Evaluation of robustness indicators using railway operation simulation

L. W. Jensen[1], A. Landex[2] & O. A. Nielsen[1]
[1]*Technical University of Denmark, Department of Transport, Denmark*
[2]*Ramboll Danmark A/S, Denmark*

Abstract

The classical way of evaluating the robustness of railway timetables is the use of microscopic simulation. This is precise and offers a high level of detail, but it also requires a high amount of work. The alternative is to use robustness indicators that directly or indirectly indicate the robustness of a railway system. However, the semantics of these are mainly unknown and indicators are therefore best for comparison of alternatives.

The paper therefore reviews and evaluates different robustness indicators against a microscopic simulation. This evaluation show that the indicators compare well to the microscopic simulation and are, to some extent, able to predict the outcome of the simulation.

1 Introduction

Robustness of railway operations has gained more and more attention as the importance of service and travel time reliability has been recognised [1, 2]. In railways, the robustness is usually evaluated using microscopic simulation of delays in the operation [3]. While this is precise, it requires a high amount of initial coding and calibration compared to macroscopic models and methods [4].

An alternative, or supplement, to the microscopic simulation, is the use of robustness indicators (or measures of robustness). A large amount of indicators exist that directly or indirectly indicate the robustness of a railway system based on its characteristics.

While the semantics of a microscopic simulation are well-known, the semantics of robustness indicators are typically not - especially when qualitatively combined.

WIT Transactions on The Built Environment, Vol 135,
www.witpress.com, ISSN 1743-3509 (on-line)
doi:10.2495/CR140271

This makes it hard to efficiently use the robustness indicators for other purposes than comparison of alternatives. So far, not much research have been done within the semantics of indicators and the comparison between indicators and simulations.

This paper therefore evaluate a number of robustness indicators by comparing with a microscopic simulation (using RailSys [5]) on the North West line in Denmark. This shows how well indicators perform compared to detailed simulation and information about semantics of these indicators. Uncovering these results will make it possible to evaluate the robustness of timetables more efficiently.

In section 2, the definitions of robustness is shortly described together with a review and description of different indicators and their suitability in the different planning phases. section 3.1 describes the methodology and data for the evaluation and analysis. The results are then given and discussed in section 3.2 and 3.3. A conclusion is given in section 4.

2 Robustness

Although timetable robustness has gained more attention, there is no exact definition of robustness throughout literature. A common definition of robustness is the ability of a timetable to absorb smaller delays [4, 6, 7] with or without light dispatching measures.

Time supplements are necessary to ensure that timetables are robust according to this definition. Running and dwell time supplements are used to absorb delays, while buffer time placed between train runs reduce delay propagation (knock-on delays). Both the amount of supplements and where they are added are crucial for the robustness of a timetable [4]. Thus a very robust timetable can be achieved simply by increasing time supplements and increasing buffer times by decreasing capacity consumption. However, the addition of excessive time supplements and low capacity consumption has an negative impact on travel times and frequency of service. There is thus a delicate balance between robustness and fast travel times. [8] extend the more common robustness definition and incorporate this element of efficiency in the following robustness definition: *"A railway system that is robust minimizes the total weighted real travel time of the passengers, in case of frequently occurring, small disturbances*".

2.1 Robustness indicators

The following sections review indicators that theoretically can indicate the robustness of a timetable either partly or completely according to the discussed definitions. Initial delays, realized passenger travel time and the size and distribution of buffer times and supplements are used as measures of an indicators ability to measure robustness. In the end of the section, the review is concluded by comparing and discussing the indicators described.

2.1.1 Capacity consumption – UIC406

The UIC 406 capacity method [9] is an analytical method used to measure capacity consumption on railway lines. The method uses the approach of compressing train paths. The capacity consumption is then expressed as the ratio between the completion time of the uncompressed and the compressed timetable. This ratio is thus an expression of the available buffer time between consecutive trains. This is easily observed as the compression method eliminates the minimum buffer time between consecutive train paths. In this context it is important to note that the division into line section is very important [10]. The longer the sections, the lesser information is contained in the ratio. This is especially important when the traffic is very heterogeneous. This can be seen by considering a fast train followed by a slow train between two overtaking/junction stations. The smallest buffer time will generally occur at the beginning of the section and the largest buffer time at the end of the section. Using the UIC 406 method on such a section will only reveal the minimum buffer time, thus neglecting the larger buffer times at the end of the section.

