UNIVERSITYOF BIRMINGHAM

A microscopic evaluation of railway timetable robustness and critical points

Solinen, Emma; Nicholson, Gemma; Peterson, Anders

Link to publication on Research at Birmingham portal

A Microscopic Evaluation of Railway Timetable Robustness and Critical Points

Emma Solinen [ab, 1], Gemma Nicholson [c], Anders Peterson [b]

[a] Trafikverket, SE-172 90 Sundbyberg, Sweden

[1] E-mail: emma.solinen@trafikverket.se, Phone: +46707247038

[b] Department of Science and Technology, Linköping University,
SE-601 74 Norrköping, Sweden

[c] Birmingham Centre for Railway Research and Education, University of Birmingham,
Edgbaston, Birmingham, B15 2TT, UK

Abstract

One method to increase the quality of railway traffic flow is to construct a more robust timetable in which trains are able both to recover from delays and the delays are prevented from propagating. Previous research results show that the indicator Robustness in Critical Points (RCP) can be used to increase timetable robustness. In this paper we present the use of a method for RCP optimization: how it can be assessed ex-post via microscopic simulation. From the evaluation we learn more about how increased RCP values influence a timetable's performance. The aim is to understand more about RCP increase at a localised level within a timetable in terms of effects to the pairs of trains that are part of the indicator. We present a case study where an initial timetable and a timetable with increased RCP values are evaluated. The ex-post evaluation includes the quantification of measures concerning train-borne delay and robustness of operations, as well as measures capturing the subsequent quality of service experienced by passengers to assess the broader effects of improved robustness. The result shows that it is necessary to use several key performance indicators (KPIs) to evaluate the effects of an RCP increase. The robustness increases at a localised level, but the results also indicate that there is a need to analyse the relationship between ex-post measures and RCP further, to improve the method used to increase RCP and thus its overall effect on timetable robustness.

1 Introduction

The idealised railway runs its trains as specified in a pre-defined timetable; however, inevitable day-to-day disturbances mean that trains cannot meet their planned times of arrival at stations and other timing points and times for departure from stations. More severe disruptions lead to trains not being able to run in their planned timeslots and/or the propagation of delay from one train to others in the network.

One method to increase the quality of railway traffic flow is to construct a more robust

timetable, i.e. a timetable in which trains are able to keep their originally planned slots despite small disturbances and without causing unrecoverable delays to other trains. A robust timetable should also be able to recover from small delays. With a more robust timetable railway traffic delays can be reduced and punctuality can be improved.

Previous research results presented in Andersson et al. (2013, 2015) show that the concept of critical points and the related ex-ante indicator Robustness in Critical Points (RCP) can be used to increase timetable robustness in a satisfying way. The general idea is that if the robustness in some points that are particularly sensitive to disturbances can be improved, the whole timetable will have an improved capability to recover from delay and provide a better quality of service.

Performance measurements are essential for railway operations: the capture, processing and reporting of performance measures is a regulatory requirement of most established railway systems worldwide. There is an ongoing necessity to improve the quantity and quality of railway services to meet the demands of customers; this needs objective measurement.

Across Europe a widely-used measure of operational performance is that of punctuality of trains arriving at their final destination. Some tolerance is given; in Sweden, for example, trains arriving at their destination station within 5 minutes of the timetabled time are considered to be punctual. In the UK a 5 minute threshold is also used, except for long distance services where the requirement is relaxed to a 10 minute threshold.

However, such punctuality measures do not give an assessment of performance at intermediate stations and do not necessarily capture well the degree of delay propagation, especially in timetables where journeys are padded with most of their running time supplements towards the destination stations. This is connected to the experience of passengers not being well reflected by such punctuality measures. Passengers making journeys between intermediate stations are not interested in punctuality of the train itself, but rather in the reliability of their journey. They may miss connections at intermediate stations without this fact being reported and measured.

Ensuring a high quality of service provision requires monitoring with a wide-ranging evaluation using KPIs (key performance indicators) considering the goals of both operators and passengers to check that improvement in one aspect of performance is not being made to the detriment of other important considerations, such as improvement in punctuality at destination stations at the expense of missed connections for passengers at intermediate stations.

Methods to improve robustness of timetables are usually assessed using microscopic, i.e. detailed, simulation. Typically, ex-ante indicators of robustness, based on, for example, traffic heterogeneity and speed, time supplements and buffers, are optimized and the resulting level of performance is assessed through simulation. Together with a balanced and thorough evaluation of performance using ex-post performance measures, micro simulation can give a precise assessment of changes in performance. However, this type of simulation is time-consuming and complex. Understanding the link between ex-ante robustness indicators and actual resultant performance may lead to improved ability to optimize operational performance and a reduced requirement for exhaustive micro simulation in the future.

In this paper we address the relationship between ex-ante indicators and ex-post measures by evaluating a timetable improved with the use of the ex-ante robustness indicator RCP. We present the first steps towards an implementation of an RCP optimization model in reality, where a macroscopically generated improved timetable is

assessed via microscopic simulation. It is adjusted to run without conflict, i.e. to become feasible, at the microscopic level and subsequently evaluated with several performance measures.

The aim is to understand more about RCP increase at a localised level within a timetable in terms of the effects on the pairs of trains that are part of the indicator. We combine the information in the RCP indicator with knowledge of the current punctuality in the initial timetable and with other ex-post measures to gain a deeper knowledge of when and how to apply an increase in RCP. The main contributions of this paper are furthering knowledge of how RCP optimization can be used in reality, and of the problems that may occur when transferring a macroscopically generated timetable into a microscopic environment. The microscopic evaluation also gives us insights into when and how to apply an increase in RCP and hints at what can be improved in the optimization model so that it smooths the implementation process, and results in better performing timetable. Understanding more about the relationship between ex-ante indicators and ex-post measures is another important contribution.

The paper is structured in the following way. First a review of related research is given in Section 2 and then a description of the macro to micro transformation made in this study is presented in Section 3. Section 4 includes the analysis of a real-world case study, followed by the evaluation results in Section 5. In Section 6 the results are discussed and finally, in Section 7, the main findings are summarized and directions for future research are given.

2   Related Work

Several aspects of railway timetable robustness have been assessed and analysed in previous research. The definition of a robust timetable is, however, ambiguous. In this paper we refer to a timetable as robust when trains are able to keep their originally planned slots despite small delays and without causing unrecoverable delays to other trains. In a robust timetable, we also require that trains have the capability to recover from small delays and that the delays are kept from propagating over the network.

Measures of robustness can be categorised in different ways, e.g. ex-ante and ex-post, by level of detail (macroscopic and microscopic), stage of planning, or train- or passenger-focussed, in the aspects of robustness that they assess.

2.1 Ex-post Robustness Measures

The most commonly used timetable robustness measures are ex-post measures, i.e. measures that are based on performance, typically of traffic. These measures cannot be calculated unless the timetable has been executed, either in real life or in an experimental environment with fictive disturbances via simulation. Typically, ex-post measures are based on punctuality, primary and secondary delays, number of violated connections, or number of trains being on-time to a station (possibly weighted by the number of passengers affected). For more examples and references, we refer to, e.g., the survey by Andersson (2014).

The most commonly used passenger-focussed measure in planning is the maintenance of scheduled connections, while generalised cost measures involving waiting time, time in-train, time buying tickets, etc. can also be quantified. Reduced variability in travel time is stated as highly preferential by passengers (Transport Focus, 2015). As highlighted in the recent literature review by Parbo et al. (2016), most studies have focussed on robust

timetable design from the point of view of traffic flow, using train-oriented measures to quantify performance. We refer also to this review as a second source of examples and references of such measures. Parbo et al. (2016) go on to consider passenger perception of railway performance, including the associated attributes that can be quantified ex-post. The authors note that the passenger perspective is considered infrequently in optimization and planning studies, despite the fact that the on-time experience of passengers can be up to ten percentage points lower than the associated train-based measure. They conclude that to adequately capture the factors important to passengers, and thus be able to reliably improve performance from their perspective, measures of service reliability and provision as well as more traditional punctuality measures should be used. This should help reduce the gap in perception between operators and passengers of the level of service provided.

In addition to references covered in the surveys by Andersson (2014) and Parbo et al. (2016), the following two references cover to some extent the perspective of passengers in their evaluation. Nicholson et al. (2015) describe the framework used for the ex-post evaluation of performance improvement achieved in the ON-TIME project (Quaglietta et al., 2016). They employ a robustness measure illustrating severity of delay and recovery time, alongside traditional punctuality measures, which make up the traffic performance-related components. Indicators relating to passenger journey time, comfort and ability to realize planned connections represent the passenger-aspect.

