Quantifying railway timetable robustness in critical points

Emma V. Andersson, Anders Peterson and Johanna Törnquist Krasemann

Linköping University Post Print

N.B.: When citing this work, cite the original article.

Original Publication:
Copyright: Elsevier
http://www.elsevier.com/

Postprint available at: Linköping University Electronic Press
http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-106703
Quantifying railway timetable robustness in critical points

Emma V. Andersson1*, Anders Peterson1 and Johanna Törnquist Krasemann1,2

1 Linköping University, Department of Science and Technology, 601 74 Norrköping, Sweden
2 Blekinge Institute of Technology, Department of Computer Science and Engineering, 371 79 Karlskrona, Sweden.

Several European railway traffic networks experience high capacity consumption during large parts of the day resulting in delay-sensitive traffic system with insufficient robustness. One fundamental challenge is therefore to assess the robustness and find strategies to decrease the sensitivity to disruptions. Accurate robustness measures are needed to determine if a timetable is sufficiently robust and suggest where improvements should be made.

Existing robustness measures are useful when comparing different timetables with respect to robustness. They are, however, not as useful for suggesting precisely where and how robustness should be increased. In this paper, we propose a new robustness measure that incorporates the concept of critical points. This concept can be used in the practical timetabling process to find weaknesses in a timetable and to provide suggestions for improvements. In order to quantitatively assess how crucial a critical point may be, we have defined the measure Robustness in Critical Points (RCP). In this paper, we present results from an experimental study where a benchmark of several measures as well as RCP has been done. The results demonstrate the relevance of the concept of critical points and RCP, and how it contributes to the set of already defined robustness measures.

*Corresponding author (emma.andersson@liu.se, +46 (0) 113 63 108)

Keywords: Railway traffic, Timetabling, Robustness measures, Delay management

This paper is based on “Introducing a new quantitative measure of railway timetable robustness based on critical points,” by E. Andersson, A. Peterson and J. Törnquist Krasemann, which appeared in 5th International Seminar on Railway Operations Modelling and Analysis, RailCopenhagen 2013.
1. Introduction

A tendency seen for quite some time is a growing demand for railway capacity. During 2011, in total 188 million railway journeys were made in Sweden, which corresponds to 11.4 billion passenger-kilometres [25]. Solely for the last five years, this means an increase of the Swedish railway passenger traffic by more than 10% (ibid.). This trend has led to an increase in the number of operating trains, which in turn has led to a high, at times even very high, capacity consumption and a congested, delay-sensitive network. Frequent delays result in high costs for the operators and the Swedish Transport Administration (Trafikverket) as well as high socio-economic costs for the overall society. Train delays are typically classified as either primary delays or secondary delays. Primary delays are associated with an initial source of disturbance such as a signal failure, or an unusual, lengthy passenger exchange at a certain station. Secondary delays (also denoted knock-on delays or consecutive delays) are caused by the interdependencies between trains where one delayed train may affect the trains surrounding it giving rise to a propagation of the delay. An important objective in railway timetable construction is, thus, to schedule trains so that the risk of delay propagation is limited.

We define a robust timetable, as a timetable in which trains are able to keep their original train slots despite small primary delays and without causing unrecoverable delays to other trains.

In order to maintain certain robustness, margin time (also referred to as buffer time, slack time or time supplements) is inserted into the timetable. In this paper we distinguish between runtime margin, added to a train’s shortest runtime between two stations, and headway margin, added to the technical minimum time separation between two consecutive trains sharing the same infrastructural resource. The purpose of the runtime margin is mainly to absorb smaller delays enabling the delayed train to recover, while both types serve to limit the risk of knock-on delays. It is also important to mention that the margin provide the dispatchers with certain flexibility when rescheduling the traffic to prevent delays from spreading further. In a study by Andersson et al. [3] where the on-time performance of several train services with comparable travel times is analysed, it is concluded that the variations in on-time performance is significantly affected by how the inserted margin can be used operationally by the dispatchers when disruptions occur. The drawbacks, however, of inserting run time margin is the increase in travel time while both types of margin time increase the consumption of network capacity (see e.g. UIC [32]).

The challenge in creating robust timetables is twofold: 1) A robustness measure that accurately captures the recoverability properties of the timetable is required, as well as 2) a method that suggests how to modify the timetable in order to increase the robustness in line with other given planning objectives. Before a timetable is actually used in practice, or executed in a simulated environment, it is difficult to predict how the traffic will react to disturbances and to what extent the delays that may occur will spread. Hence, already at this early planning stage, accurate robustness measures are important to use. There is also a need for indicators that point out where the weaknesses in the timetable are located and where margins should be inserted to achieve a higher robustness. In this paper, we focus on robustness measures that serve to assess the sensitivity to smaller delays by identifying specific weaknesses in a draft, or finished, timetable.

Previously proposed robustness measures can e.g. point out trains with a small amount of runtime margins, or sections that are heavily utilised. They are, however, not capturing the
interdependencies between different trains sufficiently and do not point out specific weakness in a timetable where margins should be inserted, or which train slots that should be modified at a certain section to increase the robustness. For highly-utilised railway networks with heterogeneous traffic, this is important. To enable this extended weakness analysis, we introduce a new concept referred to as critical points. Critical points refer to very time-sensitive dependencies between different pairs of trains at different locations in the network. In the context, which this paper is focused on, such points typically occur when trains enter a line behind an already operating train, or where trains overtake each other. We also define a measure of the robustness in a critical point, RCP. The critical points are intended to be used in the practical timetabling process to identify weaknesses in a timetable whereas RCP can provide suggestions for robustness improvements.

In the following section we present a summary of related work that describes how robustness in railway traffic timetables is measured in various ways. Then we present the concept of critical points and the proposed RCP measure. This measure, along with a selection of previously proposed measures, is then applied in an experimental study; first on a limited, fictive case for illustrative purposes and later on a real world example. The measures are analysed and the corresponding values are compared in order to study what information each measure provides and how the concept of critical points can be applied when constructing robust timetables. In the final section, we present our conclusions and provide some ideas for future research.

2. Measures of timetable robustness

Robustness in railway timetables can be quantified and measured in various ways. In this section we provide an overview of existing robustness definitions used by the research community. As we will see, many definitions incorporate information on traffic performance, which means that the timetable has to be used, or at least simulated, before its robustness can be evaluated. In our context, focusing on ex-ante measures, no such information is available. Therefore, in subsection 2.2, we will narrow our scope to robustness measures that can be computed solely based on information from a timetable.

2.1 Definitions of robustness

During the last decade several approaches have been proposed to investigate, measure, compare, improve, and optimise timetable robustness. Robustness refers to, e.g., “the ability to resist to ‘imprecision’” (Salido et al. [21]), the tolerance for “a certain degree of uncertainty” (Policella, [20]) or the capability to “cope with unexpected troubles without significant modifications” (Takeuchi and Tomii [30]).

According to Dewilde et al. [9] a robust timetable minimises the real passenger travel time in case of small disturbances. The ability to limit the secondary (i.e. knock-on) delays and ensure short recovery times is necessary, but not enough to define a robust timetable according to the authors.

Also Schöbel and Kratz [22] have defined robustness with respect to the passengers and as a robustness indicator they use the maximum initial delay possible to occur without causing any missed transfers for the passengers.
Takeuchi et al. [24] have also defined a robustness index with respect to the passengers. They mean that a robust timetable should be based on the passengers’ inconvenience, which in turn depends on e.g. congestion rate, number of transfers and waiting time.

Goverde [11] on the other hand has defined a timetable as stable (and also robust) when delays from one time period do not spread to the next period. The approaches relies on that the timetable is periodic.

Salido et al. [21] have presented two robustness definitions. The first definition is the percentage of disruptions lower than a certain time unit that the timetable is able to tolerate without any modifications in traffic operations. A disruption here refers to a delay of one single event in the execution of the timetable. The second definition is whether the timetable can return to the initial stage within some maximum time after a delay bounded in time.

As indicated by the definitions above, robustness analyses are focused on recovering capabilities and how inserted margins can be operationally utilised. Kroon et al. [14] describe a robust timetable as a timetable in which initial delays can be absorbed, few initial delays result in secondary delays for other trains and delays can quickly disappear due to light dispatching operations.

