Robustness in Swedish Railway Traffic Timetables

Emma Andersson, Anders Peterson, Johanna T. 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
A tendency seen for quite some time in the Swedish railway network is a growing demand for capacity which no longer can be accommodated. This causes congestion and delays, and the relationships between the trains and how they affect eachother are significantly harder to overview and analyse. Railway traffic timetables normally contain margins to make them robust, and enable trains to recover from certain delays. How effective these margins are, depends on their size and location as well as the frequency and magnitude of the disturbances that occur. Hence, it is important to include marigns so, that they can be used operationally to recover from a variety of disturbances and not restricted to a specific part of the line and/or the timetable. In a case study we compare the performance of a selection of passenger train services to the different prerequisites given by the timetable (e.g. available margins and their location, critical train dependencies). The study focuses on the Swedish Southern mainline between Stockholm and Malmö on which a wide variety of train services operate, e.g. freight trains, local and regional commuter train services as well as long-distance trains with different speed profiles. The analysis shows a clear mismatch between where margins are placed and where delays occur. We also believe that the most widely used performance measure, which is related to the delay when arriving at the final destination, might give rise to an unnecessarily high delay rate at intermediate stations.

Keywords
Railway traffic, Timetabling, Scheduling, Robustness, Delay management.

1 Introduction

A tendency seen for quite some time in the Swedish railway network is a growing demand for capacity and certain train slots, which no longer can be fully met. The network is becoming more and more congested and the relationships between the trains and how they affect eachother are significantly harder to overview and analyse. A key number supporting this claim is the annual passenger traffic volume, which now is at an all-time-high level. During 2009, approximately 11.1 billion passenger-kilometres were produced in the Swedish railway network [1].

During the process of constructing the national traffic timetable for 2011, the capacity limit was reached for several crucial links in the mainlines of the network. Since the Swedish railway market became fully deregulated October 1st, 2010, there are several operators competing for attractive slots along these lines. Consequently, the shortage of such slots and the competiton for the accessibility, led to that commercial and legal issues were raised.

The shortage of capacity has not only created problems at a planning level, but has also been confirmed at an operational level. In particular, the fast long-distance trains, which run through several different commuter traffic regions, suffer from the increased traffic density on the mainlines and have experienced increased puncuality problems.

One example is the fast long-distance train X2000 operating on the Southern mainline Stockholm-Malmö. In Figure 1 below, its puntuality statistics for January-September 2010 is compared to that of other trains on the Swedish network.

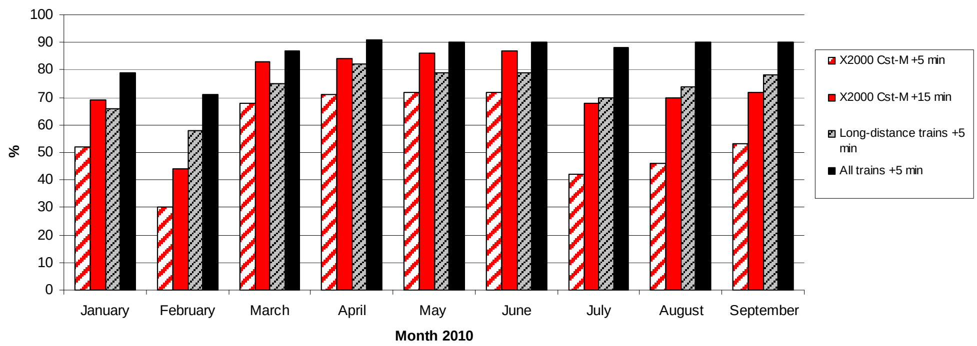


Figure 1. Punctuality statistics for the Swedish railway traffic January–September 2010. The bars show the percentage of the traffic arriving at the final destination at most 5 (15) minutes delayed, see legend. Source: SJ AB web statistics.

The punctuality during 2010 has decreased significantly, compared to 2009. The problem consists of mainly two parts; 1) Disturbances occur causing delays, 2) which in turn may cause ripple effects and consecutive delays. Consequently, there are two complementary solutions: 1) To prevent certain disturbances from occuring and 2) to design timetables which contain sufficient flexibility (e.g. margins, alternative locations for meeting and overtaking of trains if necessary) which makes it valid and executable despite that the prerequisites are slightly different from those that were given when the timetable was created.

In this paper, we describe how margins are computed and included in the timetable construction to provide a certain level of robustness. We refer to margins as all extra time that is added to the technical running time in a timetable, which includes all different types of buffer-times and time supplements.

We are interested in how these margins are utilized operationally to achieve high punctuality, and and present a case study, where the performance of a selection of passenger train services is compared to the prerequisites given by the timetable (e.g. available margins and their location, critical train dependencies). This comparison is the main contribution of this paper.

The remaining part of this paper is organised as follows. In the following section we describe the problem, define robustness and explain how time margins are computed and motivated in the current timetable planning. A brief summary of related work concludes the background part. Thereafter we present our case study, where margins and performance patterns are compared and analysed in different ways. In the final section, we present our conclusions and provide some ideas for future research.

2 Background

2.1 Problem description

Currently, as previously outlined, X2000 is very sensitive to even minor disturbances. If an X2000 between Stockholm (Cst) and Malmö (M), see Figure 2 below, experiences a smaller delay of a few minutes at the beginning of its journey it may become heavily delayed later on. It is so critical that the train times its entrances to the commuter traffic regions, that the train will pass through. To illustrate, consider Figure 3 below, where the traffic history of the X2000 service 521 for two weeks in October, 2010, is depicted. The x-axis shows, from left to right, the stations it has visited. The y-axis shows the deviation from the train's timetable, where a positive value indicates a delay and a negative value that the train is ahead of schedule and has some margin. As can be seen, for smaller delays the train arrives quite well in time to its final destination despite significant but small intermediary delays (< 10 minutes) and the reason is that it has a large margin allocated in the very end of its journey.