Warg and Bohlin (2016) present an approach to evaluate the quality of a timetable with the combined use of capacity analysis and economic assessment. A train's total runtime consists of the minimum runtime, the added runtime margin and also the delay. By weighting them in different ways we can measure the passenger perspective and observe that delays are important when economic aspects of timetable quality are considered.

A typical approach to increase robustness is to disrupt a timetable with stochastic disturbances, simulate the outcome and analyse one or more ex-post measures. Then the timetable is re-scheduled in a way that increases the robustness. This procedure is repeated several times, and after the iterative process, results in an improved timetable; see for example Kroon et al. (2008) and Fischetti et al. (2009). Such an iterative procedure is however time consuming since several timetables need to be executed before we achieve a more robust timetable. Another approach is to use ex-ante indicators to improve the timetable robustness.

2.2 Ex-ante Robustness Indicators

Ex-ante indicators, sometimes also referred to as measures, are based on timetable characteristics and can already be computed and compared at an early planning stage without knowledge of the disturbances that may occur. Generally ex-ante indicators involve margin time which can be added to the runtime and stopping time to prevent trains from arriving late despite small delays. Margin time can also be added to the headway, i.e. the time separation between two consecutive trains using the same infrastructural resource, which serves to reduce the effects of knock-on delay.

Not only the amount of margin time, but also its allocation is important. There are, however, no clear results showing how the margin time should be allocated to achieve increased robustness. For example, Vromans (2005) shows that a uniformly distributed runtime margin allocation leads to poor results when it comes to delay recovery. Vromans (2005), Kroon et al. (2007) and Fischetti et al. (2009) use the measure Weighted Average Distance (WAD) and show that it is preferable to have the runtime margin

concentrated early in a train's journey along a line. They also mention that if the disturbances occur later along the line, the runtime margin located prior to the occurrence may be of no use.

Robustness measures concerning headway margin time have been studied by e.g. Vromans et al. (2006) and Carey (1999). Vromans et al. (2006) develop the measure SSHR (sum of shortest headway reciprocals) which also considers the traffic heterogeneity and Carey (1999) studies the distribution of headway margin along train journeys and sections.

Ex-ante indicators give us information about how to improve the timetable robustness before the timetable is executed. However, before the ex-ante indicators can effectively be used for this purpose, the relationship between them and the resulting ex-post measures needs to be clarified, an area not well covered by the literature. Jensen et al. (2014) identify a gap in the understanding of the semantics of robustness indicators, i.e. the link between ex-ante indicators and performance outcome. They perform an initial study investigating the link between established ex-ante robustness indicators and the results of micro simulation on the Danish North West line. They show that more complex indicators capture changes in performance and suggest further research in the area to better understand the relationships between ex-ante indicators and ex-post measures.

2.3 Robustness in Critical Points, RCP

Due to heterogeneous traffic and interdependencies between trains, there are points in a timetable that are particularly sensitive to disturbances. These points can be described as critical points. Critical points appear in a timetable for double track lines where it is planned that a specific train starts its journey after another already operating train, or where a train is planned to overtake another train. In case of a delay in a critical point, the involved trains are likely to require the same infrastructural resource at the same time which might affect the delay propagation significantly. Interactions that occur at crossings are not included as critical points since the trains only interact for a short time, nor interactions between trains running in the opposite direction on different tracks, since they do not normally influence each other. Refer also to Andersson et al. (2013) for more details.

Each critical point is represented by a specific station and a pair of trains, the *leader* and the *follower*, which interact at this geographic location in such a way that a time-dependency occurs. The follower refers to the train that starts its journey at the critical point behind another train (denoted the leader), or is overtaken in the critical point by the other train, i.e. the leader. The robustness in critical point p is related to the following three margin parts, which are illustrated in Figure 1:

L_p – The available runtime margin time before the critical point for the leader, i.e. the runtime margin for Train 1 between stations A and B in Figure 1. With a large L_p the likelihood of the leader arriving on-time to the critical point increases.

F_p – The available runtime margin time after the critical point for the follower, i.e. the runtime margin for Train 2 between stations B and C in Figure 1. A large F_p increases the opportunity to delay the follower in favour of the leader, without causing any unrecoverable delay to the follower.

H_p – The headway margin, sometimes referred to as buffer time, between the trains'

departure times in the critical point, i.e. the headway margin between Train 1 and Train 2 at station B in Figure 1. In the critical point the trains are separated by the headway margin plus the minimum technical headway. With a large H_p the chance to keep the scheduled train order in the critical point increases, even in a delayed situation.

We compute RCP_p, the Robustness in Critical Point p, as

$$RCP_p = L_p + F_p + H_p.$$
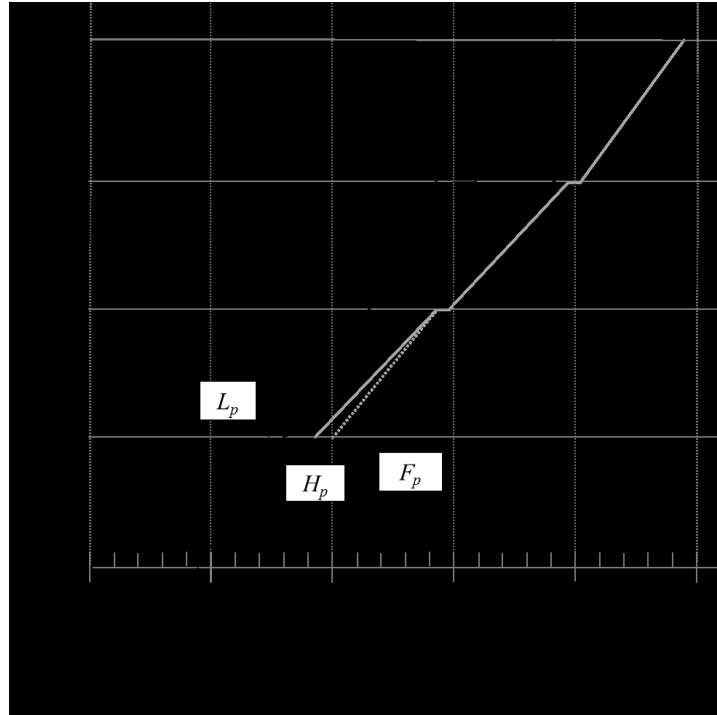
Figure 1: RCP is the sum of the three margin parts: H_p, L_p and F_p.

When the RCP values increase, train slots will be modified in a way which quickly becomes manually untraceable. To handle the modifications, Andersson et al. (2015) present a MILP (Mixed Integer Linear Programing) model, which takes an initial timetable as input, re-allocates the already existing margin time in the timetable to increase RCP and finally returns an improved timetable. Time is a continuous variable, whereas the order in which various services use infrastructural resources is captured with sets of integer (binary) variables. The model includes several physical and logical restrictions for how the timetable can be re-organized. In short, train constraints control the trains' events and ensure that runtimes are respected, whereas infrastructural constraints restrict the train order and how the trains can use the tracks, including minimum headway and clearance times.

In the model all critical points are identified and RCP_p for each point p is calculated. When optimizing the timetable the trains' event times are modified to fulfil the requirement that RCP_p always is larger than or equal to some threshold value. In the objective function the difference between the new and the planned event times is minimized, which keeps the timetable changes at a minimum. A comprehensive explanation of the MILP model is beyond the scope of this paper, and interested readers are referred to the complete description by Andersson et al. (2015).

In Andersson et al. (2015) a preliminary evaluation is presented in which an iterative optimization model is used for the simulation. In that model primary delays are given to certain pre-selected trains with some randomness. However, the inserted primary delays and the dispatching algorithms used in Andersson et al. (2015) cannot claim to represent real conditions and therefore it is also uncertain whether RCP optimization gives the same optimistic results in a more realistic environment or not.

2.4 Macroscopic Data Used for Microscopic Simulation

The MILP model referred to above is macroscopic, which means that there are some simplifications regarding train movements, infrastructure layout and dispatching strategies. Transferring a macroscopically valid timetable to microscopic simulation typically leads to some complications due to the difference in level-of-detail. A problem frequently arising is that the optimal macroscopic timetable is infeasible, i.e. not executable, at microscopic level due to details arising from train speed profiles and exact trajectories. This problem is well-known in the literature.

A comparison of several macroscopic models with a state-of-the-art microscopic model of the Dutch national timetable is made by Kecman et al. (2013), who evaluate their performance with respect to the feasibility of the solution.

Schlechte et al. (2011) present a bottom-up approach, where they start at the detailed microscopic level as it is described in the simulation tool, in their case OpenTrack, and transform the model to a macroscopic representation of the network. Running and headway times are then rounded by a special cumulative method. This method is further examined by Blanco and Schlechte (2014), who can prove it to satisfy route-wise optimality, such that the total time on each individual route is not underestimated and the corresponding (overestimating) error is minimal. Local optimality, such that the overestimating error on any sub-route does not exceed some given tolerance level, is presumably guaranteed.