2.1.2 Heterogeneity indices

Heterogeneity indices can be used to measure the distribution of trains on line sections and at stations. These indices can be used to indirectly indicate the robustness of a timetable by measuring the spread of buffer times. Heterogeneity indices are not suitable for lines with scheduled bidirectional operation (single track lines for example).

[11] proposed a number of heterogeneity indices that are able to measure the distribution and heterogeneity of trains over the hour. Two of these are SSHR and SAHR. SSHR is based on the minimum headway, h_i^-, between consecutive trains on a track section. The second measure, SAHR, is based on the arrival headway between trains, $h_{t,i+1}^A$. This measure is proposes as arrival heterogeneity seem to be more important as fast trains catch up on slow trains at the end of sections [11].

$$SSHR = \sum_{i=1}^{n} \frac{1}{h_i^-} \tag{1}$$

$$SAHR = \sum_{i=1}^{n} \frac{1}{h_i^A} \tag{2}$$

Both SAHR and SSHR is based on the absolute headway. This means that these indicators cannot be used to compare timetables where the number of trains differ. Based on [12] we therefore suggest the following heterogeneity index for a cyclic timetable where $h_{t,i}$ is the headway time at arrival or departure:

$$H = 1 - \sum_{i=1}^{n} \min\left(\frac{h_{t,i}}{h_{t,i+1}}; \frac{h_{t,i+1}}{h_{t,i}}\right) \cdot \frac{1}{n} \tag{3}$$

2.1.3 Complexity indices

Complexity indices are generally used to evaluate the complexity of infrastructure and the timetable at a station. These indices are indirect indicators of robustness.

www.witpress.com, ISSN 1743-3509 (on-line)

Several complexity indices exist that offer increasing level of detail. The indices are shortly described below. For a more thorough description of these indicators we refer to [12, 13].

The infrastructure complexity of a station, φ_n, can be calculated as the ratio between incompatible train route combinations, n_k, and number of train route combinations, n_Σ. n_λ is used to adjust for the number of routes that cannot be set consequentially:

$$\varphi_n = \frac{n_k - n_\lambda}{n_\Sigma - n_\lambda}. \tag{4}$$

The above indicator does not account for the plan of operation. This can be incorporated by using the probability of conflicts between the train routes ($p_{k,ij}$). This probability can be found by multiplying the number of trains using the 1st train route (n_i) and the 2nd train route (n_j) and dividing this number by the total number of trains squared (N^2):

$$p_{k,ij} = \frac{n_i \cdot n_j}{N^2} \tag{5}$$

$$\varphi_p = \frac{p_k - p_\lambda}{p_\Sigma - p_\lambda} \tag{6}$$

However, the above indicator does not account for the amount of time a route is blocked. This can be included by using the minimum headway times $t_{h,min,ij}$ between route combinations and the time period considered, yielding the occupation ratio, W, of the station:

$$W = \frac{1}{T} \cdot \sum \frac{n_i \cdot n_j}{N} \cdot t_{h,min,ij}. \tag{7}$$

None of the three methods above accounts for the exact timetable and the distribution of buffer times at the station. A new complexity index, φ_d, was therefore proposed by [12, 13] that makes it possible to assess the complexity of a timetable. This index is based on the probability of train delays and the buffer time between conflicting train paths. Thus the method takes the size and distribution of buffer times into account as well as expected delays. For more information, see [13].

2.1.4 Train path fix points and risk profiles

[14] proposes to use fix points to measure the complexity of a train path. Fix points are points in the timetable where a train path is dependent on another train path. Examples of fix points are: scheduled crossings and overtakings, level junctions and transition, terminal and transfer stations [14]. Fix points are thus closely related to the complexity indices described in section 2.1.3, that is stations where consecutive delays are likely to occur. A measure of train path complexity for a train or a group of trains can be the amount of kilometres or minutes of running time per fix point.

