Impact of railway disruption predictions and rescheduling on passenger delays

Ghaemi, Nadjla; Zilko, Aurelius; Yan, Fei; Cats, Oded; Kurowicka, Dorota; Goverde, Rob

DOI
10.1016/j.jrtpm.2018.02.002

Publication date
2018

Document Version
Accepted author manuscript

Published in
Journal of Rail Transport Planning and Management

Impact of Railway Disruption Predictions and Rescheduling on Passenger Delays

Nadjla Ghaemi[a], Aurelius A. Zilko[b], Fei Yan[a,*], Oded Cats[a], Dorota Kurowicka[b], Rob M.P. Goverde[a]

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

Abstract

Disruptions such as rolling stock break-down, signal failures, and accidents are recurrent occurrences during daily railway operation. Such events disrupt the deployment of resources and cause delay to passengers. Obtaining a reliable disruption length estimation can potentially reduce the negative impact caused by the disruption. To investigate the impact of the disruption length estimates on the rescheduling strategy and the resulting passengers delays, this research presents a framework consisting of three models: a disruption length model, short-turning model and passenger assignment model. The framework is applied to a part of the Dutch railway network. The results show the effects of the different disruption length estimates on the number of affected passengers, generalized travel time and number of passengers rerouting and transferring.
Keywords: railway disruption, prediction, dependence model, short-turning, passenger assignment

- Railway disruption length is modelled as a probability distribution.
- Short-turning measures are determined based on disruption length predictions.
- Consequences on passenger costs of prediction accuracy and rescheduling are assessed.
- The integrated framework is applied to a case study on a part of Dutch railway network.
- By incorporating a reliable disruption length, we can achieve a faster transition to the original timetable.

1. Introduction

Railway operations are repeatedly disturbed by events such as technical and mechanical failures of infrastructure and rolling stock, traffic accidents and malicious attacks. Railway timetables are usually designed to compensate for some delays by including buffer times. However in case of long disruptions and infrastructure unavailability, these buffer times are ineffective and a new timetable should be designed with adjusted train services.

*Corresponding author
Email address: F.Yan@tudelft.nl (Fei Yan)

The decisions regarding the rescheduling of resources need to be carefully communicated between the railway infrastructure manager, the train operators and other involved actors to ensure the safety and feasibility of the plan. To facilitate the challenging task of the traffic controllers in such cases, many countries use contingency plans designed specifically for disruption scenarios (Chu and Oetting (2013)). In The Netherlands these plans are manually designed by expert traffic controllers and are specific for each location and disruption case regardless of disruption length.

The proposed solution in the contingency plans is based on the timetable (basic hourly pattern) and the capacity of the disrupted location. The solution is instructing the traffic controllers how to deal with the disrupted traffic by determining cancelled services, short-turned or rerouted services and the services that are allowed to operate in the original timetable. Short-turnings are particularly beneficial for isolating the disrupted area, while maintaining services on both sides of the disruption. This implies short-turning the arriving trains to the last station before a disruption (on both sides) and continue service in the opposite direction. In case of short-turning, the stations where the short-turning should occur as well as the platform and departure times also need to be determined. By means of simulating the short-turning Coor (1997) concluded that this measure is most efficient in case of large disruptions.

The traffic level during a disruption can be conceptualized as a process that resembles a bathtub (Ghaemi et al. (2016)). As is shown in Figure 1 some services are cancelled due to the disruption. This reduction in train traffic starts immediately after the disruption occurs. Three phases can be identified within the disruption period. In the first phase the traffic controllers are facing lots of uncertainty regarding the disruption location, cause and most importantly the estimation of disruption length. In the Dutch railway operation, there are rough estimates for the length of different kinds of disruptions. These estimates are used to inform the passengers about the expected disruption length. Once the location is known, the traffic controllers retrieve the relevant contingency plan designed for that specific location. In case there is a need for repairing the disrupted infra, the repairmen are sent to the field to deal with the cause of disruption. The repairmen estimate the required time for resolving the problem and report it to the traffic controllers. In case the cause of disruption is resolved earlier than the informed disruption length, the operation is not resumed until the communicated time has elapsed. In case the disruption takes longer than the initially estimated length, the passengers are updated with a new disruption length. Throughout this paper the estimates that are longer than the actual disruption length are referred to as pessimistic and those that are shorter than the real length are referred to as optimistic estimates.

Once the communicated length has passed and the cause of the disruption has been removed, the traffic can resume and recover back to original level. The first and third phases are called transition phases where the operation has a transition from the original timetable to the disruption timetable and vice versa. The contingency plan corresponds to the second phase of the bathtub model where there is a stable though decreased level of traffic. Since there is no reliable disruption length estimation, the contingency plans do not provide any insight regarding the third phase where there is a transition from the disruption timetable to the original timetable. In practice the effects of different disruption length estimations on the rescheduled timetable during the three phases of disruption, and consequently the affected passengers, are remain unknown.

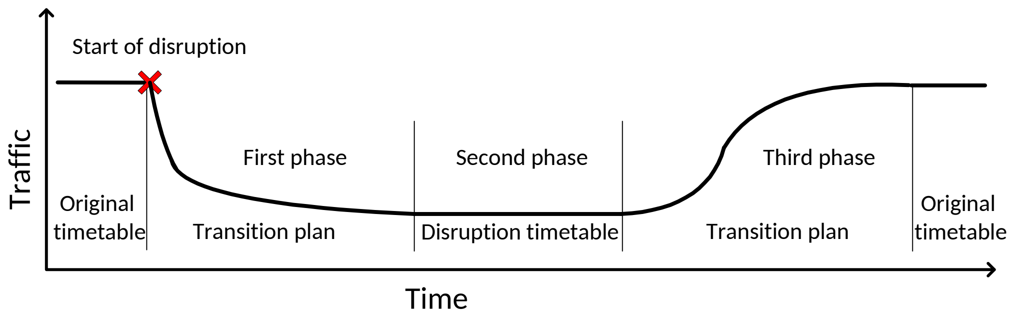


Figure 1: The service level during disruptions

While the bathtab model is widely known and used to conceptualize traffic states during disruptions, only limited research efforts have been devoted to analyzing and modeling railway disruption management. Ghaemi et al. (2017) provide a review of rescheduling models for disruptions and concluded that only a few studies considered all three phases. Examples of such models were developed by Veelenturf et al. (2016) and Nakamura et al. (2011). However, the disruption length is assumed to be known in advance and passenger delays are not taken into consideration. Meng and Zhou (2011), Yang et al. (2013) and Yang et al. (2014) model the third phase by taking into account the uncertainty of the disruption length. Besides the lack of a timetable for the first and second phase, their approaches do not explicitly model the influencing factors on the disruption length. In particular Yang et al. (2013) and Yang et al. (2014) model the disruption length as a fuzzy variable that reflects the estimation by expert judgment. Hirai et al. (2006) and Zhan et al. (2015) focus on the first phase where trains need to stop before the disruption area. However both approaches disregard the uncertainty regarding the disruption length and the consequences for passenger delay. De-Los-Santos et al. (2012) and Cats (2016) introduced indexes for measuring network robustness by measuring the effects of disruptions in terms of changes in passengers travel times. Yet the defined indexes are not suitable for real-time application where the disruption length is not yet known and might get updated frequently. Canca et al. (2016) propose a short-turning model to accommodate extra demand induced by a disruption in the tactical level, but the disruption length is not taken into account. Zhan et al. (2016) and Nielsen et al. (2012) incorporate the uncertainty of disruption length for rescheduling through a rolling horizon framework. Their approaches do not include the impacting factors on the disruption length. Moreover, the impact of the rescheduled timetable on the passengers is disregarded. Kumazawa et al. (2008) developed a rescheduling model considering passenger inconvenience. They do not provide any information regarding the length of the disruption. Cats and Jenelius (2014) analyzed the impacts of disruptions on passenger welfare using a non-equilibrium passenger loading model. They quantified the value of real-time information provision in case of disruption.

As mentioned above the contingency plans do not provide any instructions regarding the transition phases and the proposed solution is given independently of the disruption length estimation. In reality, the disruption length is very uncertain and it is unknown how long a disruption will last. Having a reliable disruption length prediction is instrumental in devising the rescheduling measures and thus for achieving a smooth and fast transition to the original timetable in the third phase of the bathtub model. To tackle this problem, Zilko et al. (2016) represent the disruption length as a probability distribution. Several determinants of disruption length are considered from which the joint distribution between disruption length and these factors is constructed with a Copula Bayesian Network. Having the joint distribution enables the traffic controller to obtain a conditional distribution of disruption length when a disruption occurs by conditioning the model on the observed values of the influencing factors.

A disruption length prediction is derived from this conditional distribution. Having a probability distribution enables the traffic controller to choose different values of prediction corresponding to different quantiles of the distribution. If the controller is optimistic about the disruption length, a lower quantile of the distribution can be chosen. Alternatively, a higher quantile of the distribution can be chosen. Without a reliable length estimation, there is no support for the traffic controllers for the third transition phase.

In this paper a framework is proposed to investigate the effects of different choices of predictions on the rescheduling measures and consequently the passengers delay. The framework integrates three components:

- Estimating the disruption length.
- Rescheduling the timetable given the estimated disruption length.
- Measuring the passenger delays based on the computed schedule.