In this paper we will use the term robustness as the timetable’s ability to handle small delays where a robust timetable is a timetable that can recover from small delays and keep the delays from spreading over the network. In a robust timetable, trains should be able to keep their originally planned train slot despite small delays and without causing unrecoverable delays to other trains.

Measures of railway timetable robustness can be divided in two groups: Ex-ante measures, which are related to the timetable characteristics, and ex-post measures which are based on the traffic performance. Measures relying on the traffic performance can not be calculated unless the timetable has been executed in real time or analytically with disturbances, or at least simulated. Measures related to the timetable characteristics can be computed and compared already at an early planning stage without knowledge of the disturbances. Fig. 1 depicts the fundamental difference between the two types of measures.

Robustness measures based on the traffic performance are by far the more common of the two types mentioned, both in research and industry. Typically, measures are based on punctuality, delays, number of missed connections, or number of trains being on time to a station (possibly also weighted by the number of passengers affected). For example Delorme et al. [8] measure the sum of secondary delays for each train in a timetable, Bük er and Seybold [5] measure punctuality, mean delay and delay variance, Larsen et al. [15] use consecutive delays and tardiness as performance indicators and Medeossi et al. [18] measure the conflict probability. All of the examples above are based on perturbing a timetable with observed or simulated disturbances.

In this paper we will only consider ex-ante robustness measures which are applicable at an early stage of the timetable construction and which can be used to determine the quality of an alternative timetable design. Consequently, we will not further consider robustness measures based on the traffic performance, nor will we consider measures that use disturbances and passenger disutility as input.
Fig. 1. Two types of robustness measures used when analysing timetable robustness; Timetable characteristics and Traffic performance.

2.2 Measures related to timetable characteristics

A commonly used expression for robustness is the amount of margin inserted in the timetable. Margins can be added to the runtime and stopping time to prevent trains from arriving late despite small delays. Headway margin is used between any two consecutive trains in the timetable which serve to reduce the knock-on delay effects. A disadvantage of margins is, however, increased travel times and increased consumption of line capacity. Therefore the robustness is often measured by the price of robustness, which is the ratio between the cost of a robust timetable and of an optimal timetable without robustness, see for example Cicerone et al. [7], and Schöbel and Kratz [22].

Not only the amount of runtime margin, but also its allocation is important. Today it is not unusual that the margin allocation is based on intuition, or is either proportional to the average disturbances or uniform over a train journey. However, Vromans [28] and Vekas et al. [27] show that this leads to poor results when it comes to delay recovery.

Among others, Kroon et al. [12], Fischetti et al. [10] and Vromans [28] have studied the allocation of margins. The authors use the Weighted Average Distance (WAD) to calculate the relative distance to the runtime margin from the start of the line to capture the allocation. Dividing the line into \( N \) sections and letting \( s_t \) denote the amount of margin associated with section \( t \), WAD can be calculated as:

\[
WAD = \sum_{t=1}^{N} \frac{2t - 1}{2N^2} \cdot s_t.
\]

WAD is a relative number between 0 and 1, where WAD = 0.5 means that the same amount of margin is placed in the first half of the considered line as in the second half, whereas WAD < 0.5 means that more margin is placed in the first half.

Both Vromans [28], Kroon et al. [12] and Fischetti et al. [10] mean that it is preferable with runtime margin concentrated early on the line (i.e. a small WAD value) so that early appearing delays do not spread further down the line. However, if the disturbances occur later on the line, the runtime margin located prior to the occurrence of the disturbance may be of no use. Therefore Vromans [28] conclude that average runtime margin should be allocated on the middle part of a line, with a slight shift to the first half of the journey.
Clearly, robustness is also gained by increasing the headway margins. Yuan and Hansen [31] have studied how to allocate headway margins spatially at railway bottlenecks. They concluded that the mean knock-on delay time for a train decreases exponentially with the size of the headway margin to the preceding train.

The distribution of headway margins is considered by Carey [6], who has developed heuristic measures both for individual trains and for complete timetables. Three headway-based measures are proposed: The percentile of the headway distribution for every train type, the percentage of trains which has a headway value smaller/larger than some target value, and the standard deviation and mean absolute deviation of the headways. A method to increase the robustness, suggested by Carey [6], is to maximise the minimum headway.

Robustness is also gained by increasing traffic homogeneity, i.e. by making speed profile and stopping pattern more similar for a sequence of trains, so that the minimum headway times fall and the buffer times increase, Salido et al. [21]. Vromans et al. [29] have studied how to make a timetable less heterogeneous and list several options: Slowing down long-distance trains, speeding up short distance trains, inserting overtakings, letting short distance trains make even shorter journeys or equalising the number of stops. The authors have measured heterogeneity by considering the smallest headway $h_i$ between each train $i$ and any consecutive train using the same track section. In an attempt to quantify the robustness at the track section, the authors summarised the reciprocals of these smallest headways. The measure SSHR (sum of shortest headway reciprocals) hence also captures the spread of the trains over time and is calculated as:

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

A disadvantage of this measure, also mentioned by the authors, is that it does not capture where the smallest headway is located. It is more crucial that the trains arrive on time than depart on time and therefore the arrival headway is of more interest. Alternatively, one can restrict the consideration to headways on arrival only. The restricted measure is called SAHR (sum of the arrival headway reciprocals). Lindfeldt [16] has analysed several heterogeneity measures and found that SSHR and SAHR show good performance when explaining secondary delays in simulations.

In the robustness calculation made by Goverde [11] the stability margin and recovery time for a periodic timetable is considered. The stability margin is the largest time increase that all events in a timetable can have at the same time, but still be stable. By a stable timetable Goverde refers to a timetable in which all planned events can be operated within the given cycle time. The recovery time between two events is the maximum time the first event can be delayed without disturbing the other event.

There are also models intended for calculating the capacity utilisation for a line, UIC [32]. As a result from these models we get information of where in the network there is congestion, and where the traffic is sensitive to disturbances. Mattson [17] analysed the relationship between train delays and capacity utilisation. It is, however, not only the amount of trains on the tracks that affects the robustness. It is also of great importance in what intervals the trains run on the tracks. As Vromans et. al. [29] have concluded; the headway between the trains needs to be equalised to achieve a more robust timetable.
Salido et al. [21] have introduced two methods to measure robustness. The first measure is the sum of a number of timetable characteristics and traffic parameters, and the second measure is the number of disruptions that can be absorbed with the available margins. These two measures are valid for single-track lines with crossings, overtakings and heterogeneous traffic and a significant amount of stations. The measures can not define whether a timetable is robust or not, but rather compare two timetables and specify which of them that is the most robust.

Shafia and Jamili [23] have extended the second robustness measure by Salido et al. to instead consider the number of non-absorbed delays when a train is affected by a certain disruption.

The common approach for the listed robustness measures is to identify weaknesses in the timetable, delay-sensitive line sections or train slots. Most of them involve either headway or runtime margins, or where the margins are distributed in the timetable. The studied measures could be useful for identifying trains with an insufficient amount of margin or a poor margin allocation. They, however, do not suggest where and how margins should be inserted to increase the robustness. Similarly, the measures incorporating headway and heterogeneity, can single out sections where the robustness seems to be low, but they do not suggest how the robustness should be increased by e.g. which trains that should potentially be moved, or given a different runtime.

3. Critical points

Through empirical observations of the Swedish timetable and traffic during 2011, we have identified some points in the timetable that are especially sensitive to delays. Trains that are delayed in these locations frequently become even more delayed and the delays are also spreading to other trains. We will further on refer to these delay sensitive locations as critical points.

