Introducing a New Quantitative Measure of Railway Timetable Robustness Based on Critical Points

Emma Andersson, Anders Peterson, Johanna Törnquist Krasemann
Department of Science and Technology, Linköping University
Postal adress: SE-601 74 Norrköping, Sweden
e-mail: {emma.andersson, anders.peterson, johanna.tornquist.krasemann}@liu.se

Abstract
The growing demand for railway capacity has led to high capacity consumption at times and a delay-sensitive network with insufficient robustness. The fundamental challenge is therefore to decide how to increase the robustness. To do so there is a need for accurate measures that return whether the timetable is robust or not and indicate where improvements should be made. Previously presented measures are useful when comparing different timetable candidates with respect to robustness, but less useful to decide where and how robustness should be inserted. In this paper, we focus on points where trains enter a line, or where trains are being overtaken, since we have observed that these points are critical for the robustness. The concept of critical points can be used in the practical timetabling process to identify 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 RCP (Robustness in Critical Points). A high RCP value is preferred, and it reflects a situation at which train dispatchers will have higher prospects of handling a conflict effectively. The number of critical points, the location pattern and the RCP values constitute an absolute value for the robustness of a certain train slot, as well as of a complete timetable. The concept of critical points and RCP can be seen as a contribution to the already defined robustness measures which combined can be used as guidelines for timetable constructors.

Keywords
Railway traffic, Timetabling, Robustness measures, Delay management

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 produced in Sweden, which corresponds to 11.4 billion passenger-kilometres. Solely for the last five years, this means an increase of the passenger transport by more than 10%. (Trafikanalys [17]) This trend has led to a high, at times even very high, capacity consumption and a delay-sensitive network. Frequent delays result in high costs for the operators and the Swedish Transport Administration as well as high socio-economic costs for the overall society.

One common approach to deal with the high capacity utilisation and occurring disturbances is to create more robust timetables. By this we mean timetables in which trains 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.

In order to maintain certain robustness, margin time (also referred to as buffer time, slack time or time supplements) is inserted into the timetable. However, margins also increase travel time and the consumption of line capacity (see e.g. UIC [21]). Fundamental challenges are therefore to decide how much margin to insert and where, since its location often dictates its effectiveness. Andersson et al. [1], who have studied several train services with comparable travel times, conclude that the services have different on-time performance as an effect of variations in how the inserted margins can be used operationally.

The challenge of creating robust timetables is twofold: 1) to measure the robustness of a given timetable and 2) to modify the timetable to increase the robustness in line with other given planning objectives. Before the timetable is actually used in practice or in a simulated environment, it is difficult to predict how the traffic will react to disturbances and how possible delays may spread. Consequently, already at this early planning stage, accurate robustness measures are desired. There is also a need for indicators that point out where the weaknesses in the timetable are and where margins should be inserted to achieve a higher robustness. In this paper we focus on robustness measures which are applicable at an early stage of the timetable construction and which can be used to determine the quality of a timetable design.

Previously proposed robustness measures can e.g. point out trains with a small amount of runtime margins or sections that are heavily utilised. They can however not indicate exactly where along a train's service runtime margins should be inserted or which train slots that should be modified at a certain section to increase the robustness. To overcome this deficiency, we introduce a new concept referred to as *critical point*; points where trains enter a line behind an already operating train or where trains overtake each other. We also define a measure of the robustness in critical points, RCP. The critical points are intended to be used in the practical timetabling process to identify weaknesses in a timetable and RCP can provide suggestions for robustness improvements.

In the following section we present a summary of related work, describing how robustness in railway traffic timetables is measured in various ways. Then we present the concept of critical points and the proposed measure of the robustness in critical points. This measure, along with a selection of previously known measures, is then applied to a real world example. The measures are calculated, analysed and compared and we discuss what information that each measure provides and how critical points can be useful when creating more robust timetables. In the final section, we present our conclusions and provide some ideas for future research.

2 Measures of Timetable Robustness

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. [13]), the tolerance for "a certain degree of uncertainty" (Policella, [11]) or the capability to "cope with unexpected troubles without significant modifications" (Yoko and Norio [19]).

According to Dewilde et al. [5] 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.

Also Schöbel and Kratz [14] have defined robustness with respect to the passengers and as an robustness indicator they use the maximum initial delay possible to occur without causing any missed transfers for the passengers.

Takeuchi et al. [16] 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 [7] 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.

Salido et al. [13] have presented two robustness definitions. The first robustness 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. 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.

In this paper we will use the term robustness as the timetables 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: Those related to the timetable characteristics, and those that are based on the traffic performance. Measures relying on the traffic performance can not be calculated unless the timetable has been executed, or at least simulated, whereas measures related to the timetable characteristics can be computed and compared already at an early planning stage. Figure 1 depicts the principle difference between the two types of measures.

To measure the traffic performance is by far the most used approach, both in research and industry. Measures are based on, for example, punctuality, number of missed connections, and number of trains being on time to a station (possibly also weighted by the number of passengers affected). An example of a robustness study for the Swedish Southern mainline is given by Peterson [12] who has studied the on-time performance en-route.

In this paper we focus on 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 or measures concerning passenger disutility.