In the remaining of the paper, the three components are described in more details in Section 2. The modeling framework is then demonstrated using an application to part of the Dutch railway network in Section 3. Section 4 concludes with practical implications and directions for future studies.

2. Framework

The three components are explained in detail under sub-sections 2.1 to 2.3 and the interaction between these components are shown in sub-section 2.4.

2.1. *The Disruption Length Model*

This modeling component has been first introduced by Zilko et al. (2016) and is adopted in this paper where this technique is integrated into a real-time prediction and mitigation framework.

The disruption length is divided into two sequential stages: the *latency time* and the *repair time*. The latency time is the length of time the mechanics need to get to the disrupted site while the repair time is the length of time they need to repair the problem.

The joint distribution between the latency time and repair time with the influencing factors is constructed using a Copula Bayesian Network. As a prototype, a Copula Bayesian Network model is constructed for disruptions caused by track circuit (TC) failures in the Netherlands. These disruptions are affected by eight influencing factors: (1) contract type, (2) distance to the nearest mechanics' workshop, (3) distance to the nearest level crossing, (4) whether or not the disruption is during the mechanics' contractual working time, (5) whether or not the temperature is above 25^{o}C, (6) whether or not the disruption occurs during rush hour, (7) whether or not there is another disruption going on at the same time, and (8) the cause of disruption.

The Copula Bayesian Network uses a Bayesian Network (BN) to 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. Figure 2(a) presents the TC disruption length model. The eleven nodes in the structure correspond to the ten variables in the model and the variable "Disruption Length" which is the sum of the latency and repair times. 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.

The joint distribution of the ten variables is constructed using *copula*. A copula is the n-dimensional joint distribution in the unit hypercube of n uniform random variables. It is a popular tool to model the dependence between variables (see, e.g. Nelsen (2006) and Joe (2014)). The theorem of Sklar states that any cumulative distribution function $(X_1, \dots, X_n)$, denoted as $F_{1,\dots,n}$, can be rewritten in terms of the corresponding copula C as

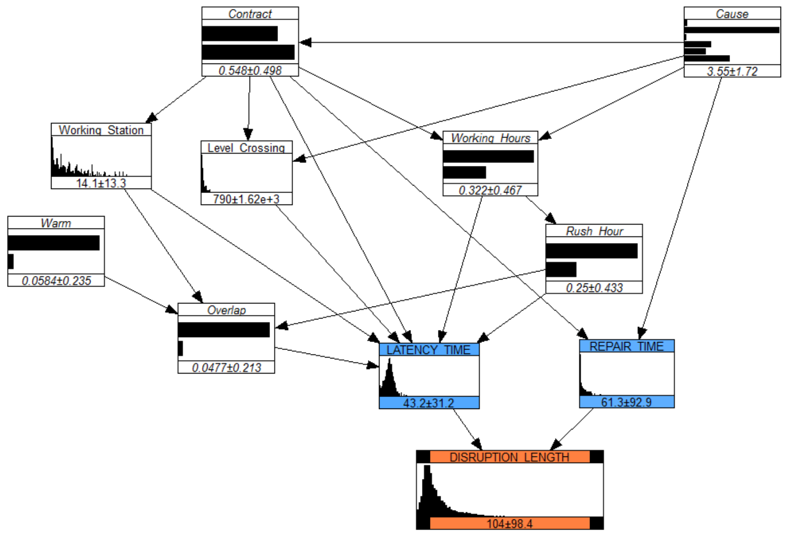
$$F_{1,\dots,n}(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)) \tag{1}$$

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

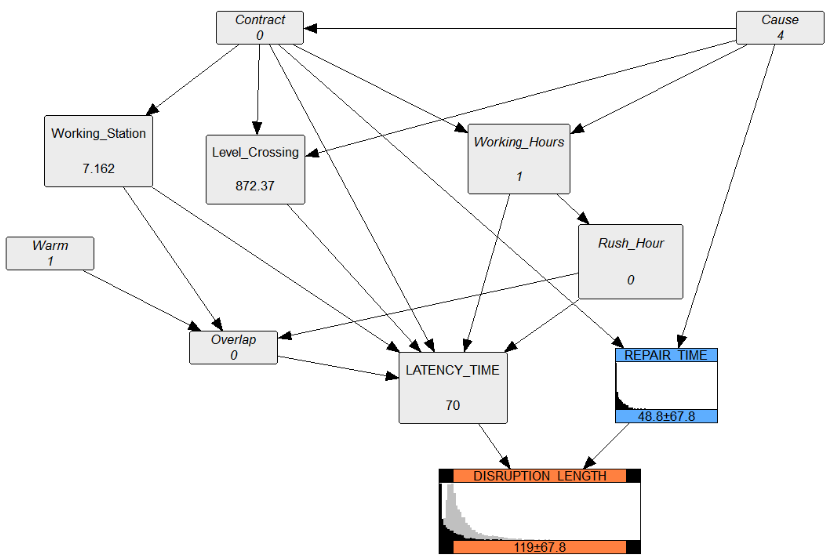
There are many different copula families. This approach used the multivariate Normal, or Gaussian, copula C_Σ to construct the TC disruption length model. This copula is defined as

$$C_\Sigma(u_1, \dots, u_n) = \Phi_\Sigma(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n)) \tag{2}$$

where Φ^{-1} denotes the inverse cumulative distribution of a univariate standard normal distribution and Φ_Σ



(a) The unconditional Track Circuit BN



(b) A conditional Track Circuit BN

Figure 2: The Track Circuit BN.

denotes the cumulative joint distribution of a multivariate normal distribution with a mean value of zero and correlation matrix Σ. The parameter Σ of the TC disruption length model corresponds to the arcs in the BN structure in Figure 2. This copula is of interest because it allows conditionalization to be computed rapidly, a very useful feature in the real-time decision making environment of the traffic control centers.

The copula parameter Σ is computed using the maximum likelihood approach using ProRail's SAP database. The constructed model, as presented in Figure 2, was validated using empirical data with the model being able to obtain a good conditional distribution of disruption length. Figure 2(a) presents the unconditional BN, i.e. the TC BN model when no information is available. Figure 2(b) presents the conditional BN when information about the influencing factors is available. Notice that the distribution of disruption length changes.

However, conditionalization on the variable Cause can only be performed after the mechanics diagnose the problem and find the cause. This time is called the "diagnosis time". Unfortunately, the diagnosis time is not available in the data and is actually included in the definition of repair time. The data does not provide any information to allow decoupling the diagnosis time from the actual repair time. In practice, usually the mechanics are given 15 minutes to diagnose the problem after they arrive at the site. Therefore, in this paper we assume that diagnosis time always takes 15 minutes and the cause is always found in this time.

2.2. The Short-turning Model

The short-turning model is designed to cope with the disruption cases with complete blockages where no train can use the infrastructure during the disruption period. In such cases, all the trains running towards the disruption location should short-turn before the blockage. The short-turning model is an extended version of the Mixed Integer Linear Program introduced in Ghaemi et al. (2016) that computes the optimal short-turning time and station for each approaching train service. We briefly discuss the optimal short-turning problem while focusing on the recovery plan which is the extended part. Prior to the problem description, it is necessary to describe the definition of services used in this approach. In this formulation a service is a trip between a departure and arrival (either with or without dwell time). Thus, a train line that may have multiple stops, consists of an ordered set of services performed by a train. Each service is denoted as $v^i_{l,n}$ where i indicates the order of the service within the operational line, l is the line number that determines the stops and n determines the time of operation.

The short-turning model is an assignment model that allocates the arriving trains to the scheduled departures in the opposite direction. In the example illustrated in Figures 3 to 5 service $v^i_{l,n}$ can either short-turn in station a' to serve $v^j_{l,m}$,$v^j_{l,o}$ or $v^j_{l,r}$ or it can continue as service $v^{i+1}_{l,n}$ and short-turn to serve $v^{j-1}_{l,m}$,$v^{j-1}_{l,o}$ or $v^{j-1}_{l,r}$ in station a. Obviously in case $v^{i+1}_{l,n}$ short-turns to serve $v^{j-1}_{l,m}$, this departure would be delayed. The reason is that the arrival time of train $v^{i+1}_{l,n}$ is scheduled after the departure of service $v^{j-1}_{l,m}$ which is shown in grey. These short-turning possibilities are shown by the red arcs for the arriving service $v^i_{l,n}$. The output of the model is the short-turning pattern which refers to the selected arc that determines the departure time and location of an arriving train.