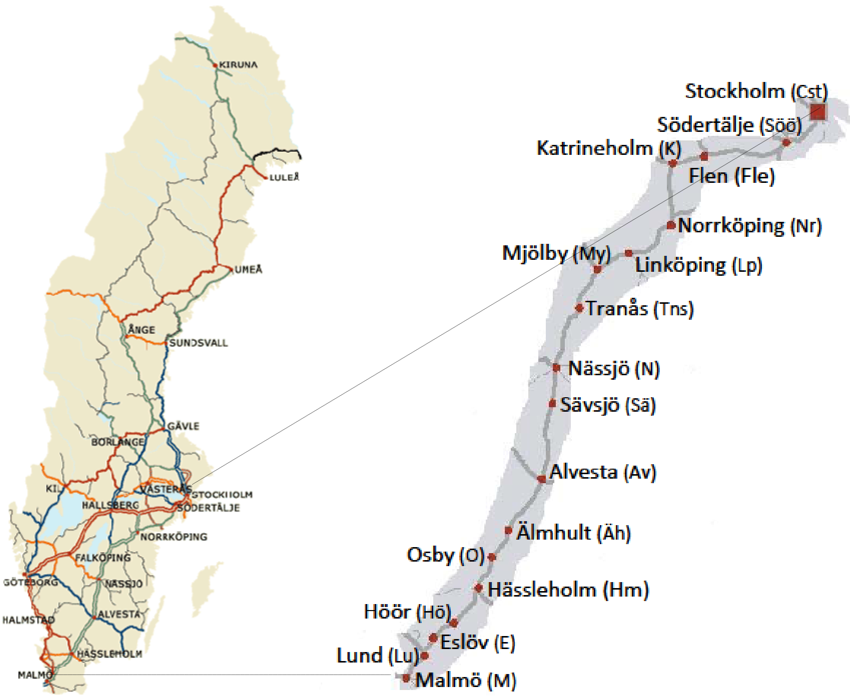


Figure 2. The Swedish railway network (left) and the double-tracked Swedish Southern mainline (right). Source: Swedish Transport Administation and SJ AB (adapted).

Since the delays typically occur already quite early on, the location of the margins is not ideal since the passengers at intermediate stops are delayed and the train disturbs surrounding traffic. The ultimate margins (or any other type of flexibility in the timetable) should instead be located so that they can be used to quickly absorb different types of disturbances and thus make the timetable robust.

In Figure 3, we can also see that there are a few days where train 521 did not reach its destination in time, which can have several explanations. To what extent a train can recover from a delay depends not only on whether its timetable contains sufficient margins. It is also affected by how the margins are used by the train dispatchers. Once a train is classified as delayed, it can be given a lower priority or be guided to wait at a side-track in favor of other trains. How the dispatchers make their decisions is not obvious and it often depends on multiple factors and the experience of the dispatchers in charge. Sometimes the dispatcher lets X2000 overtake a slow train when it is delayed and sometimes not. Therefor two trains with the same delay at one point can end up in Malmö with a very different delay.

Many questions are raised when studying and analysing the behaviour of train 521 concerning which factors that contribute most to its oscillating behaviour. When comparing with the corresponding graph for the similar X2000 service 524, which runs in the opposite direction (see Figure 4), it is clear that there is a need to structurally analyse and benchmark a set of train services along this line and investigate any potential correlation between timetable attributes and performance indicators.

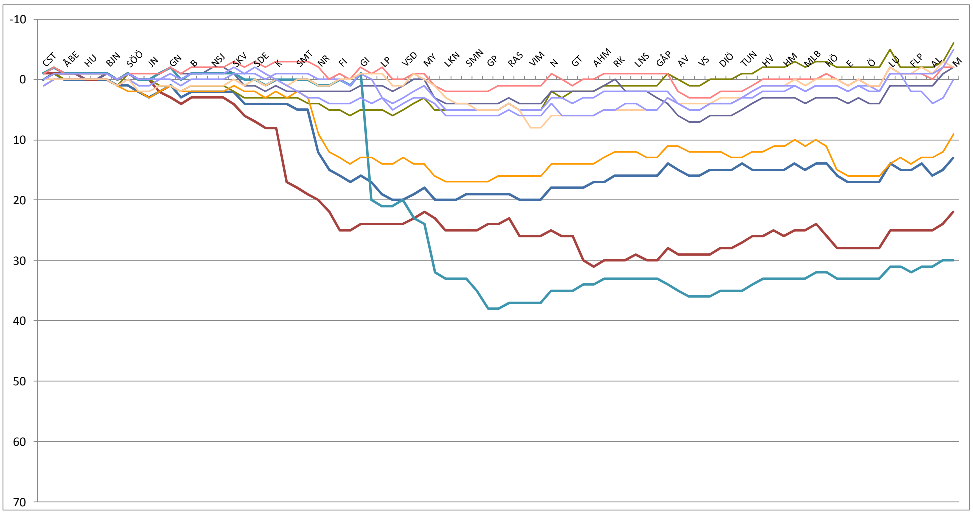


Figure 3. The X2000 service 521 and its deviation from its timetable at different locations along its journey from Stockholm (Cst) to Malmö (M). Statistics were collected during two weeks in Octobre, 2010.

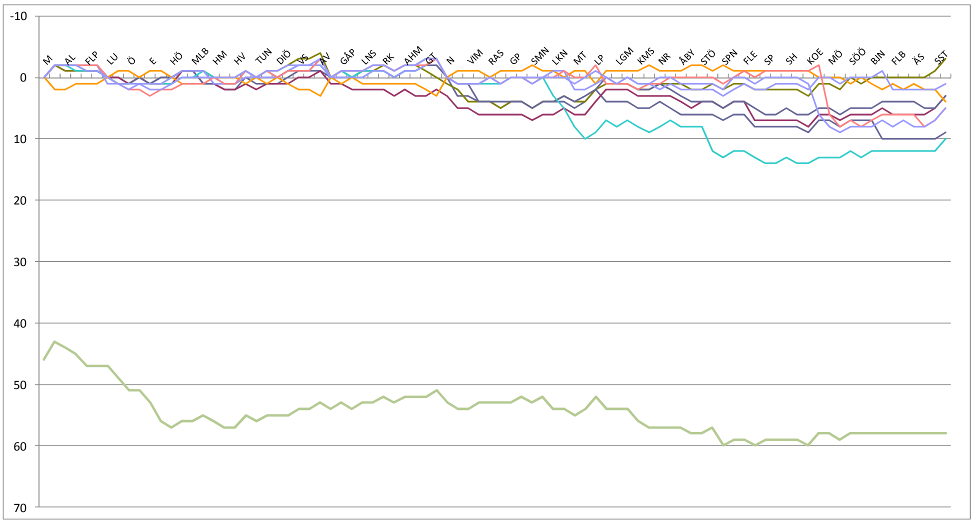


Figure 4. The X2000 service 524 and its deviation from its timetable at different locations along its journey from Malmö (M) to Stockholm (Cst). Statistics were collected during two weeks in Octobre, 2010.