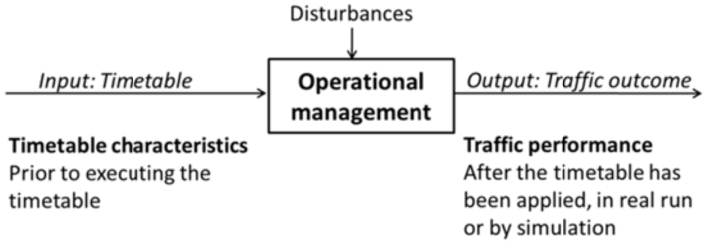

Figure 1: Two types of robustness measures used when analysing timetable robustness; Timetable characteristics and Traffic performance

2.2 Measures Related to Timetable Characteristics

A commonly used expression for robustness is the amount of margins inserted in the timetable. Margins can be added to the runtime and stopping time to prevent trains from arriving late despite small delays. Headway margins are 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. [4], and Schöbel and Kratz [14].

Not only the amount of runtime margin, but also its allocation is important. Kroon et al. [8] and Fischetti et al. [6] have used the Weighted Average Distance (WAD) of the runtime margins from the point where the train departs to capture the allocation of the margins along the line (i.e. its journey). Dividing the line into N sections and letting s_t denote the amount of margins associated with section t, WAD can be calculated as

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

WAD is a relative number between 0 and 1, where $\text{WAD} = 0.5$ means that the same amount of margins are placed in the first half of the considered line as in the second half, whereas $\text{WAD} < 0.5$ means that more margins are placed in the first half.

Both Kroon et al. [8] and Fischetti et al. [6] have concluded that it is preferable with runtime margins concentrated early on the line (i.e. a small WAD value) to be able to recover quickly from any disturbance. However, if the disturbances occur later on the line, the runtime margins located prior to the occurrence of the disturbance may be of no use.

Clearly, robustness is also gained by increasing the headway margins. Yuan and Hansen [20] have studied how to allocate headway margins 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 [3], 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 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 [3], 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, Salido et al. [13]. Vromans et al. [18] 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).

Salido et al. [13] has claimed that the flow of passengers also affects the robustness, since a large passengers exchange at a station increases the probability for disruptions.

In the robustness calculation made by Goverde [7] the critical path of activities during a time cycle in a periodic timetable was considered.

There are also models intended for calculating the capacity utilisation for a line, UIC [21]. 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 [10] 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. [18] has concluded, the headway between the trains needs to be equalised to achieve a more robust timetable.

Salido et al. [13] 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. They can not define whether a timetable is robust or not, their purpose is to compare two timetables and return which of them is more robust than the other. Shafia and Jamili [15] have advanced 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.

Table 1: A selection of research publications, where timetable characteristics and robustness measures are proposed

Publication	Timetable characteristic	Measure	Numerical example
Carey [3]	Headway	Percentage of headway larger than X The Xth percentile of distribution of headways The standard deviation of headways The mean absolute deviation of headways	none
Fischetti et al. [6]	Allocation of margins	WAD	real world
Goverde [7]	Margins (runway and headway)	Stability margin (periodic timetables) Recovery times (periodic timetables)	fictive/ real world
Kroon et al. [9]	Headway Delay-sensitive crossing movements	Amount of technically possible delay-sensitive crossing movements (headways smaller than 5 minutes)	fictive/ real world
Kroon et al. [8]	Allocation of margins	WAD	fictive/ real world
Salido et al. [13]	Runtime margins Number of trains Number of commercial stops Flow of passengers Tightest track (single-lines)	A weighted sum of timetable and traffic parameters (single-track) No. of disruptions that can absorbed with the avaliable margins (single-track)	real world
Vromans et al. [18]	Heterogenity/Headway	SSHR/SAHR	fictive/ real world
Yuan and Hansen [20]	Headway margin in bottlenecks	Amount of headway margin in bottlenecks	fictive

We summarise the robustness measures based on timetable characteristics in Table 1. In the table we also list whether the work described in each publication includes numerical examples with the use of a fictive and/or a real world timetable or no numerical examples at all. 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 and where the margins are placed in the timetable.

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*.

The critical points can be more or less critical depending on the amount of margins available in the point. Therefore we define Robustness in Critical Points, RCP, as a measure of the robustness in each of these points.

3.1 Empirical observations

When studying timetables, we have found that points in the timetable where a train is planned to be overtaken by, or to enter the network after an already operating train, are typically very sensitive to disturbances. If the already operating train is just a little bit 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 not withheld by the dispatcher but instead gets priority and will run 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. Since the two trains share the same infrastructural resource during a longer time, a conflict here has a much larger influence than a crossing.

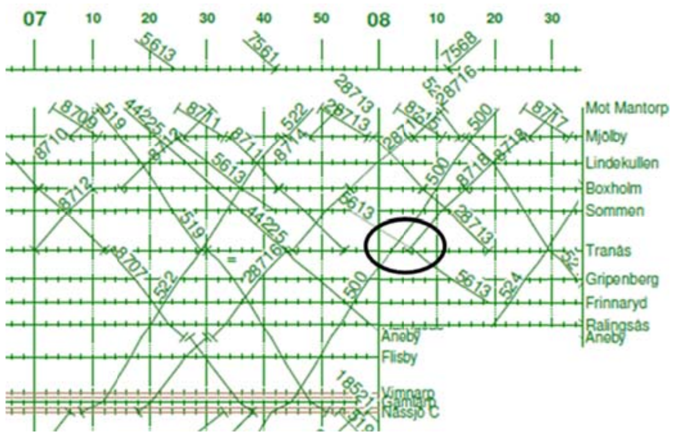

Figure 2: A snapshot from the timetable for the Swedish Southern mainline (a typical weekday in 2011) showing an example of a critical point between a fast train (train 500) and a commuter train (train 8718) in Tranås. The time is given on the x-axis and the stations on the line represent the y-axis.