If the short-turning of service $v^i_{l,n}$ occurs in station a' to serve $v^j_{l,m}$ as shown in Figure 4 then the service $v^{i+1}_{l,n}$, all of the associated short-turning patterns in station a, and $v^{j-1}_{l,m}$ should be cancelled. The passengers

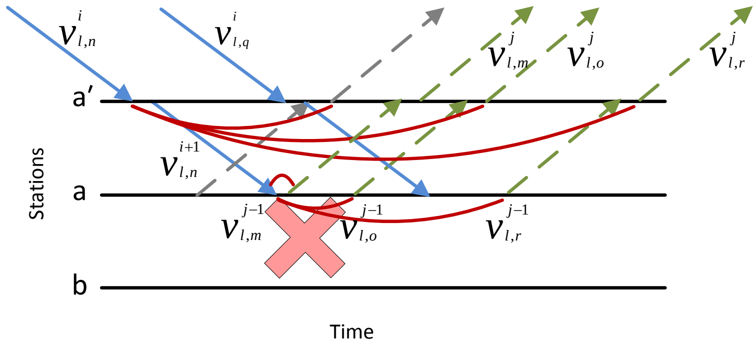


Figure 3: The possible short-turning patterns

travelling between these stations will be affected by these cancellations. Notwithstanding, early short-turning can reduce the delay propagation to the opposite stations and overall can result in less delay.

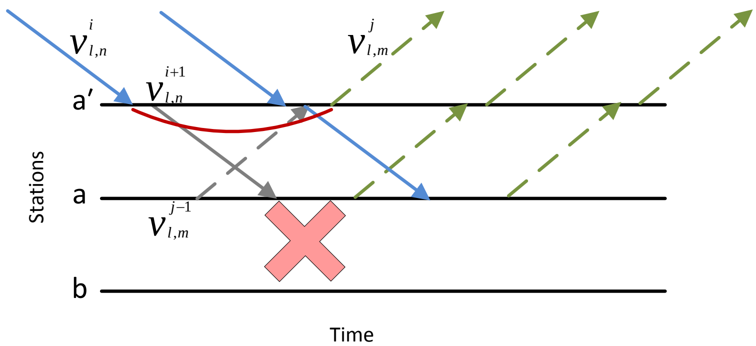


Figure 4: The cancellation resulting from short-turning in station a'

In the preprocessing phase we define the services that operate in the disruption area during the disruption period. Those services with both planned departure and arrival within the disruption location and period are cancelled. In case there is a service towards the disrupted location and departs before the start of disruption is not cancelled and it is assumed to be the last running service in the blockage section before the disruption. In case the departure of the approaching services towards the disrupted area is within the disruption period and the arrival is close to the end of disruption, then it might be better to wait until the disruption is over and then depart on the original route. The main decision for the final service is whether to continue short-turning or wait until the disruption is over and continue on the original route. This decision concerns those trains that arrive close to the end of the disruption period. In case there are more arriving services than the number of scheduled departures in the opposite direction within the disruption period, then the extra train service (due to the periodicity of the timetable, there is usually one extra train service) should wait until the disruption is over and start using the recently resolved blockage. If there are more scheduled departures in the opposite direction, then the final arriving train can either short-turn or wait until the disruption is over and continue in the same direction. Based on this decision, there would be cancellation either for the scheduled departure in the opposite direction, or the scheduled departure in the same direction. The short-turning model computes the disruption timetable for the second phase of disruption and the transitions. With a disruption length prediction we are able to plan the recovery phase where trains are able to continue their original routes and start using the track after the blockage ends.

A set of candidate transition services that might be canceled needs to be defined. As shown in Figure 5, service $v_{l,n}^{i+2}$ is cancelled since its departure and arrival is within the disruption period. The transition services ($v_{l,q}^{i+2}$ and $v_{l,r}^{j-2}$ shown by dash-dotted arrows) are planned to depart before the end of disruption and arrive after the disruption. These services are either cancelled or wait until the disruption is over and continue on their original route. In other words, for transition services the possibility for operating on the original route with a possible delay is considered. In this example, in case service $v_{l,r}^{j-2}$ is cancelled, the

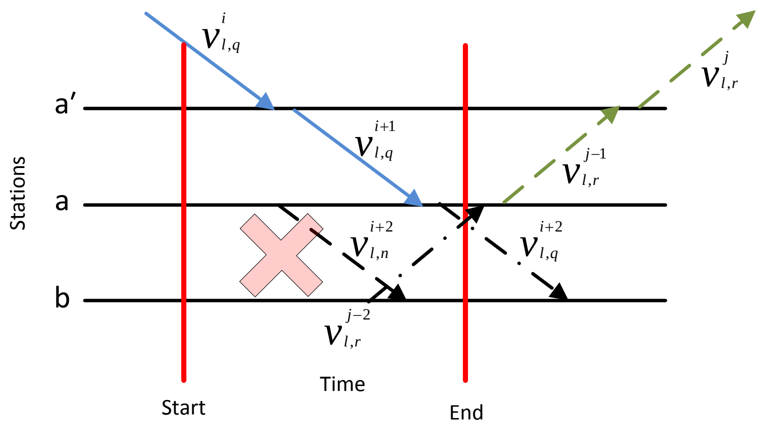


Figure 5: The transition services shown by dash-dotted arrows ($v_{l,q}^{i+2}$ in one direction and $v_{l,r}^{j-2}$ in the other direction)

following service $v_{l,r}^{j-1}$ is either performed by a short-turning in station a or should also be cancelled. But if the service $v_{l,r}^{j-2}$ is not cancelled, it should depart after the end of disruption and this would introduce a delay to all of the following services performed by this train. The decision whether to cancel service $v_{l,n}^{i}$ is modelled as a binary variable $c_{v_{l,n}^{i}}$.

The main objective of the short-turning model (3) is to minimize the departure and arrival delay ($d_{v_{l,n}^{i}}^{d}$, $d_{v_{l,n}^{i}}^{a}$) and the number of cancelled services. The penalties assigned to departure or arrival delay and a cancelled service $v_{l,n}^{i}$ are denoted by $\omega_{v_{l,n}^{i}}^{d^d}$,$\omega_{v_{l,n}^{i}}^{d^a}$ and $\omega_{v_{l,n}^{i}}^{c}$, respectively.

$$\min \sum_{v_{l,n}^{i} \in V} (\omega_{v_{l,n}^{i}}^{d^d} \cdot d_{v_{l,n}^{i}}^{d} + \omega_{v_{l,n}^{i}}^{d^a} \cdot d_{v_{l,n}^{i}}^{a} + \omega_{v_{l,n}^{i}}^{c} \cdot c_{v_{l,n}^{i}}). \tag{3}$$

2.3. Passenger Flow Distribution Model under disruptions

A passenger loading model is developed to represent how passengers are distributed over the network in the event of a disruption. The passenger flow distribution model allows assessing the impact of alternative scenarios on passengers by calculating passengers' total travel delay, transfer times and the number of transfers compared to the scheduled timetable. The disruption length model generates predictions on the length of the blockage time with updated predictions from time to time, and the short-turning model finds an adjusted timetable according to the given disruption length. Based on the given rescheduled timetable, the passenger flow model generates alternative travel routes for each pair of origin and destination (OD) and assigns passengers to selected routes from the corresponding alternative routes. Passengers face different route choice conditions during the course of the disruption. In particular, three phases, as illustrated in Figure 6. Under normal operations, route choice is based on the planned timetable. When a disruption occurs, a rescheduled timetable is generated based on the predicted disruption length, passengers are made informed of the new departure times and will thus choose their route based on the prevailing conditions. Hence, passengers choose from a new set of alternative routes based on the rescheduled timetable. Furthermore, passengers who have already boarded a train might need to reroute as a consequence of the disruption and the rescheduling. Additional updates to the disruption length predictions may result in additional rescheduling of train services and consequently the rerouting of passengers. Finally, when the disruption has ended, delays might still occur as the service recovers back to the original timetable.

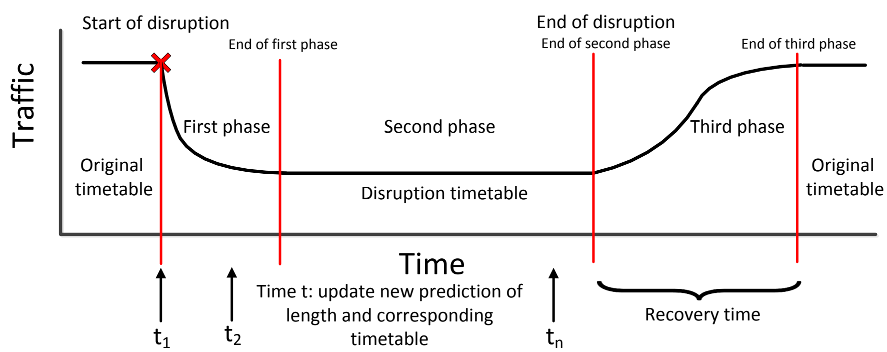


Figure 6: Transition points in the process of dynamic passenger loading

The dynamic passenger loading model consists of two steps. In the first step, alternative routes for each origin-destination (OD) pair are generated, followed by a probabilistic route choice model based on the framework of discrete random utility models. The latter determines the share of passengers that are assigned to each route. The overall workflow, including model input and output (parallelogram) and the main modules (rectangles), is depicted in Figure 7 and described in the following sub-sections.