Given the fix points, a risk profile can be created for a train or a group of trains as the amount of time supplements between fix points [14]. This can be used as an indicator of the timetable's ability to absorb delays on the different sections.

Table 1: The ability of robustness indicators to measure size and distribution, over distance (s) or time (t), of initial delays, buffer times or time supplements. PTT = passenger travel time. L = lines, S = stations, A = aggregated. I = infrastructure, TT = timetable (including I), P = plan of operation, D = delays.

	Initial delays		Time supp.		Buffer times		PTT	Applicable	Input
Indicator	Size	Dist(s)	Size	Dist(s)	Size	Dist(t)			
UIC 406					(•)			L	TT
Heterogenity						•		L & S	TT
Infra. complexity					(•)			S	I (P)
TT complexity	•	•			•	•		L & S	TT D
Fix points			•	•				L & S	P
WTTE	•	•	•	•	•	•	•	A	TT D
Simulation	•	•	•	•	•	•	(•)	A	TT D
Max-Plus	•	•	•	•	•	•		A	TT D

2.1.5 Robustness indicator including passenger travel time

[8] includes the aspect of efficiency by measuring robustness, as the weighted travel time extension (WTTE). This is calculated as the relative difference between the total weighted real travel time and the nominal (planned) travel time of passengers. To obtain the total weighted real travel time either real delays has to be used or a simulation of delays have to be conducted. The lower the value of the indicator the more robust the timetable is.

2.2 Other means of robustness evaluation

Microscopic simulation is regarded as the most precise method to measure the stability and robustness of a timetable. If the initial coding of the infrastructure has to be reduced or the data is not available, macroscopic simulation can be used [4]. This is usually relevant for early planning phases where reduced precision is accepted.

As an alternative to simulation, max-plus algebra can be used to derive the stability or robustness of a timetable. Based on running times, minimum headway times, supplements and initial disturbances the timetables ability to absorb these delays can be derived. As [4] uses this method on a macroscopic representation of the infrastructure it has some of the same disadvantages as a macroscopic simulation.

2.3 Summary and discussion

Table 1 show an overview of the different indicators described in this section. The table shows what the different indicators can capture either directly or indirectly (marked in parenthesis). Furthermore, the last column indicates the needed input. In this context infrastructure complexity is suited for early/strategic planning, fix points and UIC406 for strategic to tactical planning if the train order is known. The remaining are only suited for the tactical level (timetabling) as a timetable is required. All indicators, except the complexity indices, can be used on a macroscopic to microscopic level.

Regarding fix points and train path risk profiles, we propose that this approach can also be extended to include the minimum buffer time to the following train path if the timetable is available.

Lastly, it should be noted that other robustness indicators exist as well. These have not been considered for this short review as they are quite similar to the ones reviewed.

3 Evaluation of indicators on the North West railway line

As stated in the introduction robustness indicators are relative measures of robustness. This means that the semantics of the indicators are mainly unknown. In this section an evaluation of the indicators are therefore carried out. Section 3.1 describes the methodology of this evaluation while the results are given and discussed in section 3.2 and 3.3.

3.1 Methodology

The indicators reviewed in sections 2.1.1 to 2.1.4 are evaluated by comparing indicator results with the result of a microscopic simulation. The North West line in Denmark is used as a case. This line is currently being extended from single to double track between the Lejre (Lj) and Vipperød (Pe) (see figure 1). An increase of the line speed from 120 to 160 km/h is also part of the project. The capacity increase achieved is used to improve the service in the 2020 timetable (see figure 2). This yields the following four scenarios that together with the possibility to isolate the line from the network makes it a good case for this analysis.

1. 2010 timetable (with loko-hauled trains) on the old infrastructure
2. 2010 timetable (with loko-hauled trains) on the new infrastructure
3. 2010 timetable (with trainsets) on the new infrastructure
4. 2020 timetable (with trainsets) on the new infrastructure

The scenarios are chosen so only variable is changed at a time. Thus between scenario 1 and 2 the infrastructure is changed, between 2 and 3 the rolling stock used is changed and between 3 and 4 the timetable used is changed.