2.2 Definitions

During the last decade several approaches have been proposed to investigate, measure, compare, improve or optimize the *robustness*, which refers to, e.g., “the ability to resist to ‘imprecision’” [13], the tolerance for “a certain degree of uncertainty” [12] or the capability to “cope with unexpected troubles without significant modifications” [19].

Robustness is a consequence of the concatenated planning process. An attempt to give a hierarchical structure is given by Liebchen [8], who identifies the following steps: network design, line planning, timetabling, vehicle scheduling, duty scheduling and crew rostering. Studying the timetable construction is, in this meaning, to address the robustness issue on a mid-level.

The margins to be added to the running time of trains and between train paths to ensure that minor delays are suppressed instead of amplified, constitute one component of the railway capacity. For a given infrastructure, that also involves the following three dimensions: Number of trains, average speed and heterogeneity [18]. Increasing robustness should be done without interfering with the other capacity components.

To be able to use robustness as an objective for the planning process, a quantifiable performance measure is needed. At present, the Swedish Transport Administration measures the imprecision in the railway operations by studying trains arriving more than five minutes late to their final destination, cf. Figure 1.

The performance is reported monthly in reports, published on the authority’s webpage [16]. Clearly it is not possible to capture all aspects of robustness, by considering only one single indicator.

2.3 The Swedish Railway Traffic Timetable Construction

As in most other European countries, the Swedish railway traffic timetable is finalized every autumn and is then valid for the next year (e.g. December, 12th 2010 – December 10th, 2011). As the railway market is deregulated, it is a time-consuming process every year for the Swedish Transport Administration to unite the requests from multiple and sometimes competing operators. Freight and passenger trains are mixed and the infrastructure permits traffic in both directions although during peak-hours left-hand traffic is preferred. Depending on which part of the network that is of interest, margins are added 1) to ensure that train dependencies are not too tight making the timetable sensitive to minor deviations and 2) to adjust the theoretical running times and compensate for certain deviations and variations in driving style.

In order to handle the capacity shortage on certain stretches during the timetable construction process, The Swedish Transport Administration has defined some stretches in the Swedish railway network as over-utilized. That means, that there is a lack of capacity on the tracks during some peak-hours every day and the traffic needs to be planned differently there, than other stretches. A set of planning rules are given, according to which there has to be a minimum distance between two trains using the same track through these passages. The passages along the Southern mainline are all located in the northern end, close to Stockholm, and the minimum distances there are:

- Stockholm Central – Stockholm Södra: 2 minutes
- Stockholm Södra – Älvsjö: 3 minutes
- Älvsjö – Järna: 4 minutes
- Järna – Katrineholm: 5 minutes

Because of these minimum distances margins may be needed to synchronize the trains in the planning. There is one more rule for defining margins. The Swedish railway network has been divided into nodes and between every node there should be some extra margins. Those nodes on the Southern mainline are: Stockholm, Mjölby, Alvesta and Malmö. Between every node there has to be at least four minutes for X2000 and three minutes for other passenger trains. How to distribute these three-four minutes between the nodes depends on the timetable constructor and his/her experience and opinion.

There exists also a guideline document from 2000 (referred to as TF601) which as an example suggests that:

- If the train is running on a track other than the main track, one or two minutes can be added to the running time.
- After a stop with passenger interchange, one minute can be added to compensate for irregularities in the dispatching.

Prior to every stop where punctuality is of importance, the running time is rounded up to full minutes. This provides some margins before every station that can be used for recovering from delays. In addition to above mentioned margins, The Swedish Transport Administration is using a three percent driver allowance when calculating the running times to compensate for different driver behaviour. The calculated running times are therefore three percent longer than the theoretically shortest running times. This margin cannot be manipulated and is the same in each timetable.

To conclude, the timetable construction relies to a large extent on tacit knowledge and previous experiences. A timetable can therefore have large variations depending on which timetabling constructor that have made it.

2.4 Related work

Robustness aspects of Swedish railway operations have been addressed previously. Mattsson [11] considers the reliability in the railway system and its vulnerability to various kinds of disruptive events. He separates the delay in a scheduled and a non-scheduled part, where passengers and carriers put a much higher negative value on unscheduled delay. Furthermore Mattsson notes that the objective to maintain reliability should be related to the number of passengers and to the value of delivering the goods on time. This weighting strategy aligns with the overall governmental objective for the transportation sector, which is to "ensure economical efficiency and sustainability in the transportation system to citizens and industry all over the country" [15].

Lindfeldt [10] studies how to place crossing stations as to minimize the expected crossing delay, given some initial traffic perturbation. He also analyses different timetable patterns in a mixed traffic with a cyclic structure (Taktfahrplan). Of special interest, with respect to the up-coming deregulation of the traffic, is how primary delays affect other trains. Lindfeldt has used the standard margins, which are currently used by the Swedish authorities.

Sipilä [14] uses simulation as a method to study how changes in the timetable infect the punctuality of X2000 in the Swedish Western mainline. Four minutes are added to or subtracted from the margins in the timetable used 2009. The disturbances in the model are based on real delays along the train run and on stops with passenger transfer. Sipilä has also studied the effect of corridors for X2000 with at least five minute distance to all other trains during the run.

The Swedish Institute of Computer Science, SICS [5] is running a project about timetabling. The idea in this project is to increase flexibility by loosening the timetable constraints and postponing the detailed timetabling to an operational level (i.e. shortly before the departure). The problem addressed is that of incomplete information used in the timetable construction process. The appearance of a given timetable is a result of a long construction process, where typically several trains later will be cancelled, or will have an adjusted speed profile (due to type of engine, number of wagons etc.). In the end, crossings and overtakings are not efficiently placed. With the increased flexibility, the margins added for robustness are available in a larger geographical area. The work is, of course, of larger interest for cargo trains with less number of non-avoidable stops than passenger trains have.

Törnquist Krasemann [1] focuses on re-scheduling railway traffic during disturbances and has developed a greedy heuristic. The heuristic computes the consecutive net delay if trains compete for a slot considering the margins available for potential recovery, and gives priority (if feasible) to the train which benefits most (i.e. suffers less) from getting access to the track.

