KWT on *duration* grouped by *month*, *track* and *start* are all equal to zero, while the p-value of performing KWT on *duration* grouped by *day* is 0.1713. Based on these results, we can conclude that *month* and *day* have effects on *daily counts*, and *month*, *cause*, *track* and *start* have effects on *duration*. After that, we further tested whether the effects are different between outcomes of each influence factor and classify the outcomes with similar effects into one cluster. This is done by the optimal *k*-means clustering together with KWT. The gained clusters and corresponding within-cluster p-values is shown in Table 2.

Table 2: Within-cluster p-values of performing KWT on *daily counts* and *duration*

Responsive variables	Explanatory variables	Clusters	Within-cluster p-values
daily counts	*month*	Nov	1.0000
		Oct, Jul and Dec	0.7646
		Other months	0.5356
	day	Weekdays	0.0915
		Weekends	0.0752
duration	*month*	Jan, Feb and Mar	0.0929
		Other months	0.2367
	cause	Crew strike and modified timetable (e.g. winter timetable, summer timetable, etc.)	0.0500
		Extreme winter weather (e.g. snow and ice), flooding, defuse a bomb from World War II, catenary failure, etc.	0.0824
		Copper theft, maintenance work, track failure (e.g. slippery track), etc.	0.9889
		Collision with persons	1
		Switch failure, signal failure, excessive delays, etc.	0.7040
		Objects on tracks (e.g. animals, trees, cars, etc.), extended maintenance work, damaged rail bridge, etc.	0.7295
		Line-side fire, rolling stock breakdown, etc.	0.2525
		Large passenger demand, police investigation, etc.	0.8098
		Rescue services	1
	track	e.g. Amsterdam Centraal - Leiden Centraal, Den Dolder-Amersfoort	0.9357
		e.g. Amsterdam Centraal- Breda, Rotterdam-Vlaardingen	0.9852
		e.g. Haarlem - Leiden Centraal, Schiphol Airport - Utrecht Centraal	0.8733
		e.g. Lelystad Centrum - Schiphol Airport, Amersfoort - Schiphol Airport, etc.	0.4739
	start	3 a.m. and 4 a.m.	0.3225
		0 a.m.,1 a.m.,2 a.m. and 5 a.m.	0.4782
		6 a.m.	1
		Other hours	0.1742

Clearly, the larger the within-cluster p-value is, the more similar the effects originated from the outcomes within the same cluster will be. Note that, in Table 2, for each explanatory variable, the corresponding cluster shown at the upper side has worse effect than the cluster at the lower side. For example, during January, February and March, the disruption durations are usually longer than the ones during other months.

For *daily counts*, the effects from *month* are divided into three clusters. The first one includes only the month November, during which the number of disruptions occurring in one day are more than those of other months. Moreover, the cluster that results in secondary frequent disruptions includes July, October and December. According to the practice, we think such a clustering result is reasonable. Usually, during autumn like in November and October, strong winds, rain and leaf fall lead to more frequent track blockages and slippery tracks. During winter like in December, snow and ice results in more frequent switch and catenary failures, while in summer like in July, high temperature and prolonged rainfall brings more frequent disruptions such as buckled track and line-side fires (Network Rail, 2016a, 2016b).

For *duration*, the effects from *month*, *cause*, *track and start* are divided into two clusters, nine clusters, four clusters and four clusters respectively. For example, considering disruption causes, crew strike or updating modified timetable usually lead to the longest duration compared to other causes.

Besides performing KWT on *daily counts* and *duration*, with the outcomes of *month*/*day* as predictors, we perform MLR on *cause*, *track* and *start* separately and find that all outcomes of the temporal factors, i.e. *month* and *day*, are proven to have distinct effects at 5% significance level. However, unlike our anticipation, we found that *cause* does not affect *track* and *start* significantly. One possible reason is that the disruption causes that have the characteristics of *track-variation* and *start-variation* only account for a very small proportion of the historical data.

