Technical University of Denmark

DTU

Micro-simulation based analysis of railway lines robustness

Cerreto, Fabrizio

DTU Library
Technical Information Center of Denmark

Micro-simulation based analysis of railway lines robustness

Fabrizio Cerreto

Department of Transport, Denmark's University of Technology
Bygningstorvet 116B, 2800 Kgs. Lyngby, Denmark
E-mail: facer@transport.dtu.dk, Phone: +45 45256548

Abstract

Railway Undertakers and Railway Infrastructure Managers have a variety of parameters to measure robustness of timetables: this paper examines empirical data collected from Nederlandse Spoorwegen on the heavily occupied railway line between The Hague and Rotterdam in The Netherlands. The results show that the robustness indicator examined are affected by infrastructure changes in different ways.

A micro-simulation of the line is used to determine the robustness of different railway infrastructure scenarios measuring the capacity consumption, as well as the share of trains affected by amount of delays in case of disturbance. The individual robustness indicators' sensitivity to primary delays is estimated through regression analysis and either the second derivative or the amplification as a function of the train frequency. A skimming method is used for the selection and simulation of a delayed train course that can represent the total train delay along the line. Dispatching measures in case of disturbance are excluded by applying First-In First-Out rule in simulations.

The benefits of modifications to the track infrastructure, the timetable, and the signaling system, in terms of consecutive delays reduction, are estimated by inserting a range of primary delays. The research highlighted the need of a step further than currently planned in the infrastructure development to improve the line's robustness.

This is significant for the relationship between IMs and RUs, as the same infrastructure or planning/scheduling improvements could be measured in a different way from each other contractor, with economic impact on the infrastructure use agreements.

Keywords

Stability, Robustness, Micro simulation, Timetable, Railway infrastructure

1   Introduction

Investments in railways usually require massive resources from both Infrastructure Managers (IM) and Railway Undertakers (RU): alignment modification and signaling system upgrades on one hand, and rolling stock on the other, should be carefully designed and examined. This is why every modification should be benchmarked and measured on actual results. The relevance of robustness of operation is rising, especially on densely occupied networks, typically European suburban railways, and a method is needed to benchmark its improvement given by modifications.

IM and RU can choose on a variety of indices to assess robustness, which are affected by traffic volume increase and depend on primary delays in different ways: this paper analyses the relation between some of this indices, the traffic volume, and the primary delays. Being based on micro-simulation it is a resources consuming process, so we propose, in addition,

an effective method to reduce the computational load with reasonable approximation.

Railway systems are often characterized by high traffic density and heterogeneous circulation, which is sensitive to disturbances: high service frequency requires short headways and minimal buffer times between following trains, which has considerable influence on delays propagation. Traffic density and occurrence of disturbances in railway operation are often in positive correlation (Wiklund, 2002), and the extent of disruptions is also strongly affected by the traffic volume. An environment capable to absorb disturbances is needed to keep the service quality while increasing the production volume: a system where we can cope with disruptions and unexpected events without significant modifications to the operation is referred to as robust (Takeuchi & Tomii, 2005). Robustness is directly related to the timetable planning, and depends on the infrastructure and rolling stock characteristics, among others. Some of the tools to gain robustness in the timetabling process are the running time supplements and the buffer times in headways between trains; the reduction of heterogeneity in trains' characteristics is also indicated as a method to rise robustness by Salido et al. (2008).

Railway Undertakers' and Infrastructure Managers' different business target lead the first ones to favor buffer times between trains, as the running time supplements increase the scheduled running times and drop the service appeal to the passengers; on the other hand, Infrastructure Managers make profits from the sale of train slots, which availability is cut by the introduction of buffer times. Many parameters are available to measure robustness, according to the purpose of the performance analysis, and, to our knowledge, there is no literature on the negotiation on robustness performance: different KPIs better suit either RUs or IMs point of view. The international development of the railway markets is shining a light on the need to integrate the railway systems across different countries, legislations and technological infrastructures. This is why the need for a shared method to evaluate the performance in robustness is rising: interoperability on one side, and the diversity of railway systems require an approach adaptable to different contexts.

Different techniques to evaluate railway robustness have been proposed in the last years, some of them based on analytical approaches, others based on empirical models and what-if analysis. Vroman, Dekker, & Kroon (2006) proposed the relation between a timetable homogeneity and the propagation of delays due to interdependencies between trains, while the inequality of headway time dispersion was proposed by Carey & Carville (2000). A what-if analyses was also suggested to evaluate the effects of simulated incidents on a railway line: Salido, Barber, & Ingolotti (2008) assessed robustness in terms of average delay per train and settling time, to evaluate different solutions of their rescheduling model.