Robustness in timetabling is, of course, of academic interest also abroad. The following overview is delimited to the most recent publications.

Delorme et al. [3] are studying the stability of the timetable at station level. They develop an optimization model which is based on delay propagation and use a shortest path problem resolution. Their model can be used for optimizing and evaluating an existing timetable.

Cicerone et al. [2] analyse the recoverable robustness timetables, that is, to what extent it is possible to absorb small delays by, if necessary, applying given limited recovery capabilities. The quality for the robustness in the timetable is computed as a ratio between the cost of the recoverable robust timetable and that of a non-robust timetable. Their event-based model is generic, with respect to the structure of the timetable.

Khan and Zhou [6] have formulated a stochastic optimization model for slack-time allocation and develop a heuristic framework in which they decompose the timetabling problem into one single problem for each individual train.

Fischetti et al. [4] evaluate four different methods to improve the robustness of a train timetable, given as an optimal solution to a problem accounting for the efficiency of the infrastructure usage. Minimising the cumulative delay over all events and scenarios seems to give the best robustness. The model is, according to the authors, very time-consuming to apply for realistic instances. Some experiments with simplified versions of their model have, however, given promising results for large-scale instances.

The models proposed by Kroon et al. [7] and Liebchen et al. [9] differ from the previous discussed models in the sense that they assume a cyclic timetable. Both models are tested for real-world applications in the Netherlands, and Germany, respectively. Kroon et al. minimise the average weighted delays of the trains, whereas the objective proposed by Liebchen et al. also accounts for transfer passengers.

3 Case Study and Robustness Analysis

Since the number of trains is increasing continuously and dependencies become more difficult to overview, there is a need to perform an in-depth investigation of the performance, in relation to the prerequisites given by the timetable (e.g. available margins and their location, critical train dependencies). The focus in this paper is on robustness which serves to limit the effects of more frequent and smaller delays and disturbances in the railway traffic.

We have studied the traffic on the nearly 630 km long Swedish Southern mainline, a double track line between Malmö and Stockholm, see Figure 2 above. This is one of the most important railway lines in Sweden and, via the bridge connection from Malmö to Copenhagen in Denmark, also of international interest. Long-distance trains are connecting the important end point markets in Malmö/Copenhagen and Stockholm. The twin cities Linköping and Norrköping, some 200-250 km South of Stockholm are the two largest cities passed along the line. There are also a large number of connections at major junctions in Hässleholm, Alvesta and Nässjö. Connecting possibilities of various importances are also offered in Lund, Eslöv, Mjölby, Linköping, Norrköping, Katrineholm, Flen and Södertälje. Between Katrineholm and Stockholm the line connects with the Western mainline (Gothenburg-Stockholm), which is not included in this case study.

The traffic on a typical weekday on the Southern mainline consists of 13 fast trains of type X2000 in each direction operate between Stockholm and Malmö/Copenhagen at 200 km/h, two long-distance trains in each direction with the maximum speed of 160 km/h. Interregional trains operate on parts of the line; North of Katrineholm, between Katrineholm and Linköping and between Malmö and Alvesta. Three commuter train systems are located on this line; Malmö-Höör (Skånetrafiken), Tranås-Norrköping (Östgötatrafiken) and Gnesta–Stockholm (Storstockholms lokaltrafik). The commuter train services have cyclic timetables and typically with 20-40 minutes intervals and have a maximum speed of 140-180 km/h.

3.1 Train selection

The focus of this case study is on long-distance passenger train services that operate between Stockholm and Malmö. Today there are 15 trains per direction and day that are roughly divided into three categories:

- X2000 regular (11 southbound trains and 13 northbound trains)
- X2000 fast (two southbound trains)
- IC (two trains in each direction)

X2000 operates according to one of two possible traffic structures: The first includes stops at all large stations (regular) while the second only allows stops at a few stations (which we refer to as *the fast X2000*). The fast X2000 runs twice a day in one direction only, from Stockholm, in the afternoon. The selected train services of every category are shown in Table 1. 201 and 202 are IC services and 503 is a fast X2000 service. The remaining services are regular X2000. During two weeks in October, traffic data concerning these trains collected (i.e. Monday to Friday, 4th to 15th of October, 2010). We will refer to these as *test week 1 and 2*.

	Southbound trains					Northbound trains			
	201	503	521	537	539	202	524	530	534
Departure time Stockholm	08:25	17:06	06:21	14:21	15:21				
Departure time Malmö						08:17	06:14	09:14	11:14
Travel time	5h 21m	4h 10m	4 h 25m	4 h 25m	4 h 25m	5h 18m	4 h 24m	4 h 25m	4 h 25m
Stop	Söö, Nr, Lp, N, Sä, Av, Hm, Lu	Lp, Hm, Lu	Söö, Nr, Lp, My, N, Av, Hm, Lu	Flb, Nr, Lp, My, N, Av, Hm, Lu	Flb, Nr, Lp, N, Av, Hm, Lu	Lu, Hm, Av, Sä, N, Lp, Nr, Söö	Lu, Hm, Av, N, Lp, Nr, Flb	Lu, Hm, Av, N, My, Lp, Nr, Söö	Lu, Hm, Av, N, My, Lp, Nr, Söö
Margins (minutes)	20	13,1	17,9	19,5	21,2	24,6	17,4	17,5	18,9
Punctuality (test week 1 and 2)	90%	40%	60%	70%	50%	80%	70%	70%	70%

Table 1. Selected trains and data. Data were collected from the timetable for 2010, and punctuality statistics were collected during test week 1 and 2.

The fast X2000 has the shortest travel time between Stockholm and Malmö, only four hours and ten minutes. Regular X2000 has a travel time of four hours and 25 minutes while IC has a travel time of approximately 5 hours and 20 minutes.

In the next three figures the margins and actual delays for test week 1 and 2 are shown. Trains with a larger delay than 20 minutes are excluded since they are of less interest in this case study. The margins today are designed so that trains can recover from approximately 20 minutes delays. It is not realistic to have larger margins because it increases the travel time too much.

For practical reasons, the figures only contain southbound trains. Figures for trains in direction Malmö – Stockholm are shown in Appendix 1. In these figures the delays are shown as positive numbers for an easy comparison with the margins.