Based on aforementioned results, we give Table 3 to show the statistically significant influence factors (explanatory variables) of each random variable (response variable).

Table 3: The influence factors of each random variable

Response variable	**Explanatory variables**			
daily counts	*month*	*day*		
cause	*month*	*day*		
track	*month*	*day*		
start	*month*	*day*		
duration	*month*	*cause*	*track*	*start*

4 Disruption scenario simulation

In the previous section, the influence factors of each random variable are identified, which provides insight on the characteristics of current disruption practice. However, rather than recognizing the characteristics only, it is also important to take actions against disruptions, and to evaluate the actions beforehand with the performance indicator of system vulnerability reduction. For this purpose, a Monte Carlo simulation that is capable of measuring the system vulnerability is established, where the system vulnerability is quantified as the total disruption durations. Based on the simulation, the baseline scenario that reflects the current disruption practice is constructed and measured with its resulting system vulnerability. Moreover, the mitigation scenarios are constructed by implementing different mitigation strategies separately on the baseline scenario, and the corresponding system vulnerabilities are measured afterwards. Clearly, the system vulnerability difference between the mitigation scenario and the baseline scenario is the vulnerability reduction due to the mitigation strategy.

In this section, details of Monte Carlo simulation for our case are given. We first introduce how to generate the distributions of random variables, along with the sampling method. After that, we describe how to construct the baseline scenario and the mitigation scenarios. Finally, the simulation procedure is given.

4.1 Distribution generation and sampling

Although *daily counts* and *duration* (accurate to minutes) are numerical variables, they are not continuous but discrete. Therefore, we treat each particular value of *daily counts/duration* as one outcome. Here, an outcome of *daily counts* indicates an exact number of disruptions happening in one day and an outcome of *duration* indicates the exact minutes a disruption lasts. Clearly, all possible outcomes are mutually exclusive in one independent observation, so are the outcomes of *cause*, *track* and *start*. Thus, the categorical distribution $f\left(x|p\right)=\prod_{i=1}^{n}p\left(i\right)^{[x=i]}$ is used to describe the possible outcome of a random variable with the probability of each outcome separately specified. Here, x refers to the outcome and $\left[x=i\right]$ equals to 1 if the outcome i is true, and otherwise it equals to 0. Moreover, $p\left(i\right)$ represents the occurring probability of outcome i, which is calculated as the relative frequency of outcome i observed in historical data with $\sum_{i=1}^{n}p\left(i\right)=1$, assuming n mutually exclusive outcomes.

For each random variable, the calculation of p depends on historical disruption data. For example, according to the results of independence tests, we know that each month/day has a distinct effect on *track*. Thus, we need to calculate p under any possible combinations of month and day. As there are 12 months and 7 days of a week, there will be 84 combinations in total. For the combination like ‘January and Monday’, the probability of each outcome of *track* is the relative frequency of the outcome observed in the data of disruptions that occurred in January and exactly Monday. Here, one outcome of *track* refers to a particular track like ‘Amsterdam Centraal - Leiden Centraal’. In this paper, we only chose the tracks that have been disrupted as the outcomes of *track*. In other words, the tracks that have never been affected by disruptions are excluded. Such a principle is also applied to decide the outcome range of other random variables.

In the simulation, one pseudo-random number indicates one outcome of the random variable, and it is produced according to the given distribution of the random variable. The procedure of producing pseudo-random number is called sampling. In this paper, it is realized by $j^{*}=\arg\min_{j}\left(\sum_{i=1}^{j}p\left(i\right)-r_{\mathrm{u}}\geq 0\right)$ where r_{u} is produced from a uniform $\left(0,1\right)$ distribution and j^{*} represents the pseudo-random number. In such case, $\left[x=j^{*}\right]=1$ and $\left[x=i\right]=0,\forall i\neq j^{*},i,j^{*}\in\left\{1,\cdots,n\right\}$. In this paper, such a sampling method is realized by a build-in function of MATLAB.