Critical points refer to very time-sensitive dependencies between different pairs of trains at different locations in the network. These locations are points in the timetable where a train is planned to be overtaken by, or to enter the network after, an already operating train. Typically these are very sensitive to disturbances and if the already operating train is just slightly delayed, it will disturb the entering train. The situation deteriorates if the delay of the operating train is so large, that the entering train is prioritised by the dispatcher and released just in front of the operating, delayed train. In such a situation the delayed operating train can not use its own margin time for recovering and could be even further delayed. Although this kind of prioritisation seems fair, it is not necessarily the best solution [1]. Since the dependency between the trains remains during a significant part of their journey, the described situation has a large impact on the trains’ performance and it is strongly influenced by the arrival constraints in the critical point.
In Fig. 2 we can see an example of a critical point. The figure depicts a cut of the Swedish Southern mainline during the morning peak period. In Tranås, a commuter train (train 8718) is scheduled to start at 8.05, just after a fast long-distance train (train 500) has passed at 8.02. This means that if the long-distance train is just a few minutes delayed there will be a conflict with the commuter train; the two trains want to use the tracks at the same time. Today the train dispatchers in Sweden have a guideline saying that a train on time should be prioritised before a delayed train in a conflict [26]. However, as a rule of thumb, the train dispatcher holds the commuter train up to two minutes in favour of delayed long-distance trains. If the long-distance train is more than five minutes late it has to run after the slower commuter train and its delay will increase until there is a possibility for overtaking or the commuter train has reached its final destination.

These situations are frequently appearing in the Swedish railway network and their negative effects for the traffic performance can easily be observed. When long-distance trains end up after a slower train in a critical point, their delays often increase to a level from which they can never recover and they might continue to spread the delays in the network for a long time. Fig. 3 shows an example of a critical point at station B, where train 2 is planned to enter the line after the already operating train 1. In the figure, train 1 is delayed at station A and ends up in a conflict with train 2 in the critical point. In the scenario to the right, train 2 runs according to its schedule, i.e. the dashed and solid grey lines coincide. Depending on the train dispatcher’s decision we get two alternative scenarios; either train 1, or train 2, is prioritised. We can see that the decision significantly influences the delay for train 1, whereas train 2 will arrive on time to the end station E in both scenarios.
Because of the large impact the critical points have on the traffic performance it is important to have a timetable robustness measure that can capture this effect of interdependencies between the trains.

### 3.1 Defining a critical point

A critical point is defined both in time and space, and there are always two trains involved. These trains are planned to run in the same direction and on the same line section track. There are two situations which we consider to be critical: The time and location when 1) a train is planned to enter a line after an already operating train, or 2) a train is planned to overtake another train. The situations are similar when it comes to dependencies between the trains in a delayed situation. We will further on refer to these two trains involved as either the entering, or the operating train. In an overtake, the overtaken train corresponds to the entering train.

If there are two trains planned to enter the network after an already operating train, we delimit ourselves to only consider the relationship between the first entering train and the operating train as a critical point. We also exclude origin stations from the critical points, since we assume that trains depart from their origin station on time in this context.

The procedure of identifying critical points in a timetable starts by selecting all trains that enter the network somewhere on the line and all trains that are being overtaken. The locations where one of these two types of events occurs will be the location of the critical points. The entering or overtaken train represents train 2 in Fig. 3. To find train 1, we search for the closest already operating train before train 2 in the critical point that runs in the same direction and at the same track. This search results in a geographical location of the critical point and we have identified the corresponding two trains involved.

### 3.2 Robustness in critical points

Since delays in critical points often result in increasing and spreading of the delays, it is important to have a high robustness in the points. A high robustness will provide the train dispatcher with good...
possibilities to solve a disturbed situation. As a measure of the robustness in critical points we introduce Robustness in Critical Points (RCP). RCP is the sum of the following three parts which are also illustrated in Fig. 4:

i) The available runtime margin for the operating train before the critical point. By available margin we generally refer to the accumulated amount of margin time from the previous point in the timetable where the train had a fixed departure time. With a large amount of runtime margin for the operating train before the critical point the possibility to arrive on time to the critical point increases.

ii) The available runtime margin for the entering train after the critical point. By available margin we generally refer to the accumulated amount of margin time to the next point in the timetable where the train has a fixed arrival time. With a large amount of runtime margin for the entering train after the critical point, the possibility to delay this train in favour of the other increases.

iii) The headway margin between the trains’ departure times in the critical point. The headway margin is calculated as the total headway time minus the minimum headway time. The total headway time is the planned time distance between the trains in the timetable and the minimum headway time is the technically possible minimum time distance between the trains. With a large headway margin the possibility for the operating train to run ahead of the entering train increases, even when delayed.
Fig. 4. The three parts of margins included in RCP.

By available runtime margin we generally refer to the accumulated amount of runtime margin time before or after the critical point. In the example, there are no other trains in the timetable except for the two trains involved in the critical point, which means that the available runtime margin is easy to calculate. If there are other trains in the timetable which the two trains involved in the critical point have to relate to, the total accumulated runtime margin time might not be able to be fully used operationally. If there are other trains scheduled close before and/or after the two considered trains, these can not deviate from their planned slots to the full extent of their margin without interacting with other traffic. Therefore the available runtime margin consists of the accumulated runtime margin possible to use operatively without disturbing any other train.

If the operating train has a stop in the critical point, before the entering train is planned to start, the dwell time margin can be added to RCP. By dwell time margin we refer to the margin time added to the minimum stop time. This margin influences RCP in the same way as the runtime margin for the operating train before the critical points. If the train is delayed it can use its dwell time margin to recover in time for the critical points.

RCP is a measure of the maximum flexibility in a critical point and it consists of the total amount of margin available. With a larger RCP value, the dispatcher will have higher prospects to handle conflicts in an effective way. The three different parts in RCP could however have different importance for the robustness depending on which actions the train dispatchers are permitted to
take in a conflict situation. The runtime margin before a critical point could for example be more useful than the runtime margin after a critical points since the train dispatchers might not want to delay a train that runs according to schedule in favour of a delayed train. An option could be to weight the three parts according to their operational utility and the flexibility that is provided. However, we will then lose the benefit of having a non-relative measure giving the robustness in exact number of seconds. Therefore, in this paper we have chosen to give the three parts equal weight.

When calculating RCP we assume that the train dispatchers are allowed to re-schedule the two trains involved in the critical point in an operational conflict. It is for example possible to operationally re-schedule several trains in a conflict situation, which would result in a higher RCP value. However, this will soon lead to a chain of reactions, hard to grasp both for the timetable constructors and train dispatchers, and therefore we restrict RCP to only consider the two trains explicitly involved in the conflict.

4. Experimental benchmark analysis
In order to analyse what type of information the previously proposed robustness measures provide and their applicability, we have implemented a selection of the robustness attributes and measures presented in previous research. We have also implemented our proposed measure, RCP, to see the differences and compatibility between the measures. The measures are first implemented in a small fictive timetable and later in a real-world instance from the Swedish Southern mainline.

4.1 Robustness measures and timetable instance
We have selected seven robustness measures for the benchmark, which are listed and described below:

1) **Number of trains per section and hour (NoT)**
   - NoT gives a good picture of the traffic density, and helps in identifying bottlenecks. It is also an indicator of the capacity usage. The measure is dependent of number of tracks at each section. When comparing NoT for several sections, the number of available tracks at each section must be equal or the number of trains could be divided with number of tracks at the sections. Robustness attributes of this type are used by Salido et al. [21].

2) **Total amount of runtime margin for each individual train (TAoRM)**
   - TAoRM is strongly connected to the capability to recover from delays. Robustness measures of this type are used by Salido et al. [21].

3) **Maximum runtime difference per partial stretch (MRD)**
   - MRD serves to capture the heterogeneity in the traffic by comparing the runtime, including margins and commercial stops, between the fastest and the slowest train. We divide the line into partial stretches which are naturally bounded by the traffic structure. There are too few trains that operate on the whole line and one line section is a too short measuring distance; hence the use of partial stretches. This measure is inspired by the work of Vromans et al. [29].

4) **Sum of Shortest Headway Reciprocals (SSHR)**
   - S SHR captures both the heterogeneity and the size of the headways and was suggested by
Vromans et al. [29], for further details see Section 2.2.

5) Weighted Average Distance (WAD)
WAD is a measure of the distribution of the runtime margins, and has been used by Kroon et al. [12] and Fischetti et al. [10], for further details see Section 2.2.

6) Percentage of headways equal to or less than the minimum value (PoH)
Poh is a measure of the occurrence of planned short headway times. The measure is inspired by the work of Carey [6] and Kroon et al. [13].

