Postprint

N.B. When citing this work, cite the original published paper.

Validation of a simulation model for capacity evaluation of double-track railway lines

Anders Lindfeldt

Royal Institute of Technology, Department of Transport Science,
100 44 Stockholm, Sweden, e-mail: anders.lindfeldt@abe.kth.se

Abstract

Capacity is partially a matter of quality versus quantity. Increasing capacity utilization gives more scheduled waiting time in the timetable as well as more delays when trains are operated. A common way to analyse capacity is to first make a timetable and then use it in a simulation where trains are perturbed. This way it is possible to establish the scheduled waiting time and the expected delays of the trains at the same time as market demands on traffic, e.g. frequency of different train services, are considered. However, setting up simulations are often time consuming and requires a lot of input data.

A method is developed with the objective to improve and speed up capacity analysis of double track railway lines. It can be used for both timetable generation and timetable simulation. The focus of this paper is to present the model and validate that it can simulate timetables accurately. The model has a macroscopic representation of the railway infrastructure and rolling stock while timetables and perturbations are modelled in more detail. When resolving conflicts, the dispatching algorithm can consider several trains and take into account how the dispatching decision affects the possibility to make good decisions at the next station.

The validation is performed using the micro simulation tool RailSys as reference. The validation covers a wide range of scenarios of different traffic density, timetable heterogeneity, inter station distance and primary delays. The validation shows good results and the conclusion is that despite the simplicity of the model, it is accurate enough to use in capacity analysis.

Keywords

Railway, simulation, validation, delay, capacity

1 Introduction

When capacity utilization increases several things happens. On the tactical level, more trains imply more restrictions when the timetable is constructed, especially if it is a single track line or a double track line with a heterogeneous traffic mix. As traffic becomes denser scheduled waiting time, i.e. longer planned travel times, are unavoidable when more frequent crossings and overtakings are necessary. On the operational level, trains scheduled closer together increase the risk of secondary delays (knock-on delays) at the same time as it may also become more difficult to recover from delays.

When trying to describe and measure a complex phenomenon such as capacity, it is important to have a good measure that considers all relevant aspects of the property it is trying to describe, or it risks being less useful. Besides having a good definition, it should preferable also be easy to measure. One definition of maximum capacity is when the point is reached where the cost of scheduling one more train slot becomes higher than the

benefit. Comparing costs and benefits can be achieved by, for example, applying socio economics to put monetary values on delays and frequency of service etc. Trafikverket (2014). Another definition that is commonly used in practise is to restrict the number of train paths so that e.g. punctuality does not drop below a certain limit. More allowances can be introduced in the timetable to reduced delays and improve punctuality Sipilä (2010), but making travel times longer on the other hand, is not good from a market perspective when for example trains are competing with other modes of travel.

Abril et al. (2007) give a good introduction to the concept of capacity and how it is affected by different properties including infrastructure, traffic and operating parameters. They also give different methods how it can be calculated, including analytical methods, optimization methods and simulation methods. They state that while analytical and optimizing methods can be used to estimate theoretical capacity, simulation is an attractive way to obtain practical capacity measures where the trade-off between capacity and reliability, i.e. delays, can be included. However, there also exists a lot of work developing efficient methods for making optimal timetables with respect to both train running times and measures of timetable robustness, e.g. Fischetti (2009).

Salido et al (2012) continues the work presented in Abril et al (2007) and develops both fast analytical methods and a simple rescheduling algorithm to estimate the impact of a short disruption on secondary- and total delay. Several factors are varied including traffic density, heterogeneity and allocation of running time allowance. However, primary delays are modelled in a simplified manner as only one primary delay is inserted at a time. The methods are good to establish how fast a timetable can recover from a disruption, which is indeed one definition of robustness, but is not designed to analyse real life operation with multiple disturbances.