2.3.1. Alternative route generation

Given a timetable, either the original or the outcome of the rescheduling model, the alternative route module generates a set of alternatives from which individuals travelling between a given pair of OD will

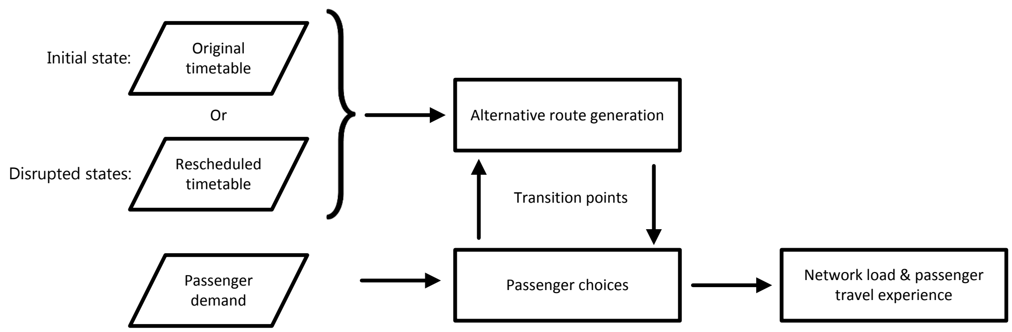


Figure 7: Passenger loading model under disruption

choose from. The choice-set is generated by iteratively searching for routes with an increasing number of transfers. A forward search algorithm is applied where transfer alternatives further downstream are examined by considering all scheduled train trips and their corresponding stopping pattern and scheduled arrival and departure times. For indirect alternatives, the transfer time must be within a user-defined acceptable range, $[\gamma_{trans}^{min}, \gamma_{trans}^{max}]$: satisfying the minimum transfer time – to ensure a sufficient time between train arrival and the next train departure, and a maximum transfer time – to avoid excessively long transfer times. In addition, indirect alternatives that induce a detour that exceeds by a user-defined ratio of γ_{detour}^{max} are removed from the choice-set as well as alternatives that are dominated by other alternatives when considering the number of transfers, in-vehicle time, transfer time and service level (i.e. intercity vs. regional). For each of the alternatives obtained in this process, the following attributes are stored along with the route itinerary: total travel time, number of transfers, total in-vehicle time and transfer time. These attributes are then used in the following choice step.

2.3.2. Passenger assignment

A multinomial logit (MNL) choice model is applied for route choice, to calculate the share of passengers travelling along each route alternative, shown in equation (4).

$$P_{ijk} = \frac{exp(-\theta\widetilde{t}_{ijk})}{\sum_{k\in R_{ij}} exp(-\theta\widetilde{t}_{ijk})} \tag{4}$$

P_{ijk} is the share of passengers that choose route k when travelling between i to j and R_{ij} is the set of alternative routes from i to j. Route k consists of an ordered set of legs denoted by a sequence of stations, $k = (s_{k,1}, s_{k,2}, ..., s_{k,|k|})$ and $s \in S$ where S is the set of stations in the network. $\widetilde{t}_{ijk}$ is the total generalised cost of route k for a given OD pair ij. θ is the logit scale factor for route choice. More specifically, passenger generalised travel time consists of waiting time, in-vehicle time, transfer time and other fixed penalties, which can be described as follows.

$$\widetilde{t}_{ijk} = \lambda^w \cdot t^w_{s_{k,1}} + \sum_{v=1}^{|k|-1} \lambda^{in} \cdot t^{in}_{s_{k,v},s_{k,v+1}} + \sum_{q=2}^{|k|-1} \lambda^{tr} \cdot t^{tr}_{k,q} + \beta^{tr} \cdot N^{tr} + \beta^{re} \cdot N^{re} \tag{5}$$

t^w_s, $t^{in}_{s_1,s_2}$ and t^{tr}_s are the initial waiting time, in-vehicle time and transfer time, respectively, and λ^w, λ^{in} and λ^{tr} are the corresponding weights. $N^{tr} = |k| - 1$ and N^{re} stand for the number of transfer and rerouting decisions, and β^{tr} and β^{re} are penalty terms for each transfer and rerouting respectively, represented in time equivalent units. The well-known IIA property of the MNL model is partially counteracted by the filtering rules which result with a choice-set comprising of distinctive alternatives where the most correlated paths are either eliminated due to dominance rules or merged into hyper-paths as described in Cats et al. (2016). By calculating the choice model probabilities, the passenger flow on each diachronic time-dependent network link can be calculated by equation (6), where d_{ij} is the number of passengers from i to j.

$$f_{ijk} = d_{ij} \cdot P_{ijk} \tag{6}$$

The actual nominal travel time can be obtained as follows:

$$t_{ijk} = t^w_{s_{k,1}} + \sum_{v=1}^{|k|-1} t^{in}_{s_{k,v},s_{k,v+1}} + \sum_{q=2}^{|k|-1} t^{tr}_{k,q} \tag{7}$$

The dynamic passenger flow distribution model under a disruption is conducted as shown in Figure 7 by performing the following sequence of steps:

- **Step 1:** Alternative route generation: Given a passenger OD demand matrix and scheduled timetable, find alternative routes for each pair of stations and calculate passenger's total in-vehicle time and transfer time for each route. This is an initialization phase.

- **Step 2:** Passenger route choice: Simulate passenger generation and train movements. Progress simulation clock from the beginning of the disruption, and calculate the generalised cost for each passenger route departing on each minute. Use a logit model to obtain the proportion of each route, and then assign passengers to routes.

- **Step 3:** Alternative route update: When the prediction length is updated, search for new alternative routes for each passenger OD based on the corresponding rescheduled timetable.

- **Step 4:** Passenger assignment update: Reroute passengers assigned in Step 2 with routes in the disruption area, and update the corresponding travel time and routes information. Then assign new passengers to routes until the next prediction is generated.

- **Step 5:** Transition and normal operations: Repeat step 4 until the last prediction is made, then simulate the model until a certain pre-defined time in order to have a fair comparison among prediction scenarios.

  Passenger loading results are assessed by calculating the following outputs: passenger total generalised travel time, total passenger nominal travel time, the number of passengers, average passenger transfer time, the number of average transfer, average in-vehicle time, average waiting time, average train load and link load.

2.4. The Interaction Between the Models

The three models interact in a dynamical fashion, i.e. interaction occurs every time new information becomes available. New information can be input concerning the observed influencing factors in the disruption length model or when we learn that the previous disruption length prediction was too short. The crosses in the time diagram in Figure 8 illustrate when during the disruption period the interaction might occur.

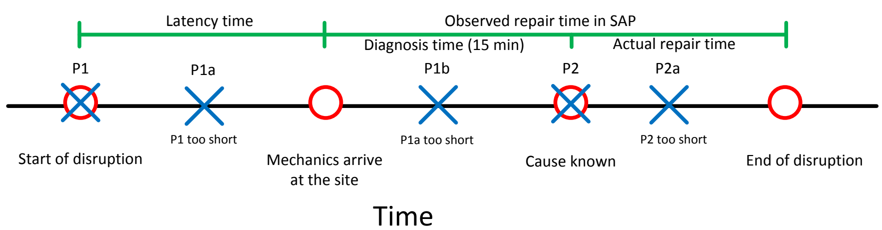


Figure 8: The time diagram of a railway disruption.

Figure 8 shows an example of a disruption. When a disruption occurs, only the influencing factors of the latency time are typically known. The unconditional BN is conditionalized based on this information. Predictions made from this conditional BN model are called the "P1"predictions where, mainly, the latency time is predicted using the additional sources of information, potentially yielding more accurate predictions. "P1" suggests that the disruption ends within a certain time period marked by "P1a". The start and end time of the disruption period that are predicted by "P1" are communicated to the short-turning model. Based on the disruption location, the relevant timetable and the disruption period, a disruption timetable is computed and passed on to the passenger flow distribution model. To measure and compare the impact of alternative disruption management scenarios the following key performance indicators (KPIs) are computed and stored:

1. The total number of passengers being affected during the disruption period.

2. The total experienced generalized travel time corresponding to equation (5) of all passengers considered in the experiment.

3. The total number of passengers reroutings and transfers.

This interaction is shown in Figure 9.

When the predicted length "P1" has elapsed, it might be realized that the disruption is not over yet. If the prediction is too short, the disruption is still unresolved even after the predicted disruption ends. This situation occurs when the prediction is too "optimistic", i.e. the chosen quantile of the conditional distribution of disruption length is too low for the case. If this happens, the prediction is updated by approximating a new conditional distribution of disruption length on the information that the disruption length is longer than the predictions. This is done via sampling the original conditional distribution on the quantiles higher than the prediction. In this paper, these "revised" predictions are denoted alphabetically in orderly fashion. For instance, a "P1" prediction is updated to "P1a", "P1b", and so on. With each prediction update the cycle shown in Figure 9 repeats and the mentioned statistics are computed and stored.

Fifteen minutes after the arrival of the mechanics to the disruption site, they report the diagnosis about the cause of disruption. Knowing the cause of disruption, the BN is further conditionalized. The new conditional BN is used to produce the "P2" predictions. Similarly with each prediction update, the short-turning model computes the disruption timetable and passenger assignment model computes and stores the total number of affected passenger, generalized travel time, and total number of reroutings and transfers.

3. Experiments

The framework that is applied on a part of the Dutch railway network is depicted in Figure 10. The disruption length model is constructed using a computationally-efficient software called UNINET which was developed at Delft University of Technology and is available at www.lighttwist.net/wp/uninet. The short-turning model is implemented in MATLAB R2016a and YALMIP (Löfberg (2012)) is used to construct the MILP and Gurobi is used as the solver. The passenger flow distribution model is constructed in MATLAB 2014. In the experiment, we consider a complete blockage in the railway segment between stations Utrecht and Houten. The blockage is caused by a track circuit failure.