4.2 Scenarios construction

The baseline scenario is the scenario that is capable of reflecting the current disruption practice. For this purpose, the probabilities that are used to form the distribution of each random variable are generated from the historical disruption data, exactly as we mentioned in the previous section.

As to the mitigation scenarios, they are based on the baseline scenario, but somehow modified by particular mitigation strategies. In this paper, we propose the strategies that can lower the disruption probabilities of tracks. Generally, such strategies can be realized by different actions regarding the disruption causes. For example, if *switch failure* is one of the causes resulting in frequent disruptions on a specific track, updating the track with improved switches could be a solution to reduce disruptions, i.e. reduce vulnerability. The following is a description of how we modify the baseline scenario in terms of the strategies.

During simulation, the strategy effect is reflected by modifying the probabilities of outcomes of random variables that are affected by the strategy. Under our proposed strategy, the random variables *track* and *daily counts* are affected. Moreover, we still consider the magnitudes of strategies by defining magnitude as the reduced proportion of disruption probability. The value of magnitude ranges from 0.1 to 1.

Suppose in the baseline scenario the disruption probability of track q is $p_{\text{track}}\left(q\right)$ and we perform d magnitude of the aforementioned strategy on track q. Then, under the mitigation scenario, the disruption probability of track q turns to $p^{*}_{\text{track}}\left(q\right)$ that can be computed by $p^{*}_{\text{track}}\left(q\right)=\left(1-d\right)\cdot p_{\text{track}}\left(q\right)$. If we have n_{track} tracks in total, the mitigated disruption probability of a track except track q can be computed by $p^{*}_{\text{track}}\left(j\right)=p_{\text{track}}\left(j\right)+d\cdot p_{\text{track}}\left(q\right)\cdot\dfrac{p_{\text{track}}\left(j\right)}{1-p_{\text{track}}\left(q\right)},\forall j\neq q,$ where $p_{\text{track}}\left(j\right)$ is the disruption probability of the track in the baseline scenario with $j\in\left\{1,\cdots,n_{\text{track}}\right\}$. This means that to gain the probability of track j under the mitigation scenario, i.e. $p^{*}_{\text{track}}\left(j\right)$, we add the reduced probability of track q, $d\cdot p_{\text{track}}\left(q\right)$, to the initial probability of track j, $p_{\text{track}}\left(j\right)$, according to the proportion of $p_{\text{track}}\left(j\right)$ versus the initial probabilities of all tracks except track q. Note that under any possible combinations of month and day, the corresponding p_{track} is updated according to the magnitude specified.

In addition, as the disruption probability of track q is reduced, the number of disruptions during a certain period must be less. This leads to a decreased number of disruptions happening in one day. Thus, the probabilities of outcomes of *daily counts* also need to be modified.

Suppose in the baseline scenario, the probability of i disruptions happening in one day is $p_{\text{daily}}\left(i\right)$. Then, the expected value of p_{daily} is $E_{\text{daily}}=\sum_{i=1}^{n_{\text{daily}}} i\cdot p_{\text{daily}}\left(i\right),$ if the maximum value of *daily counts* is n_{daily}. Here, E_{daily} represents the average number of disruptions happening in one day. Due to the mitigation strategy, E_{daily} is reduced to E^{*}_{daily} that can be calculated by $E^*_{\text{daily}} = E_{\text{daily}} \cdot \left(1 - d \cdot p_{\text{track}}\left(q\right)\right)$. Hence, p_{daily} should be modified to realize E^*_{daily}. We define each modified $p_{\text{daily}}\left(i\right)$ as $p^*_{\text{daily}}\left(i\right)$. In fact, there are many solutions of p^*_{daily} to realize a given E^*_{daily}. However, in this paper, we assume that for any i that is smaller than E_{daily}, $p^*_{\text{daily}}\left(i\right) > p_{\text{daily}}\left(i\right)$, while for any i that is equal or larger than E_{daily}, $p^*_{\text{daily}}\left(i\right) \le p_{\text{daily}}\left(i\right)$. For the first case, we define $\mu_s\left(i\right) = p^*_{\text{daily}}\left(i\right) - p_{\text{daily}}\left(i\right), \forall i < E_{\text{daily}}$. For the second case, we define $\mu_b\left(i\right) = p^*_{\text{daily}}\left(i\right) - p_{\text{daily}}\left(i\right), \forall i \ge E_{\text{daily}}$. Finally, we formulate the following model to get the values of any $\mu_s\left(i\right)$ and $\mu_b\left(i\right)$.