7) Robustness in Critical Points (RCP)
RCP is our proposed measure, based on the occurrence of critical points and the amount of margin in each point. The calculation of the measure consists of two parts; identification of the critical points and calculation of the corresponding RCP value. Since the studied timetables in Sweden do not explicitly include dwell time margin for the train stops, the calculation of RCP is only based on runtime and headway margins.

In Appendix A, the pseudo code for identifying critical points (Algorithm A.1) and calculating RCP (Algorithm A.2) is provided. Algorithms for computing the robustness measures enumerated as 1–6 above are given by Andersson [2].

4.2 Small fictive example
We illustrate the selected robustness measures with a small fictive example, given by the timetable in Fig. 5 (graphical format) and in Table 1 (table format). The runtime margin is also explicitly presented in Table 1, and we define two partial stretches, Station A–C and Station C–F, respectively. In this example there are no dwell time margins.

There is only one critical point in the timetable, namely at station C, where train 2 enters the network with three minutes scheduled headway to train 1. The minimum headway at station C is set to three minutes.
Fig. 5. Example timetable for robustness measures calculation.

Table 1. The timetable from Fig. 5 expressed in table format.

<table>
<thead>
<tr>
<th>Station</th>
<th>Timetable</th>
<th>Runtime margin (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Train 1</td>
<td>Train 2</td>
</tr>
<tr>
<td>A</td>
<td>Dep. 08:08</td>
<td>-</td>
</tr>
<tr>
<td>B</td>
<td>Arr. 08:16</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Dep. 08:16</td>
<td>-</td>
</tr>
<tr>
<td>C</td>
<td>Arr. 08:23</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Dep. 08:23</td>
<td>08:26</td>
</tr>
<tr>
<td>D</td>
<td>Arr. 08:30</td>
<td>08:37</td>
</tr>
<tr>
<td></td>
<td>Dep. 08:30</td>
<td>08:37</td>
</tr>
<tr>
<td>E</td>
<td>Arr. 08:37</td>
<td>08:47</td>
</tr>
<tr>
<td></td>
<td>Dep. 08:39</td>
<td>08:48</td>
</tr>
<tr>
<td>F</td>
<td>Arr. 08:46</td>
<td>08:57</td>
</tr>
</tbody>
</table>

The result of the robustness calculation is shown in Table 2. The headway between train 1 and 2 at station C is equal to the minimum headway which gives PoH equal to 12.5%.

NoT shows the number of trains at the tracks per hour and in this small example it is easy to see that section C–F has a higher utilisation than section A–C. We can also see that the SSHR value is much higher for section C–D than for any other section. This is an effect of the short headway between train 1 and train 2. When the headway between the three trains is more equal, the SSHR value decreases cf. section D–F.

TAoRM shows that train 2 has the best prerequisites for arriving on time despite some delay. Train 1 has a smaller WAD value than the other trains which relates to the fact that the train has more of its margin in the beginning of the line.
When it comes to MRD, the first partial stretch, between A – C, has a much lower value than the second. At the first stretch, the two trains are running with almost the same speed and make no stops. At the second stretch train 2 enters the line. It runs with a lower speed and makes more stops than the other trains, and hence the difference between the trains’ runtime is higher.

There is only one critical point, at station C when train 2 enters the line. The headway between the trains is 180 seconds and the minimum headway at the station is 180 seconds, i.e. the headway margin is 0 seconds. Train 1 has 60 seconds margin before the critical point, and train 2 has 60 seconds margin after the critical point. The RCP value is 120 seconds. This will practically mean that train 1 can depart from station A up to 120 seconds late, but still be able to run before train 2 at station C.

Table 2. The values of the robustness measures for the fictive example

<table>
<thead>
<tr>
<th>Section</th>
<th>Partial stretch</th>
<th>MRD (seconds)</th>
<th>TAoRM Train</th>
<th>RCP (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-B</td>
<td>A-C</td>
<td>60</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>B-C</td>
<td>C-F</td>
<td>240</td>
<td>2</td>
<td>150</td>
</tr>
<tr>
<td>C-D</td>
<td></td>
<td></td>
<td>3</td>
<td>120</td>
</tr>
<tr>
<td>D-E</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-F</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Modified timetable
To achieve a higher RCP value in a critical point we can modify the timetable and increase the three parts of margins in the RCP measure. The first modification is to increase the headway margin between the trains in the critical point by giving train 2 one minute later departure time from station C. The second modification is to increase the runtime margin for train 1 before the critical point by giving train 1 one minute earlier departure time from station A. The third modification is to increase the runtime margin for train 2 after the critical point by giving train 2 one minute later arrival time to station D (and on the continuing stations as well). The modified timetables are shown in Table 3.

Table 3. The three modifications of the example timetable

<table>
<thead>
<tr>
<th>Station</th>
<th>Modification 1</th>
<th>Modification 2</th>
<th>Modification 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Train 1</td>
<td>Train 2</td>
<td>Train 3</td>
</tr>
<tr>
<td>A</td>
<td>Dep. 08:08</td>
<td>-</td>
<td>08:28</td>
</tr>
<tr>
<td>B</td>
<td>Arr. 08:16</td>
<td>-</td>
<td>08:36</td>
</tr>
<tr>
<td>C</td>
<td>Arr. 08:23</td>
<td>-</td>
<td>08:44</td>
</tr>
<tr>
<td></td>
<td>Dep. 08:23</td>
<td>08:27</td>
<td>08:46</td>
</tr>
<tr>
<td>D</td>
<td>Arr. 08:30</td>
<td>08:37</td>
<td>08:54</td>
</tr>
<tr>
<td></td>
<td>Dep. 08:30</td>
<td>08:38</td>
<td>08:54</td>
</tr>
<tr>
<td>E</td>
<td>Arr. 08:37</td>
<td>08:48</td>
<td>09:01</td>
</tr>
<tr>
<td></td>
<td>Dep. 08:39</td>
<td>08:49</td>
<td>09:03</td>
</tr>
<tr>
<td>F</td>
<td>Arr. 08:46</td>
<td>08:58</td>
<td>09:10</td>
</tr>
</tbody>
</table>

All three modifications of the timetable give an identical increase of RCP from 120 seconds to 180 seconds, but the modifications influence the other robustness measures in different ways, see Table
4. In modification 1 all other robustness measure values have remained the same, except for the SSHR value, which has decreased for section C–F and PoH which has decreased to 0%.

Modification 2 has increased TAoRM and decreased WAD for train 1. The SSHR value has decreased at section A–B and MRD has decreased at partial stretch A–C, which indicates a better robustness.

Modification 3 has increased the TAoRM and decreased WAD for train 2. The SSHR value has increased at section C–D but decreased for section D–F. MRD has increased at partial stretch C–F because of the added runtime for train 2.

Table 4. Increased (↑) or decreased (↓) robustness in the three modifications according to the other robustness measures.

<table>
<thead>
<tr>
<th></th>
<th>Modification 1</th>
<th>Modification 2</th>
<th>Modification 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>PoH</td>
<td>↑</td>
<td>↑</td>
<td></td>
</tr>
<tr>
<td>NoT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSHR</td>
<td>↑</td>
<td>↑</td>
<td>↓/↑</td>
</tr>
<tr>
<td>MRD</td>
<td>↑</td>
<td>↓</td>
<td></td>
</tr>
<tr>
<td>TAoRM</td>
<td>↑</td>
<td></td>
<td>↑</td>
</tr>
<tr>
<td>WAD</td>
<td>Decreased value</td>
<td>Decreased value</td>
<td></td>
</tr>
</tbody>
</table>

The WAD-value has decreased with modification 2 and 3 but if this has increased or decreased the timetable robustness is hard to judge.

All modifications increase RCP with the same amount of margin but we can see in Table 4 that some of the other robustness measures indicate that modification 3 is not as good for the robustness as the other modifications. More specifically, the values for SSHR and MRD have increased at some sections, indicating that the robustness problem is just moved rather than removed. However, the best strategy to increase RCP is highly depending on the original timetable construction. In this small example, modification 1 and 2 seem to be the best choices when increasing RCP but with another timetable, modification 1 or 2 could have influenced the other measures in another way. We therefore argue that RCP can be seen as a contribution to the already existing robustness measures, to be used combined with them to increase the overall robustness.