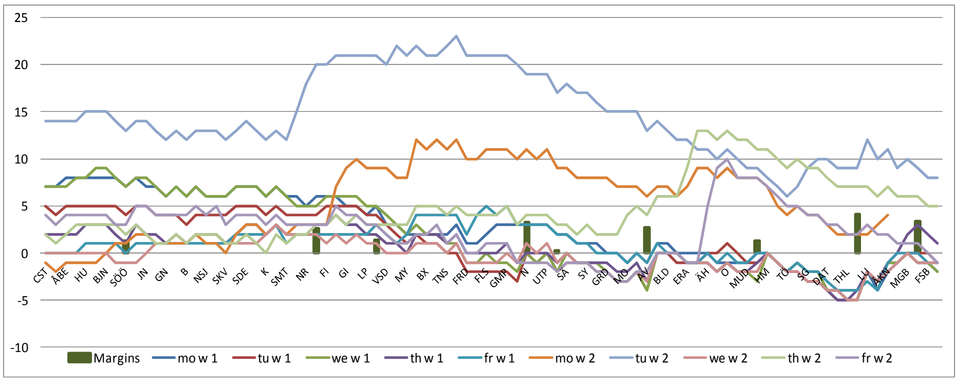


Figure 5. The IC train 202; margins and deviation from its timetable at different locations along its journey from Stockholm (Cst) to Malmö (M). Statistics were collected during test week 1 and 2.

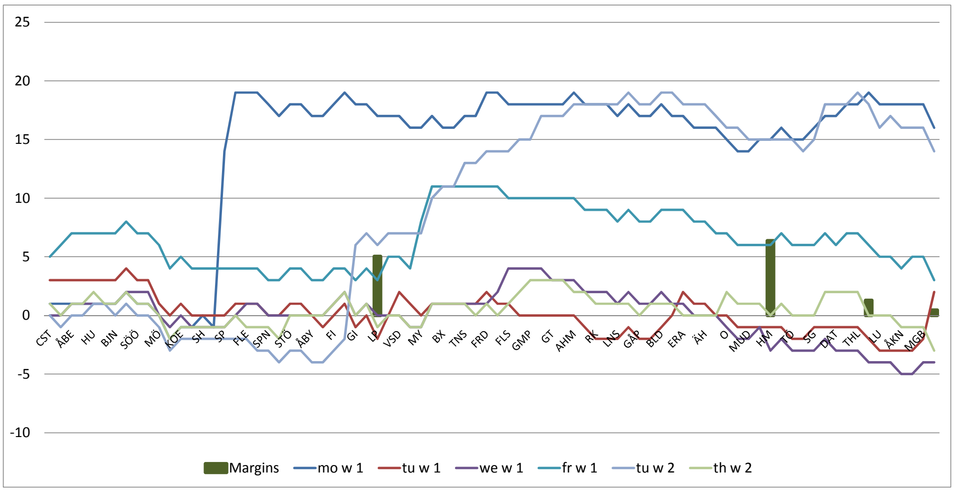


Figure 6. The fast X2000 service 503; margins and deviation from its timetable at different locations along its journey from Stockholm (Cst) to Malmö (M). to Statistics were collected during test week 1 and 2.

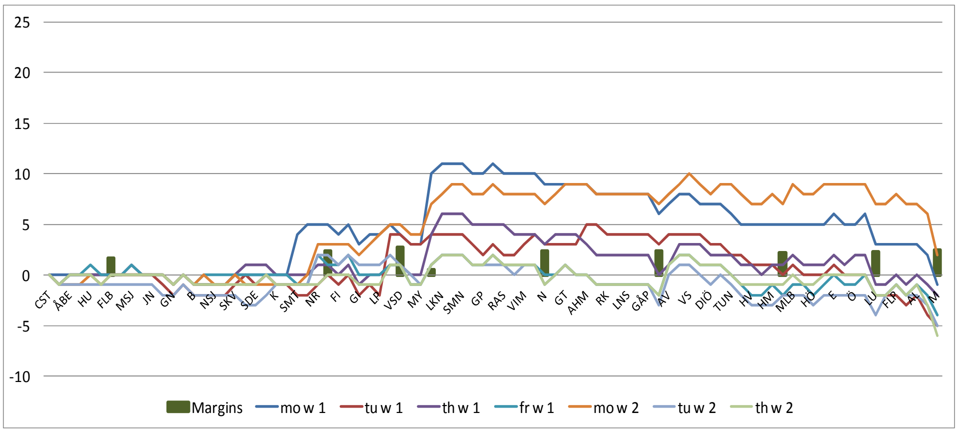


Figure 7. The regular X2000 service 537; margins and deviation from its timetable at different locations along its journey from Stockholm (Cst) to Malmö (M). to Statistics were collected during test week 1 and 2.

3.2 Two by two comparisons

In this part, delay recovering and margin utilization are studied. Today there is seldom more than 20 minutes margin placed between Stockholm and Malmö. Therefore it is not possible for trains with a larger delay than 20 minutes to fully recover. Trains suffering from severe disturbances will therefore not be included in the analysis, according to the reasoning in part 3.1.

One way to get a good picture of how the margins fill their purposes is to compare the trains two by two. Three comparisons are made:

- Northbound X2000 regular – Southbound X2000 regular. Two trains of the same type with the same stopping pattern but with different directions. Compare Figure 7 and train 530 in Appendix 1.
- IC – X2000 regular. Two trains in the same direction but of different type. Compare Figure 5 and Figure 7.
- X2000 regular – X2000 fast. Two trains of the same type in the same direction but with different stopping pattern. Compare Figure 6 and Figure 7.

The collected data is reported at each intermediate station with a one-minute precision, which explains the discrete steps in the graphs. In reality, the difference might be only a few seconds. The margins, on the other hand, are assigned with a precision in seconds. Therefore the trains can seem to recover delays that are larger than the available margins.

Northbound X2000 regular – Southbound X2000 regular

The most significant difference between two trains of the same type in different directions is the margins and their location Southbound X2000 services have some more margins than northbound, see Table 1. Southbound services have most of the margins in the second half of their journey. Southbound trains being delayed during the run therefore have better possibilities to recover before they reach Malmö. In Figure 7 all trains, even the ones that have a intermediary delay of 10 minutes, are on time in Malmö. Northbound trains, on the other hand, have their margins in the first half; see e.g. train 530 in Appendix 1. They can recover from delays in the beginning of the journey, but if the delays remain or occur when passing Östgötatrafiken commuter traffic system, the probability to arrive in Stockholm on time is small. Train 530 is delayed several days within or after this section and it has more difficulties to recover and arrives late in Stockholm.