$$\text{minimize} \quad \sum_{i<E_{\text{daily}}} i \cdot u_s\left(i\right) + \sum_{i \ge E_{\text{daily}}} i \cdot u_b\left(i\right) \tag{11}$$

$$\text{subject to} \quad \sum_{i<E_{\text{daily}}} i \cdot u_s\left(i\right) + \sum_{i \ge E_{\text{daily}}} i \cdot u_b\left(i\right) \ge -d \cdot p_{\text{track}}\left(q\right) \cdot E_{\text{daily}} \tag{12}$$

$$\sum_{i<E_{\text{daily}}} i \cdot u_s\left(i\right) + \sum_{i \ge E_{\text{daily}}} i \cdot u_b\left(i\right) < 0 \tag{13}$$

$$\sum_{i<E_{\text{daily}}} u_s\left(i\right) + \sum_{i \ge E_{\text{daily}}} u_b\left(i\right) = 0 \tag{14}$$

$$u_s\left(i\right) \le 1 - p_{\text{daily}}\left(i\right), \qquad \forall i < E_{\text{daily}} \tag{15}$$

$$u_s\left(i\right) \ge 0, \qquad \forall i < E_{\text{daily}} \tag{16}$$

$$u_b \ge -p_{\text{daily}}\left(i\right), \qquad \forall i \ge E_{\text{daily}} \tag{17}$$

$$u_b\left(i\right) \le 0, \qquad \forall i \ge E_{\text{daily}} \tag{18}$$

with $i \in \left\{1, \cdots, n_{\text{daily}}\right\}$. Here, n_{daily} represents the maximum value of i, which can be defined as the maximum number of disruptions happened in one day before.

The formula displayed in equation (11) represents the value of E^*_{daily} minus E_{daily}, which should be close to $-d \cdot p_{\text{track}}\left(q\right) \cdot E_{\text{daily}}$. Therefore, we set the inequalities (12) and (13) to reach this purpose. In addition, to ensure that $\sum_i p^*_{\text{daily}}\left(i\right) = 1,$ equation (14) is set. As $p^*_{\text{daily}}\left(i\right)$ should not be negative, equation (15) and (17) are set. Moreover, according to the aforementioned assumption about $\mu_s\left(i\right)$ and $\mu_b\left(i\right)$, we set equation (16) and (18).

Above, the method of modifying relevant distributions regarding the mitigation strategy of lowering the disruption probabilities of tracks is introduced, which can also be applied to other mitigation strategies like lowering the disruption probabilities of causes. Since the method is universal, we only show the case of mitigating disruption probabilities of tracks in section 5.

4.3 Simulation procedure

In this paper, we construct a yearly disruption scenario that is formed by 365 daily disruption scenarios. To calculate the average yearly vulnerability of the entire system, we repeat 10000 times to simulate the yearly scenario. Below is the procedure of simulation.

Step 1: Set the Monte Carlo repetition counter $\alpha = 1$.

Step 2: For each day during January to December, according to the distribution of *daily counts* under the current month and day of week, we first produce one pseudo-random number r_{daily} that indicates the number of disruptions happening in the day. Then, we produce r_{daily} pseudo-random numbers from the distributions of *cause* under the current month and day of week. In like manner, we produce r_{daily} pseudo-random numbers for *track* and *start* respectively. In other words, we decide the cause, affecting track and start hour for each of the r_{daily} disruptions. After that, we produce r_{daily} pseudo-random numbers from different distributions of *duration*. Each of these distributions corresponds with the current month and the cause, affecting track and start hour of each disruption. In other words, we simulate the duration of a disruption according to the corresponding values of *month*, *cause*, *track* and *start*.