Mattson (2007) studied the interferences between trains under different capacity utilization values through a micro-simulation tool: this appeared to be the most precise way to analyze secondary delays, even though it needed very detailed input and is a time consuming process. The detailed input and output give flexibility to the method, making it adaptable to different contexts: signaling systems, operation rules, rolling stocks, and infrastructure alignments can be modelled through the micro-simulation of any guided transport system. Pellegrini, Marlière, & Rodriguez (2014) found out that the simulation of speed variation dynamics is too hard to compute in re-scheduling problems, compared to the accuracy gain; nevertheless, it is effective for estimation of actual train interactions through the signaling system: the robustness increment given by better acceleration and braking performance can be investigated though micro-simulation to compare different kinds of rolling stock.

The methods proposed measure robustness of singular solutions, but still their behavior

is to be understood as a function of train frequency: in this paper we propose an evaluation of the link between some robustness indicators and the boost of train frequency. The results show that some indicators are more affected by the traffic volume increase than others: general conclusions about the disruptions propagation and fade are shown in the last section.

The method that we propose is described in section 2, where the parameters analyzed, the micro-simulation tool, the procedures to examine the traffic volume influence and to reduce the computational load are explained. In section 3 we show the application of the method to the railway corridor between The Hague and Rotterdam, in The Netherlands: we described with good approximation the total delay as a function of the primary delay as a quadratic parabola; a linear relation was found between the fading time and the primary delay, that revealed independent from the traffic volume; furthermore, the capacity consumption was shown not to be an exhaustive indicator of timetables' robustness. General conclusion for the method proposed and the possible further research are given in section 4.

2   Methods

Previous approaches tended to measure robustness parameters to compare different timetables (Salido, Barber, & Ingolotti, 2008): they defined a timetable's robustness as the percentage of random disruptions lower than a given value that could be tolerated with no modification in traffic operations. Our study is focused on robustness appraisal in relation to the traffic volume: we put in relation the extent of given disruptions with their effect on the operation, comparing the trend of some robustness parameters as a function of the primary delay and of the traffic volume. We identified some relevant robustness parameters, listed below, and propose hereby their evaluation through the generation of several timetables. The behavior of each parameter can be defined by a compact index, calculated against primary delays through a what-it analysis. The indices describe the increase of disruptions impact as a function of the primary delay. These metrics will be studied as a function of the traffic volumes to assess the feasibility of different railway system configurations, such as infrastructure alignments, rolling stocks, signaling systems. All the indices should be normalized to compare their connection with the traffic volumes and the primary delays: for this reason, every index should be set equal to 1 for the original timetable, operated on the basic scenario. This allows to compare the different dependencies of robustness indices on the traffic volume for every scenario.

Being based on micro-simulation, the method would be high resources consuming, and a procedure to reduce the amount of iterations needed is proposed. Furthermore, the simulation process can be simplified by the exclusion of dispatching: this is a reasonable assumption on relatively short railway lines, where dispatching is usually realized at far ends rather than at intermediate stations. The First In First Out rule can be applied at junctions: even if local trains swaps can take place according to the amount of primary delay, the study focuses on the sensitivity of the robustness parameters against the primary delay and the individual peculiarities of each simulation are taken into account as part of the general phenomenon.

2.1 Robustness parameters
Our method proposes a mixture of analytical and micro-simulation derived parameters. We chose parameters suitable to answer these questions to assess the robustness:
- In what degree is the line exploited?

- In what measure and mode does a disruption affect the line?
- What is its individual impact on trains?

The parameters chosen are explained below.

Capacity consumption

This parameter can be measured through the compression method from the International Union of Railways (UIC, 2004). It gives a measure of the line use and the average headway and buffer time of the timetable. The line capacity consumed by a timetable can be visualized on a space-time graph by moving the courses together as close as possible, and can be calculated as the ratio of the line exploitation divided by the total time period,

$$\eta = \frac{t_e}{t_p},\qquad(1)$$

where η is the capacity consumption, t_e is the time of line exploitation, and t_p is the period duration. The line exploitation t_e can be measured through the compression method by micro-simulation, gathering the train paths all together with no modification to trains order and running times. The route occupancy is defined as the minimum time that can be computed between two courses of the same service. This work is feasible only on simple and short line sections, where no takeovers are scheduled: in other cases a critical buffer chain should be found, combining several train sequences by computer-based tools. The International Union of Railways recommends maximum values of capacity consumption for peak hours and daily periods to keep satisfying operating quality; this values come from experience, and the parameter depends on too many factors to be considered an exhaustive KPI.