4.3 Real-world example

The above listed measures were also used on a real-world case with data for the Swedish Southern mainline provided by Trafikverket. The Southern mainline is a double-track line between Stockholm and Malmö, see Fig. 6. This is one of the most busy railway lines in Sweden and, via the bridge connection from Malmö to Copenhagen in Denmark, also of international interest. Long-distance trains are connecting the important end point markets in Malmö/Copenhagen and Stockholm. On parts of the line there are also regional trains, commuter trains and freight trains, sharing the tracks with the long-distance passenger trains. An analysis of the on-time performance during the autumn 2011 was presented by Peterson [19].

In this example we will use the southern part of the line, a ca. 200 km long stretch between Malmö (M) and Alvesta (AV).
Fig. 6. The Swedish railway network (left) and the double-tracked Swedish Southern mainline (right). Source: Swedish Transport Administration and SJ AB (modified).

For our calculation, we have chosen the morning period 05.45–07.15, on September 8th, 2011. This is a typical Swedish autumn weekday and in a time period where we can follow one fast long-distance train, train 500, the stretch M–AV. The graphical timetable for the chosen line and time period can be seen in Fig. 7. The four partial stretches are enumerated from 1 to 4 on the right hand side. In the figure also the 14 identified critical points are illustrated and named from A to N.
A RailSys model [4] provided by Trafikverket (The Swedish Transport Administration) was used to calculate the minimum headway required at each specific section. At some points, the given (true) timetable does not fulfil this minimum headway requirement. Small differences can be explained by an inappropriate level-of-detail in the RailSys-analysis. In our analysis, however, we also detected a few points in the timetable where the scheduled headway was significantly smaller than the minimum required headway. These headway violations indicate that the timetable cannot be executed as planned, even if all trains are perfectly on time. This fact will also influence some of the robustness measure values.

When calculating MRD we need to define some longer partial stretches of the line. We have chosen the following four partial stretches which are naturally bounded by traffic pattern:

1) Malmö (M) – Lund (LU)
2) Lund (LU) – Höör (HÖ)
3) Höör (HÖ) – Hässleholm (HM)
4) Hässleholm (HM) – Alvesta (AV)

**Result from the real-world calculation**

The overall robustness measure, PoH, is 4%, which means that there are several headway values that are equal or less than the minimum headway values. If this is an acceptable value is hard to tell, but the fact that it exits headways values that are smaller than the minimum values will result in operational disturbances for the trains.

Table 5 presents the values for TAoRM and WAD. Some of the trains have no runtime margin at all, which means that they have no possibilities to recover from delays at this part of the Southern mainline. For those trains it is not possible to calculate WAD, and we denote this by ‘-’ in the corresponding entry. Many of the trains continue their journeys outside of the studied time/space network, therefore TAoRM and WAD may be misleading. But if we only consider this part of the Southern mainline it is reasonable that trains that operate the same distance should have approximately the same amount of margins. It is, however, hard to conclude if the timetable is robust or not only by looking at the amount of margin and where they are allocated. The required amount of margin and their allocation is highly correlated to the occurrence and magnitude of the disturbances the margin should be able to handle. In the timetable construction phase the disturbance distribution is unknown and therefore hard to include in the process.

Table 5. The values of the robustness measures for the Swedish Southern mainline example – TAoRM and WAD

<table>
<thead>
<tr>
<th>Train</th>
<th>TAoRM (seconds)</th>
<th>WAD</th>
<th>Train</th>
<th>TAoRM (seconds)</th>
<th>WAD</th>
</tr>
</thead>
<tbody>
<tr>
<td>45517</td>
<td>0</td>
<td>-</td>
<td>11312</td>
<td>122</td>
<td>0.391032</td>
</tr>
<tr>
<td>45513</td>
<td>0</td>
<td>-</td>
<td>11319</td>
<td>122</td>
<td>0.680328</td>
</tr>
<tr>
<td>484</td>
<td>244</td>
<td>0.549180</td>
<td>11321</td>
<td>338</td>
<td>0.596154</td>
</tr>
<tr>
<td>500</td>
<td>424</td>
<td>0.372003</td>
<td>11323</td>
<td>182</td>
<td>0.795410</td>
</tr>
<tr>
<td>522</td>
<td>144</td>
<td>0.475198</td>
<td>11325</td>
<td>60</td>
<td>0.187500</td>
</tr>
<tr>
<td>524</td>
<td>244</td>
<td>0.440530</td>
<td>11327</td>
<td>6</td>
<td>0.550000</td>
</tr>
<tr>
<td>11004</td>
<td>48</td>
<td>0.911765</td>
<td>11333</td>
<td>60</td>
<td>0.375000</td>
</tr>
<tr>
<td>11006</td>
<td>356</td>
<td>0.479383</td>
<td>1504</td>
<td>154</td>
<td>0.733333</td>
</tr>
<tr>
<td>11008</td>
<td>122</td>
<td>0.391032</td>
<td>1505</td>
<td>309</td>
<td>0.428317</td>
</tr>
<tr>
<td>11010</td>
<td>170</td>
<td>0.320752</td>
<td>1506</td>
<td>324</td>
<td>0.676852</td>
</tr>
<tr>
<td>11012</td>
<td>165</td>
<td>0.364577</td>
<td>1507</td>
<td>199</td>
<td>0.561307</td>
</tr>
<tr>
<td>11014</td>
<td>60</td>
<td>0.125000</td>
<td>1703</td>
<td>110</td>
<td>0.494652</td>
</tr>
<tr>
<td>11015</td>
<td>62</td>
<td>0.229839</td>
<td>1705</td>
<td>110</td>
<td>0.494652</td>
</tr>
<tr>
<td>11017</td>
<td>122</td>
<td>0.275313</td>
<td>1707</td>
<td>110</td>
<td>0.494652</td>
</tr>
<tr>
<td>11019</td>
<td>233</td>
<td>0.582082</td>
<td>1712</td>
<td>50</td>
<td>0.892857</td>
</tr>
<tr>
<td>11021</td>
<td>122</td>
<td>0.275313</td>
<td>1714</td>
<td>110</td>
<td>0.462567</td>
</tr>
<tr>
<td>11023</td>
<td>366</td>
<td>0.551698</td>
<td>1716</td>
<td>110</td>
<td>0.462567</td>
</tr>
<tr>
<td>11025</td>
<td>60</td>
<td>0.078947</td>
<td>6100</td>
<td>38</td>
<td>0.700000</td>
</tr>
<tr>
<td>11029</td>
<td>148</td>
<td>0.377764</td>
<td>6160</td>
<td>103</td>
<td>0.512483</td>
</tr>
<tr>
<td>1204</td>
<td>371</td>
<td>0.707659</td>
<td>7140</td>
<td>25</td>
<td>0.534483</td>
</tr>
<tr>
<td>1205</td>
<td>77</td>
<td>0.776696</td>
<td>42734</td>
<td>193</td>
<td>0.500000</td>
</tr>
<tr>
<td>1206</td>
<td>73</td>
<td>0.149128</td>
<td>44721</td>
<td>675</td>
<td>0.559722</td>
</tr>
<tr>
<td>1207</td>
<td>471</td>
<td>0.395406</td>
<td>69472</td>
<td>90</td>
<td>0.375000</td>
</tr>
<tr>
<td>1209</td>
<td>60</td>
<td>0.375000</td>
<td>69474</td>
<td>90</td>
<td>0.375000</td>
</tr>
<tr>
<td>1254</td>
<td>73</td>
<td>0.192990</td>
<td>69501</td>
<td>119</td>
<td>0.625000</td>
</tr>
</tbody>
</table>
Table 6 comprises NoT between 06 and 07 in the morning and we can identify the most utilised section as AL–MGB. However, at this particular line section there are four, parallel tracks instead of two, henceforth the section is not that heavily utilised compared to the other sections with two tracks. When looking at the solid double-track (north of AL), section ÅK–BLV is the most utilised section in terms of traffic volume per track.

The largest SSHR value is on the other hand found on section LU–FLP, where the traffic is dense with small headway values. This indicates that the robustness should be increased on section LU–FLP.