Step 3: After the simulation for every day of one year, update the Monte Carlo repetition counter $\alpha = \alpha + 1$. If $\alpha < 10000$, go to step 2, otherwise end the simulation.

5 Experiments

5.1 Vulnerability analysis of baseline scenario

In this paper, we choose the yearly disruption duration as the indicator of vulnerability. Clearly, the longer the duration is, the more serious the vulnerability will be. Based on the simulation results of baseline scenario, we filter the top 10 vulnerable tracks according to the yearly disruption durations. Details about the ten tracks can be found in Table 4.

Table 4: Information of the top 10 vulnerable tracks

Rank	Track	Yearly duration (min)	Percentage of yearly duration (%)	Cumulative percentage (%)
1	Schiphol Airport - Utrecht Centraal	13922	3.50	3.50
2	Amsterdam Centraal - Schiphol Airport	12581	3.16	6.66
3	Amersfoort - Schiphol Airport	11376	2.86	9.52
4	Rotterdam Centraal - Utrecht Centraal	10967	2.76	12.28
5	Den Haag Centraal - Utrecht Centraal	10309	2.59	14.87
6	Lelystad Centrum - Schiphol Airport	9900	2.49	17.36
7	Den Haag HS - Rotterdam Centraal	9280	2.33	19.69
8	Leiden Centraal - Schiphol Airport	8706	2.19	21.88
9	Rotterdam Centraal - Schiphol Airport	7795	1.96	23.84
10	's-Hertogenbosch - Utrecht Centraal	7574	1.90	25.74

In Table 4, the third column refers to the yearly disruption duration of each top 10 vulnerable track. The fourth column indicates the percentage of yearly disruption duration of each top 10 vulnerable track versus the yearly disruption durations of all tracks. The fifth column refers to the cumulative percentage of yearly disruption duration from the current track to the most vulnerable track. We can see that the top 10 vulnerable tracks account for 25.74% disruption durations to the entire railway system in one year.

Moreover, to see the geographical distribution, we highlight the top 10 vulnerable tracks (red color) in the railway networks (shown in Figure 1, the coordinates of tracks and stations are provided by Dekkers, 2016). We can see that most of these tracks are in the western areas of the Dutch railways. These areas are the places with high-density populations and large demand for daily motilities. It is not hard to imagine that long disruption durations there can lead to lots of passengers affected. Therefore, it is necessary to propose some mitigation strategies on these tracks for shortening the disruption durations. One way could be to lower the disruption probabilities of these tracks.