Goverde et al. (2016) present a performance-based railway timetabling framework integrating timetable construction and evaluation on three levels: microscopic, macroscopic, and a corridor fine-tuning level, where each performance indicator is optimized or evaluated at the appropriate level. The integration of macro and micro level is further studied by Bešinović et al. (2016) and Bešinović et al. (2017), who have developed an iterative scheme for adjusting train running and minimum headway times until a feasible and stable timetable has been generated at the microscopic level.

The method for transforming the macroscopic timetable to a microscopic level used in this paper is described in Section 3.

3 Macroscopic to Microscopic Transformation

To evaluate a timetable where RCP values have been optimized, as is described by

Andersson et al. (2015), we assess ex-post measures using the microscopic railway simulation tool RailSys (RMCon, 2015, Bendfeldt et al., 2000). RailSys is the standard simulation tool used by the Swedish Transport Administration (Trafikverket), which has provided the infrastructure model. For simplicity, we refer to the improved timetable as *optimized* throughout this paper.

The first step in the evaluation in RailSys is to transform the macroscopically generated timetable into a microscopic format by adding more details and making necessary adjustments. For example, the infrastructure description used in the optimization model is aggregated. Stations are treated as nodes with one single stopping point, regardless of track layout. This means that the trains' runtimes at a station are identical, regardless of track use and stopping location.

The input data that are gathered from the timetable construction at Trafikverket are insufficient when it comes to track use, a consequence of track use having no impact for the runtimes. In the optimization model all trains must have a track allocated to each event at all sections and if there are conflicts in the track use, the trains are routed to different tracks, without affecting runtimes. In RailSys the level of detail is higher and a train's runtime will differ depending on whether the train has to stop at a side track where the possible speed is lower or not. Therefore, when the optimized timetable is imported in RailSys, runtimes for some trains will be changed, leading to a changed amount of runtime margin time and possibly also infeasibility. In the optimization model, all tracks at a station are treated equally, even though some tracks are connected to a platform and some are not. In RailSys we get a warning if a train with passenger exchange stops at a track without a platform, which indicates that the timetable is to some extent infeasible.

At some stations the tracks have several stop boards depending on e.g. train length and platform connection. Since the optimization model handles each station as a node with only one place to stop, the trains' runtimes might change when the stopping place needs to be corrected to fulfil demands such as train length and platform connection. This problem is illustrated in Figure 2 using Kimstad station (Kms) as an example. Here the centre of the station is located 1 km south of the platforms at (2) in Figure 2. The aggregated runtimes used as input to the optimization model are based on running to the centre. For commuter trains that stop at the platforms the runtimes north and south of Kms will then be invalid in RailSys.
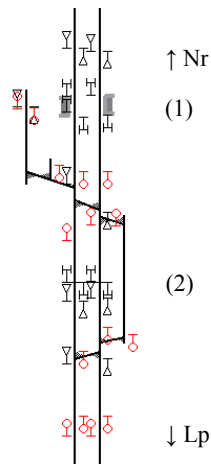

Figure 2: Kimstad station (Kms) where the platforms

(1) are located 1 km north of the station centre (2).

Figure 3 illustrates an example of a train's timetable that has to be modified in RailSys to make it feasible. In the optimization model the timetable is feasible, but it its infeasible when inserted in RailSys. The platform at Vikingstad, where the stop has to take place, is a distance from the station centre in a similar way as illustrated by Kimstad station in Figure 2. This means that the runtime between Linköping and Vikingstad becomes negative in RailSys; see the upper table in Figure 3. To make the timetable feasible the stop in Vikingstad can manually be scheduled 26 seconds later than requested and the runtime margin time after Vikingstad can be decreased to keep the arrival time at the next station after Vikingstad (Mantorp). The lower table in Figure 3 shows how the timetable has changed after it has been made feasible in RailSys.

Station	Rout	Platform	SchedArr	SchedDep	Fix	RTAllow [s]
Norrköping C	[N...	F-2	8:20:59	8:21:00	☐	46
Fiskeby	[N...	Z-3		8:24:12	☐	
Kimstad	[Fi...	N-2	8:32:00	8:32:01	☐	16
Gistad	[K...	Z-1		8:37:03	☐	
Linghem	[Gi...	N-1	8:41:37	8:41:38	☐	51
Linköping C	[L...	E-1	8:48:00	8:50:00	☐	-26
Linköpings Godsbangård	[L...	Z-2		8:51:19	☐	
Vikingstad	[L...	E-3	8:56:21	8:56:22	☐	112
Mantorp	[Vs...	N-1	9:03:00	9:03:01	☐	27
Mjölby	[M...	N-1	9:09:00	9:09:01	☐	

Station	Rout	Platform	SchedArr	SchedDep	Fix	RTAllow [s]
Norrköping C	[N...	F-2	8:20:59	8:21:00	☐	46
Fiskeby	[N...	Z-3		8:24:12	☐	
Kimstad	[Fi...	N-2	8:32:00	8:32:01	☐	16
Gistad	[K...	Z-1		8:37:03	☐	
Linghem	[Gi...	N-1	8:41:37	8:41:38	☐	51
Linköping C	[L...	E-1	8:48:00	8:50:00	☐	0
Linköpings Godsbangård	[L...	Z-2		8:51:25	☐	
Vikingstad	[L...	E-3	8:56:47	8:56:48	☐	86
Mantorp	[Vs...	N-1	9:03:00	9:03:01	☐	27
Mjölby	[M...	N-1	9:09:00	9:09:01	☐	

Figure 3: In RailSys the timetable is at first infeasible with negative runtime between Linköping C and Vikingstad (upper table) but it is made feasible by scheduling the stop in Vikingstad 26 seconds later (lower table).

Also the headway may differ between the macro and micro representation of the network. In RailSys the minimum headway time between two trains depends on track use, signalling placement and also the trains' speed. Locations of signals also affect the track occupancy and hence the minimum headway time. To make the timetable feasible in the microscopic model, departure and/or arrival times for some trains at some stations must be

slightly changed in similar way as illustrated in Figure 3.

Table 1: Example of how some RCP values (in seconds) in the initial timetable have changed between the macroscopic and microscopic model.

Critical point	Macroscopic values				Microscopic values			
	H_p	F_p	L_p	RCP_p	H_p	F_p	L_p	RCP_p
CP9	140	29	134	303	164	46	93	303
CP15	20	4	39	63	60	0	43	103
CP16	80	270	64	414	55	270	56	381
CP23	254	50	33	337	332	52	24	408
CP33	13	0	0	13	18	0	0	18

Since most of the minimum runtimes and headway times differ between RailSys and the optimization model, the RCP values will also differ. In Table 1 we present some examples for critical points in the macroscopic and microscopic model, respectively. This is a selection of critical points from the case study (see Section 4); microscopic values for all critical points are shown in Table 2. Some points, such as CP15, receive a higher RCP value and some points a lower value, as for CP16. In most cases the summarized RCP value differs less than its components. For most critical points the difference is however quite small and respective magnitudes are kept.

When the timetable is optimized for better robustness subject to $RCP_p \geq e.g. 360$ s, departure and arrival times for some trains are changed. Since the difference from the initial timetable, as part of the objective, is being minimized, the changes are kept small. Typically runtimes and headways are adjusted when rescheduling the trains. When the optimized timetable is inserted into RailSys, where the minimum runtimes and headway times are given in more detail, the timetable easily becomes infeasible. In particular, freight trains have a different performance in the optimization model and in RailSys. The macroscopic model treats all freight trains the same, whereas in reality, and in the RailSys model, freight trains have a larger diversity in performance characteristics than passenger trains. For example, the freight trains' different speed profiles often results in longer or shorter minimum headway times than the general values used in the macroscopic model, which might lead to infeasibilities. To keep the RCP values at a minimum of 360 seconds some manual adjustments in the timetable for these trains must be made.

For points such as CP23, the optimization model would not need to improve the RCP value if the value from the microscopic model (408 s) were used instead of the macroscopic value (337 s). Now the RCP value is increased even though the value is in fact large enough in the initial timetable.

4   Case Study

A suitable approach to assess the robustness achieved after optimization is to simulate an initial and an optimized timetable and compare the result. This approach involves perturbing timetables with stochastic disturbances and using dispatching algorithms to re-schedule the trains in real-time. In the simulation, trains run using all of their scheduled runtime in the case of no disturbances. If a train gets delayed it can run faster than in the undisrupted case since there are runtime margins in the timetable. This can sometimes lead to trains realising a shorter journey time than the planned one if the train departs late from a timing point.