MRD shows much larger values for HM–HÖ and HÖ–LU, than the other partial stretches. This has to do with south-bound trains being overtaken at these stretches and these trains have a much longer runtime than fast long-distance train with no stops. There are no north-bound trains that are being overtaken at these stretches which result in less runtime differences. The large MRD values indicate that the robustness should be increased for HM–HÖ and HÖ–LU.

Many of the critical points are located at section LU–FLP which also is indicated by the SSHR measure. Consequently, both these measures point out LU–FLP as a section in need of increased robustness.

Table 6. The values of the robustness measures for the Swedish Southern mainline example – NoT, SSHR, MRD and RCP

<table>
<thead>
<tr>
<th>Section</th>
<th>NoT (6 a.m. – 7 a.m.)</th>
<th>SSHR</th>
<th>Partial stretch</th>
<th>MRD (seconds)</th>
<th>Critical point</th>
<th>RCP (seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AV-BLD</td>
<td>2</td>
<td>0.000896</td>
<td>HM-AV 150</td>
<td>A  581</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BLD-VS</td>
<td>2</td>
<td>0.000890</td>
<td>HÖ-HM 346</td>
<td>B  171</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VS-ERA</td>
<td>2</td>
<td>0.002038</td>
<td>LU-HÖ 614</td>
<td>C  503</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERA-DIÖ</td>
<td>3</td>
<td>0.002375</td>
<td>M-LU 240</td>
<td>D  22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIÖ-ÅH</td>
<td>3</td>
<td>0.002842</td>
<td>AV-HM -</td>
<td>E  433</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ÅH-TUN</td>
<td>4</td>
<td>0.027722</td>
<td>HM-HÖ 1282</td>
<td>F  113</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TUN-O1</td>
<td>5</td>
<td>0.008213</td>
<td>HÖ-LU 1153</td>
<td>G  210</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O1-O</td>
<td>5</td>
<td>0.07902</td>
<td>LU-M 240</td>
<td>H  61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O-HV</td>
<td>5</td>
<td>0.011069</td>
<td></td>
<td>I  325</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HV-MUD</td>
<td>5</td>
<td>0.013794</td>
<td></td>
<td>J  274</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MUD-HM2</td>
<td>5</td>
<td>0.013898</td>
<td></td>
<td>K  -8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HM2-HM</td>
<td>5</td>
<td>0.013773</td>
<td></td>
<td>L  724</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HM-MLB</td>
<td>10</td>
<td>0.053911</td>
<td></td>
<td>M  140</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MLB-TÖ</td>
<td>12</td>
<td>0.046173</td>
<td></td>
<td>N  -103</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TÖ-HÖ</td>
<td>11</td>
<td>0.049377</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HÖ-SG</td>
<td>14</td>
<td>0.056208</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SG-E</td>
<td>14</td>
<td>0.057451</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-DAT</td>
<td>14</td>
<td>0.058070</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAT-Ö</td>
<td>14</td>
<td>0.058427</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ö-STB</td>
<td>15</td>
<td>0.068269</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STB-THL</td>
<td>14</td>
<td>0.074221</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
When calculating RCP, the available amount of runtime margin is to a large extent bounded by other traffic, as mentioned in Section 3.2.

The RCP values at the critical points “K” and “N” are negative. In these points the headway is smaller than the minimum headway value, as provided by the RailSys model, and there is no runtime margin to cover for this. It means that the critical points themselves produce disturbances which have to be taken care of by other margins. These are some of the points that influence the PoH values negatively.

The negative value at point “K” may be explained as a rounding error, but at point “N”, RCP is –103 seconds. Since the headway is smaller than the minimum headway, it is impossible to run as close to the other train as the timetable shows.

The other critical points have a positive RCP value which means that there are some margins in the points. Many of the RCP values are high and the train dispatcher has several marginal minutes to use if disturbances occur.

**Discussion**

In the example from the Swedish Southern mainline some of the previously proposed measures can be used to identify trains with an insufficient amount of runtime margins, as well as where along the line most of the margin is allocated. They can also indicate sections that are more utilised than others and where an increased robustness could be needed. It is, however, hard to draw conclusions about how the timetable should be modified to achieve a higher robustness from these measures. For example the largest values for TAoRM, SSHR and MRD are found on section ÅK-BLV, LU-FLP and HM-HÖ, respectively. This does not give a clear view of the problem since they indicate poor robustness at different sections. Even if we, with the previous known measures, can get knowledge of areas where there is a lack of robustness, we do not get any suggestions of which trains we should modify to increase the robustness. That can be done by the use of critical points, which will point out specific locations in the timetable and network that could be modified to achieve a higher robustness.

To increase the robustness in the timetable with respect to the critical points, the first step should be to increase RCP at points with negative values, such as “K” and “N” in the example. Then the timetable in itself will be executable without constructing any delays. In a second step, it is recommendable to increase any low RCP value, for example at points “D” and “H” in the example. However, when increasing the RCP value, also other robustness measures will be affected. When, for example, adding runtime margin before and/or after the critical points, TAoRM will increase and the MRD and SSHR values will be affected in a way that could decrease robustness at other sections, see the modification of the small fictive example in Section 4.2.
5. Conclusions and future research

This paper discusses several ways to measure railway timetable robustness. Previously proposed measures can be used to, for example, identify trains with a small amount of runtime margin or sections that are heavily utilised. They are, however, not capturing the interdependencies between different trains sufficiently and do not point out specific weaknesses in a timetable where margins should be inserted, or which train slots that should be modified at a certain section to increase the robustness. For highly-utilised railway networks where very little symmetry is found and where the traffic is highly heterogeneous, time-critical dependencies between trains are important to identify and handle.

Our attention has been drawn to the points in timetables where trains are planned to enter a line, or to overtake another train. We believe that these points are critical for the robustness. The number of critical points, their locations and the corresponding RCP values constitute a useful measure of the robustness of either a certain train slot or a complete timetable. Critical points can easily be used in the timetabling process to identify weaknesses. The RCP measure provides the timetable constructors with substantial suggestions for where improvements should be made and which service to modify. However, when modifying a timetable to achieve higher RCP values, also other robustness indicators may be affected. The concept of critical points and RCP can be seen as a contribution to the already defined robustness measures.

Several aspects regarding the use of critical points and RCP should be further analysed, among them is how overtaking possibilities near a critical point affect the corresponding robustness. Another possible extension to the measure is to account for the runtime difference between the two trains involved in a critical point. The concept described in this paper is primarily targeting double-track lines and it is for future research to adopt it to single-track lines.

A possible way of using the proposed measure in a larger scale is to apply it in an optimisation model, maximising the timetable robustness in critical points, given some set of permitted adjustments. Such a model can be designed to, e.g., restrict the occurrence of critical points, set a minimum tolerance level for RCP values, or maximise the margins in the most critical points. Also this is part of our future work.

Acknowledgements

This research was conducted within the research project “Robust Timetables for Railway Traffic”, which is financially supported by grants from VINNOVA, Trafikverket and SJ AB. The authors are grateful for all data provided by Trafikverket and SJ AB. We would finally also like to thank the anonymous reviewers for their comments and suggestions that increased the quality of this paper.

References


Appendix A

This Appendix includes the pseudo code for identifying critical points and calculating RCP. Sets and parameters used can be seen in Table A1 and Table A2.

The railway network is divided into sections. A section can be either a station or a line section and contains one or several tracks.

For every train $i$ there is an event list $S_i$ containing the $|S_i| = e_i^{\text{train}}$ events for this train. Each event describes the trains’ activity on one specific section, e.g. for train $r$ that runs on 10 line sections and 5 station sections, there are $e_r^{\text{train}} = 15$ events. Analogously, there is a list $K_j$ representing all $|K_j| = e_j^{\text{section}}$ events occurring at section $j$. Each event is associated both to a train $i$ and to a section $j$, i.e. for every event $s$ there exists a section $j$ and a train $i$, such that $s \in K_j$ and $s \in S_i$.

Every event $s \in S_i$ has a planned start time $t_{i,s}^{\text{start}}$, and a planned end time $t_{i,s}^{\text{end}}$, along with a minimum occupation time $d_{i,s}$. For a line section $d_{i,s}$ is the minimum runtime, and for a station section $d_{i,s}$ is the minimum duration time.