The microscopic simulation is conducted using RailSys. Perturbations for the simulation are generated on the basis of track occupation data (RDS data) from the 2010 timetable to obtain a calibrated simulation model. In this context the

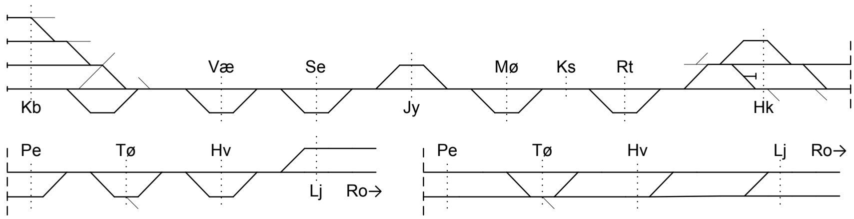

Figure 1: Schematic track plan before and after extension. Lower left part: base scenario. Lower the right part: future scenario.

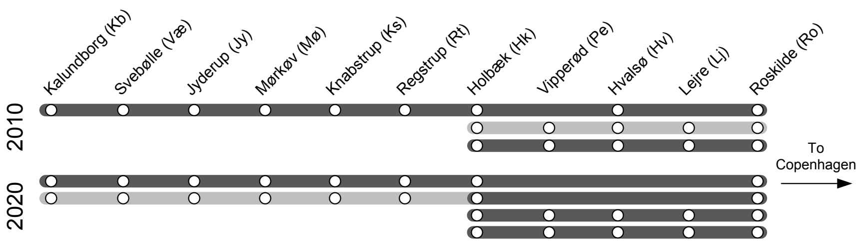

Figure 2: Line diagram in 2010 and 2020. Light grey lines: peak hours only.

North West line is considered as a closed system where trains enter with a certain amount of input delay at Roskilde station and exit with a certain amount of exit delay. In the calibration, input delay distributions are assigned on stations between Roskilde and Kalundborg so the exit delay at Roskilde is equal to the recorded delay. In the simulation of the future infrastructure the calibrated distributions are used to obtain a realistic result. For all simulations the robustness is measured both quantitatively, as punctuality and average delay, and qualitatively by analysing the quantitative results. Especially the stability of the system is important. Following [4]'s definition of local stability, the timetable can be considered robust or stable if the exit delay is smaller than the input delay.

For the calculation of robustness indicators minimum headway times, buffer times, time supplements and minimum running times are collected from the RailSys model. For the calculation of complexity indices at crossing stations only through-going train routes have been considered.

3.2 Indicator results

Figure 3 shows some of the indicator results for the four scenarios. From the UIC 406 calculation it is seen that the amount of buffer time is significantly lower in scenario 1 compared to scenario 2-4 on the single track section between Lejre and Vipperød. This is expected as this is the section that is upgraded to double track.