The evaluation concerns two macroscopically generated timetables, one initial timetable and one optimized. The first step is to transform these macroscopically

generated timetables into a microscopic format, which requires some manual adjustments as described in Section 3. The two microscopically feasible timetables can then be assessed via simulation in RailSys.

The evaluation of the initial versus the optimized timetable is divided into two experiments with different primary delay distributions. In the first experiment we have used the same primary delays as the ones in Andersson et al. (2015). This mainly captures the differences in dispatching strategies between the macroscopic MILP model and the microscopic simulation tool. In the second experiment we have used primary delays collected from statistics on the particular railway line that we study.

4.1 Timetable Evaluation

The timetable instance used for the evaluation covers the main part of the Swedish Southern mainline between the stations K and Hm. The selected time period is between 8 a.m. and 11 a.m. and the total number of trains in that time period is 79. This line is circa 400 km long and includes both fast long-distance traffic as well as commuter trains and freight trains. In this timetable instance 33 critical points have been identified according to the method presented in pseudo code in Andersson et al. (2013). Figure 4 shows the timetable – the triangles illustrate the 33 identified critical points. Upward pointing triangles illustrate critical points for northbound trains and downward pointing triangles illustrate critical points for southbound trains. Most of the critical points appear in Nr, My, N and Av where commuter trains and other passenger trains start their journeys. The critical points are both locations where trains start their journey after an already operating train (e.g. CP29) and locations where a train overtakes another train (e.g. CP33).
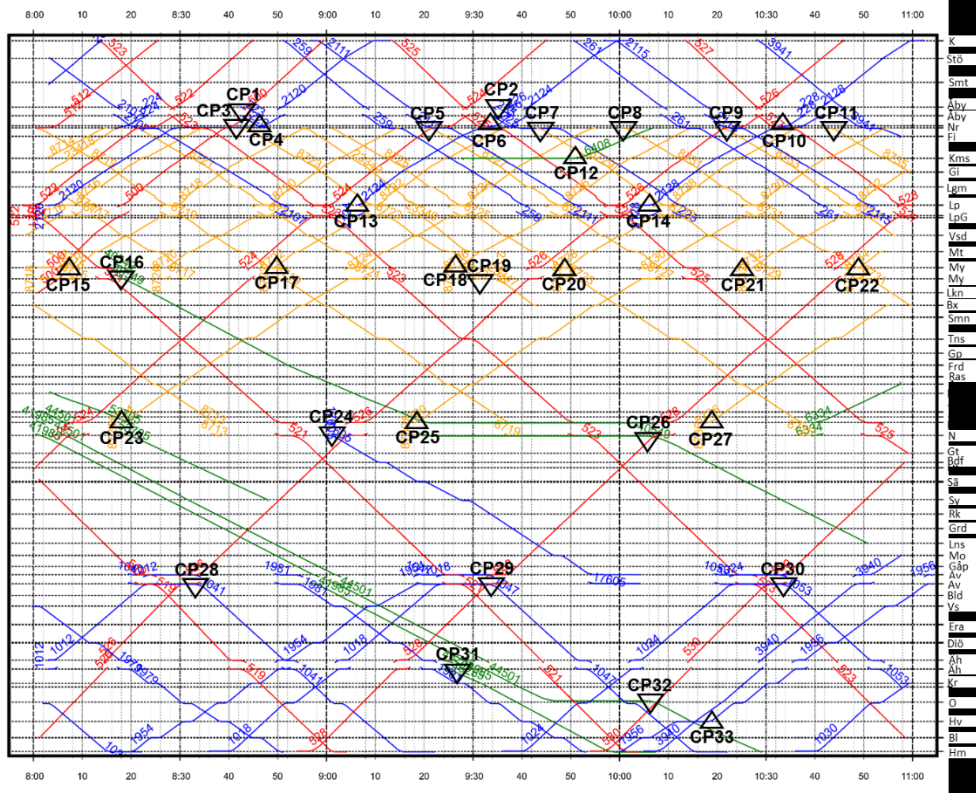
Figure 4: The selected initial timetable instance with 33 critical points.

In some points the headway between the trains is large, which indicates that the RCP value in this point is also high. However, this does not show the full picture. For example, CP16 seems like it should have a lower robustness than CP20 if we only compare the headways in Figure 4 and H_p in the initial timetable in Table 2. But if we compare the initial RCP values in Table 2 we can see that CP16 has a higher total value than CP20 because of the runtime margin for the following train (F_p). This indicates that the total robustness as measured by RCP is higher in CP16 than in CP20. In Table 2 we can see that RCP varies considerably amongst the critical points, between 18 and 1238 seconds in the initial timetable. For this timetable instance the maximum possible value for a RCP increase is 380 seconds. To achieve higher RCP values than 380 for all critical points, the trains' total runtimes have to be allowed to increase, which is assumed not to be desirable. This is explained further in Andersson et al. (2015). In that paper the minimum RCP value in the optimization model was chosen to 360 seconds (6 minutes), and the output timetable from that optimization is the optimized timetable that is evaluated in this paper. Therefore, in the optimized timetable all RCP values are at least 360 seconds but due to the modifications made, the RCP values for other critical points have also increased or decreased. That there are no critical points with very low values indicates that the timetable should be more robust.

Table 2: RCP values in the microscopic model. Points refer to the numbering in Figure 4.

Critical point	Initial timetable				Optimized timetable			
	H_p	F_p	L_p	RCP_p	H_p	F_p	L_p	RCP_p
CP1	763	47	21	831	780	48	26	854
CP2	241	60	21	322	247	113	29	389
CP3	341	36	26	403	333	35	26	394
CP4	224	117	73	414	360	117	228	705
CP5	219	46	0	265	245	17	99	361
CP6	328	14	67	409	325	50	67	442
CP7	322	37	29	388	343	36	29	408
CP8	678	37	240	955	688	36	113	837
CP9	164	46	93	303	198	17	150	365
CP10	328	29	67	424	375	28	67	470
CP11	333	36	26	395	341	36	26	403
CP12	354	22	13	389	346	56	13	415
CP13	215	100	153	468	228	109	235	572
CP14	228	95	179	502	228	111	119	458
CP15	60	0	43	103	193	0	214	407
CP16	55	270	56	381	75	285	0	360
CP17	170	0	79	249	232	0	202	434
CP18	1076	0	109	1185	1074	0	109	1183
CP19	1187	0	51	1238	1210	0	26	1236
CP20	212	0	68	280	218	0	166	384
CP21	1074	0	109	1183	1068	0	109	1177
CP22	151	0	80	231	170	0	251	421
CP23	332	52	24	408	373	51	0	424
CP24	324	88	55	467	341	88	75	504
CP25	514	52	132	698	595	51	46	692
CP26	661	67	67	795	666	62	67	795
CP27	331	52	132	515	502	51	0	553
CP28	401	44	57	502	401	43	57	501
CP29	221	44	129	394	243	0	123	366
CP30	221	44	129	394	250	8	129	387
CP31	14	35	0	49	33	204	123	360
CP32	45	149	2	196	36	307	17	360
CP33	18	0	0	18	227	0	181	408

4.2 Dispatcher Parameters for Microscopic Simulation

For the dispatching, we have chosen parameter values that imitate real-world dispatcher decisions as closely as possible. Several parameters can be used to control the quality of the train dispatcher and also how the dispatcher should prioritize trains in conflict. In Sweden there is a dispatching rule that always gives priority to the train on-time in a conflict situation. However, this rule is not always applied and most of the time the train dispatchers try to solve conflicts with a larger perspective in mind. This means, for example, that a delayed fast long-distance train can be given a higher priority when it is in conflict with another train that is on-time, up to a certain point. When the negative consequences for the on-time train become too large if it is held back in favour of a delayed fast train, the on-time train will typically be given higher priority. In RailSys it is possible to assign different priorities to different train categories depending on how delayed they are. The chosen parameter value for the example above is that the fast long-distance train has highest priority compared to all other trains until it is delayed by 6 minutes. When the train is more than 6 minutes delayed its priority is decreased and the train then gets the same priority as other passenger trains that are running on-time. When compared to real dispatching decisions on the Southern mainline, this priority setting results in the simulated dispatching being executed in a way that is close to reality. Other

parameters that affect the dispatching are, for example, minimum lateness for different kinds of routing choices and lengths of conflict prediction time. The parameters have also been set so as to closely represent real traffic dispatcher decisions. This is a large difference compared to the re-scheduling algorithm used in Andersson et al. (2015), where the model relies on optimal dispatcher decisions. From that, we can assume that the results from the evaluation presented in this paper do not give the minimum possible delays, as in an optimization model. The outcome from the RailSys simulations in the second experiment can be assumed to be more representative of the real world impact.