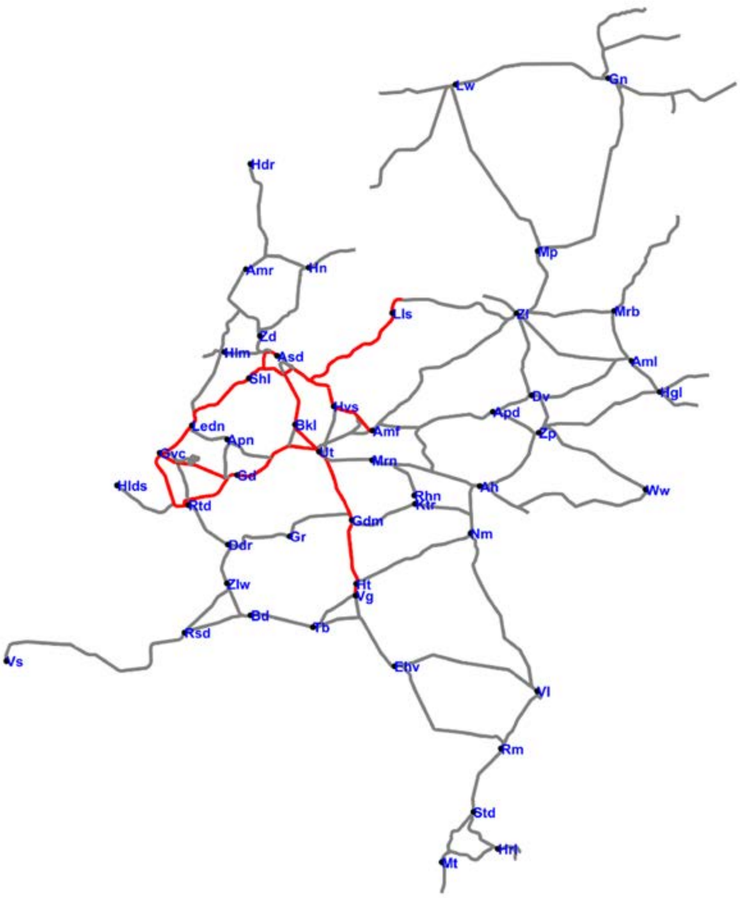

Figure 1: The top 10 vulnerable tracks (red color) in the Dutch railways

5.2 Vulnerability analysis of mitigation scenarios

For each of the top 10 vulnerable tracks that are derived from the baseline scenario, we lower its disruption probability from 10% off to 100% off, respectively. In other words, the reduced proportion of disruption probability ranges from 0.1 to 1. Thus, 100 mitigation scenarios are constructed of each a mitigation strategy with a specific magnitude is implemented on a particular track. The effectiveness of each mitigation strategy is quantified with the indicator of percentage of vulnerability reduction. The percentage of vulnerability reduction is the rate of difference between the baseline scenario vulnerability and the mitigation scenario vulnerability versus the baseline scenario vulnerability. Here, the vulnerability of either baseline scenario or mitigation scenario refers to the yearly disruption durations of all tracks in the railway system. Figure 2 displays the effectiveness of each mitigation strategy. The points in each successive line represent the percentages of vulnerability reduction by reducing the disruption probabilities of different tracks under a particular mitigation magnitude.

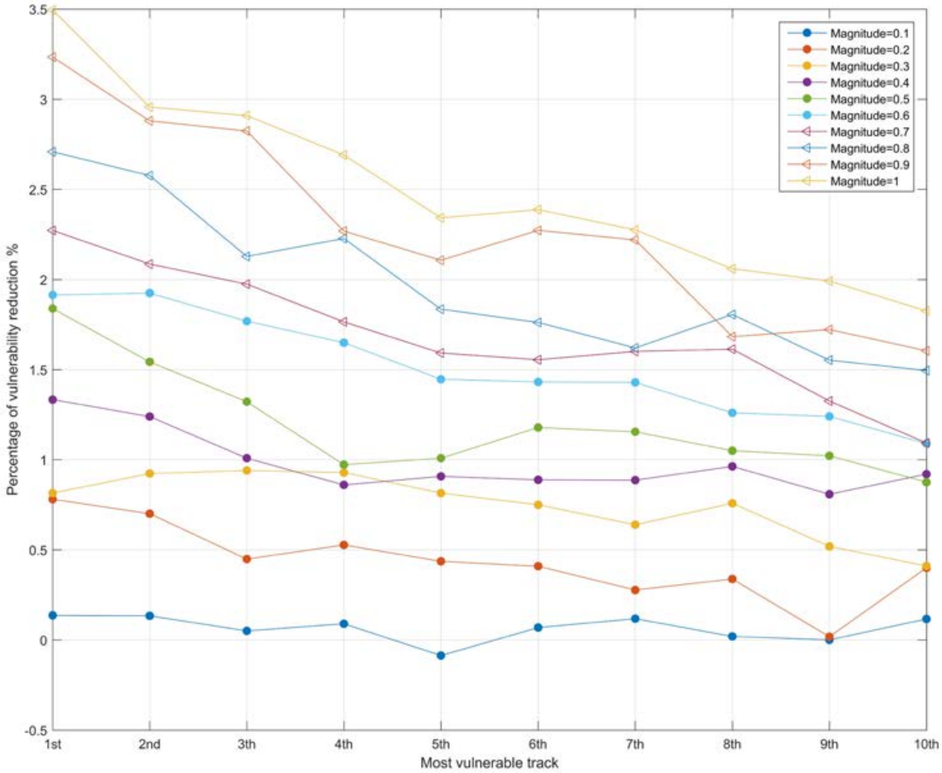

Figure 2: Percentage of vulnerability reduction by mitigating the top 10 vulnerable tracks under ten different mitigation magnitudes