Total delay

The total delay d represents the magnitude of the disruption impact on the line: it is the total deviation from the scheduled timetable and can be calculated as the sum of every train's delay

$$d = \sum_{i,s} d_{is} \ \forall d_{is} \geq t_d,\qquad(2)$$

with d_{is} being the delay of train i registered at station s (difference between real and scheduled time), and t_d being the delay threshold, under what we considered the train punctual.

The total delay measured can be approximated to a square parabola as function of the primary delay, so we propose its second derivative to address the effect amplification of disruptions as their bulk increases. We can find the approximated parabola parameters a, b, and c

$$d(p_d) = a \cdot p_d{}^2 + b \cdot p_d + c,\qquad(3)$$

where p_d is the primary delay, through the regression method. The resulting total delay sensitivity index is so determined:

$$i_d = d''(p_d) = 2a.\qquad(4)$$

Settling time

With appropriate distribution of running time supplements and buffer times between trains, the railway traffic should recover from delay after the expiration of the primary cause of delay. The time lapse within which all the trains moving on the network get back on schedule can be considered an indicator of the timetable stability, meaning its ability to recover from disturbances. It clearly depends on both running time supplements and buffer times: the first allow trains to recover from delay individually, while the seconds reduce the interferences to the following trains.

The settling time growth as a function of primary delays is not regular, and another method is needed to compare measurements from different timetables. Our suggestion is to gage the curves amplification compating them with the original timetable through linear regression. Let us define $S_t = \{s_{t_1}, s_{t_2}, ..., s_{t_n}\}$ the array of settling time values measured on timetable t, after the generation of primary delays from 1 to n minutes; if we call the original timetable "a", the measured array is $S_a = \{s_{a_1}, s_{a_2}, ..., s_{a_n}\}$; then we define the array $\overline{S}_t = m_t \cdot S_a = \{m_t \cdot s_{a_1}, m_t \cdot s_{a_1}, ..., m_t \cdot s_{a_n}\}$: this is the curve to be associated to the timetable t, taking m_t as its multiplication factor. We calculate the amplification factor m_t minimizing the difference between the real measured curve and the multiplied one: it is each timetable's settling time indicator of sensitivity to primary delays.

$$m_t := \min\left(\left(m_t \cdot p_{a_1} - p_{t_1}\right)^2 + \left(m_t \cdot p_{a_2} - p_{t_2}\right)^2 + \cdots + \left(m_t \cdot p_{a_n} - p_{t_n}\right)^2\right)$$
$$= \sum_{j=1}^{n}\left(m_t \cdot p_{a_j} - p_{t_j}\right)^2. \tag{5}$$

If we take the original timetable curve as the basic effect, the multiplication factor m_t can be meant as an indication of how many times the effect is amplified in higher frequency timetables.

Number of delayed trains and average delay per train

These parameters focus on the effect of circulation hitches on separate trains: the passengers' point of view is kept in consideration in this way, as we measured what they experience individually. We propose the same regression method valid for the settling time to calculate the sensitivity of the share of delayed trains to the primary delay.

The average delay per train d_t is calculated dividing the total delay by the number of trains delayed n

$$d_t = \frac{d}{n}. \tag{6}$$

The average delay per train can reasonably be approximated to a straight line, when measured as a function of primary delay. For this reason our sensitivity index is the gradient of the approximated line.

2.2 Traffic volume

The indices behavior can be studied with regard to the increase of traffic volume. For this reason we developed some timetables starting from the original one, with limited modifications at train characteristics. Stop patterns and trains order keep equal for all the trains and new trains scheduled should only be copies of existing trains in the timetable. Starting from an original timetable, which will be the reference in the following steps, we

need to develop new timetables increasing the traffic volume. These timetables are built by stepwise frequency increases: we can add new services to the current schedule, keeping the existing trains order.

The original timetable could contain different categories of train. The share of train categories within a timetable is one of its peculiar characteristics. For this reason, the ratio between categories consistence should be kept equal, or at list on the same scale, in all the timetables. Defined the array containing the number of trains of each category in the timetable t, $C_t = \{c_{t_1}, c_{t_2}, ..., c_{t_z}\}$, and the total traffic volume of each timetable

$$v_t = \sum_{l=1}^{z} c_{t_l}, \tag{7}$$

we should keep as regular as possible the share $\frac{c_{t_q}}{\sum_{l=1}^{z} c_{t_l}} = \frac{c_{t_q}}{v_t}$ for each category q in the total traffic volume in every timetable t.

The traffic volume can be increased up to consume the whole capacity, in which buffer times between services are nulled and running time supplements are reduced to shrink the heterogeneity.