In the UIC code 406 (2004), a methodology for capacity evaluation is developed with the intention of creating an international standard. The fundamental feature of the proposed method is to compress an existing timetable, in time, to calculate the infrastructure occupation time. By adding buffer times to the infrastructure occupation time timetable stability is ensured and capacity consumption achieved. The analysed railway line is divided into appropriate sections for which the compression is made separately. The sections are chosen to match changes in traffic patterns and infrastructure standard. If the compression indicates free capacity, it should be attempted to enter a new train path to determine if the unused capacity can be used or if should be considered as lost. It is emphasised that capacity depends on the type of traffic on the line, and that the results only are valid for the specific timetable analysed. However, by analysing an existing timetable, market needs are considered. Lindner (2011) discusses the applicability of the UIC code 406 method for evaluating capacity. A question that is raised is how the enrichment process, where additional train paths are entered into the timetable if possible, should be done. More train paths can be added if paths are allowed to be bundled, which however might not be acceptable from a market perspective. UIC 406 is also analysed in depth in Landex (2008). He also discusses how socioeconomic utility depends on the amount of supplements in the timetable, i.e. scheduled waiting time. Because operational delay has a higher time value than scheduled waiting time, there exists an optimal amount of supplements that maximize the socioeconomic utility.

In Lindfeldt (2009) a combinatorial model, TVEM, is developed. By varying starting times of the different train patterns, a large set of timetables are created using a sequential scheduler. The first steps are to schedule cyclic train services, normally passenger trains. In the final stage, the remaining capacity is used to schedule as many train paths as possible, normally freight trains. The quality of the train paths is controlled by limiting the

scheduled waiting time. A similar approach is presented in Meng (2013) where a rolling horizon optimization algorithm is used to measure capacity by iteratively schedule more freight trains in a timetable with passenger trains.

Sogin (2011) analyse the effect traffic density and heterogeneity on a single track freight network. The analysis is performed with a micro simulation software called Rail Traffic Controller, RTC, and the measure of performance is delay of the freight trains in min per 100 train miles. The delay includes both times for meets and passes, i.e. they are not planned in advance, and are calculated by RTC. Hence, it is not possible to distinguish between scheduled waiting time and operational delay. In the paper this is not a problem since the focus lies on freight trains that are not operated according to a timetable.

Lindfeldt (2010) uses the experimental design Latin Hypercubes to perform a simulation study where response surface models are used in the evaluation of additional operational delay. Factors in the experiment are distance between overtaking stations, speed and frequency of train services as well as primary delay levels. In a similar study, Lindfeldt (2011) combines a sequential timetable generator with the micro simulation software RailSys Radtke (2001) to perform an extensive full factorial design experiment. Scheduled waiting time and operational delay are evaluated as functions of distance between overtaking stations, traffic density, timetable heterogeneity and primary delays.

Sipilä (2014) uses RailSys to generate and simulate timetables on a single track railway line. There is no native support in RailSys to quickly generate several timetables. However, the paper presents a method where RailSys simulations are used to generate timetables. This is achieved by systematically designing entry delays in such a way that they give trains their desired departure times. No other primary delays are applied in the simulation and the results are then used as timetables in a second step where they are simulated with stochastic primary delays.

It is clear from the literature review that railway capacity is a concept widely used but hard to define because of the great complexity of railway operation. Many studies are focused on either scheduling or delay estimation, but few combine them and try to define capacity. Optimization methods are used to calculate optimal timetables with respect to both train running times and robustness. However, to calculate operational delays of close to real life scenarios, simulation seems to be the only viable method. Many simulation studies on the other hand either simulate rail traffic operated without a timetable, which is less useful when passenger traffic is analysed, or rely on timetables being provided as input, which often requires time consuming manual work. This often limits the number of timetables that are used and since the timetable is absolutely crucial to train operation, it is hard to make a general statement about the capacity.

This paper presents a new model for simulation of timetables on double track railway lines. There are already many different simulation tools available for this purpose, but the intention of the model presented here is that it will be integrated into a larger model for capacity analysis. The capacity model will first generate timetables and then simulate them. However, before the new capacity model can be put to use, a necessary step is to validate the simulation model, which is the focus of the present paper.