Two local train lines are considered by the short-turning model: line 16000 which runs between Utrecht (Ut) and s'Hertogenbosch (Ht) and line 6000 which operates between Ut and Tiel (Tl). Due to the disruption, these trains have the possibility to short-turn either at the station Geldermalsen (Gdm) or the latest at station Houten (Htn).

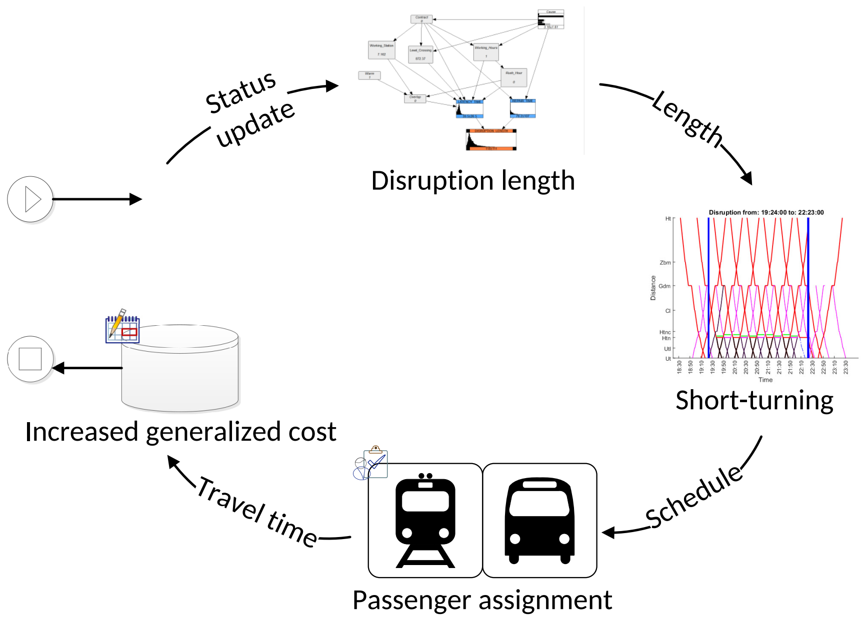


Figure 9: The interaction between the models

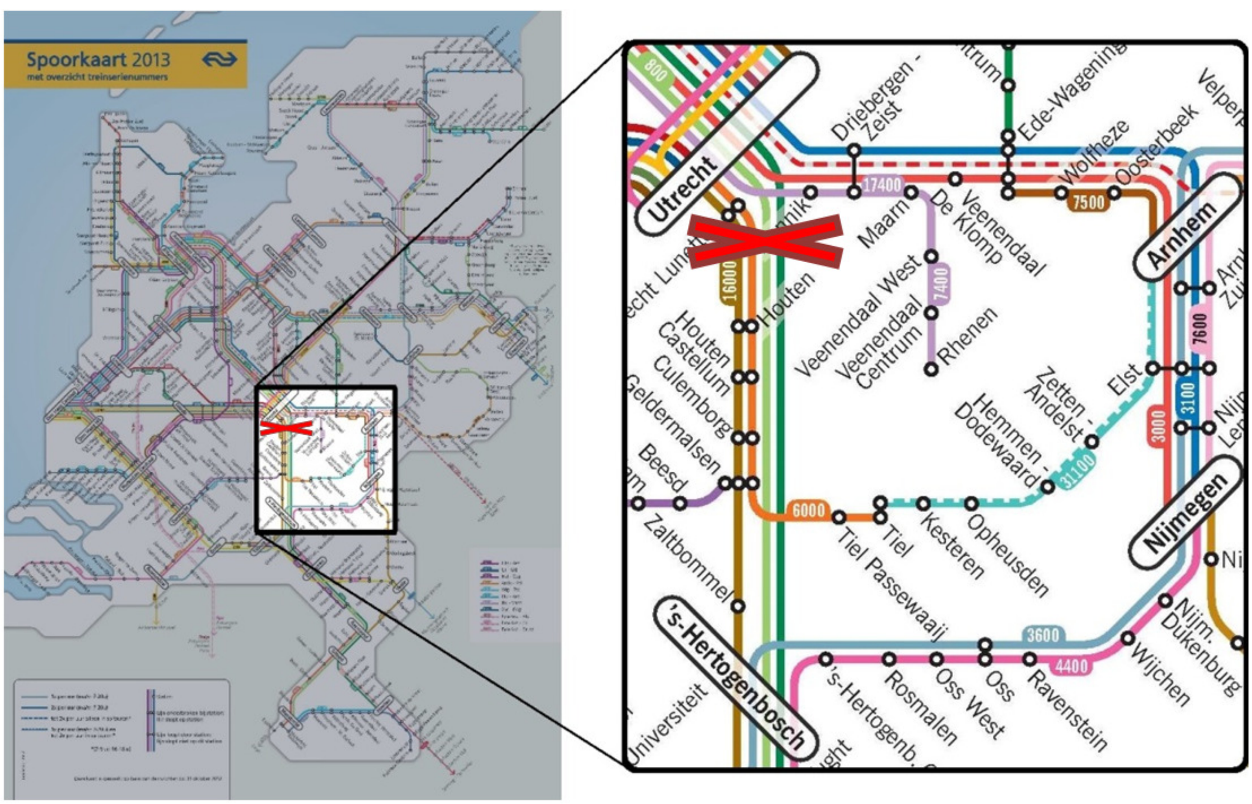


Figure 10: The area the passenger flow model considers in the experiment. The illustration is adapted from NS' main train service map. Source: The Dutch railways (NS).

To study the effect of different choices of quantiles of the conditional distribution of predicted disruption length, different disruption length predictions are examined. For each prediction, when a successive prediction is made, the new prediction is chosen to be the value of the new conditional distribution corresponding to the same quantile as in the previous prediction. For instance, when a prediction using the quantile 50 (median) is updated, the new prediction is also taken to be the median of the updated conditional distribution. In principle, this does not necessarily have to be the case. This choice is made to narrow down the space of possible "combinations" of prediction scenarios.

Note that the "P1" predictions are only valid until new information regarding the cause is available from which the updated "P2" predictions can be made. This means that the "P1" predictions and the entailed disruption timetables are only used until at most 15 minutes after the mechanics arrive at the disrupted site.

The "P2" predictions are updated until the actual disruption ends. When this happens, no new timetable is computed but, instead, the last "P2" timetable is run until the "P2" predicted end of disruption. This choice is made to penalize a prediction that is too "pessimistic", i.e. a prediction that surpasses the realized disruption length. In contrast, choosing a of lower quantile is undesirable because it is likely to be too optimistic and thus results in many "revised" predictions. This is not attractive from the passengers' point of view who, will presumably perceive the communicated information as unreliable. Additionally, having many "revised" predictions is not practical from a logistical point of view. In practice every time a new prediction is made, aside from the train traffic, the traffic controllers must also revise the rolling stock and crew assignments. Therefore, train operators are not inclined to choose the lower quantile predictions. Therefore, in this experiment, we only consider the 25% quantile as the representative of these scenarios for comparison. The remaining quantiles that are considered in the experiments are 50%, 75%, 85%, 90%, as well as the mean.

The short-turning model computes the disruption timetable for lines 16000 and 6000 through stations Ut, Htn, Houten Castellum (Htnc), Culemborg (Cl), Gdm, Tiel Passewaaij (Tpsw), Tl, Zaltbomel (Zbm) and Ht. In the short-turning model, the main parameters are the penalties for arrival, departure delays and cancelled services. Since the frequency of the services in Houten is either 16 or 14 minutes in both directions, and with each cancelled service the travellers need to wait around a quarter of an hour for the next train, the cancellation penalty ($\omega^c_{v^i_{l,n}}$) is set to 1000 seconds. Arrival and departure delays ($\omega^{d^a}_{v^i_{l,n}}$ and $\omega^{d^d}_{v^i_{l,n}}$) are equally penalized by 1. Moreover the minimum short-turning time is assumed to be around 7 minutes (420 seconds). For the choice of minimum short-turning time we refer to the study by Chu and Oetting (2013). A norm of 3 minutes is considered for the minimum headway.

Given a disruption timetable, the passenger flow model computes the passengers traffic. Due to data availability limitations, the passenger flow model considers in this case study only the passenger-trips for which both origin and destination are within the case study area, i.e. the loop shown in Figure 10. Data about the number of daily passengers between all pairs of these OD stations was obtained from the Netherlands

Railways (*Nederlandse Spoorwegen*/NS). The daily distribution of passenger demand was specified based on data made available by NS which manifests the conventional morning and afternoon peaks.

Passengers travelling between Utrecht Centraal and Houten stations have two alternatives: to detour via Arnhem and Nijmegen or to take the public bus service between the two stations. The travel time with bus between Utrecht Centraal station and Houten is about 35 minutes while a regular train service would have taken only 9 minutes. On the other hand, the detour via Arnhem and Nijmegen is also not very attractive due to the tremendous detour it induces. For passengers travelling to Houten from Utrecht, this detour takes almost 2 hours. To fairly compare the different choices of predictions, we monitor the train traffic and passengers flow for a fixed period of six hours in all scenarios. Two actual disruption cases are chosen and examined. Also the consequences of the same disruption cases if they would have occurred during a different time of the day are investigated. Thus, we analyze four case studies which are explained in detail in the following sections.