2.3 Micro-simulation and reduction of computational load

Our simulation was based on the micro-simulation tool OpenTrack. The program uses continuous computation of train motion equations and simulates the interaction between trains through discrete processing of signal boxes state. Given user defined infrastructure, rolling stock, and timetable databases, it is possible to calibrate the simulation through a performance parameter individually set for every train. This is a crucial parameter that influences the analyses output: it rules the percentage of train's maximum tractive effort used and the percentage of max allowed speed that the train will reach either in ordinary or delayed condition. Though we can reasonably assume that a delayed train driver tries to stick back to the timetable running at the maximum performance and speed available, it is hard to model the standard behavior. It is clear that higher performance parameter values for the standard operation reduce the running time margin, affecting the capability of one train to recover from delays along its path, increasing the follow-up delays.

Micro-simulation's massive computation load is well known (Mattsson, 2007), so we also propose a method to reduce the amount of scenarios to simulate and the resources needed, which we called the skimming method. The parameters' behavior against primary delays depends on the specific hindered train: according to its own scheduled running time supplement and to the margin time in the following train headway, the same disruption could affect different shares of trains and generate different amount of delays. For this reason, the disruption should be simulated against every train to measure its effect on the timetable, meaning considerable resources employment.

Our skimming method proposes to only apply a very detailed analysis of one parameters to the original timetable, measuring the effects of the same disruption given individually to each train. The analyses is not extended to all the trains, timetables and scenarios under test: it is rather the selection basis for the most representative train, with respect to the effect of disruptions. The loss of information due to the reduction of simulation should be contained, so we needed an indicator of approximation goodness. We suggest the total delay to be the parameter to compare the impact of disruptions affecting different trains, as it synthetically represents the overall hindrance phenomenon through its magnitude.

The total delay on the line is measured as function of primary delay separately for each train. We calculate, then, the average total delay among all the trains given a primary delay

and choose the most representative one comparing its behavior with the average. If we define the array of total delay values associated to each train primary delay from 1 to n minutes $D_c = \{d_{c_1}, d_{c_2}, \ldots, d_{c_n}\}$ and the analogous average total delay array $\overline{D} = \{\bar{d}_1, \bar{d}_2, \ldots, \bar{d}_n\} = \left\{\frac{\sum_{c=1}^{C} d_{c_1}}{C}, \frac{\sum_{c=2}^{C} d_{c_2}}{C}, \ldots, \frac{\sum_{c=n}^{C} d_{c_n}}{C}\right\}$, the course c selected to be the most representative one is the one which total delay array is the closest to the average total delay $\overline{D}$:

$$c: \min\left(\left(\frac{\sum_{c=1}^{C} d_{c_1}}{C} - d_{c_1}\right)^2 + \left(\frac{\sum_{c=1}^{C} d_{c_2}}{C} - d_{c_2}\right)^2 + \cdots + \left(\frac{\sum_{c=1}^{C} d_{c_n}}{C} - d_{c_n}\right)^2\right) =$$
$$\min \sum_{j=1}^{n} \left(\frac{\sum_{c=1}^{C} d_{c_j}}{C} - d_{c_j}\right)^2. \tag{8}$$

The load reduction can be estimated through the ratio between the number of simulations needed before and after the skimming method.

We can define the array of the primary delay values generated $P_d = \{p_{d_1}, p_{d_2}, \ldots, p_{d_n}\}$, where n is the number of primary delay values generated. In addition, we need the array of the traffic volume of each test timetable, measured as the number of trains in the time period subject of study, $V = \{v_1, v_2, \ldots, v_{n_{tt}}\}$, where n_{tt} is the number of the test timetables. We should run a simulation for every amount of primary delay, given to every trains, on every scenario studied: the number of simulations needed is

$$n_{si} = n \cdot \sum_{j=1}^{n_{tt}} v_j \cdot n_{sc}, \tag{9}$$

with n_{sc} being the numbers of scenarios to be tested. In the skimming method we require a detailed analysis of the original timetable on the original scenario, and a shrunk analysis of one delayed course for every timetable on each scenario. The number of simulations needed with the skimming method is

$$n_{si}^* = n \cdot v_1 + n \cdot n_{tt} \cdot n_{sc} = n(v_1 + n_{tt} \cdot n_{sc}). \tag{10}$$

The first addend refers to the detailed analysis to select the most representative train, while the second is the result of the shrunk analyses giving primary delay only to the selected train. We can rate the relative computation saving η comparing the amount of simulations needed in the two cases:

$$\eta = \frac{n_{si} - n_{si}^*}{n_{si}} = 1 - \frac{n_{si}^*}{n_{si}} = 1 - \left(\frac{v_1}{\sum_{j=1}^{m} v_j \cdot n_{sc}} + \frac{n_{tt}}{\sum_{j=1}^{m} v_j}\right) = 1 - \left(\frac{v_1}{\sum_{j=1}^{m} v_j \cdot n_{sc}} + \frac{1}{\bar{v}_j}\right), \tag{11}$$

where $\bar{v}_j = \frac{\sum_{j=1}^{m} v_j}{n_{tt}}$ means the average traffic volume per timetable.