At each station there are alternative tracks that can be used in the dispatching, e.g. in a delayed situation the dispatcher can re-route a train to a side track and let another train pass for an unplanned overtaking, if there is a train with higher priority that should run ahead. However, even though it is technically possible to reach a side track located on one side of the double-track from both of the tracks, trains are seldom re-routed to a side track accessed only via crossing the opposite direction of the mainline. Therefore, in the simulation model the alternative side tracks are limited to only be used by trains that do not have to cross opposite trains' paths to reach the tracks. This is a limitation compared to the re-scheduling model used in Andersson et al. (2015) where there were no limitations in track usage.

4.3 Primary Delays and Perturbations

For the case study three types of perturbations are used:

- Entry perturbations – the initial disturbances the trains may have when they enter the line, e.g. *train T* is 5 minutes late when it starts its journey at s*tation A*.
- Dwell time perturbations – the disturbance that may occur if a scheduled stop takes a longer time than planned, e.g. *train T* stops at *station B* 2 minutes longer than planned.
- Line perturbations – the disturbance that may occur on the line during the train run, e.g. the runtime for *train T* is increased by two minutes between *station A* and *station B*.

With these perturbations we can capture the trains' delays when they enter the line and also disturbances that are likely to appear during the run.

In the first experiment the disturbances are based on controlled entry and line perturbations in a similar way as in Andersson et al. (2015). Six trains, on average, are delayed in each simulation run, three involved in a critical point and three not involved in a critical point. We use the same set of trains with possibility of delay and the same locations where they receive their delays. Also, the magnitude of the disturbances is between 5 and 10 minutes, just as in Andersson et al. (2015).

For the second experiment data for the entry perturbations have been collected from the punctuality statistics from the Southern mainline in 2014. It is possible to get the data in the same empirical format as used in RailSys. As an example, the disturbance distribution for train 522 when the train enters the line in Linköping is; 42 trains delayed by 59 seconds, 7 trains by 120 seconds, 6 trains by 240 seconds, 2 trains by 360 seconds, 5 trains by 660 seconds and 4 trains by 960 seconds.

The dwell time and line perturbations are based on previously gathered information from Nelldal et al. (2008) of how much longer a stop for a train of a certain category may take and also how long disturbances on the line usually are. This information has then been used to create perturbations which in turn have been calibrated by comparing the

model output and punctuality statistics from 2014.

To get statistically significant data the number of simulation replications is 500 in the first experiment and 1000 in the second experiment. The reason for only having 500 runs in the first experiment is that the randomness is limited due to a small number of possible delayed trains. Also the magnitude and location of the delay is highly limited which reduces the need for a large number of simulation runs.

5 Results from the Simulations

In this section we first formulate the evaluation measures, and then give the results of the two experiments.

5.1 The Evaluation Measures

For the evaluation of the timetable quality we have chosen to assess the timetable robustness with measures of punctuality and delays at commercial stops from the perspective of the whole timetable's performance, and, for experiment 2, with measures of arrival deviation and journey time at selected stations for the pairs of trains involved in each identified critical point.

In the first experiment we repeat the evaluation made in Andersson et al. (2015), but with a different dispatching strategy that is more close to reality. Therefore the selected timetable robustness measures for experiment 1 match those presented in that paper:

- Total delay at end station, minutes (TD)
- The share of trains delayed less than 3/5 min at end station, % (TD+3 and TD+5)
- Total delay at planned commercial stop, minutes (TDS)
- The share of trains delayed less than 3/5 min at planned commercial stops, % (TDS+3 and TDS+5)

For experiment 2 we study the trains' overall punctuality and delays, but we have additionally calculated three critical point-specific measures for the pairs of trains involved in the critical points. These measures involve both of the trains in a critical point, the actual station and one or more subsequent stations:

- the commercial station stop directly before the critical point, *p-1*
- the commercial station stop at the critical point, *p*
- the commercial station stop directly after the critical point, *p+1*
- the destination station, *d*

Some details of the CP-specific localised measures are given below:

The first measure is punctuality per service involved in critical points, CPP^l and CPP^f measures; for the leader/follower in each critical point, the percentage corresponding to the number of times that the service arrived at its destination, *d*, within a 5 minute threshold. The mean value over all simulations is given.

The second measure, the 85th percentile of lateness of services involved in critical points, is computed for the leader and follower, $T_{s,l}^{85}$ and $T_{s,f}^{85}$, respectively, at stations *s*. The term lateness is distinct from delay; it may take both negative (for arrivals earlier than scheduled) and positive (for arrival delays with respect to the schedule) values. It is defined at stations *p-1*, *p*, *p+1* and *d* for the leader and at stations *p+1* and *d* for the

follower.

The third measure is the 85th percentile of percentage of scheduled journey time of services involved in critical points, ($J_{j,l}^{85}$ and $J_{j,f}^{85}$) in a disturbed scenario is computed as a percentage of the scheduled journey time between:

- station *p-1* and *p+1* for the leader
- station *p-1* and *d* for the leader
- station *p* and *p+1* for the follower
- station *p* and *d* for the follower

We have chosen to examine three of the critical points, in which interesting local effects of a range of magnitudes can be observed, in the context of CP-specific localised measures, namely CPs 17, 27 and 33. Table 3 shows the services involved in these critical points and the stations corresponding to those outlined above.

Table 3: Services and stations at which measures are defined for CPs 17, 27 and 33. For CP27 the station *p+1* and *d* for the leader is the same, Lp.

Critical point	Service number Leader/follower	Station stop before CP (leader)	Critical point station	Station stop after CP leader/follower	Destination station leader/follower
CP17	524/8722	N	My	Lp/Mt	K/Nr
CP27	528/8736	Av	N	Lp/Any	Lp/Bx
CP33	3940/1956	Hm	Hv	Av/O	Mo/Av

The corresponding results of the CP-specific localised measures for the three chosen critical points are given in Table 5 and Table 6.

5.2 Results from the First Experiment

The overall result shows that the timetable robustness is improved when RCP values increase, even without optimal rescheduling algorithms. There are, however, some differences in the results, e.g. the decrease in TD is larger in the microscopic model but the decrease in TDS is larger in the macroscopic model. In Table 4 we can compare the results from the macroscopic and microscopic experiments and see that the robustness values between the initial timetable and the optimized timetable is improved at both the macroscopic and microscopic level.

Table 4: Results from the first experiment at macroscopic level (from Andersson et al. 2015) and microscopic level.

	Initial timetable		Optimized, $RCP_p \geq 360$	
	Macroscopic	Microscopic	Macroscopic	Microscopic
TD (minutes)	16.0	23.7	14.4	18.8
TD+3 (%)	96.8	96.7	97.1	97.0
TD+5 (%)	98.9	98.0	98.9	98.4
TDS (minutes)	115.2	105.8	98.7	92.8
TDS+3 (%)	79.7	80.9	82.8	82.4
TDS+5 (%)	88.0	88.4	88.9	89.5

In the optimization model used in Andersson et al. (2015) the objective was to minimize the delays at the end station. The result of this objective can be seen in Table 4

since the measures involving TD are better in the macroscopic experiment. However, the dispatching algorithm in the microscopic experiment tries to handle all conflicts with a more local perspective which results in the measures involving TDS being better in the microscopic experiment.

5.3 Results from the Second Experiment

In the second experiment we try to imitate real world conditions with both dispatching and initial disturbances. The aggregated punctuality for all trains combined indicates small changes between the initial and the optimized timetable (see columns 3 and 4 of Table 5). However, when considering individual trains there are some more significant differences; aggregated measures do not provide sufficient information on the differences between initial and optimized timetable.

Trains involved in a critical point with a low RCP value in the initial timetable show significant improvement in the optimized timetable. For example, train 500 with very few scheduled stops realises an increased punctuality from 78.3 % to 84.7 %. Trains 3940 and 1956, which are involved in CP33, both have an increased punctuality at their end stations Av and Mo; see Figure 5. The punctuality for train 1956 is however decreased at the first two stops after Hm which is an effect of the runtime margin having been re-allocated from before to after the critical point in Hv in the optimization model to fulfil $RCP_p \geq 360$ for CP33. We can see that the recovery is much larger after Hv with the optimized timetable compared to the initial.