In Figure 2 we can see an example of a critical point. The figure depicts a cut of the Southern mainline during the morning peak period. In Tranås (TNS), 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. As a rule of thumb, the train dispatcher holds the commuter train up to two minutes in favour of delayed long-distance trains. If train 500 is more than five minutes late it has to run after the slower train 8718 and its delay will increase until there is a possibility for overtaking or the commuter train has reached its final destination.

It is easy to identify the negative effects of critical points on the traffic performance. Figure 3 depicts the en-route punctuality for train 500 before and after the critical point in TNS described above. We can identify three groups of performance cluster reflecting the train dispatchers' strategies. Trains in group 1 have a small delay in TNS and therefor they are allowed to overtake the commuter train soon after TNS. This group of trains can however not recover much from their delays since they do not have that much runtime margin allocated in this area.

Trains in group 2 have a larger delay in TNS than the trains in group 1. They are allowed to overtake the commuter train further along the line, but will end up after another commuter train, 8720, in Linköping (LP). This will give rise to a growing delay subsequent to that critical point.

Trains in group 3 have a large delay in TNS and will have to run after the commuter train during its entire journey to Norrköping (NR). The delays for this group of trains grow with more than 5 minutes.

As we can see by the different trains delay pattern in Figure 3, it is not always clear which trains that will end up in which group. Some trains with a small delay in TNS can end up in group 2 and some trains with a larger delay can end up in group 1. It depends on how the train dispatcher prioritises in each individual case.

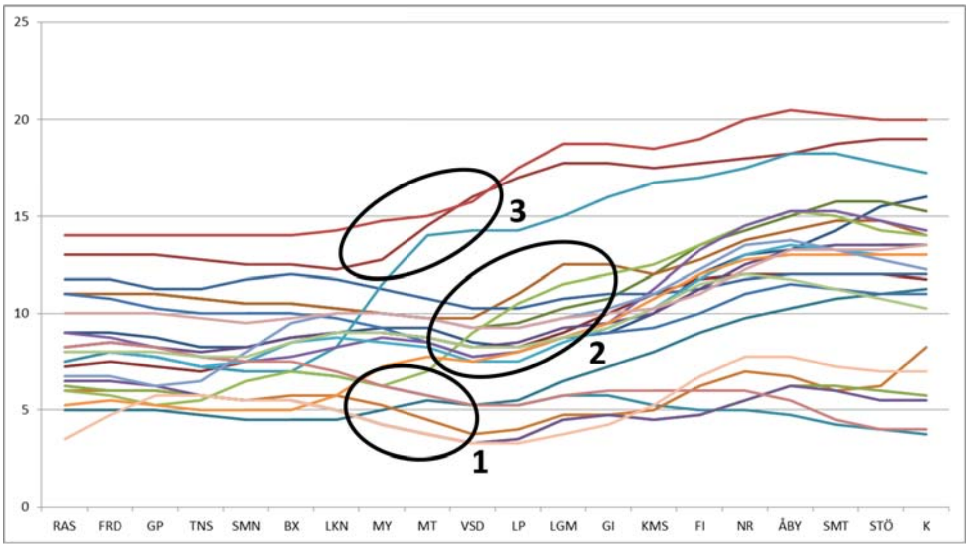

Figure 3: Punctuality statistics reflecting the effect of the critical point in Tranås (TNS) for train 500. (Trains with a delay smaller than 5 minutes are not shown, as they are not affected by the conflict) The x-axis gives the space defined by stations and the y-axis gives the delay in minutes.

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. (In the example above the train dispatcher deviates from the guideline and withhold the train on time a few minutes in favour of the delayed train.) The prioritisation of trains on time is a good general rule but there are many situations, like the example above, where it could be better overall to prioritise the delayed train. In this particular case, train 500 continues all the way to Stockholm, and have several subsequent train slots to catch, whereas the local commuter trains end in NR.

3.2 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 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 and we will further on refer to the two trains involved as either the entering or the operating train. In an overtaking, the overtaken train corresponds to the entering train.

If there are two trains planned to enter the network after an already operating train, only the relationship between the first entering train and the operating train is considered 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.

Figure 4 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. 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 greatly influences the delays for train 1, whereas train 2 will arrive on time to the end station E in both scenarios.