The equation (11) shows an inverted hyperbolical saving as a function of the number of scenarios under test, meaning that the more scenarios, the better saving. The same relation is valid between the average traffic volume of the timetables and the saving, while it as a negative linear trend against the ratio between the original timetable and total traffic volume generated in all the timetables. In other words, the first term within the parenthesis quantifies the computational load of the first deep analyses to select a representative course

compared to the total load of a complete analyses applied to every scenario; the second term quantifies the saving of the mere reduction in simulations needed.

2.4 Applicability

The method described, can compare different scenarios simulating the railway system in its entirety: different models of infrastructure, rolling stock, timetable, operation sets of rules, and signaling system can be tested and benchmarked. Furthermore, the effects of several modifications can be studied either individually or giving shape to combined changes to assess the combined benefits.

Different railway-like transport systems can be modelled in the micro-simulation tool, so the method can be applied, for instance, to metros, people movers, and other guided systems, making accuracy and flexibility the strengths of this method.

3 Application: the Oude Lijn in the Netherlands

The proposed method was used to evaluate the benefits of major works that are taking place on a Dutch densely occupied railway corridor. The current timetable runs 11 trains/h between The Hague and Rotterdam. The results are discussed in following sections.

The railway is undergoing an infrastructure upgrade in Delft: a viaduct in the city center will be replaced by a tunnel. It is arranged to host four tracks, though the last two will be built in a second phase. The five set up scenarios represent respectively the current 2-tracked viaduct in Delft, the new railway tunnel in Delft with ceiling speed of through running trains increased from 100 km/h to 140 km/h, the planned quadrupled tunnel to Delft Zuid and an hypothetical extension of quadruple tracks to Rotterdam; besides, a signaling system modification is studied, being applied on the current viaduct infrastructure.
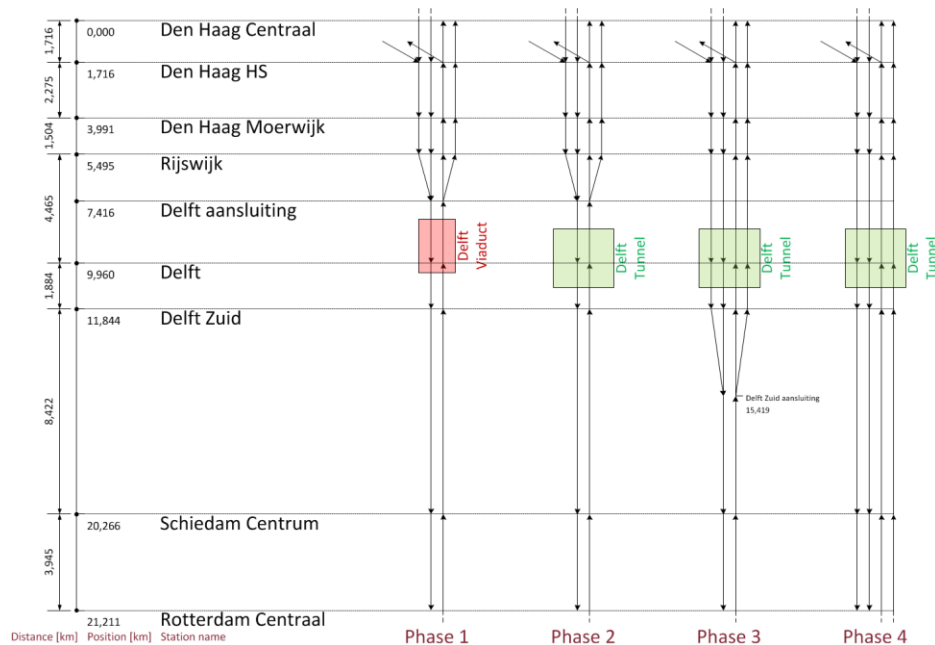

Figure 1: Track layout of each infrastructure scenario on the railway corridor

Table 1: Infrastructure scenarios under comparison

Scenario	Junction stationing	Tracks in Delft	Speed limit in Delft (km/h)	Status
Phase 1	North Delft	2 (Viaduct)	100	Current state
Phase 2	North Delft	2 (Tunnel)	140	Under construction (2015)
Phase 3	South Delft	4 (Tunnel)	140	Planned
Phase 4	None	4 (Tunnel)	140	Hypothetical
Signaling	North Delft	2 (Viaduct)	100	Hypothetical

The infrastructure scenarios' individual characteristics are summarized in Table 1 and Figure 1. In this case study we only considered the southbound traffic, pushing the northbound traffic and the interaction between opposite flows to further studies.