Safety rules restrict the time distance $h_j$ separating any two consecutive trains using line section $j$. If a line section has multiple block sections, it might be simultaneously used by two trains, $i_1$ and $i_2$, running in the same direction, provided that they are separated by a minimum headway for that particular section. The direction of a train is controlled by the binary parameter $r_i$, and we require $r_{i_1} = r_{i_2}$.

Further, there is need for two infrastructure parameters; $o_j$ is a binary parameter that indicates whether section $j$ is a line section ($o_j = 1$) or a station section ($o_j = 0$) and $p_{i,s}$ gives the track number that train $i$ is planned to use during event $s \in S_i$.

We can calculate the amount of runtime margin for train $i$ at event $s$ as

$$m_{i,s}^{\text{runtime}} = t_{i,s}^{\text{end}} - t_{i,s}^{\text{start}} - d_{i,s} \quad \forall i \in T, s \in S_i.$$

Table A1. Sets used.

<table>
<thead>
<tr>
<th>Set</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T$</td>
<td>- set of trains</td>
</tr>
<tr>
<td>$C$</td>
<td>- set of sections</td>
</tr>
<tr>
<td>$S_i$</td>
<td>- set of events for train $i \in T$</td>
</tr>
<tr>
<td>$K_j$</td>
<td>- set of events for section $j \in C$</td>
</tr>
</tbody>
</table>

Table A2. Parameters used.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_{i,s}^{\text{start}}$</td>
<td>- Initial start time for train $i$ at event $s$</td>
</tr>
<tr>
<td>$t_{i,s}^{\text{end}}$</td>
<td>- Initial end time for train $i$ at event $s$</td>
</tr>
<tr>
<td>$d_{i,s}$</td>
<td>- Minimum occupation time for train $i$ at event $s$</td>
</tr>
<tr>
<td>$h_j$</td>
<td>- Minimum headway at section $j$</td>
</tr>
<tr>
<td>$e_i^{\text{train}}$</td>
<td>- Number of events for train $i$</td>
</tr>
</tbody>
</table>
Algorithm A.1. Identify the critical points, $P_{t_i,j}$

Let $P_{t_i,j}$ denote the critical point at section $j$ involving trains $i$ (the train that enter the network or is being overtaken at section $j$) and $\hat{i}$ (the operating or overtaking train at section $j$). Let $t_{i,x} \in \mathcal{C}$ denote the section where event $s$ of train $i$ is scheduled to, and $\omega_{i,s}$ the corresponding number in the event list for that section.

Since there are two different ways to identify a critical point depending on whether it is a location where a train is entering the line or a location where a train is planned to overtake another train, the algorithm is divided into two steps.

Step 1: Find the critical points for trains entering the network.

In step 1 $u$ is a section where a train enters the double-track line, $\alpha$ identifies a train that is already operating on the line and $\varepsilon$ is a stop flag ending the search for a critical point if there is another train entering the network at the same section before the entering train.

\[
\text{for } i \in T \text{ where } \tau_{i,1} > 1 \text{ and } \tau_{i,1} < \vert S_i \vert \text{ and } e_i^{\text{train}} > 1. \\
\alpha = 0; \varepsilon = 0; \ u = 0; \\
\text{if } (y_{i,1} > 2) \\
\text{ while } (u = 0) \\
\text{ for } s \in S_j \text{ where } s > 1 \text{ and } \alpha_{i,x} = 1 \text{ and } y_{i,x} = 2 \text{ and } u = 0. \\
\text{ Let } u = \tau_{i,s}; \\
\text{ end for } \\
\text{ if } (y_{i,1} = 2) \text{ Then let } u = \tau_{i,2}; \\
\text{ if } (u > 0) \\
\text{ while } (\varepsilon = 0 \text{ and } \alpha = 0) \\
\text{ for } v \in \{ 1, \ldots, \vert K_u \vert \} \text{ where } v < \omega_{i,s} \text{ and } \varepsilon = 0 \text{ and } \alpha = 0. \\
\text{ Let } \omega_{i,s} \text{ denote the index of the event on section } u \text{ corresponding to event } s \text{ of train } i, \ s \in S_i, \\
\text{ Let } \omega_{i,x} \text{ denote the index of the event on section } u \text{ corresponding to event } x \text{ of train } i \in T \text{ and } x \in S_i. \\
\text{ Let } \omega_{i,x} = \omega_{i,s} - v; \\
\text{ if } (\varepsilon = 0 \text{ and } \alpha = 0 \text{ and } x \leq 2) \text{ Then let } \varepsilon = 1; \\
\text{ else if } (\varepsilon = 0 \text{ and } \alpha = 0 \text{ and } r_i = r_i \text{ and } e_i^{\text{train}} > x \text{ and } x > 2 \text{ and } p_{i,x} = p_{i,x}) \\
\text{ Then let } \alpha = i; \\
\text{ end for } \\
\text{ if } (\varepsilon > 0) \text{ Then let } P_{i,\alpha,u-1} = 1; \\
\text{ end } \\
\text{ end while } (\varepsilon = 0 \text{ and } \alpha = 0) \\
\text{end for } \\
\text{end for } \]

Step 2: Find the critical points when a train overtakes another train

In step 2, $\alpha$ identifies a train that is overtaking another train.

\[
\text{for } i \in T, s \in S_i \text{ where } \tau_{i,1} < \tau_{i,s} \text{ and } e_i^{\text{train}} > \tau_{i,s} \text{ and } \alpha = 0 \text{ and } t_{i,s}^{\text{end}} > t_{i,s}^{\text{start}}. \\
\text{ Let } \alpha = 0; \\
\text{for } u_{i,x} \in K_{i,x} \text{ where } u_{i,x} > \omega_{i,s} \text{ and } \alpha = 0 \text{ and } t_{i,s}^{\text{end}} > t_{i,s}^{\text{end}}. \\
\text{ Let } u_{i,x} \text{ denote the index of the event on section } \tau_{i,s} \text{ corresponding to event } x \text{ of train } i \in T \text{ and } x \in S_i. \\
\]
\begin{algorithm}
\caption{Calculation of $RCP_{i,i,j}$ in each critical point $P_{i,i,j}$ (RCP)}
\begin{algorithmic}
\State The calculation of $RCP_{i,i,j}$ in a critical point $P_{i,i,j}$ is divided into four steps: 1) calculate $\gamma_{i,i,j}$, the headway margin between train $i$ and $i$ at section $j$, 2) calculate $\mu_{i,i,j}$, the available runtime margin for the entering train $i$ after section $j$, 3) calculate $\delta_{i,i,j}$, the available runtime margin for the operating train $i$ before section $j$, and 4) sum the three parts of RCP, $RCP_{i,i,j} = \gamma_{i,i,j} + \mu_{i,i,j} + \delta_{i,i,j}$.

We use $\theta_i^{\text{end}}$, $\theta_i^{\text{start}}$, and $\theta_i^{\text{margin}}$ to capture the possible start time, end time and runtime margin for train $i$ and $i$, respectively, and denote $j$ as the next or previous stop for the calculated train $i$ and $i$. Quantities $t_{i,s}$ and $\omega_i$ are defined as in Algorithm A.1.

\textbf{Step 1: Calculate the headway margin ($\gamma_{i,i,j}$) at the critical points.}
\For{$i, i \in T, j \in C$ where $P_{i,i,j} > 0$.}
\State Let $s$ denote the event for train $i$ at section $j$ and let $\hat{s}$ denote the event for train $i$ at section $j$, $s \in S_p, \hat{s} \in S_c$.
\State Let $\gamma_{i,i,j} = t_{i,s+1} - t_{i,s+1} - h_i$.
\EndFor

\textbf{Step 2: Calculate the runtime margin for the entering train ($\mu_{i,i,j}$) after the critical point.}

In step 2 $\alpha$ identifies the next train after train $i$ at the sections after the critical point and $\varepsilon$ is a stop flag interrupting the summation of margin time.