3.1. Case Study 1

The first case study is based on an incident which occurred on Thursday, 10 July 2014. The incident started at 14:22 and had the information listed in Tables 1.

Contract type	OPC
Working station distance	7.1620
Level Crossing distance	872.372
Working Time	yes
Warm	yes
Rush Hour	no
Overlapping disruption	no
Cause	a setting problem caused by heat

Table 1: The initial information of the disruption case 1

Moreover, the real observed latency and repair time are 70 and 73 minutes, respectively. This means the total disruption length is $70 + 73 = 143$ minutes.

Table 2 presents the "P1" predictions which are presented in terms of the length (in minutes) and the predicted end of disruption in time. Notice that in the scenario corresponding to the 25% quantile, in total there are 8 predictions that are generated throughout the disruption. The predictions in Table 2 are used by the short-turning model to produce the disruption timetable whose cyclic characteristics are shown in Table 3. This Table contains the number of cancelled services ($\#C$), the number of delayed services ($\#D$), the total train delay in minute (Del), and the number of short-turned services ($\#ST$).

At 15:32, the mechanics arrive at the site. After 15 minutes of diagnosis time, the cause of a TC failure is identified and, at 15:47, the "P1" predictions are updated to the "P2" predictions. These "P2" predictions are presented in Table 4. The results of the short-turning model are shown in Tables 5 and 6.

Table 2: The P1 predictions for Case Study 1.

Qtl	P1		P1a		P1b	
(%)	Length	Time	Length	Time	Length	Time
25	49	15:11	72	15:34	97	15:59
50	81	15:43	144	16:46		
75	143	16:45				
85	205	17:47				
90	254	18:36				
Mean	118	16:20				

Table 3: The results of the short-turning model for P1 in Case Study 1.

Qtl	P1				P1a				P1b			
(%)	#C	#D	#ST	Del.	#C	#D	#ST	Del.	#C	#D	#ST	Del.
25	12	2	3	6	20	0	5	0	28	0	7	0
50	20	9	5	24	40	0	10	0				
75	40	0	10	0								
85	56	0	14	0								
90	68	0	17	0								
Mean	32	0	8	0								

Table 4: The P2 predictions for Case Study 1.

Qtl	P2		P2a		P2b		P2c		P2d	
(%)	Length	Time	Length	Time	Length	Time	Length	Time	Length	Time
25	85	15:47	102	16:04	118	16:20	135	16:37	150	16:52
50	95	15:57	134	16:36	179	17:21				
75	133	16:35	240	18:22						
85	160	17:02								
90	194	17:36								
Mean	119	16:21	188	17:30						

Table 5: The results of the short-turning model for P2 in Case Study 1.

Qtl	P2				P2a				P2b			
(%)	#C	#D	#ST	Del.	#C	#D	#ST	Del.	#C	#D	#ST	Del.
25	24	0	6	0	28	0	7	0	32	0	8	0
50	24	9	6	24	36	0	9	0	48	0	12	0
75	36	0	9	0	64	0	16	0				
85	44	0	11	0								
90	52	0	13	0								
Mean	32	0	8	0	52	0	13	0				

Table 6: The remaining results of the short-turning model for P2 for Case Study 1.

Qtl	P2c				P2d			
(%)	#C	#D	#ST	Del.	#C	#D	#ST	Del.
25	36	0	9	0	40	0	10	0

Each timetable is used when the prediction is still “valid”, i.e. it has not been changed. For instance, the disruption timetable generated with the P1a prediction of the 50% quantile is used only for four minutes between 15:43 and 15:47. At 15:47, the prediction is updated to P2 and a new disruption timetable is constructed.

These disruption timetables are used by the passenger flow model to compute passenger route choice and the resulting passenger distribution over train services for six hours between 14:22 and 20:22. In this period, there are 22163 passengers who are traveling in the case study area. We measure the impact on the passengers for each choice of quantile. The results are presented in Table 7.

Table 7: The impact of different predictions to the passengers in Case Study 1.

Qtl (%)	Excess (minute)	# Affected Passengers	Inc. Orig (%)	Inc. Bench (%)	# Rerouting 1	# Rerouting 2	# Transfers
25	7	8948	17.71	0.26	595	4	3597
50	36	11469	20.53	2.67	466	1	4015
75	97	16560	24.33	5.90	247	0	4671
85	17	9791	17.56	0.13	5	0	3775
90	51	12854	20.16	2.35	0	0	4126
Mean	45	12299	20.33	2.49	282	0	4136
Real	0	8374	17.4026	0	0	0	3773

The last row of Table 7 shows the benchmark case with the true disruption length. In this case, the true end time of disruption is already known when the disruption starts at 14:22[1]. Each scenario is compared to this case to measure the increase in the impact of the prediction on the passengers with respect to the ideal situation.

The second column of Table 7 shows the difference (in minute) between the true end of disruption and the last P2 prediction when the blocked railway section between Utrecht and Houten is opened for train operation. The third column presents the total number of passengers traveling during the blockage of the section. The fourth and fifth column provide the increase (in %) in the total generalized travel time with respect to the normal situation without disruption and the benchmark, respectively. In the sixth and seventh column, the total number of passengers who have to reroute once or twice in each scenario are provided. The number of transfers performed by the passengers can be found in the last column.

The benchmark case represents the best possible situation. In this case, fewer passengers are affected and no passengers have to be rerouted since the initially provisioned information is accurate. Unsurprisingly, the increase in the generalized travel time with respect to the no disruption situation is also the lowest.

In general, the longer the difference between the true end of disruption and the last P2 prediction is, the more passengers are affected. However, this does not necessarily translate to a higher total generalized travel time. Notice that the increase in the generalized travel time is higher in the 25%-quantile scenario than in the 85%-quantile scenario even though the difference between the prediction and the realized value is only 7 minutes in the former and 17 minutes in the latter. The eight predictions in the 25%-quantile scenario cause many passengers to reroute due to the frequent updates of the disruption timetable. Consequently, the total generalized travel time is penalized severely. In the 85%-quantile scenario, much fewer passengers need to

[1]This is the best possible situation but, of course, is not realistic.

reroute due to the pessimistic prediction.

Notice that the P2 predicted end time of the disruption of the 75%-quantile scenario (at 16:35, see Table 4) is 10 minutes shorter than the actual end time. Because of this slightly too optimistic prediction, the predicted end time is updated to P2a that is at 18:22. This new prediction is dramatically too long and consequently disrupts the late afternoon peak demand period. As a result, this scenario is the worst as indicated by the number of affected passengers and the increase of the total generalized travel time.

3.2. Case Study 1A

In order to investigate the effect of the disruption's time of occurrence on the choice of predictions, an artificial disruption is considered in this case study. The exact same disruption as in Case Study 1 is assumed to occur later on the same day, at 19:24. The realizations of the latency and repair time of this artificial disruption are taken from the values of the computed conditional distribution of latency and repair time which correspond to the same quantiles as the realizations of Case Study 1. In this case, the latency and repair time are 89 and 73 minutes, respectively. The total length is 162 minutes and the disruption ends at 22:06.

The P1 predictions and the short-turning results are presented in Table 8 and 9.

Table 8: The P1 predictions for Case Study 1A.

Qtl	P1		P1a		P1b		P1c	
(%)	Length	Time	Length	Time	Length	Time	Length	Time
25	54	20:18	77	20:41	102	21:06	127	21:31
50	86	20:50	149	21:53				
75	149	21:53						
85	209	22:53						
90	258	23:42						
Mean	122	21:26						

Table 9: The results of the short-turning model for P1 in Case Study 1A.

Qtl	P1				P1a				P1b				P1c			
(%)	#C	#D	#ST	Del.	#C	#D	#ST	Del.	#C	#D	#ST	Del.	#C	#D	#ST	Del.
25	19	0	3	0	23	2	4	6	31	0	6	0	39	0	8	0
50	27	0	5	0	43	2	9	2								
75	43	2	9	2												
85	55	5	12	39												
90	73	2	15	8												
Mean	35	9	7	15												

The P2 predictions are made at 21:08, 15 minutes after the mechanics' actual arrival time at 20:53. These predictions are presented in Table 10. The corresponding results of the short-turning model are represented in Table 11 and 12.

As before, the predictions are used by the short-turning model to produce the disruption timetables which are used by the passenger flow model to attain the distribution of passengers traffic. Between 19:24 and

Table 10: The P2 predictions for Case Study 1A.

Qtl	P2		P2a		P2b		P2c		P2d	
(%)	Length	Time	Length	Time	Length	Time	Length	Time	Length	Time
25	104	21:08	121	21:25	137	21:41	154	21:58	169	22:13
50	114	21:18	153	21:57	198	22:42				
75	152	21:56	259	23:43						
85	179	22:23								
90	213	22:57								
Mean	138	21:42	207	22:51						

Table 11: The results of the short-turning model for P2 in Case Study 1A.