According to data from test week 1 and 2 most of the delays occur nearby the Östgötatrafiken commuter system, in Alvesta (AV) and Hässleholm (HM) and this can be seen for trains in both directions. See Figure 7 and Appendix 1. Compared to how the margins are placed for southbound trains, southbound trains have better possibilities to recover from the delays.

IC – X2000 regular

X2000 and IC have different speed restrictions between Malmö and Stockholm. Regular X2000 services depart every hour according to a cyclic timetable. The few IC services that run between Malmö and Stockholm are placed to meet the largest travel demand and when there is room in the timetable. As a consequence, the southbound and northbound IC services have much variation in the margins.

The collected data for IC shows that delays occur on several places along the line. Delays for X2000 services occur more frequently on some places. This is easily shown in the graphs for train 202 and 530 in Appendix 1. For 530 almost all delays are concentrated between Linköping and Stockholm while the delays for 202 are more evenly distributed along the line. With this information it should be easier to place the margins in a suitable way to cover for the delays. Overall, IC services have a higher punctuality than X2000 services. The IC services have more margins and the margins also seem to be used in a better way.

The main reason why IC services are recovering better than regular X2000 services is that IC services have most of the margins placed after the locations where the largest delays occur. X2000 services have much of their margins placed in a way that they are restricted to be used before the significant delays occur.

In addition, X2000 services have a large margin between Stockholm Södra (Sst) and Stockholm Central (Cst) and it is not possible to recover from all delays in the last stretch before Stockholm central. IC services have at lot of margins all the way from Södertälje (Söö) to Stockholm central which increases the chance to arrive in Stockholm on time.

X2000 regular – X2000 fast

It is interesting to investigate the difference between two train services of the same type but with different stopping patterns. To enable an X2000 service to run faster between Stockholm and Malmö compared to the regular X2000, both stops and margins are removed. This has shown to influence the punctuality significantly. During test week 1 and 2, only 40 % of the fast X2000 services were on time in Malmö. If a fast X2000 service becomes more than five minutes delayed on-route, it seems very difficult to make it arrive at the final destination on time. If we compare train 503 on Friday week 1 and Tuesday week 2 in Figure 6 with train 537 on Monday week 1 and Monday week 2 in Figure 7, they all are approximately ten minutes late in Mjölby (My). However, after Mjölby their delays start to differ. The 537 trains can recover from the delays and be on time in Malmö. The 503 trains on the other hand have more problems with recovering. On Friday week 1, the 537 can recover quite well but the delay could just as well grow as for the train on the Tuesday week 2.

3.3 In-depth single train analysis

In order to investigate further how the timetable prerequisites may affect the performance the trains, a single train has been studied in more detail. The selected train is 537, a regular X2000 service from Stockholm to Malmö. The margins for 537 are placed as shown in Figure 8. If the train is on time to e.g. Norrköping, it cannot use the unused margins later on the run. If it is to early, the train have to wait and leave Norrköping on time for its departure.

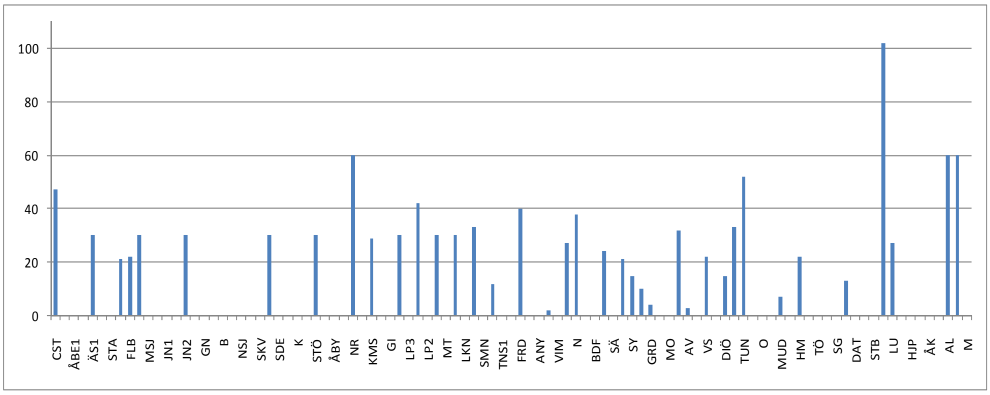


Figure 8. Placement of the margins for regular X2000 service 537 (seconds)

In Figure 9 all trains that with some kind of dependencies to train 537 are shown. Arrows pointing left, towards the train, represent the first trains after 537 that will be affected if 537 is delayed. Arrows pointing right, away from the train, represent connections that cannot be kept if 537 is delayed. As in previous comparisons, only small delays are interesting, and hence, only trains that are affecting or affected within 20 minutes are therefore considered. The length of the arrows represents how delayed train 537 can be at the given point on the line before the dependencies become critical (given by the x-axis). That is, if 537 has a delay smaller than 10 minutes when leaving Norrköping (Nr), then train 8765 is not affected.