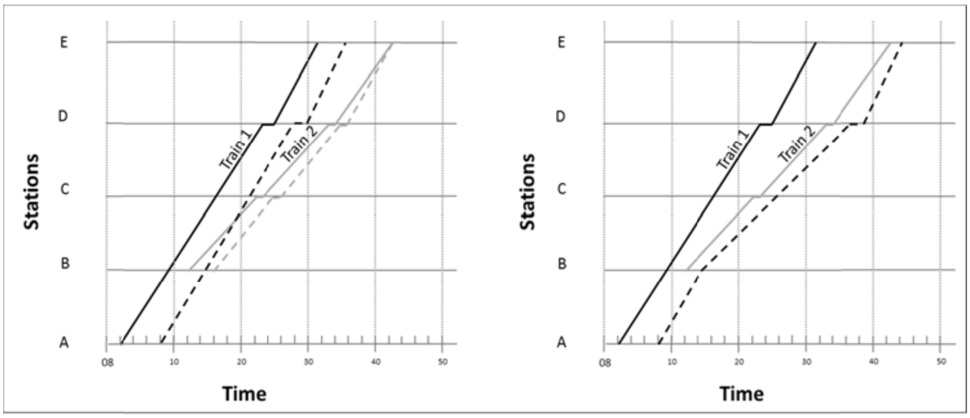

Figure 4: Two different scenarios when the delayed train 1 runs either before or after the initially punctual train 2. The continuous lines are the planned timetable and the dashed/dotted lines represent the outcome with delays. In the scenario to the right, train 2 runs according to its schedule, i.e. the dashed and solid grey lines coincide.

The procedure of identifying critical points in a timetable starts with the search for all trains that enter the network somewhere on the line and for all trains that are being overtaken. The locations where one of these two events occurs will be the location of the critical points. The entering or overtaken train represents train 2 in Figure 4. 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 the corresponding two trains involved.

3.3 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 parameters which are also illustrated in Figure 5:

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. The value might, however, be bounded by other traffic, see Section 3.4. With a large amount of runtime margins 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. The value might, however, be bounded by other traffic, see Section 3.4. With a large amount of runtime margins 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 in the critical point. The headway margin is calculated as the total headway time minus the minimum headway time. With a large headway margin the possibility for the operating train to run ahead of the entering train increases, even when delayed.

RCP is a measure of the maximum flexibility in a critical point and it consists of the total amount of margins available. With a larger RCP value, the dispatcher will have higher prospects to handle conflicts in an effective way.

When calculating RCP we delimit ourselves to only consider allowing rescheduling of the two trains involved in the critical point. It is for example possible to operationally reschedule 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 involved in the conflict.

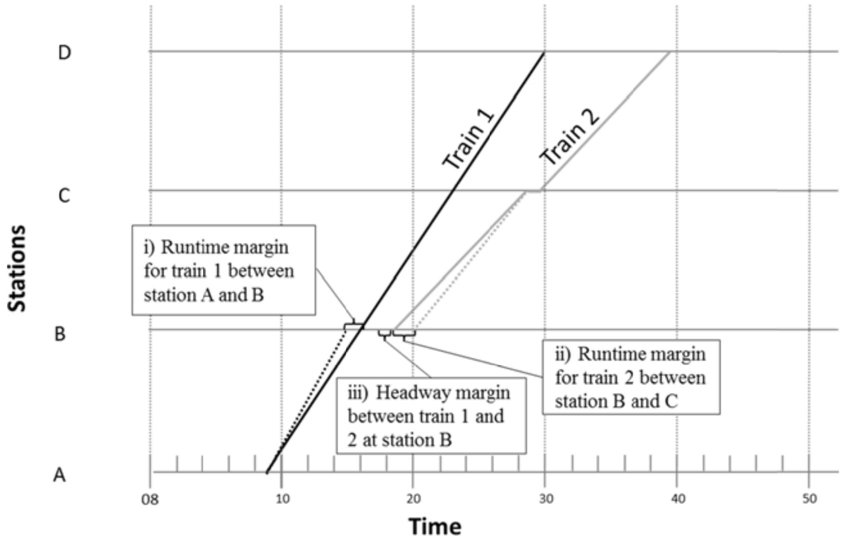

Figure 5: RCP consists of three parts; the runtime margin for train 1 between station A and B, the runtime margin for train 2 between station B and C and the headway margin between train 1 and 2 at station B.

3.4 Example of How to Calculate the Available Runtime Margin in RCP

As described above, RCP consist of three parts; the headway margin between the two trains at the critical point, the available runtime margin for the operating train from the previous planned stop to the critical point and the available runtime margin for the entering train from the critical point to the next planned stop. If there are no other trains close to the involved trains in a critical point, the total amount of runtime margins from the previous stop to the next stop can be used in RCP. But if there are other trains close to the involved trains, the total amount of the runtime margins are not possible to use in RCP. This is illustrated in Figure 6 and Figure 7.

Critical point "M" is the location in the timetable where fast long-distance train, 500, is planned to overtake the slower regional train, 11006, in HM. One part in RCP is the available runtime margin for train 500 from the previous planned stop, LU. At that stretch train 500 has the total amount of runtime margin of 310 seconds. This means that train 500 could depart up to 310 seconds late from LU but still be on time at the critical point in HM. However, this is not operationally possible since there is another train running close after train 500. If train 500 is 310 seconds late at LU train 1204 will be disturbed. Therefore the available amount of runtime margin for train 500 has to be decreased. When departing from LU, train 500 can be at most 9 seconds late without disturbing train 1204 somewhere on the line, therefore the available runtime margin for train 500 in RCP is 9 seconds.