Qtl	P2				P2a				P2b			
(%)	#C	#D	#ST	Del.	#C	#D	#ST	Del.	#C	#D	#ST	Del.
25	31	0	6	0	35	2	7	6	39	2	8	6
50	35	0	7	0	43	9	9	24	55	2	12	8
75	43	9	9	15	73	2	15	10				
85	47	4	10	38								
90	55	12	12	69								
Mean	39	9	8	15	55	5	12	31				
Real	47	0	10	0								

01:24, 7102 passengers are traveling in the case study area. Notice that there are fewer passengers in this set-up than the previous one due to the different time of the day under consideration. Table 13 summarizes the impact on passengers for each quantile.

In comparison to Case Study 1, the increase in the total generalized travel time with respect to the normal situation is higher. This is because the disruption is longer than in the previous case study due to the longer latency time.

The benchmark case still represents the best situation with fewer passengers being affected. The total generalized travel time is the lowest in this scenario and no passengers are rerouted.

The longer the last P2 prediction is, the more passengers are affected by the disruption. For this reason, the 75%-quantile scenario yields the highest total generalized travel time. Notice that as in Case Study 1, the P2 prediction of this scenario is 10 minutes shorter than the actual end time of the disruption. Consequently, the prediction is updated to P2a, which is then too long.

The nine predictions in the 25% quantile scenario cause many passengers to reroute. In this scenario, there is a total of 1487 rerouting activities including a considerable amount of passengers who need to change their plans more than twice. As a result, the total generalized travel time of this scenario is the second largest due to the heavy penalty associated with rerouting.

3.3. Case Study 2

In this case study, we consider another real TC disruption at the same location which occured on Saturday, 18 October 2014 and started at 19:24. The incident had the information shown in Table 14.

Table 12: The remaining results of the short-turning model for P2 in Case Study 1A.

Qtl	P2c				P2d			
(%)	#C	#D	#ST	Del.	#C	#D	#ST	Del.
25	43	9	9	33	47	2	10	10

Table 13: The impact of different predictions to the passengers in Case Study 1A.

Qtl	Excess	# Affected	Inc. Orig	Inc. Bench	# Rerouting		# Transfers
(%)	(minute)	Passengers	(%)	(%)	1	2	
25	7	4966	32.68	7.47	438	195	1353
50	36	5563	31.09	6.18	479	8	1430
75	97	6563	35.12	9.45	83	0	1512
85	17	5180	29.46	4.86	52	0	1379
90	51	5843	32.15	7.05	0	0	1461
Mean	45	5733	27.16	3.00	91	0	1306
Real	0	4812	23.4516	0	0	0	1291

Contract type	OPC
Working station distance	7.1620
Level Crossing distance	872.372
Working Time	no
Warm	no
Rush Hour	no
Overlapping disruption	no
Cause	a cable problem

Table 14: The initial information of the disruption case 2

The observed latency and repair time are 47 and 88 minutes, respectively, with a total length of 135 minutes.

The P1 predictions and the short-turning results for this case study are presented in Tables 15 and 16.

Table 15: The P1 predictions for Case Study 2.

Qtl	P1		P1a	
(%)	Length	Time	Length	Time
25	54	20:18	77	20:41
50	86	20:50		
75	149	21:53		
85	209	22:53		
90	258	23:42		
Mean	122	21:26		

Table 16: The results of short-turning model for P1 for Case Study 2.

Qtl	P1				P1a			
(%)	#C	#D	#ST	Del.	#C	#D	#ST	Del.
25	19	0	3	0	23	2	4	6
50	27	0	5	0				
75	43	2	9	2				
85	55	5	12	39				
90	73	2	15	8				
Mean	35	9	7	15				

Fifteen minutes after the mechanics' actual arrival time at 20:11, the P2 predictions are made. Tables 17 and 18 presents these predictions and the short-turning model. Table 19 summarizes the impact on passengers for each choice of quantile.

Table 17: The P2 predictions for Case Study 2.

Qtl	P2		P2a		P2b		P2c	
(%)	Length	Time	Length	Time	Length	Time	Length	Time
25	67	20:31	94	20:58	120	21:24	151	21:55
50	104	21:08	173	22:17				
75	173	22:17						
85	237	23:21						
90	280	00:04						
Mean	142	21:46						

Table 18: The results of short-turning model for P2 for Case Study 2.

Qtl	P2				P2a				P2b				P2c			
(%)	#C	#D	#ST	Del.	#C	#D	#ST	Del.	#C	#D	#ST	Del.	#C	#D	#ST	Del.
25	23	0	4	0	27	9	5	33	35	2	7	4	43	2	9	6
50	31	0	6	0	47	2	10	18								
75	47	2	10	18												
85	65	7	13	102												
90	77	8	16	79												
Mean	43	0	9	0												
Real	39	2	8	2												

Table 19: The impact of different predictions to the passengers in Case Study 2.

Qtl	Excess	# Affected	Inc. Orig	Inc. Bench	# Rerouting		# Transfer
(%)	(minute)	Passengers	(%)	(%)	1	2	
25	16	4562	14.43	0.64	426	32	1219
50	38	5052	17.11	2.99	252	32	1506
75	38	5052	16.11	2.11	0	0	1437
85	102	6246	19.30	4.92	0	0	1558
90	145	6792	21.43	6.79	0	0	1693
Mean	7	4351	14.89	1.04	2	0	1385
Real	0	4182	13.7099	0	0	0	1309

The benchmark case represents the best possible situation. With the least number of affected passengers with no need for rerouting, the total generalized travel time is the lowest of all scenarios.

Notice that the final predicted end time of disruption of the 50%-quantile and 75%-quantile scenario are the same. Consequently, the same number of passengers are affected in both cases. However, the 50%-quantile scenario has three predictions while the 75%-quantile scenario has only two. Consequently, many passengers need to be rerouted and more transfers need to be performed in the former. This results with higher total generalized travel time in the case of the 50%-quantile scenario. The pessimistic 90%-quantile scenario disturbs the most number of passengers because of a P2 prediction that is too long. Consequently, the total generalized travel time is the largest, making this scenario the worst performing one in this case study.

3.4. Case Study 2A

Similarly to Case Study 1A, the incident in Case Study 2 is also considered to occur at a different time of the day. In this case study, we assume this hypothetical incident to occur on the same day at 14:22. In this case, because the incident occurs during the weekend, the prediction lengths do not change from Case Study 2; only the time-dependent passenger demand generation process is adjusted. Tables 20 and 21 present the P1 predictions and the corresponding short-turning results. The P2 predictions and the short-turning results are presented in Tables 22 and 23.

Table 20: The P1 predictions for Case Study 2A.

Qtl	P1		P1a	
(%)	Length	Time	Length	Time
25	54	15:16	77	15:39
50	86	15:48		
75	149	16:51		
85	209	17:51		
90	258	18:40		
Mean	122	16:24		

Table 21: The results of the short-turning model for P1 in Case Study 2A.

Qtl	P1				P1a			
(%)	#C	#D	#ST	Del.	#C	#D	#ST	Del.
25	16	0	4	0	20	2	5	2
50	24	0	6	0				
75	40	0	10	0				
85	56	0	14	0				
90	68	2	17	4				
Mean	32	2	8	4				

Table 22: The P2 predictions for Case Study 2A.

Qtl	P2		P2a		P2b		P2c	
(%)	Length	Time	Length	Time	Length	Time	Length	Time
25	67	15:29	94	15:56	120	16:22	151	16:53
50	104	16:06	173	17:15				
75	173	17:15						
85	237	18:19						
90	280	19:02						
Mean	142	16:44						

The impact of different scenarios on the passengers are presented in Table 24. Notice that more passengers are affected by the disruption in comparison to Case Study 2. This is due to the disruption occurring during the day time when more passengers are traveling.

Like the previous three case studies, the scenario with the true disruption length is the best performing one in terms of the total generalized travel time. The least number of passengers are affected and none of

Table 23: The results of short-turning model for P2 for Case Study 2A.

Qtl	P2				P2a				P2b				P2c			
(%)	#C	#D	#ST	Del.	#C	#D	#ST	Del.	#C	#D	#ST	Del.	#C	#D	#ST	Del.
25	20	0	5	0	24	9	6	15	32	0	8	0	40	2	10	2
50	28	0	7	0	48	0	12	0								
75	48	0	12	0												
85	64	0	16	0												
90	76	0	19	0												
Mean	36	9	9	33												
Real	36	0	9	0												

Table 24: The impact of different predictions to the passengers in Case Study 2A.

Qtl	Excess	# Affected	Inc. Orig	Inc. Bench	# Rerouting		# Transfers
(%)	(minute)	Passengers	(%)	(%)	1	2	
25	16	9031	12.31	1.22	723	120	3249
50	38	10928	15.87	4.43	491	46	3725
75	38	10928	14.60	3.28	0	0	3703
85	102	16359	19.95	8.11	0	0	4549
90	145	19044	23.25	11.08	0	0	5069
Mean	7	8294	11.10	0.13	4	0	3128
Real	0	7736	10.9567	0	0	0	3128

them have to change their travel plans.