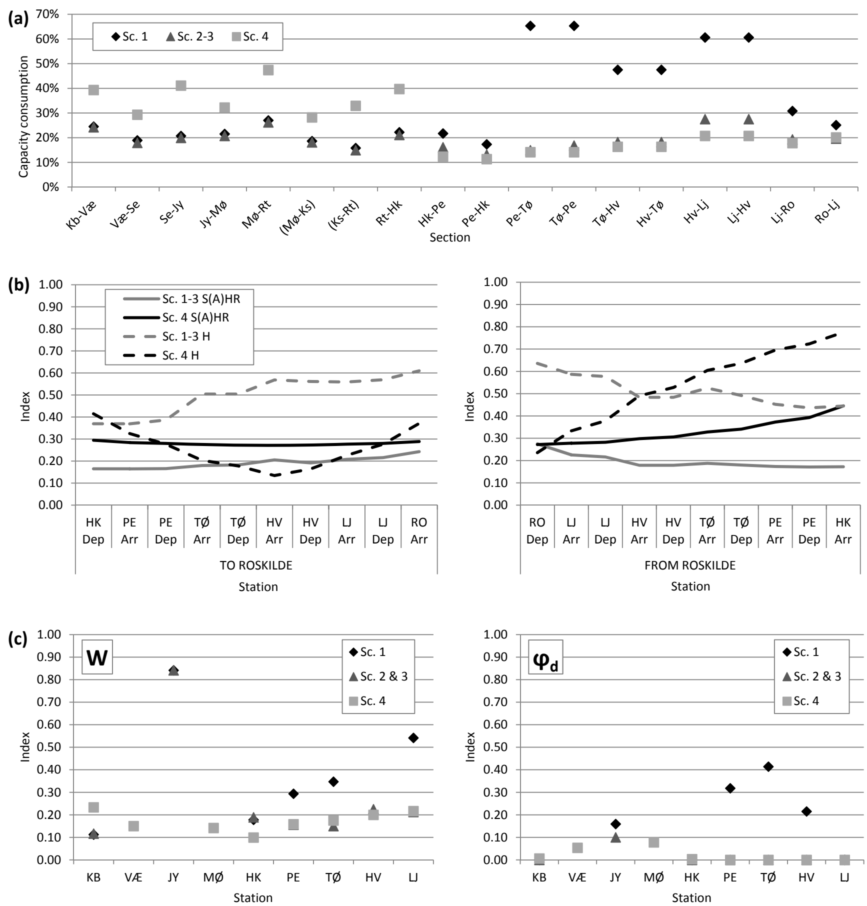

Figure 3: (a) UIC406 capacity consumption (b) Heterogeneity indices (c) Complexity indices.

The increase in trains from scenario 1 to 4 is also seen as the buffer times are smaller on the unaltered single track between Holbæk and Kalundborg.

The heterogeneity index, H, (figure 3b) show that the (relative) heterogeneity in scenario 4 is lower than in scenario 1-3 towards Roskilde and approximately the same in the other direction. The $SAHR$ index show that the heterogeneity of scenario 4 is higher than in the other scenarios. This is a result of the higher amount of trains operated and thus lower headways times.

In figure 3c the complexity indices, W and φ_d, are shown for the four scenarios. Values of φ_n and φ_p are 1 for Kalundborg and the crossing and transition stations. The values of φ_n and φ_p drop to 0.5 in scenario 2-4 for the crossing stations that is situated on the upgraded line. This is caused by the separation of the two directions.

φ_p for Holbæk, where some of the trains terminate, show that the complexity is increased in scenario 4 due to the number of increase in trains. The more detailed complexity indices W and φ_d show that the crossing stations and the transition station Vipperød are of highest concern. This is especially the case with φ_d where these are the only stations where the value is above 0. For Hvalsø-station W indicates no difference among scenarios.

From the analysis of the train patterns, with the use of fix points and risk profiles, the amount of running time supplements are particularly interesting. Scenario 1-3 are operated with high amounts of running time supplements in the excess of 20% in average. This is much higher than Rail Net Denmark's recommended 5%. Some of the reason for this is scheduled waiting time. However, in scenario 4 the running time supplements are much lower. 6% for the outbound (from Roskilde) Kalundborg trains and only 3% for the inbound. For the Holbæk trains the running time supplement is 8.6% which is in line with Rail Net Denmark's recommendation of 9% at 160 km/h.

The conclusion to be drawn from the robustness indicators are: the crossing stations are of concern in terms of delay propagation and the amount of running time supplements are very low in scenario 4. The single track section between Vipperød and Lejre makes scenario 1 vulnerable in terms of delay propagation. However, there is a large amount of running time supplement. Scenario 4 seems very vulnerable as the amount of supplements are low and the available buffer time between Kalundborg and Holbæk has been decreased.

3.3 Comparision with simulation results

The delay data for the 2010 timetable show that trains enter the North West line with a fairly high amount of delay, but exits with less and more acceptable delay. Following the definition of stability by [4] scenario 1 is stable as the output delay is smaller than the input delay. The simulation of scenario 1 shows that delays increases at the stations, especially the crossing/transition stations between Lejre and Holbæk. This is also identified by φ_d. However, the simulation also show that the timetable is able to quickly absorb delays due to the high running time supplement. As the infrastructure is upgraded in scenario 2 and 3, but the timetable is the same, the simulation of these scenarios shows even higher robustness and stability. However, considering the robustness definition by [8] these scenarios are probably not robust as they contain a high amount of unused supplements (inefficient) leading to scheduled delay of passengers (long travel times). Thus the scenarios are stable according to the definition by [4], but lack efficiency.