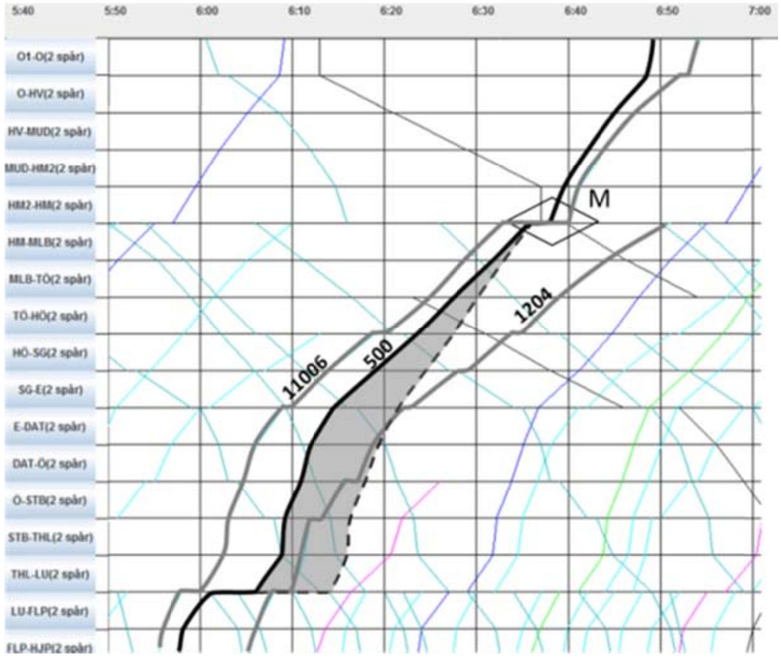

Figure 6: The total amount of runtime margin for train 500 between LU and HM (310 seconds)

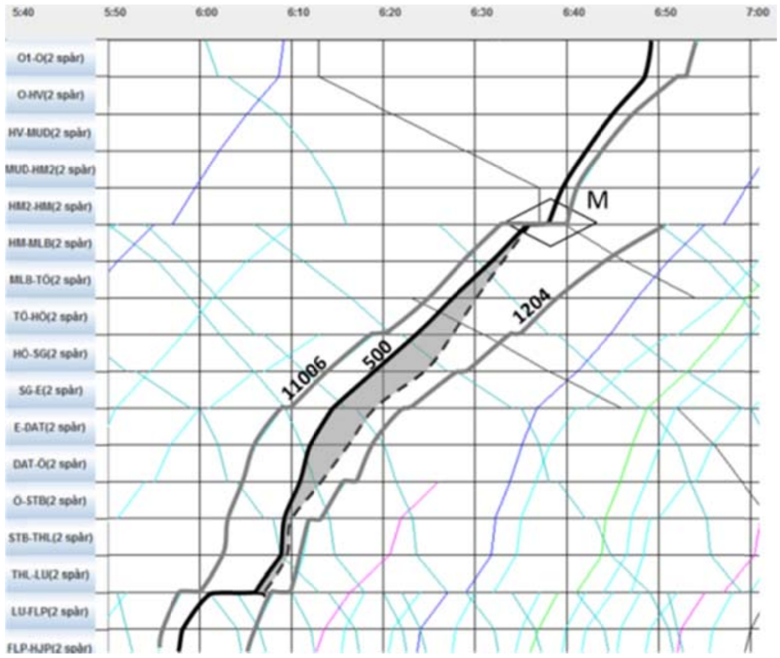

Figure 7: The available amount of runtime margin for train 500 between LU and HM (9 seconds)

The same principle applies when calculating the available runtime margins for the entering train after the critical point. If there are no other trains close after the entering train, the available runtime margin corresponds to the total amount of runtime margin until the next planned stop. If there is a train close after the entering train the available runtime margin consists of the amount of margin until the next location where the entering train has to be on time to not disturb the other train.

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.

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. The measure is also dependent of number of tracks at each section. When comparing NoT for several sections, the number of available tracks at each sections must be equal. Robustness attributes of this type are used by Salido et al. [13].

2) **Total amount of runtime margins for every train, TAoRM**
   TAoRM is strongly connected to the capability to recover from delays. Robustness measures of this type are used by Salido et al. [13].

3) **Maximum runtime difference per partial stretches, MRD**
   MRD serves at capturing 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, f, which are naturally bounded by the traffic structure. There are too few trains that operate on the whole line and a one line section gives a too short measuring distance; hence the partial stretches. This measure is inspired by the work of Vromans et al. [18].

4) **Sum of Shortest Headway Reciprocals, SSHR**
   SSHR captures both the heterogeneity and the size of the headways, and was suggested by Vromans et al. [18], for further details see Section 2.2.

5) **Weighted Average Distance, WAD**
   WAD is a measure for the distribution of the runtime margin, and has been used by Kroon et al. [8] and Fischetti et al. [6], for further details see Section 2.2.

6) **Percentage of headways equal to or less than the minimum value, PoH**
   PoH is a measure for the occurrence of planned short headway times. The measure is inspired by the work of Carey [3] and Kroon et al. [9].

7) **Robustness in Critical Points, RCP**

RCP is our proposed measure, based on the occurrence of critical points and the amount of margins in each point. The calculation of the measure consists of two parts; identification of the critical points and calculation of RCP.