When large parts of the runtime margin are re-allocated from one location to another in a trains' timetable to fulfil $RCP_p \geq 360$ s, the trains' ability to recover from delays might vary considerably. In Figure 6 we can see the average lateness for train 528 where almost 3 minutes of runtime margin time is re-allocated from before N to after N in the optimized timetable to increase RCP in CP22 located between N and Lp. Therefore the lateness in the intermediate stations Av and N is much higher in the optimized timetable. Compared to the small increase in punctuality at the end station, the negative intermediate effect should outweigh it, in favour for the initial timetable. To assess robustness, therefore, the intermediate effects also need to be considered along with other KPIs; we should not only consider end station punctuality.
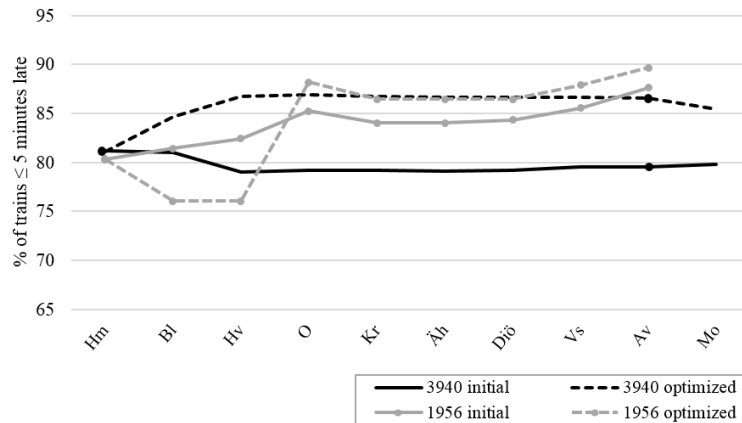
Figure 5: Punctuality for train 3940 and 1956. The trains start their journeys in Hm and are involved in CP33 located in Hv; the intermediate stops are marked with a point.
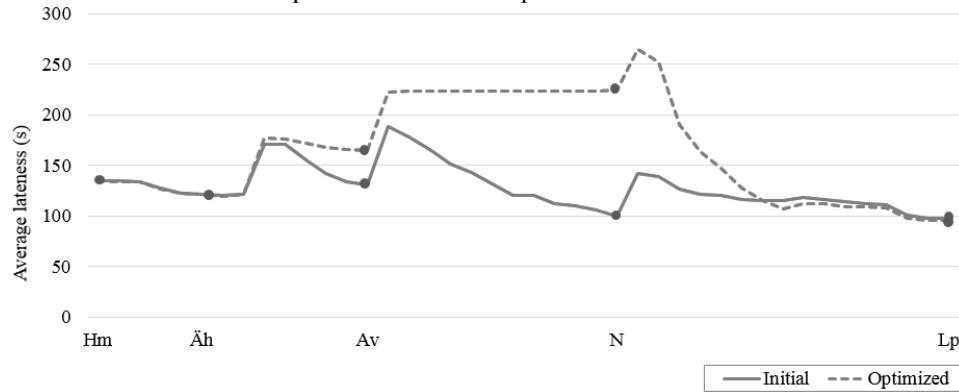

Figure 6: Average lateness for train 528 with the initial and optimized timetable. The train travels from Hm to Lp, the intermediate stops are marked with a point and the critical point CP22 is located between N and Lp.

For some trains the punctuality between the initial and optimized timetable does not differ, even though the RCP values concerning these trains are higher in the optimized timetable. For example, this is the case for train 526 where RCP in CP20 is increased from 280 to 384 seconds, as can be seen in Table 2. The reason for this is that train 526 has a high punctuality in My which implies that there is rarely a conflict even in the initial timetable. Solely considering the RCP value is hence insufficient to predict punctuality.

Figure 7 shows the relationship between change in RCP and punctuality, respectively for all trains involved in the critical points. The measures clearly have a positive correlation, but with individual deviations.
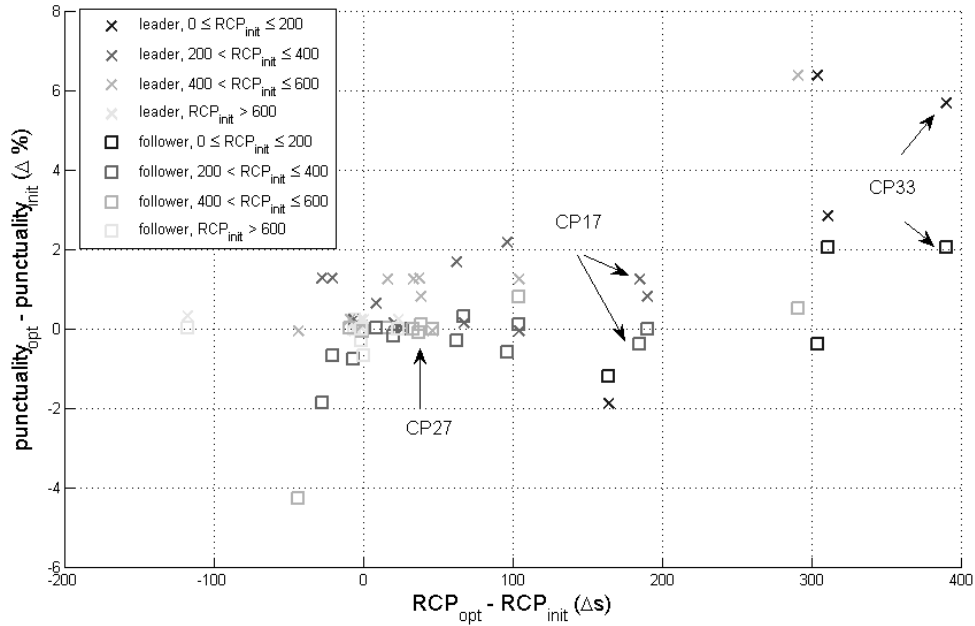
Figure 7: Change in punctuality at the final station against change in RCP value between initial and optimized timetables for experiment 2.

The CP-specific measure values for the three chosen CPs, namely CPs 17, 27 and 33 introduced in Subsection 5.1, are given in Table 5 and 6. The results corresponding to arrival lateness and actual/scheduled journey time in these CPs obtained over all simulations are shown in Figure 8 to Figure 10. In the first six of the subplots, (a)–(f), the arrival lateness in each simulation for both the initial and optimized timetables is plotted. The data have been sorted into ascending order of lateness; in this way the 85th percentile (marked with a vertical dashed line) and the overall picture of lateness across the simulations can be visualised. Similarly the sorted ratio of the actual/scheduled journey time is plotted in subplots (g)–(j).

Table 5: Critical point punctuality and 85th percentile of arrival lateness in critical points for CPs 17, 27 and 33.

| | Critical point | Critical point punctuality, CPP (%) | | 85th percentile of arrival lateness (s) | | | | | | | |
| | | | | T_{p-1}^{85} | | T_{p}^{85} | | T_{p+1}^{85} | | T_{d}^{85} | |
		Init	Opt	Init	Opt	Init	Opt	Init	Opt	Init	Opt
leader	CP17	85.7	87.0	353	377	325	226	249	150	282	267
follower		98.2	97.8					131	121	46	79
leader	CP27	94.1	94.9	304	346	224	396	121	123	121	123
follower		97.5	97.6					156	157	60	60
leader	CP33	79.8	85.5	397	401	404	252	1032	254	995	291
follower		87.6	89.7					288	217	229	187

Table 6: 85th percentile of actual/scheduled journey time in critical points 17, 27 and 33.

Critical	journey time actual/scheduled (%) of 85th percentile

	point	(scheduled journey time, s)							
		$J^{85}_{p-1:p+1,l}$		$J^{85}_{p-1:d,l}$		$J^{85}_{p:p+1,f}$		$J^{85}_{p:d,f}$	
		Init	Opt	Init	Opt	Init	Opt	Init	Opt
leader	CP17	99.5 (2879)	98.4 (3004)	103.7 (5694)	102.7 (5737)				
follower						100.0 (316)	100.0 (316)	98.6 (2879)	98.8 (2819)
leader	CP27	100.3 (4920)	99.9 (4960)	100.3 (4920)	99.9 (4960)				
follower						92.9 (718)	92.9 (718)	98.1 (2227)	98.2 (2227)
leader	CP33	108.7 (2519)	101.9 (2580)	124.9 (3038)	106.3 (3039)				
follower						126.9 (353)	100.0 (351)	100.5 (2759)	100.4 (2716)

In CP17 we see evidence that the leader, train 524, performed slightly better in the optimized timetable at the critical point station and at the following and final stations, the arrival lateness has decreased as can be seen in Table 5 and the punctuality has increased as can be seen in Figure 7. The effect of optimizing the timetable for the follower, train 8722, was less significant. This train recorded very similar lateness values over all simulations at both the station after the critical point and the final station, where, in fact, the lateness was slightly worse, as reflected in the measure T^{85}_d for the follower. The follower also experienced broadly similar journey times across all simulations, the main difference being that there were fewer instances of actual journey times being more than 100% of scheduled between My and Mt. The leader (524) performed better in terms of consistency of journey time for both of the journeys assessed, N – Lp and N – K.