In terms of scenario 4, the combination of low running time supplements and a high amount of input delay at Roskilde station makes the scenario less robust than the others. The timetable is still stable as the output delay is a bit smaller than the input delay. As expected from the indicators the problem is the Kalundborg trains, where the timetable is unstable on the trip from Kalundborg to Holbæk (delay is added). The trains to and from Holbæk on the contrary performs very well, but is also operated with a higher amount of running time supplement as described above.

The following indicators compared well to the results of the simulations: W, φ_d, UIC406 and train path risk profiles. A link between results of the simulation and the heterogeneity indices may also exist, but is difficult to identify due to the difference in running time supplements. φ_n and φ_p captures the change when going from single to double track and the increase in trains at Holbæk (φ_p), but otherwise none. These two indicators are therefore more valuable in infrastructure planning than in timetabling.

4 Conclusion

In this paper we have reviewed six types of indicators that theoretically can measure robustness of timetables either directly or indirectly based on the initial delays, time supplements, buffer times and realised travel times. A comparison of indicator results with a microscopic simulation showed that the more detailed complexity indices, the UIC 406 results and the train path risk profiles performed well and captured robustness changes. However, it was not possible to uncover semantics of the indicators. Yet, the results seem to suggest that the timetable complexity, φ_d, for a station should be no more than approx. 0.01 if it is to be considered robust.

For future research more case studies and scenario analysis have to be conducted to undercover semantics of indicators in more detail. This would make it possible to recommend values of indicators to IMs and RUs for timetable and infrastructure planning purposes.

Acknowledgements

The authors would like to thank The Danish Council for Strategic Research for making it possible to conduct this research. Also many thanks to Rail Net Denmark for providing the infrastructure, timetable and delay data.

References

[1] van Oort, N., *Service reliability and urban public transport design*. Ph.D. thesis, Delft, the Netherlands, 2011.

[2] Chang, J.S., Assessing travel time reliability in transport appraisal. *Journal of Transport Geography*, **18(3)**, pp. 419 – 425, 2010.

[3] Nielsen, O.A., Landex, A. & Frederiksen, R.D., *Passengers route choices in delayed rail networks*, Springer, volume 46 of *Schedule-based Modelling of Transportation Networks: Theory and Applications*, 2008.

[4] Hansen, I.A. & Pachl, J. (eds.), *Railway Timetable & Traffic - Analysis, Modelling, Simulation*, 2008.

[5] Bendfeldt, J., Mohr, U. & Müller, L., Railsys, a system to plan future railway needs. *Computers in railways*, pp. 249–255, 2000.
[6] Goverde, R.M., *Punctuality of railway operations and timetable stability analysis*. Ph.D. thesis, Delft, the Netherlands, 2005.
[7] De-Los-Santos, A., Laporte, G., Mesa, J.A. & Perea, F., Evaluating passenger robustness in a rail transit network. *Transportation Research Part C: Emerging Technologies*, **20(1)**, pp. 34–46, 2012.
[8] Dewilde, T., Sels, P., Cattrysse, D. & Vansteenwegen, P., Improving the robustness in railway station areas. *European Journal of Operational Research*, **235(1)**, pp. 276–286, 2014.
[9] UIC, Capacity (UIC Code 406). Technical report, International Union of Railways (UIC), 2004.
[10] Landex, A., *Methods to estimate railway capacity and passenger delays*. Ph.D. thesis, Lyngby, Denmark, 2008.
[11] Vromans, M.J.C.M., *Reliability of railway systems*. Ph.D. thesis, 2005.
[12] Landex, A. & Jensen, L.W., Measures for track complexity and robustness of operation at stations. *Journal of Rail Transport Planning & Management*, **3(1)**, pp. 22–35, 2013.
[13] Jensen, L.W. & Landex, A., Measuring robustness of timetables in stations using a probability distribution. *RailCopenhagen*, 2013.
[14] Schittenhelm, B. & Landex, A., Development and application of Danish key performance indicators for railway timetables. *RailCopenhagen*, 2013.