As in Case Study 2, the difference between the predicted end of disruption and the truth is 38 minutes in both the 50%-quantile and the 75%-quantile scenario. However, the increase in the generalized travel time is higher in the former case. This is due to the more frequent prediction updates so many passengers have to reroute and slightly more transfers are needed. Consequently, the total generalized travel time of this scenario is penalized more.

The very pessimistic 90% quantile scenario performs the worst in terms of the total generalized travel time. The dramatic difference between the predicted end of disruption and the truth means a lot of passengers are affected by the disruption. As a result, the total generalized travel time becomes very high.

4. Conclusions and Future Work

To measure the impact of a disruption length prediction on the passengers, a cost function needs to be defined. In this paper, we choose the total generalized travel time as our cost function. The impact is primairly measured using the weighted total travel time of all passengers which takes into the account the waiting time, the in-vehicle time, the transfer time, the number of transfers, and the number of reroutings.

The impact of the uncertainty in the disruption length on the train traffic and passengers was investigated in terms of changes in this cost function. On one hand, when the prediction is pessimistic, more passengers are affected by the disruption which increases the total generalized travel time. On the other hand, when the prediction is optimistic, while less passengers are affected, this choice leads to more frequent updates of the prediction and hence more passengers have to change their travel plans and reroute.

In the experiment when a prediction is updated, the new prediction is chosen to be the value of the new conditional distribution corresponding to the same quantile. Moreover, no train operations are resumed before the predicted end of the disruption even when the disruption has actually ended. The consequence of these choices is that a prediction that is slightly shorter than the truth results with high costs. Because the disruption ends not long after the predicted time, the updated prediction becomes too long. As a result, more passengers are affected and the total generalized travel time increases. This situation is illustrated in the 75%-quantile scenario in Case Study 1.

Rescheduling measures to mitigate the effects of train disruptions are conventionally devised or optimized assuming that the disruption length is known a-priori or by discarding the disruption length. The latter has to be predicted by service providers and the quality of the prediction has important consequences for the desired rerouting of train traffic and the resulting rerouting of passenger flows. In this paper, we present an integrated framework for predicting disruption length using a Bayesian network approach, determining train short-turning measures using a mixed integer linear program and a probabilistic multi-stage passenger load distribution model.

In a series of case studies, we have shown how different choices of predictions of disruption length would affect the passengers in terms of the generalized travel time. On one hand, when the prediction is too optimistic, many passengers have to be rerouted which increases the inconvenience and, hence, the total generalized travel cost. On the other hand when the prediction is too pessimistic, more passengers are affected which results in a higher total generalized travel cost. Similarly when the prediction is just slightly shorter than the truth, more passengers are also affected. This is evident specially with pessimistic prediction that leads to keeping the track blocked much longer than necessary, significantly beyond the end of disruption. The experiments provide insights about the impact of the predictions on the passengers delays, suggesting that starting with pessimistic predictions and thereafter gradually switching to the optimistic predictions might result in a better prediction. Notwithstanding, we cannot conclude based on a few case studies which value of the conditional distribution of disruption length is the “best” for a prediction. Moreover, different quantiles were found to lead to the least passenger costs under different scenarios. To draw such a conclusion, many more case studies need to be performed. The close-to-optimal solution can be found by testing a large number of points from the solution space, i.e. by generating many different realizations of disruption length from the conditional distribution. For each realization, the short-turning model and the passenger flow model are run to compute the effect of different choices of disruption length predictions.

Moreover when a prediction is updated, in the case studies we took the same quantile of the conditional distributions as the predictions to simplify the problems. In principle, this does not need to be the case. For instance, the choice of quantile in the “P2” prediction might depend on the quantile that realizes the latency time in the “P1” prediction. The realization of the latency time might indicate how fast/slow the mechanics work on the specified incident which might be a useful information to produce a more accurate P2 prediction. Other possibilities could be considered as well.

The function of the total generalized travel time specified in this study can be extended in future studies. During peak demand periods and disruptions it is especially relevant to consider vehicle capacity constraints and the vehicle type. The bus service between Utrecht and Houten provides significantly less seat capacity than the train service. Moreover, different vehicle types provide different levels of comfort to the passengers. An Intercity train is generally more comfortable than a Sprinter train or a bus service. This can be accommodate in equation (5) by adding a weight corresponding to the vehicle type to the second term of the equation (5).

Choosing generalized travel time as the primary evaluation/selection criterion means the impact of the uncertainty in disruption length is only measured from the passengers' point of view. With this cost function, an optimistic prediction is not attractive only because it causes inconvenience to the passengers who would have to reroute. Note that we have not captured all costs associated with an optimistic prediction that the traffic control faces. However, from an operational point of view, having a lot of timetable updates is not practical due to the logistical issues that need to be carried out. For instance, the rolling stock and crew assignments need to be reorganized accordingly with every update. Future studies may incorporate additional aspects into the cost function in order to better reflect the aspects influencing real-time decision making in the context of disruption management.

Note that the effect of different predictions depends considerably on the location and the time of the incident. This is evident in the results of the case studies investigated in this paper. This is presumably even more pronounced in denser areas where there is a greater hierarchy among different locations, such as the Amsterdam area. Aside from the greater number of alternative rail services, passengers may switch to other modes of public transport, consisting of metro, tram and bus, which contribute to network redundancy.

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 and André Duinmeijer from ProRail for the supervision and access to the data set.

References

Canca, D., Barrena, E., Laporte, G., Ortega, F. A., 2016. A short-turning policy for the management of demand disruptions in rapid transit systems. Annals of Operations Research 246 (1), 145–166.

Cats, O., 2016. The robustness value of public transport development plans. Journal of Transport Geography 51, 236 – 246.

Cats, O., Jenelius, E., 2014. Dynamic vulnerability analysis of public transport networks: Mitigation effects of real-time information. Networks and Spatial Economics 14 (3), 435–463.

Cats, O., West, J., Eliasson, J., 2016. A dynamic stochastic model for evaluating congestion and crowding effects in transit systems. Transportation Research Part B: Methodological 89, 43 – 57.

Chu, F., Oetting, A., 2013. Modeling capacity consumption considering disruption program characteristics and the transition phase to steady operations during disruptions. Journal of Rail Transport Planning & Management 3, 54–67.

Coor, G. T., 1997. Analysis of the short-turning strategy on high-frequency transit lines. Tech. rep., Massachusetts Institute of Technology.

De-Los-Santos, A., Laporte, G., Mesa, J. A., Perea, F., 2012. Evaluating passenger robustness in a rail transit network. Transportation Research Part C: Emerging Technologies 20 (1), 34 – 46.

Ghaemi, N., Cats, O., Goverde, R. M. P., 2017. Railway disruption management challenges and possible soluion direction. (Submitted for publication) Public Transport.

Ghaemi, N., Goverde, R. M. P., Cats, O., Aug 2016. Railway disruption timetable: Short-turnings in case of complete blockage. In: 2016 IEEE International Conference on Intelligent Rail Transportation (ICIRT). pp. 210–218.

Hirai, C., Tomii, N., Tashiro, Y., Kondou, S., Fujimori, A., 2006. An algorithm for train rescheduling using rescheduling pattern description language R. Computers in Railways X, WIT Press 88, 551–561.

Joe, H., 2014. Dependence Modelling with Copulas. CRC Press.

Kumazawa, K., Hara, K., Koseki, T., 2008. A novel train rescheduling algorithm for correcting disrupted train operations in a dense urban environment. Computers in Railways XI, WIT Press 103, 565–574.

Löfberg, J., 2012. Automatic robust convex programming. Optimization methods and software 27 (1), 115–129.

Meng, L., Zhou, X., 2011. Robust single-track train dispatching model under a dynamic and stochastic environment: A scenario-based rolling horizon solution approach. Transportation Research Part B 45 (7), 1080–1102.

Nakamura, T., Hirai, C., Nishioka, Y., 2011. A practical train rescheduling algorithm using three predetermined factors. Tech. rep.

Nelsen, R. B., 2006. An Introduction to Copula. Springer, New York.

Nielsen, L. K., Kroon, L., Maróti, G., 2012. A rolling horizon approach for disruption management of railway rolling stock. European Journal of Operational Research 220 (2), 496 – 509.

Veelenturf, L. P., Kidd, M. P., Cacchiani, V., Kroon, L. G., Toth, P., 2016. A railway timetable rescheduling approach for handling large-scale disruptions. Transportation Science 50 (3), 841–862.

Yang, L., Zhou, X., Gao, Z., 2013. Rescheduling trains with scenario-based fuzzy recovery time representation on two-way double-track railways. Soft Computing 17 (4), 605–616.

Yang, L., Zhou, X., Gao, Z., 2014. Credibility-based rescheduling model in a double-track railway network: a fuzzy reliable optimization approach. Omega 48, 75 – 93.

Zhan, S., Kroon, L. G., Veelenturf, L. P., Wagenaar, J. C., 2015. Real-time high-speed train rescheduling in case of a complete blockage. Transportation Research Part B: Methodological 78, 182 – 201.

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

Zilko, A. A., Kurowicka, D., Goverde, R. M. P., 2016. Modeling railway disruption lengths with copula bayesian networks. Transportation Research Part C: Emerging Technologies 68, 350 – 368.