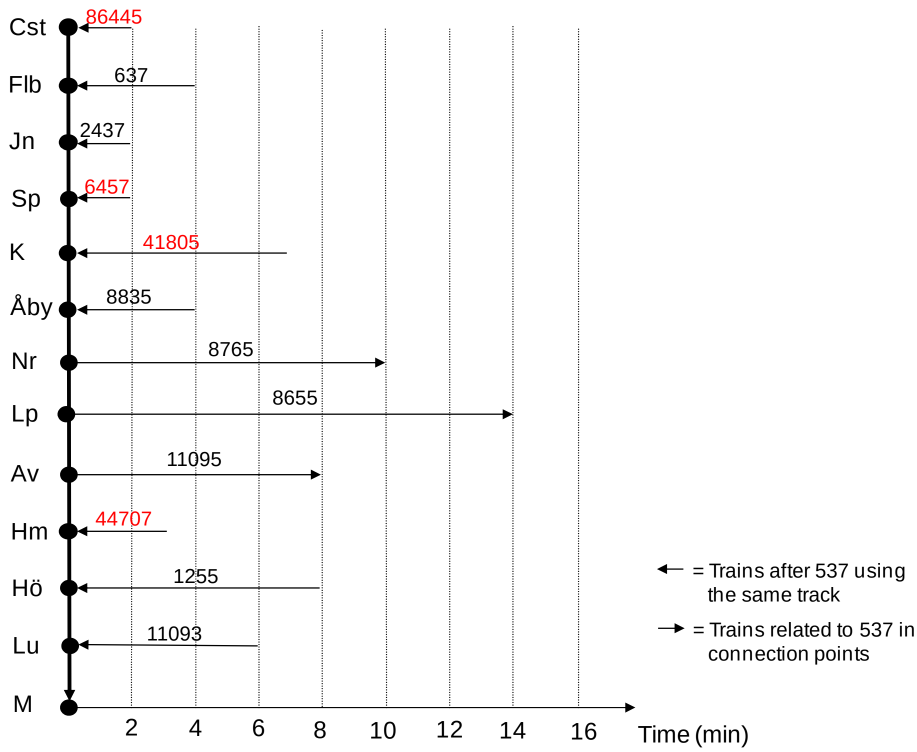


Figure 9. Trains that are affecting or affected by regular X2000 service 537.

Trains with red numbers are freight trains that are running after 537. They have a considerably lower speed than X2000. If 537 is delayed too much the freight train will get in front of it and 537 will lose more time, until an overtaking can take place. For the two test weeks when 537 was studied close, this problem rarely occurred because the freight trains very seldom were on time. If there is a conflict between a freight train and a few minutes late 537 the train dispatcher often chooses to let the freight train wait for 537 to pass. This gives the freight train some delay but it has more margins than 537. This overtaking is often planned to take place somewhere on the stretch and with a delayed X2000 service it practically means that the planned overtaking just take place somewhere else.

The conflict between a delayed X2000 service and a slower passenger train also often occurs, e.g. commuter and interregional trains like 1255 and 11093 in Figure 9. If 537 is just a few minutes delayed, the slower train waits to let the X2000 service pass. The slower train will use its own margins to recover. However, if 537 is significantly delayed or if the slower train will become further delayed due to waiting, the train dispatcher will let it go in front of 537 and increase the delay for 537 even further.

One matter of interest is to arrive on time to the intermediate connection points. We can see in Figure 9 that in both Norrköping (Nr), Linköping (Lp) and Alvesta (Av) there are connecting trains within 14 minutes. If 537 is more than 8-14 minutes delayed, the connecting train will leave before 537 has arrived. Passengers will probably not make their connections also for delays of 6-12 minutes, since the transfer time could be a few minutes. Then the train dispatcher has to make a consideration whether to let the connecting train wait and also become delayed, or let it depart on time and let the passengers from 537 wait for the next connecting train.

4 Discussion and Conclusions

The empirical analysis of how the margins in the Swedish railway traffic timetable are used in practise to recover from delays has provided several interesting insights. The data collection was carried out during two representative weeks in October of 2010 where no major incidents or any other special circumstances occurred and affected the behaviour of the studied traffic significantly.

Despite that normal conditions were ruling during this time period, the system gives a slight instable impression which supports the hypotheses provided by practitioners and researchers about the correlation between a high capacity utilisation and vulnerability, see e.g. [11]. Furthermore, a computation of the performance of the trains according to the definition of punctuality used by the Swedish Transport Administration (i.e. arriving at the final destination with a delay of five minutes or less, see Table 1) shows that northbound trains perform better than the southbound. An in-depth analysis, however, provides opposite conclusions since the data contains all delays and makes no distinction between a smaller and larger delays, i.e., if train A arrives six minutes delayed every Monday while train B arrives with a delay of 25 minutes, their reported performance with respect to punctuality will be the same. Hence, there should be a distinction when collecting statistics between severe delays and smaller, yet significant delays.

The results from the in-depth analysis also indicate that:

- Southbound services have their margins placed in a better way than northbound services. They have better possibilities to recover.

- IC services can recover from delays better than X2000 services. Their margins are larger, and placed in a better way.

- The fast X2000 services have too small margins and these are not effectively located in the timetable. These trains are delayed frequently and when so, they have difficulties in recovering. Only 40 % of service 503 was on time during the two test weeks.

- X2000 services are affecting and are affected by several trains on-route. A few minutes delay results in a train dispatching decision which can be either in favour of 537 or not. The freight trains are of less concern since they are seldom on time. X2000 services often have higher priority and can overtake the freight trains somewhere on the line. A conflict with other slower trains, e.g. commuter trains, is more difficult to manage. The train dispatcher has to decide which train that will become additionally delayed; the slow train because of an overtaking, or the X2000 train, which has to run after the slow train and not be able to run at maximum speed which results in a longer running time than planned for.

We have also made some observations about the influence of the different dispatching decisions. The dispatcher cannot know what will be the best decision for the future situation since this can change from time to time. The most satisfying situation is to have all trains arriving on time to the stations they have passenger interchange or any other critical activity such as a crossing.

As pointed out earlier, data is currently only collected by the Swedish Transport Administration for trains with respect to how they arrive at their final destination. It is also interesting how punctual the trains arrive at every intermediate station where they have a passenger interchange. The margins should be placed in a way to improve the possibilities for arriving on time everywhere, not only at the end station.

In practice, it is clearly not easy to acheive robustness, or to make optimal use of it in operational mode. This is a well-known challenge in most European countries with dense networks, but the Swedish deregulation complicates the task further. With several train operators the Swedish Transport Administration have to consider not only how to fit the trains in the system but also how to construct the timetable coequal to all operators. In any case, the need to analyse how margins are used in practise and whether their effect corresponds to the anticipated one is evident. It is also evident that the margins need to be flexible and cost-effective since the need for margins is depending on the conditions of the railway network, which in turn varies with e.g. seasonal changes (cf. Fig. 1)

This study has generated insights and pinpointed the need to conduct such studies in more depth. The conclusions drawn will be useful input to the extended study we have planned to conduct on the new timetable for 2011. The aim is to develop potential timetable margin insertion strategies and evaluate them with the use of an optimisation approach in a controlled environment. An optimised use of margins gives a better understanding for how efficient they actually are. The margin insertion strategies will also be further discussed and evaluated.

5 Acknowledgements

This research was conducted within the research project “Robust Timetables for Railway Traffic”, which is financially supported by and conducted in co-operation with VINNOVA (The Swedish Governmental Agency for Innovation Systems), Trafikverket (The Swedish Transport Administration) and SJ AB.