Of the three critical points examined in more detail, CP27 showed the most similar results between the initial and optimized timetables for the localised lateness and journey time measures. The locations where differences of any significance between initial and optimized timetables were observed are stations *p-1* (Av) and *p* (N) for the leader, where the 85th percentile lateness value was slightly worse for the optimized timetable, 346s vs 304s and 396s vs 224s, respectively. The margin re-allocation that led to the changed RCP value in CP27 is in fact an effect of the RCP increase in CP22, where train 528 also is involved. This shows that increased robustness in one critical point might lead to decreased robustness in another point. However, the actual/scheduled journey time value decreased very slightly between N and Lp for the leader (100.3% vs 99.9%).

In CP33 T^{85}_d falls significantly for the leader (3940) at its final station, Av, from 995 s to 291s between initial and optimized timetables, while a less significant decrease is seen for the follower at its final station. From Figure 10 we see that overall fewer instances of severe lateness were recorded at the final station for train 3940, but that for simulation indices >916 the lateness value recorded was worse for the optimized timetable than the initial one. We may conclude that the optimized timetable performed on the whole better than the initial one in terms of lateness at the final station for train 3940; taking the 85th percentile measure captures this, but if we had taken, for example, the 95th percentile, this value alone would suggest that the initial timetable performed better than the optimized one. This indicates that measuring values at a selection of percentiles, say 75th, 85th and 95th may be prudent. In terms of journey times for the services involved in CP33, a passenger travelling from Hv to Av with train 3940 (see Figure 10 (g)) has a 1 minute longer scheduled journey time (2580 s vs 2519 s) as a result of the optimization, but on

average passengers get a much more stable journey since the train is late less frequently and also less severely late on average. This is likely to be perceived positively by passengers, who typically highly value reliable journey provision.
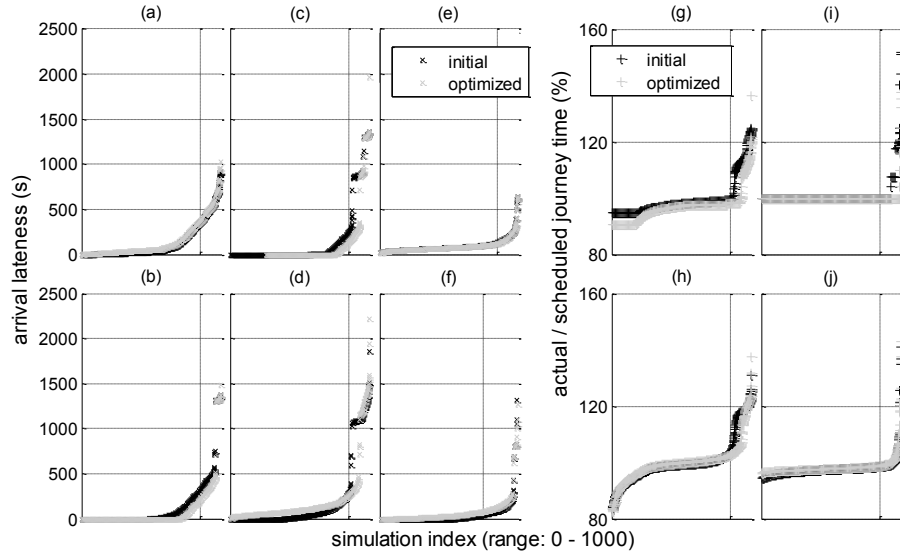

Figure 8: For CP17, the lateness of arrival for the leader, train 524 at (a) N, (b) My, (c) Lp, (d) K and for the follower, train 8722, at (e) Mt and (f) Nr; the actual journey time/scheduled journey time for the leader between (g) My and Lp and (h) My and K, and for the follower between (i) My and Mt, and (j) My and Nr.
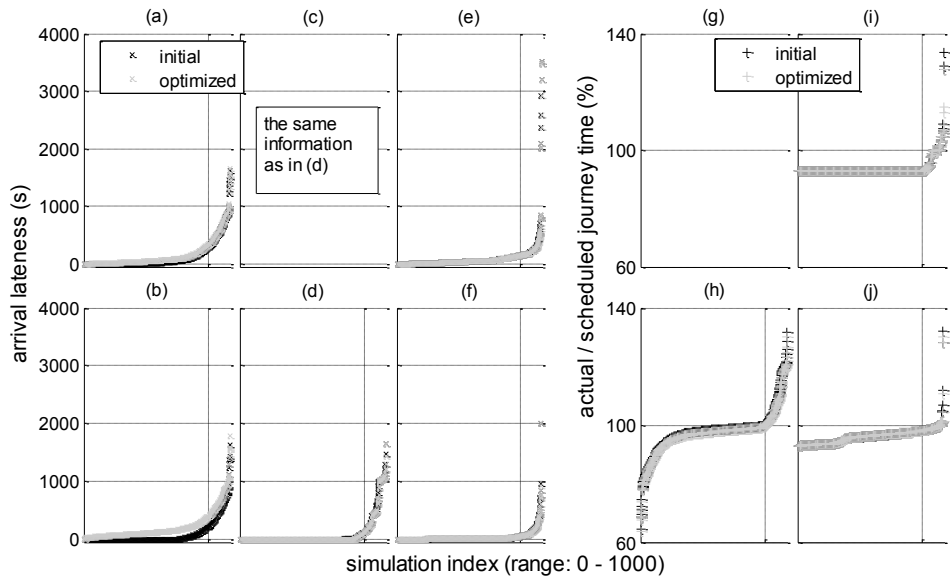

Figure 9: CP27, the lateness of arrival for the leader, train 528 at (a) Av, (b) N, (c) [intentionally blank], (d) Lp and for the follower, train 8736 at (e) Any and (f) Bx; the actual journey time/scheduled journey time for the leader between (g) [intentionally

blank] and (h) N and Lp, and for the follower between (i) N and Any, and (j) N and Bx.
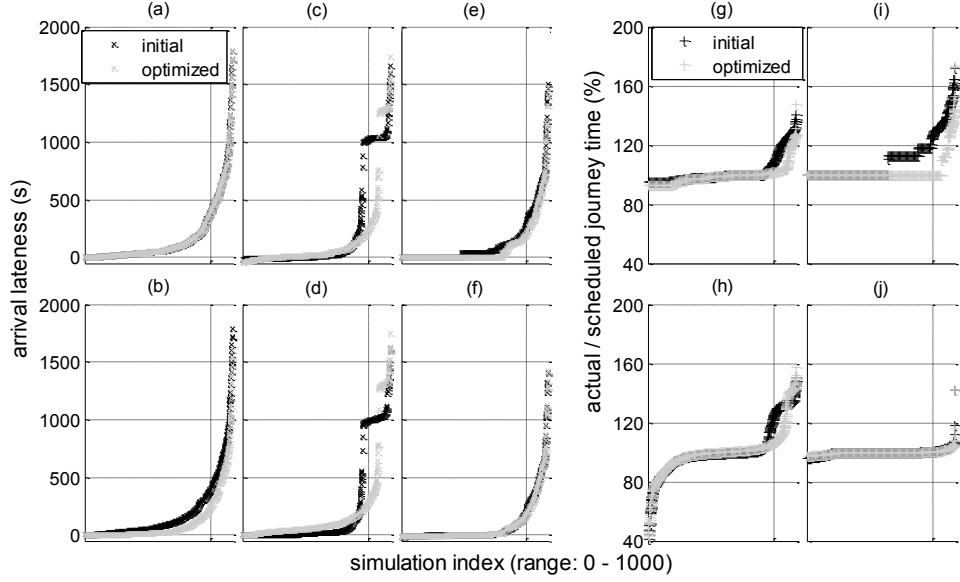

Figure 10: CP33, the lateness of arrival for the leader, train 3940 at (a) Hm, (b) Hv, (c) Av, (d) Mo and for the follower, train 1956 at (e) O and (f) Av; the actual journey time/scheduled journey time for the leader between (g) Hv and Av and (h) Hv and Mo, and for the follower between (i) Hv and O, and (j) Hv and Av.

6   Discussion

Actions taking in the optimal dispatching strategy always have the primary objective in mind. Since the objective in Andersson et al. (2015) is to minimize the end station delays, all decisions taken will lead to a minimum end station delay. In real-world dispatching situations, this is not the case, however. A more suitable method when trying to imitate real-world scenarios may be to use microscopic simulation with appropriately configured dispatching options where the dispatchers cannot always foresee the consequences for the end station and therefore solve the conflict with a more local perspective.

In the second experiment real-world disturbances are used to disturb the trains and it is possible to analyse how the ex-ante indicator RCP performs in an environment that is close to reality. At first glance, the robustness does not seem to have improved much, since the overall punctuality is almost the same, but when studying individual critical points and the involved trains some interesting observations can be made. As expected, we cannot draw conclusions about the semantics of the RCP indicator as we only have one timetable variation to test.