A total of five timetables were built to test the infrastructures: the current timetable with 11 trains/h was called "A" and the traffic volume was stepwise increased up to 18 trains/h in timetable "E". The performance parameter was updated in timetables D and E to fit shorter running times for local trains and reach better homogeneity among train paths; in every case the minimum time supplement was satisfied. We chose a set of primary delays of every integer value from 1 to 10 minutes: $P_d = \{1\ min, 2\ min, ..., 10\ min\}$, with $n = 10$: these can be considered typical daily disruptions, due to, among others, boarding at stations, minor failures at the rolling stock or at the infrastructure.

The skimming method was applied to reduce the amount of simulations needed. The original timetable was timetable A, and the reference infrastructure scenario was Phase 1, with the following results:

$$n_{si} = 3550, n_{si}{}^* = 250, \eta = 89{,}86\%.$$

The measured total delay resulting from primary delays given to every train is shown in the graph below: the average curves and the curve of the most representative course are highlighted.
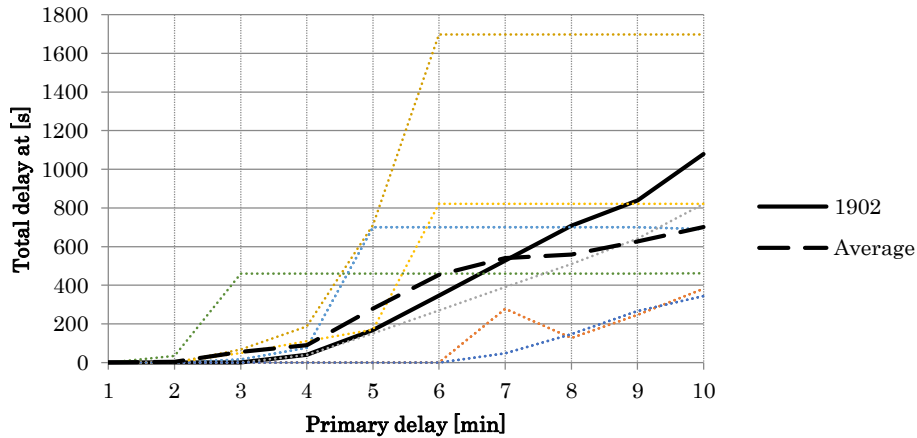

Figure 2: Total delay resulting from each trains' primary delay. The average and the selected train's total delay curves are highlighted

The resulting robustness indices are summarized in the table below for every infrastructure scenario.

Table 2: Calculation of performance indices for the five infrastructure scenarios

Infrastructure scenario	Timetable j Traffic volume v_j (trains/h)	A 11	B 12	C 14	D 16	E 18
Phase 1	Capacity consumption (%)	80,8	86,1	98,9	98,9	100,0
	Total delay	1,00	1,06	1,70	2,30	3,38
	Settling time	1,00	0,99	1,16	1,21	1,71
	Number of delayed trains	1,00	1,00	1,39	1,64	2,54
	Average delay per train	1,00	0,98	0,87	0,90	0,96
Phase 2	Capacity consumption (%)	70,3	71,9	89,4	88,1	96,7
	Total delay	0,95	1,05	1,61	2,27	2,58
	Settling time	1,00	0,98	1,14	1,19	1,61
	Number of delayed trains	1,00	1,00	1,35	1,56	2,35
	Average delay per train	0,84	0,80	0,71	0,72	0,78
Phase 3	Capacity consumption (%)	58,6	60,3	76,1	76,1	86,1
	Total delay	0,96	1,05	1,59	1,84	1,88
	Settling time	1,00	0,98	1,15	1,19	1,33
	Number of delayed trains	1,00	1,00	1,35	1,52	1,68
	Average delay per train	0,84	0,81	0,72	0,69	0,89
Phase 4	Capacity consumption (%)	60,8	63,1	70,6	73,9	78,3
	Total delay	1,11	1,16	1,52	1,82	2,26
	Settling time	1,00	0,99	1,12	1,16	1,37
	Number of delayed trains	1,00	1,00	1,34	1,52	1,41
	Average delay per train	0,77	0,75	0,68	0,63	0,79
Signaling	Capacity consumption (%)	60,3	65,0	83,3	82,5	88,9
	Total delay	1,04	1,07	1,61	1,85	2,20
	Settling time	1,00	0,99	1,17	1,20	1,32
	Number of delayed trains	1,00	1,00	1,53	1,74	2,00
	Average delay per train	1,02	0,98	0,84	0,83	0,76