From Figure 2, we can see that all mitigation strategies lead to positive mitigation effects except the one that lowers 10% disruption probability of the 5th vulnerable track Den Haag Centraal - Utrecht Centraal. Under this strategy, the percentage of vulnerability reduction is -0.09%, which means the system vulnerability increases although we initially intended to decrease it. A reason could be that the mitigation magnitude is too small for this track since the reduced vulnerability through mitigating this track is not enough to offset the increased vulnerability by other tracks due to duration variation. However, because "-0.09%" is very close to zero, we can also think that the system vulnerability is at the same level as the one of the baseline scenario.

In addition, we can see that at the same mitigation magnitude, the effectiveness of one strategy does not strongly depend on how vulnerable the track is that this strategy implements on. For example, at the mitigation magnitude of 0.4, the effectiveness of mitigating the 10th vulnerable track is better than the ones of mitigating the 9th, 7th, 6th, and 4th vulnerable tracks. One reason could be that the differences between the percentages of yearly durations of these tracks is small (less than 0.86%), thus the difference on the effectiveness of mitigating these tracks is also small. In such a case, if there were big gaps between the financial costs of mitigating these tracks, it would be beneficial to choose the least cost strategy since the resulting vulnerability reduction is not much different.

However, when compare the 10th vulnerable track with the most vulnerable track, we can find that in each mitigation magnitude the effectiveness of mitigating the most vulnerable track is better than mitigating the 10th vulnerable track. One reason could be that the difference between the percentages of yearly durations of the two tracks is large enough, i.e. 1.6%. When decision makers need to decide which of the two tracks should be mitigated while the financial costs are not much different, it is definitely wise to choose to mitigate the more vulnerable track.

6 Conclusions

In this paper, based on historical disruption data of the railways in the Netherlands, we establish a simulation model based on which system-based vulnerability can be measured and mitigation strategies can be evaluated. There are three highlights of this paper. First, independence tests for random variables that are included in the simulation model were performed, which can help to gain clear insights on the characteristics of the current disruption practice and ensure that the established simulation can reflect the reality well. Second, we focus on the railway system at a national level and consider different influence factors towards the system. Based on the constructed baseline scenario, besides vulnerable tracks, we can also derive the vulnerable months, vulnerable days, vulnerable causes, etc. of the entire system, although we did not show this in this paper. The last but not the least, based on the established simulation model, it is possible to evaluate different strategies from different aspects once the necessary data like the information of investment costs is available.

Acknowledgements

This work is partly supported by the program of China Scholarship Council (No. 201507000056).

References

Adjetey-Bahun, K., Birregah, B., Châtelet, E., and Planchet, J. L., 2016. “A model to quantify the resilience of mass railway transportation systems”, *Reliability Engineering & System Safety*, vol. 153, pp. 1-14.

Anderson, J. A., 1984. “Regression and ordered categorical variables”, *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 46(1), pp. 1-30.

Angeloudis, P., and Fisk, D., 2006. “Large subway systems as complex networks”, *Physica A: Statistical Mechanics and its Applications*, vol. 367, pp. 553-558.

Caschili, S., Medda, F.R., Reggiani, A., 2015. “Guest editorial, Resilience of Networks”, *Transportation Research Part A: Policy and Practice*, vol. 81, pp. 1-3.

Cats, O., and Jenelius, E., 2014. “Dynamic vulnerability analysis of public transport networks: mitigation effects of real-time information”, *Networks and Spatial Economics*, vol. 14(3-4), pp. 435-463.