The above listed measures were used on a real-world case which includes the southern part of the Swedish Southern mainline, between Malmö (M) and Alvesta (AV). This part of the Southern mainline is nearly 200 km long and consists of double-track (close to M there are four tracks). The line is one of the most congested lines in Sweden where fast long-distance trains share the tracks with slower freight, regional and commuter trains. This heterogeneous traffic leads to difficulties in the timetable construction phase, as well as in the operational phase. For our calculations, we have chosen the morning period 05.45–07.15, on the 8th September 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 whole way M–AV. The graphical timetable for the chosen line and time period can be seen in Figure 8. In the figure also the 14 identified critical points are illustrated and enumerated from A to N.

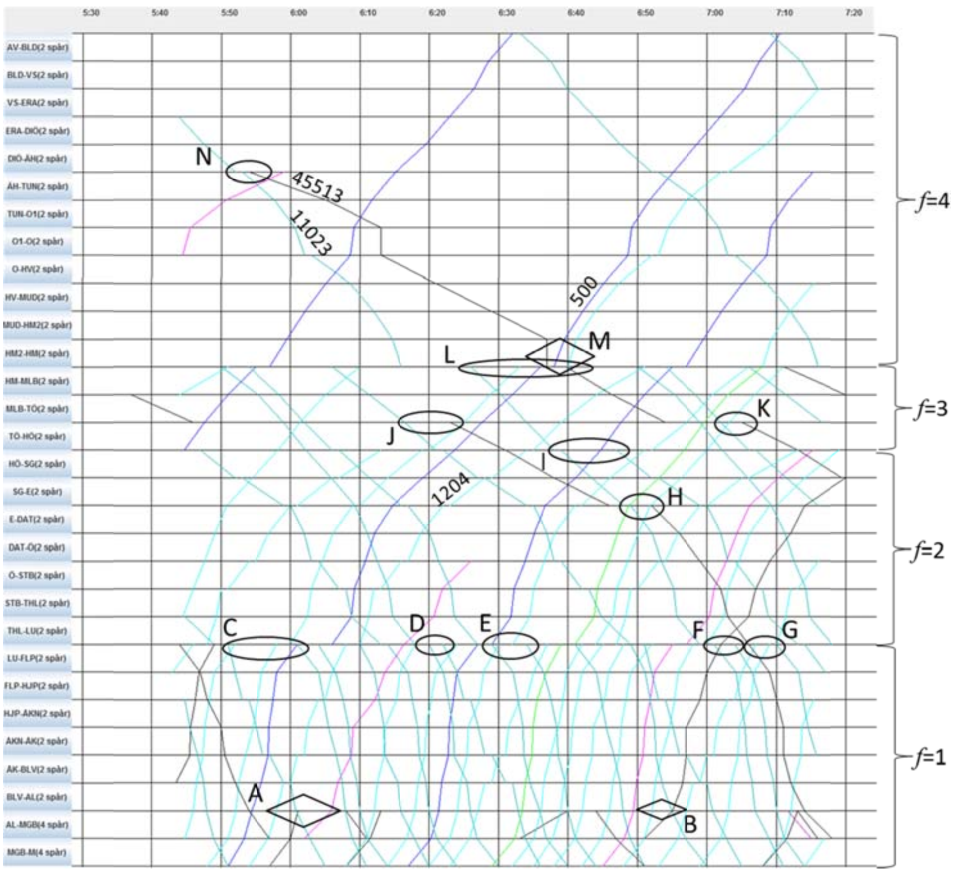

Figure 8: The graphical timetable from the Swedish Southern mainline, between Alvesta (AV) and Malmö (M) including the critical points. Critical points for south going trains are represented by circles and for north going trains they are represented by diamonds.

A RailSys model (Bendfeldt et al. [2]), provided by the Swedish Transport Administration was used to calculate the minimum headway at every section. One interesting observation when comparing the RailSys data and the established timetable is that there are headways in the timetable below the minimum headways given by RailSys. If the difference is small it can be explained as marginal error of the calculation but if the difference is large it will have an influence on the trains' runtimes. The fact that the timetable headways sometimes are smaller than the minimum value (the RailSys value) will influence some of the robustness measures.

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)

4.2 Result from the Robustness Measure Calculation

The overall robustness measure, PoH, is 4 %, which means that there are several headway values that have no or even a negative margin. If this is an acceptable value is hard to tell, but the fact that negative headway values exist will result in operational disturbances for the trains.

Table 2 presents the values for TAoRM and WAD. Some of the trains have no runtime margins 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 margins and where they are allocated. The required amount of margins and their allocation is highly correlated to the occurrence and magnitude of disturbances. In the timetable construction phase the disturbance distribution is unknown and therefore hard to include in the process.