The operation of the 2-tracked tunnel will permit an increase of the train frequency from 11 trains/h up to 12 trains/h per direction, while the extension of quadruple tracks from Rijswijk through Delft will enable the operation of up to 16 trains/h.

The maximum UIC 406 leaflet track capacity consumption will be reduced from currently 80% to 70% and finally 60% for the different infrastructure and basic timetable scenarios. The total delay of southbound trains as function of the primary delay is well described by quadratic parabolic functions. The sensitivity of the infrastructure and timetable to knock-on delays does not change significantly after the operation of 12 trains/h per direction on initially 2 tracks in the new tunnel in Delft.

The independency of the average delay per train sensitivity against the traffic volume brought us to a broad regression analyses for every scenario, in which all the timetables were grouped together. On average, the simulation results indicate that for every minute of primary delay at The Hague, for primary delays higher than 3 minutes, the following trains experience around 37 seconds consecutive delay on average. In general, the settling time of train delays in any of the infrastructure scenarios is not really affected by the train

frequency.

3.1 Discussion

The skimming method allowed a reduction of work load of almost 90%, opening the possibility of dramatic reduction of computation time and resources needed. The assumed absence of dispatching only affected few simulations with order alterations in the 2-tracked stretches. Considerable impact on the performance came, instead, from dispatching at the final station, with particular regard to the Phase 4 scenario: the tracks layout and allocation within the station generated several interdependencies and itinerary conflicts between arriving trains, resulting in worse robustness on the line. The capacity at stations is a known complex problem (Jensen & Landex, 2013): its consequences on lines and on networks should not be disregarded. In this particular case we can state that in Phase 4 scenario Rotterdam central station would be the bottleneck, being a critical robustness sink.

The results highlight the different sensitivity of the parameters to primary delays and to traffic volume.

The capacity consumption seems to be quite an incomplete indicator for robustness: it is affected by the performance parameter and the running time supplements. In all the scenarios the capacity consumption did not increase from a traffic volume of 14 trains/h to 16 trains/h and in some cases it decreased; the same happened for the last traffic volume boost, showing restricted increases. This is due to the change of the performance parameter in the schedule: trains can easily be compressed together with tighter schedules, and no evidence in the capacity consumption, as shown by timetables C, D, and E; on the other hand, the time supplement availability to recover from delays shrinks to the minimum, and disruptions' effects grow. In the specific case study the best reduction in capacity consumption was observable in Phase 3, which was the only one to tolerate up to 16 trains/h keeping safe in 75 % limitation of peak hour capacity utilization (UIC, 2004).

The total delay was the parameter most impacted by traffic volumes increase: for the current infrastructure – Phase 1 – its sensitivity to primary delays more than triplicated comparing a timetable witch 18 trains/h with the current 11 trains/h. It is also noticeable that sensible differences between the simulated scenarios could only be appreciated with 16 trains/h or more.

The settling time showed more stability against the traffic volume increase: its sensitivity index kept below 1,25 up to 16 trains/h. In addition it seemed more independent on the upgraded infrastructure, as all the scenario gave very similar sensitivity values up to 16 trains/h.

Surprisingly, the share of trains involved in the disruption appeared extremely stable with small traffic growth. From 14 trains/h, though, the sensibility spread more than the other indices, up to rather high values. This could be explained by the nature of the case study: the main differences between infrastructure scenarios were after the point of disruption, and delays were measured at the end of the line; the timetables A and B were almost identical, and it is understandable that infrastructure improvement like the partial extension of a 4-tracked stretch would not enhance the ability of single trains to recover from their own delay. At the same time the average delay per train seemed totally blind to the traffic volume. Every scenario showed that the grade of regressed line would slightly drop in higher traffic volumes. Moreover, the 0 delay point of the regressed line was linked to lower primary delay values when traffic increased. This can be interpreted as a rise of interaction between trains with the traffic increase, so that more trains are hindered by the previous train, but in a smaller amount.