Cats, O., Yap, M., and van Oort, N., 2016. “Exposing the role of exposure: Public transport network risk analysis”, *Transportation Research Part A: Policy and Practice*, vol. 88, pp. 1-14.

Dekkers, J., 2016. “Railway stations and railway tracks”, SPINLab, VU Amsterdam.

Deng, Y., Li, Q., and Lu, Y., 2015. “A research on subway physical vulnerability based on network theory and FMECA”, *Safety science*, vol. 80, pp. 127-134.

Derrible, S., and Kennedy, C., 2010. “Characterizing metro networks: state, form, and structure”, *Transportation*, vol. 37, pp. 275-297.

D’Lima, M., and Medda, F., 2015. “A new measure of resilience: An application to the London Underground”, *Transportation Research Part A: Policy and Practice*, vol. 81, pp. 35-46.

Golightly, D., Dadashi, N., Sharples, S., and Dasigi, M., 2013. “Disruption management processes during emergencies on the railways”, *International Journal of Human Factors and Ergonomics*, vol. 2, pp. 175-195.

Hong, L., Ouyang, M., Peeta, S., He, X., and Yan, Y., 2015. “Vulnerability assessment and mitigation for the Chinese railway system under floods”, *Reliability Engineering & System Safety*, vol. 137, pp. 58-68.

Johansson, J., Hassel, H., and Cedergren, A., 2011. “Vulnerability analysis of interdependent critical infrastructures: case study of the Swedish railway system”, *International journal of critical infrastructures*, vol. 7(4), pp. 289-316.

Kurant, M., and Thiran, P., 2006. “Extraction and analysis of traffic and topologies of transportation networks”, *Physical Review E*, vol. 74, pp. 036114.

Long, J. S., 1997. “Regression models for categorical and limited dependent variables”, *Advanced quantitative techniques in the social sciences*, vol. 7, pp. 50-52.

Mattsson, L. G., and Jenelius, E., 2015. “Vulnerability and resilience of transport systems–A discussion of recent research”, *Transportation Research Part A: Policy and Practice*, vol. 81, pp. 16-34.

Mood, A. M., Graybil, F.A., and Boes, D.C., 1974. *Introduction to the Theory of Statistics*, 3rd Edition, McGraw Hill.

Murray, A. T., Matisziw, T. C., and Grubesic, T. H., 2008. “A methodological overview of network vulnerability analysis”, *Growth and Change*, vol. 39(4), pp. 573-592.

Network Rail, 2016a. Summer weather - Delays explained. Retrieved November 15, 2016, from http://www.networkrail.co.uk/timetables-and-travel/Delays-explained/Winter-weather/

Network Rail, 2016b. Winter weather - Delays explained. Retrieved November 15, 2016, from http://www.networkrail.co.uk/timetables-and-travel/delays-explained/summer-weather.aspx

Reggiani, A., Nijkamp, P., and Lanzi, D., 2015. “Transport resilience and vulnerability: the role of connectivity”, *Transportation Research Part A: Policy and Practice*, vol. 81, pp. 4-15.

Rodríguez-Núñez, E., and García-Palomares, J. C., 2014. “Measuring the vulnerability of public transport networks”, *Journal of transport geography*, vol. 35, pp. 50-63.

Rose, A. Z., 2009. *Economic resilience to disasters*. CARRI Report No. 8, Community and Resilience Institute.

Schipper, D., Gerrits, L., and Koppenjan, J. F., 2015. “A dynamic network analysis of the information flows during the management of a railway disruption”, *European Journal of Transport and Infrastructure Research*, vol. 15(4), pp. 442-464.

Wang, H., and Song, M., 2011. “Ckmeans. 1d. dp: optimal k-means clustering in one dimension by dynamic programming”, *The R Journal*, vol. 3(2), pp. 29-33.
Zilko, A. A., Kurowicka, D., and Goverde, R. M. P., 2016. “Modeling railway disruption lengths with Copula Bayesian Networks”, *Transportation Research Part C: Emerging Technologies*, vol. 68, pp. 350-368.