Table 2: The values of the robustness measures for the Swedish Southern mainline example – TAoRM and WAD

	TAoRM (seconds)	**WAD**			**TAoRM (seconds)**	**WAD**
Train:				Train:		
45517	0	-		11312	122	0.391032
45513	0	-		11319	122	0.680328
484	244	0.549180		11321	338	0.596154
500	424	0.372003		11323	182	0.795410
522	144	0.475198		11325	60	0.187500
524	244	0.440530		11327	6	0.550000
11004	48	0.911765		11333	60	0.375000
11006	356	0.479383		1504	154	0.733333
11008	122	0.391032		1505	309	0.428317
11010	170	0.320752		1506	324	0.676852
11012	165	0.364577		1507	199	0.561307
11014	60	0.125000		1703	110	0.494652
11015	62	0.229839		1705	110	0.494652
11017	122	0.275313		1707	110	0.494652
11019	233	0.582082		1712	50	0.892857
11021	122	0.275313		1714	110	0.462567
11023	366	0.551698		1716	110	0.462567
11025	60	0.078947		6100	38	0.700000
11029	148	0.377764		6160	103	0.512483
1204	371	0.707659		7140	25	0.534483
1205	77	0.776696		42734	193	0.500000
1206	73	0.149128		44721	675	0.559722
1207	471	0.395406		69472	90	0.375000
1209	60	0.375000		69474	90	0.375000
1254	73	0.192990		69501	119	0.625000
11255	17	0.655709		86111	21	0.500000
11308	326	0.534893		91016	2	0.850000
11310	122	0.391032		91324	59	0.375000

In Table 3 we can see 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 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 volumes per track.

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

MRD shows much larger values for HM–HÖ and HÖ–LU, than the other partial stretches. This has to do with southbound 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 northbound 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 3: The values of the robustness measures for the Swedish Southern mainline example – NoT, SSHR, MRD and RCP

	NoT (6 a.m. - 7 a.m.)	SSHR
Section:		
AV-BLD	2	0.000896
BLD-VS	2	0.000890
VS-ERA	2	0.002038
ERA-DIÖ	3	0.002375
DIÖ-ÄH	3	0.002842
ÄH-TUN	4	0.022772
TUN-O1	5	0.008213
O1-O	5	0.007902
O-HV	5	0.011069
HV-MUD	5	0.013794
MUD-HM2	5	0.013898
HM2-HM	5	0.013773
HM-MLB	10	0.053911
MLB-TÖ	12	0.046713
TÖ-HÖ	11	0.049377
HÖ-SG	14	0.056208
SG-E	14	0.057451
E-DAT	14	0.058070
DAT-Ö	14	0.058427
Ö-STB	15	0.068269
STB-THL	14	0.074221
THL-LU	16	0.075605
LU-FLP	25	0.204270
FLP-HJP	26	0.187605
HJP-ÅKN	26	0.173962
ÅKN-ÅK	26	0.172535
ÅK-BLV	27	0.160587
BLV-AL	26	0.159788
AL-MGB	(31)	0.175271
MGB-M	(27)	0.127105

MRD (seconds)	
Partial stretch:	
HM-AV	150
HÖ-HM	346
LU-HÖ	614
M-LU	240
AV-HM	-
HM-HÖ	1282
HÖ-LU	1153
LU-M	240

RCP (seconds)	
Critical point:	
A	581
B	171
C	503
D	22
E	433
F	113
G	210
H	61
I	325
J	274
K	-8
L	724
M	140
N	-103

When calculating RCP, the available amount of runtime margins is in large extent bounded by other traffic, as illustrated in Section 3.4.

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. This means that the critical points themselves produce disturbances which have to be taken care of by other margins.

The negative value at point K may be explained by as a rounding error, but at point "N", RCP is –103 seconds. It is of course impossible to run as close to the other train as the timetable shows. In practice, the timetable should be interpreted so, that as soon as train 11023 has passed ÄH, train 45513 can depart from the station. The small headway also means that if train 11023 is just a little bit delayed, train 45513 should be prioritised according to the Swedish train priority guide lines. In such case train 11023 will end up after the slower freight train and get even more delayed.

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.

4.3 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 margins are 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 conclusion of where margins should be inserted 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. 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 a picture 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. With critical points we instead identify specific locations in the network and the two trains involved 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 point "K" and "N" where the value is negative. Then the timetable in itself will be executable without constructing any delays. In a second step, it is of course also recommendable to increase any low RCP value, for example at points "D" and "H". However, when increasing the RCP value, also other robustness measures will be affected. When, for example, adding runtime margins 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.

5 Conclusions and Future Research

This paper discusses several ways to deal with robust timetables for railway traffic. In the timetable construction phase, it is difficult to predict how the traffic will react to disturbances and how the delays may spread. To enable operative flexibility, however, robustness issues must be considered already at this early planning stage. 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.

Previously proposed measures can be used to, for example, identify trains with a small amount of runtime margins or sections that are heavily utilised. They can, however, not indicate exactly where along a train's service, runtime margins should be inserted, or which train slots that should be modified at a certain section to increase the robustness.

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, the location pattern and the RCP values constitute a useful robustness measure, which can be applied both for a certain train slot and a complete timetable. Critical points can easily be used in the timetabling process to identify weaknesses in a timetable. 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. Therefore, the concept of critical points and RCP can be seen as a contribution to the already defined robustness measures which combined can be used as guidelines for timetable constructors.

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 RCP value. If there is a possibility for a delayed operating train to overtake the entering train further down the line, the critical point apparently is less critical. Also the runtime difference between the two trains involved in a critical point should have some influence on how critical the point is considered to be.