\For{$i, i \in T, j \in C$ where $P_{i,i,j} > 0$. Let $s$ denote the event for train $i$ at section $j$.}
\State Let $\varepsilon = 0, j = 0$;
\For{$c \in S_i$ where $t_{i,c} > j$ and $\varepsilon = 0$.}
\State If ($\varepsilon = 0$ and $\omega_i < t_{i,c} > t_{i,c}^{\text{start}}$ and $\varepsilon = 0$).
\State Then let $j = t_{i,c}$; Also let $\varepsilon = 1$;
\EndFor
\EndFor
\For{$g \in C$ where $g > j$ and $g < j$.}
\State Let $w$ denote the event for train $i$ at section $g$, $w \in S_i$.
\State If ($\varepsilon = 0$)
\State Then let $\theta_i^{\text{margin}} = \theta_i^{\text{start}} + m_{i,w}$;
\State Also let $\theta_i^{\text{end}} = t_{i,s} + \theta_i^{\text{margin}}$;
\For{$u \in \{1, \ldots, |g|\}$ where $\varepsilon = 0$.
\State Let $x \in S_i$ denote the event for train $i$ at section $j + u$.
\State Let $\alpha = 0$;
\For{$v_{n,b} \in K_{j+u}$ where $v_{n,b} > \omega_{i,x} \& \varepsilon = 0 \& \alpha = 0$.}
\State Let $v_{n,b}$ denote the index of the event on section $j + u$ corresponding to event $b$ of train $n$, $n \in T, b \in S_n$.
\State If ($r_i = r_n \& \alpha = 0 \& \varepsilon = 0$)
\State Then let $\alpha = n$;
\State If ($\varepsilon = 0$ and $\theta_i^{\text{start}} > t_{i,b}^{\text{start}} - h_{j+u}$)
\State Then let $\theta_i^{\text{margin}} = \theta_i^{\text{start}} - h_{j+u} - t_{i,b}^{\text{start}}$;
\State Also let $\mu_{i,i,j} = \theta_i^{\text{margin}}$;
\EndIf
\EndFor
\EndFor
\EndIf
\EndIf
\EndIf
\EndIf
\EndFor
\State Let $\theta_i^{\text{start}} = t_{i,s} + \theta_i^{\text{margin}} + \sum_{q \in (1, S_i)} d_{i,q}$.
\end{algorithmic}
\end{algorithm}
Let \( \theta_i^{\text{end}} = \theta_i^{\text{start}} + d_{i,c} \).

If \( \theta_i^{\text{start}} \leq s_{a,b}^{\text{start}} - h_{j+u} + \varepsilon = 0 \)

Then let \( \mu_i,j = \theta_i^{\text{margin}} \);

If \( \theta_i^{\text{start}} > s_{a,b}^{\text{start}} - h_{j+u} + \varepsilon = 0 \)

Then let \( \theta_i^{\text{margin}} = s_{a,b}^{\text{start}} - h_{j+u} - t_{i,x}^{\text{start}} \);

Also let \( \mu_i,j = \theta_i^{\text{margin}} \);

If \( \theta_i^{\text{margin}} < 0 \) then let \( \varepsilon = 1 \);

If \( j + u = j - 1 \) and \( \theta_i^{\text{end}} > s_{a,b}^{\text{start}} - h_{j+u} + \varepsilon = 0 \)

Then let \( \theta_i^{\text{margin}} = s_{a,b}^{\text{start}} - h_{j+u} - t_{i,x}^{\text{start}} \);

Also let \( \mu_i,j = \theta_i^{\text{margin}} \);

Let \( \varepsilon = 1 \);

end for

Let \( a = 0 \);

end for

if \( \theta_i^{\text{margin}} = 0 \) then let \( \mu_i,j = 0 \);

end for

\[ \text{Step 3: Calculate the runtime margin for the operating train} \ (\delta_{i,j}) \ \text{before the critical point.} \]

In Step 3, \( \alpha \) identifies the next train after the train \( i \) at the sections before the critical point and \( \varepsilon \) is a stop flag interrupting the search when we have reached a final value at \( \delta_{i,j} \).

for \( i, j \in T, j \in C \) where \( P_{i,j} \geq 0 \). Let \( \hat{s} \) denote the event for train \( i \) at section \( j \).

Let \( \varepsilon = 0 \);

Let \( j = 0 \);

Let \( \theta_i^{\text{margin}} = 0 \);

for \( c \in 1..|S_i| \) where \( c < \hat{s} \) and \( \varepsilon = 0 \).

if \( \varepsilon = 0 \) and \( \alpha_{t_{i,c}} = 0 \) and \( (t_{i,\hat{s}-c} > t_{i,\hat{s}-c} \| \hat{s} - c = 1) \)

Then let \( j = t_{i,\hat{s}-c} \);

Also let \( \varepsilon = 1 \);

end for

for \( g \in (1..|j - f|) \). Let \( w \in S_i \) denote the event for train \( i \) at section \( j - g \).

if \( \varepsilon = 0 \)

Then let \( \theta_i^{\text{margin}} = \theta_i^{\text{margin}} + m_i^{\text{runtime}} \);

Also let \( \theta_i^{\text{end}} = t_{i,\hat{s}}^{\text{start}} \);

for \( u \in 1..|g| \) where \( \varepsilon = 0 \). Let \( x \in S_i \) denote the event for train \( i \) at section \( j - u \).

Let \( v_{n,b} \) denote the index of the event on section \( j - u \) corresponding to event \( b \) of train \( n \), \( n \in T, b \in S_n \).

Let \( \alpha = 0 \);

for \( v_{n,b} \in K_{j-u} \) where \( v_{n,b} > \omega_{i,x} \) and \( \varepsilon = 0 \) and \( \alpha = 0 \).

if \( \eta = r_n \) and \( \alpha = 0 \) and \( \varepsilon = 0 \)

Then let \( \alpha = n \);

Also let \( \theta_i^{\text{start}} = t_{i,x}^{\text{start}} + \theta_i^{\text{margin}} \);

if \( \theta_i^{\text{start}} \leq s_{a,b}^{\text{start}} - h_{j-u} + \varepsilon = 0 \)

Let \( \delta_{i,j} = \theta_i^{\text{margin}} \);

if \( \theta_i^{\text{start}} \geq s_{a,b}^{\text{start}} - h_{j-u} + \varepsilon = 0 \)

Then let \( \theta_i^{\text{start}} = t_{i,x}^{\text{start}} + \theta_i^{\text{margin}} \);

Also let \( \delta_{i,j} = \theta_i^{\text{margin}} \);

Also let \( \delta_{i,j} = \theta_i^{\text{margin}} \);

if \( \theta_i^{\text{margin}} < 0 \) then let \( \varepsilon = 1 \);

end for

end for

end for

end for
if $\theta_i^{margin} = 0$ Then let $\delta_{i,i} = 0$; end for

Step 4: Calculate $RCP_{i,i}$ for every critical point.

for $i, i \in T, j \in C$ where $P_{i,i,j} > 0$.

Let $RCP_{i,i,j} = \gamma_{i,i,j} + \mu_{i,i,j} + \delta_{i,i,j}$;

end for
Fig. 1. Two types of robustness measures used when analysing timetable robustness; Timetable characteristics and Traffic performance.

Fig. 2. Example of a critical point between a fast train (train 500) and a commuter train (train 8718) in Tranås.

Fig. 3. Two different scenarios when the delayed train 1 runs either before (leftmost scenario), or after, the initially punctual train 2.

Fig. 4. The three parts of margins included in RCP.

Fig. 5. Example timetable for robustness measures calculation (graph).

Fig. 6. The Swedish railway network (left) and the double-tracked Swedish Southern mainline (right). Source: Swedish Transport Administration and SJ AB (modified).

Fig. 7. The graphical timetable from the Swedish Southern mainline, between Alvesta (AV) and Malmö (M) including the critical points.

Table 1. The timetable from Fig. 5 expressed in table format.

Table 2. The values of the robustness measures for the fictive example

Table 3. The three modifications of the example timetable

Table 4. Increased (↑) or decreased (↓) robustness in the three modifications according to the other robustness measures.

Table 5. The values of the robustness measures for the Swedish Southern mainline example – TAoRM and WAD

Table 6. The values of the robustness measures for the Swedish Southern mainline example – NoT, SSHR, MRD and RCP