When studying the measure punctuality at the end station the overall robustness is slightly higher in the optimized timetable, but when analysing the punctuality at the intermediate stops the results differ. Since the runtime margin has been re-allocated in the RCP optimization, some trains have lost all runtime margin in some sections. This means that the trains cannot recover from a delay in these sections in the same way as in the initial timetable. If the punctuality decrease there is small, then it might be an acceptable

loss to make in order to achieve a greater robustness for the overall punctuality level. For some trains the punctuality at some intermediate stations decreases by up to 3–4 percentage points in the optimized timetable, which indicates that the runtime margin is required for the robustness at their initial locations and it should not be re-allocated. It could therefore be of interest to prevent the optimization model from re-allocating all runtime margin from one location and to instead keep some of the runtime margin at that location to retain some of the capability to recover there. There is a need for adding more constraints in the optimization model to restrict how the model is allowed to re-allocate margin time.

For some points, such as CP22, the interaction between the leader and the follower is short for this particular timetable instance. This means that the leader only has to run after the follower for a short time and will therefore not receive a large secondary delay in the case of a disturbance. For these situations it is not very important to seek a high RCP value and the runtime margin might be of better use elsewhere, especially if the intermediate consequences of an RCP increase are as negative as is shown for train 528, the leader, in CP22.

Also, for some critical points where the RCP is increased in the optimized timetable, the punctuality for the involved trains does not increase. This is because the punctuality for the involved trains is already at such a high level in the initial timetable, so a conflict rarely occurs. These two examples indicate that the combination of the ex-ante indicator RCP with the ex-post measure punctuality is valuable in RCP optimization (and timetable optimization in general) to choose where and how to apply the optimization.

7   Conclusions and Future Research

In this paper we have presented the first steps towards a real implementation of a RCP optimization model. A macroscopically generated improved timetable has been assessed via microscopic simulation and subsequently evaluated with several performance measures. The results indicate that it is necessary to use several KPIs to effectively evaluate the effects of an RCP increase. If we only look at the punctuality we will get a limited view of the result from which it is hard to draw relevant conclusions. For most of the critical points with a higher RCP value in the optimized timetable, the punctuality increases, but there are also trains that are unaffected and trains that are affected negatively at some stations. The punctuality is highly related to where the initial disturbances occur. If too much runtime margin is allocated to the beginning of a train's journey, this will already be consumed when disturbances occur later. We can get more detailed information by measuring the 85th percentile lateness of the leader and follower in each critical point but the measure must be used carefully and perhaps even assessed more, based on precisely which percentile is selected for reporting.

When assessing robustness with other KPIs such as journey time and lateness we can see that an RCP increase led to a slightly longer journey time for some trains on some parts of the line. However, since the results also show that lateness and the risk of becoming more delayed decrease, the longer journey time can be readily endured.

Our experiments have verified the first steps towards a real world implementation and given examples of results. It is for the future to validate these results for other data sets, which is needed for significant results. From the fact that some trains are unaffected by the RCP increase and some trains even perform worse we can conclude that there is a need for further studies into how the optimization model should be allowed to re-allocate margin time. Future work should include analysis of the relationship between RCP and

ex-post measures and the results should be used to set up rules for the advisable re-allocation of margin time and re-distribution of H_p, L_p and F_p. The optimization model should also be updated with more detailed runtimes, headway times, etc. to give a closer representation of reality. For example, since the freight trains have a larger diversity in performance characteristics it might be a good idea to group them in to clusters with similar characteristics instead of using the same general values for all of them.

It may be of interest to measure the consequences on connectivity and involve statistics on passenger numbers to see how many passengers are affected by the RCP increase. Then it might be possible to accomplish the improved robustness we get with increased RCP with the broadest positive consequences for both operators and passengers.

Acknowledgement

References

Andersson, E.V., 2014. *Assessment of robustness in railway traffic timetables*, Licentiate Thesis No. 1636, Linköping University, Linköping, Sweden.

Andersson, E.V., Peterson, A., Törnquist Krasemann, J., 2013. "Quantifying railway timetable robustness in critical points", *Journal of Rail Transport Planning & Management 3*, pp. 95–110.

Andersson, E.V., Peterson, A., Törnquist Krasemann, J., 2015. "Reduced railway traffic delays using a MILP approach to increase robustness in critical points", *Journal of Rail Transport Planning & Management* 5, pp 110–127.

Bendfeldt, J.P., Mohr, U., Müller, L., 2000. "RailSys, a system to plan future railway needs. *Computers in railways VII*, pp. 249-255.

Bešinović, N., Goverde, R.M.P., Quaglietta, E., Roberti, R., 2016. "An integrated micro-macro approach to robust railway timetabling", *Transportation Research Part B 87*, pp. 14–32.

Bešinović, N., Goverde, R.M.P., Quaglietta, E., 2017. "Microscopic Models and Network Transformations for Automated Railway Traffic Planning", *Computer-Aided Civil and Infrastructure Engineering 32*, pp. 89–106.

Blanco, M., Schlechte, T., 2014. "Analysis of micro–macro transformations of railway networks", in: *Operations Research Proceedings 2013*, D. Huisman et al. (ed.), Springer, pp. 37–42.

Carey, M., 1999. "Ex ante heuristic measures of schedule reliability". *Transportation Research Part B 33*, pp. 473–494.

Fischetti, M., Salvagnin, D., Zanette, A., 2009. "Fast approaches to improve the robustness of a railway timetable", *Transportation Science 43,* pp. 321–335.

Goverde, R.M.P., Bešinović, N., Binder, A., Cacchiani, V., Quaglietta, E., Roberti, R., Toth, P., 2016. "A three-level framework for performance-based railway timetabling", *Transportation Research Part C 67*, pp. 62–83.

Jensen, L. W., Landex, A., Nielsen, O. A., 2014. "Evaluation of robustness indicators using railway operation simulation", In: Brebbia, C. A., Tomii, N., Tzieropoulos, P., Mera, J. M. (eds.) *Computers in Railways XIV*, WIT Press Southampton, pp. 329–339.

Kecman, P., Corman, F., D'Ariano, A., Goverde, R.M.P., 2013. "Rescheduling models for

railway traffic management in large-scale networks", *Public Transport* 5, pp. 95–123.

Kroon, L., Dekker, R., Vromans, M., 2007. "Cyclic railway timetabling: A stochastic optimization approach", In: Geraets et al. (Eds.) *Algorithmic methods for Railway Optimization*, pp. 41‒66, Springer Berlin Heidelberg.

Kroon, L., Maroti, G., Helmrich, M., Vromans, M., Dekker, R., 2008. "Stochastic improvement of cyclic railway timetables", *Transportation Research Part B 42*, pp. 553–570.

Nicholson, G.L., Kirkwood, D., Roberts, C., Schmid, F., 2015. "Benchmarking and evaluation of railway operations performance", *Journal of Rail Transport Planning & Management 5*, pp. 274–293.

Nelldal, B-L., Lindfeldt, O., Sipilä, H., Wolfmaier, J., 2008. "Förbättrad punktlighet på X2000 – analys med hjälp av simulering", Research report, KTH, Stockholm (in Swedish).

Parbo, J., Nielsen, O. A., Prato, C. G., 2016. "Passenger perspectives in railway timetabling: a literature review", *Transport Reviews* 36:4, pp. 500–526

Quaglietta, E., Pellegrini, P., Goverde, R.M.P., Albrecht, T., Jaekel, B., Marlière, G., Rodriguez, J., Dollevoet, T., Ambrogio, B., Carcasole, D., Giaroli, M., Nicholson, G., 2016. "The ON-TIME real-time railway traffic management framework: A proof-of-concept using a scalable standardised data communication architecture", *Transportation Research Part C: Emerging Technologies* 63, pp. 23-50.

RMCon, 2015. Manual RailSys Version 10, Hanover, Germany.

Schlechte, T., Borndörfer, R., Erol, B., Graffagnino, T., Swarat, E., 2011. "Micro–macro transformation of railway networks", *Journal of Rail Transport Planning & Management 1*, pp. 38–48.

Transport Focus, Office of Rail and Road, 2015. "Train punctuality: the passenger perspective", Available at http://www.transportfocus.org.uk/research-publications/publications/train-punctuality-the-passenger-perspective/

Vromans, M., 2005. *Reliability of railway systems*. PhD thesis, ERIM, Erasmus University Rotterdam, The Netherlands.

Vromans, M., Dekker, R., Kroon, L., 2006. "Reliability and heterogeneity of railway services", *European Journal of Operational Research 172*, pp. 647‒665.

Warg, J., Bohlin, M., 2016. "The use of railway simulation as an input to economic assessment of timetables", Accepted for publication in*: Journal of Rail Transport Planning & Management.*