Through the results collected we could state that the robustness of operation will not be

improved by the new tunnel in Delft until it is provided with four tracks. Good results were also reached through the bare signaling system upgrade on the current infrastructure: sensitivity reduction of parameters against primary delay could be obtained through this measure, measurable similar to the advantage given by the 4-tracked section extension. Furthermore, comparison of all the indicators' behavior as functions of frequency in all the infrastructures showed that the 4 tracks section extension would reduce the Total Delay by reducing the number of trains hindered, while a closer interaction between the trains would be gained by the signaling works, hindering more trains by a lower amount. By implementing ETCS Level 1 with braking curve supervision instead of changing track layout, the robustness indicators would improve as in the scenario with the extension of the complete line with 4 tracks to Delft Zuid without ETCS.

Gathering the information from the indices we could state that rising the traffic volume, the settling time seems to be marginally affected, while the total delay raises. It means that the disruption should take effect in the same time lapse. At the same time we found that the average delay per train is independent from the increase of traffic volume, that may mean that disruptions spread among trains in packed timetables, rather than increasing the amount individual delay; the two pieces of information match, meaning that the total delay increases because more trains are contained in the same settling time, each of them is hindered by the same amount.

4   Conclusions and further studies

This paper presented an effective and economic method to benchmark infrastructural and operational scenarios.

The method proposed highlighted the necessity of a further step in new infrastructure building in Delft: real benefits to robustness will be achieved only by the extension of the 4-tracked section. Similar results could be obtained by new a signalling system implementation, although it would be not feasible for just one line on the network.

The method allowed the comparison of different infrastructure scenarios and showed the efficacy of some actions to improve the operation stability, rather than others. The flexibility given by a micro-simulation based approach suits to benchmark and compare different infrastructures, rolling stocks operational rules and railway-like transport system. In addition, we proposed an effective procedure to reduce the micro-simulation's typically heavy computational load to make the analyses lean; the skimming method could be improved and adapted to other contexts to reduce the computing time, which would open the gates to the use of micro-simulation in real-time problem solving such us re-scheduling.

This research's implications include the availability of new negotiation tools between Infrastructure Managers and Railway Undertaker; the benchmarking is needed to measure improvements from different solutions. The paper shows the lack of information of the capacity consumption indicator about robustness, even though a correlation between the line exploitation and timetable's robustness evidently exist. The relation should be examined in depth together with the link between capacity consumption, headways between trains, running time supplements and robustness.

Acknowledgements

References

Abril, M., Barber, F., Ingolotti, L., Salido, M., Tormos, P., & Lova, A. (2008). An Assessment of Railway Capacity. *Transportation Research Part E, 44*, 774-806.

Carey, M., & Carville, S. (2000). Exact heuristic measures of schedule reliability. *Journal of Operational Research Society, 51*, 666-682.

Cerreto, F. (2011). *Stability and robustness analysis of timetable, infrastructure and operations in a densely occupied corridor. Master thesis final report.* Rome.

Friebel, G., Ivaldi, M., & Vibes, C. (2010). Railway (De)Regulation: A European Efficiency Comparison. *Economica, 77*, 77-91.

Hooghiemstra, J., & Teunisse, M. (1998). The Use of Simulation in the Planning of the Dutch Railway Services. In *Proceedings of the 1998 Winter Simulation Conference* (pp. 1139-1145).

Jensen, L. W., & Landex, A. (2013). Measuring Robustness of Timetables in Stations using a Probability Distribution. *5th International Seminar on Railway Operations Modelling and Analysis - RailCopenhagen2013* (pp. 429-447). Copenhagen, Denmark: IAROR.

Mattsson, L. G. (2007). Railway Capacity and Train Delay. In *Critical Infrastructure* (pp. 129-150). Berlin: Springer Heidelberg.

Pellegrini, P., Marlière, G., & Rodriguez, J. (2014). Optimal train routing and scheduling for managing traffic perturbations in complex junctions. *Transportation Research Part B*, 58-80.

Salido, M. A., Barber, F., & Ingolotti, L. (2008). Robustness in Railway Transportation Scheduling. *World Congress on Intelligent Control and Automation*, (pp. 2880-2885). Chongqing, China.

Takeuchi, Y., & Tomii, N. (2005). Robustness Indices for Train Rescheduling. *1st International Conference on Railway Operations Modelling and Analysis RailDelft2005*. Delft.

UIC. (2004). *Capacity (UIC code 406)*. Paris, France: International Union of Railways (UIC).

Vroman, J. C., Dekker, R., & Kroon, L. G. (2006). Reliability and heterogeneity of railway services. *European Journal of Operational Research, 172*, 647-665.

Wiklund, M. (2002). *Järnvägstransportsystemets sårbarhet – struktur för modellformulering och metodutveckling*. VTI-meddelande 932-2002.