References

[1] Banverket, Banverkets Sektorsrapport 2009, 2010.

[2] 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.

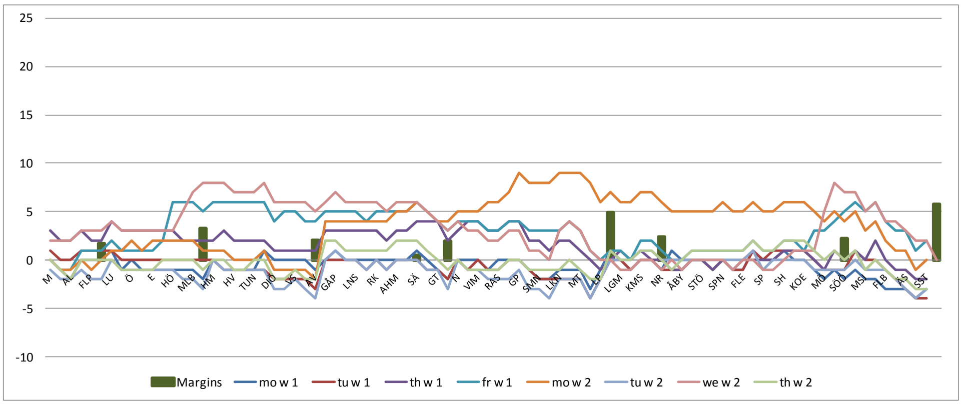
[3] Delorme, X., Gandibleux, X., Rodriguez, J., “Stability evaluation of a railway timetable at station level”, European Journal of Operational Research 195, 780–790, 2009.

[4] Fischetti, M., Salvagnin, D., Zanette, A., ”Fast approaches to improve the robustness of a railway timetable”, *Transportation Science* 43(3), 321–325, 2009.

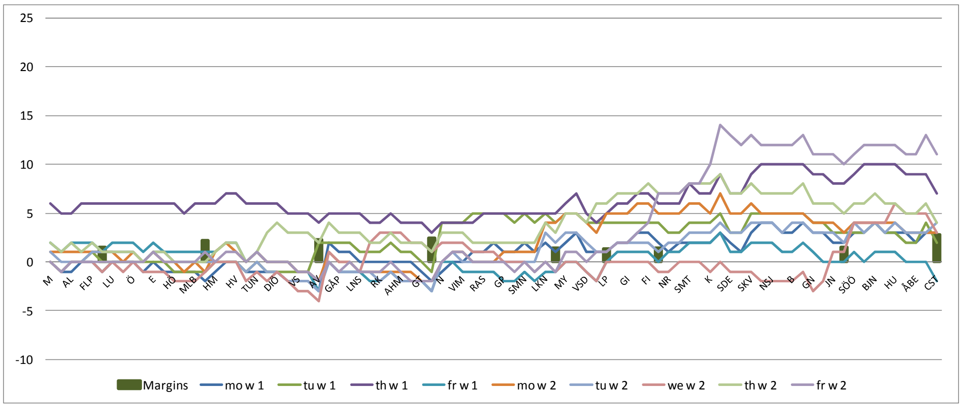
[5] Forsgren, M., Aronsson, M., Kreuger, P., “Tågtrafikplanering med successive tilldelning” (”Planning of railway operations with successive assignment”), SICS technical report T2009:04, Swedish Institute of Computer Science, Kista, Sweden, 2009.
[6] Khan, M.B., Zhou, X., “Stochastic optimization model and solution algorithm for robust double-track train-timetabling problem”, *IEEE Transactions on Intelligent Transportation Systems* 11(1), 81–89.
[7] 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.
[8] Liebchen, C., “The first optimized railway timetable in practice”, *Transportation Science* 42, 420–435, 2008.
[9] Liebchen, C., Schachtebeck, M., Schöbel, A., Stiller, S., Prigge, A., ”Computing delay resistant railway timetables”, *Computers & Operations Research* 37, 857–868, 2010.
[10] Lindfeldt, O., ”Railway operation analysis”, PhD Thesis, TRITA-TEC-PHD 10-001, Department of Transport and Economics, Royal Institute of Technology, Stockholm, Sweden, 2010.
[11] Mattsson, L.-G., “Railway capacity and train delay relationships”, in: A.T. Murray and T.H. Grubesic (Eds.), *Critical Infrastructure: Reliability and Vulnerability*, Springer-Verlag, Heidelberg Berlin, Germany, 129–150, 2007.
[12] 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.
[13] Salido, M.A., Barber, F., Ingolotti, L., “Robustness in railway transportation planning”, Proceedings of the World Congress on Intelligent Control and Automation (WCICA), Art. No. 4594481, 2833–2837, 2008.
[14] Sipilä, H., “Tidtabellsläggning med hjälp av simulering” (”Timetabelling by means of simulation”), Research report, TRITA-TEC-RR 09–007, Department of Transport and Economics, Royal Institute of Technology, Stockholm, Sweden, 2010.
[15] Swedish Government, “Mål för framtidens resor och transporter” (”Objectives for future travel and transportation”), Prop. 2008/09:93, in Swedish, 2009.
[16] Swedish Railway Authoritites, Punctuality Statistics. Published monthly on: http://www.trafikverket.se/Foretag/Trafikera-och-transportera/Jarnvag/Folja-upp-trafik/Punktlighetsstatistik/ (in Swedish).
[17] Törnquist Krasemann, J., “Design of an effective algorithm for fast response to the re-scheduling of railway traffic during disturbances”, *submitted for publication*, 2010.
[18] UIC, “Capacity”, Leaflet UIC Code 406, 1st Edition, 2004.
[19] Yoko, T., Norio, T., “Robustness indices for train rescheduling”, presentation at: *1st International Seminar on Railway Operations Modelling and Analysis*, Delft, The Netherlands, June 8–10, 2005.

Appendix 1

Graphs for a northbound X2000 and IC services during test week 1 and 2.



The IC service 202; margins and deviation from its timetable at different locations along its journey from Malmö (M) to Stockholm (Cst). Statistics were collected during test week 1 and 2.



The regular X2000 service 530; margins and deviation from its timetable at different locations along its journey from Malmö (M) to Stockholm (Cst). Statistics were collected during test week 1 and 2.