The definition of critical points described in this paper is primarily targeting double-track lines. However, we believe that there are similar points in a timetable for single-track lines that are extra sensitive for delays. How to define them is a part of future research.

A possible way of analysing and using the proposed measure in a larger scale is to apply it in an optimisation model to maximise the timetable robustness in critical points, given the permitted adjustments. Such a model can be designed to, e.g., restrict the magnitude of the critical points, maximise the margins in the most critical point, or minimise the number of critical points in the timetable.

6 Acknowledgements

This research was conducted within the research project "Robust Timetables for Railway Traffic", which is financially supported by grants from VINNOVA (The Swedish Governmental Agency for Innovation Systems), Trafikverket (The Swedish Transport Administration) and SJ AB. The authors are grateful for all data provided by Trafikverket and SJ AB.

References

[1] Andersson, E., Peterson, A., Törnquist Krasemann, J., "Robustness in Swedish railway traffic timetables", In: Ricci et al. (eds.), *Proceedings of RailRome 2011*, University of Rome La Sapienza and IAROR, 2011

[2] Bendfeldt J-P, Mohr U, Muller L (2000) RailSys, a system to plan future railway needs. In Allan et al. (eds.): Computer in Railways VII, 249-255, WIT Press, Southampton. doi: 10.2495/CR000241

[3] Carey, M., "Ex ante heuristic measures of schedule reliability", Transportation *Research Part B* 33:473-494, 1999

[4] Cicerone, S., D'Angelo, G., Di Stefano, G., Frigioni, D., Navarra, A., "Recoverable robust timetabling for single delay: Complexity and polynomial algorithms for special cases", *Journal of Combinatorial Optimization* 18:229-257, 2009

[5] Dewilde, T., Sels, P., Cattrysse, D., Vansteenwegen, P.,"Defining robustness of a railway timetable", In: Ricci et al. (eds.), *Proceedings of RailRome 2011*, University of Rome La Sapienza and IAROR, 2011

[6] Fischetti, M., Salvagnin, D., Zanette, A., "Fast approaches to improve the robustness of a railway timetable", *Transportation Science* 43:321-335, 2009

[7] Goverde, R., "Railway timetable stability analysis using max-plus system theory", *Transportation Research Part B* 41:179-201, 2007

[8] Kroon, L., Dekker, R., Vromans, M., "Cyclic railway timetabling: A stochastic optimization approach", *Railway Optimization* LNCS 4359: 41-66, 2007

[9] Kroon, L., Maróti, G., Helmrich, M.R., Vromans, M., Dekker, R., "Stochastic improvement of cyclic railway timetable", *Transportation Research Part B* 42:553–570, 2008

[10] Mattson, L.-G., “Railway capacity and train delay relationships”, In: Murray, A., Grubesic, T.H. (eds.), *Critical Infrastructure: Reliability and Vulnerability*, pp 129-150, Springer-Verlag, 2007

[11] Policella, N., *Scheduling with uncertainty: a proactive approach using partial order schedules*. PhD thesis, Dipartimento di Informatica e Sistemistica, Università degli Studi di Roma, Rome, Italy, 2005

[12] Peterson,A.,”Towards a robust traffic timetable for the Swedish Southern Mainline”, In: Brebbia, C.A., Tmil, N., Mera, J.M, Ning, B., Tzieropoulos, P. (eds.), *Proceedings of COMPRAIL 2012*, WIT Transactions on The Built Environment, 2012

[13] Salido, M.A., Barber, F., Ingolotti, L., “Robustness in railway transportation scheduling”, In: 2008 7th World Congress on Intelligent Control and Automation, Chongqing, China: 2880-2885, 2008

[14] Schöbel, A., Kratz, A., “A bicriteria approach for robust timetabling”, In: Ahuja et al. (eds.), *Robust and online large-scale optimization*, LNCS, vol. 5868:119-144, 2009

[15] Shafia, M., A., Jamili, A., Measuring the train timetables robustness, In: Proceedings of the 2nd International Conference on Recent Advances in Railway Engineering, Teheran, Iran, 2009

[16] Takeuchi, Y., Tomii, N., Hirai, C., *Evaluation method of robustness for train schedules*, Quarterly Report of Railway Technical Research Institute, Vol. 48, No.4, pp 197-201, 2007

[17] Trafikanalys, *Person- och godstransport på järnväg, 2011 kvartal 4* (”*Passenger and cargo transportation on railway, 4th Quarter, 2011*”, in Swedish), Report Statistik 2012:2, Trafikanalys, Swedish National Road Administration, Stockholm, Sweden, 2012

[18] Vromans, M., Dekker, R., Kroon, L., “Reliability and heterogeneity of railway services”, *European Journal of Operational Research*, 172:647-665, 2006

[19] Yoko, T., Norio, T., “Robustness indices for train rescheduling”, Presentation at: 1st International Seminar on Railway Operations Modeling and Analysis, Delft, The Netherlands, June 8–10, 2005

[20] Yuan, J., Hansen, I.A. “Closed form expression of optimal buffer times between scheduled trains at railway bottlenecks”, In: Proceedings of the 11th international IEEE conference on intelligent transportation systems, Beijing, China, pp 675-680, 2008

[21] UIC, Capacity, Leaflet UIC Code 406, 1st Edition, 2004