2 Method

The simulation model requires modelling of several sub modules, including infrastructure, rolling stock, timetable and primary delays. Figure 2 gives an overview of the simulation model with the sub modules and their key components.

The infrastructure model has a macroscopic representation and the only parameters required are the number of tracks at each station and the distances between stations. The number of tracks at stations is needed to determine how trains can overtake other trains. The inter station distance is not really used in the simulation but is needed when drawing graphical timetables.

Rolling stock is also modelled in a simplified manner with only a few parameters. Train running times as well as supplements for deceleration and acceleration is required to model the rolling stock as there is no dynamic calculation of train runs. It is also possible to define different minimum headway, arriving and departing stations, for different train types. Minimum headway is also dependent on the infrastructure, e.g. on signal block lengths, overlaps and station layouts. However, since these parameters do not vary in the fictitious infrastructure models used in this paper, all block sections have equal lengths, all stations have the same layout etc., it is fair to model minimum headway as dependent on train type only.
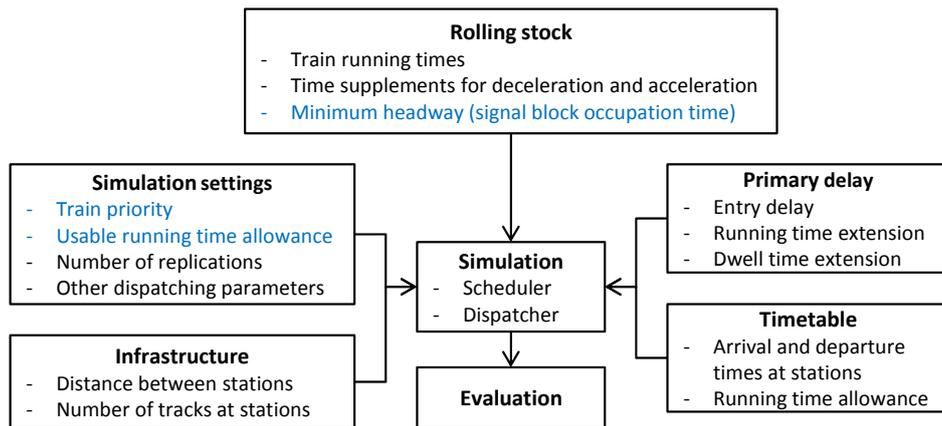
Figure 2: Input and key components of the simulation model. Blue parameters are subject to calibration in section 3.

Entry delays, running time extensions and dwell time extensions are modelled by stochastic distributions, empirical or theoretical. If a train receives an entry delay, it is added to the requested departure time at the origin. Running time extensions and dwell time extensions are applied to a train after its train path from a station to the next has been fixed, Figure 3. Stops with passenger exchange are modelled by a minimum dwell time and a scheduled dwell time. If trains are late, they can stop shorter time than scheduled but not shorter than the minimum dwell time. Dwell time extensions are added to the minimum dwell time. The way of modelling stops and dwell time extensions is the same as used in RailSys.

The actual simulation is made up by two main functionalities, a scheduler and a dispatcher. The scheduler calculates train arrival times at a station given desired departure

times at the previous station. Train running times are dependent on line section, train type and if the train is starting from a standstill at the previous station and if it is stopping at the next. Conditions that apply are that trains must arrive in the same order as they departed and that headway at departure and arrival is not smaller than the minimum headway. If minimum headway at departure is not satisfied, the departure of the later train is delayed. If trains are not arriving in the same order as they departed or if minimum headway at arrival is not satisfied, the speed of the later train is reduced.

The scheduler is used by the dispatcher to calculate delays for different possible scenarios. A more detailed description of the dispatching algorithm is presented in the next section. Other important model assumptions are listed below.

- The infrastructure is a double-track line where each track is exclusively dedicated to traffic in one direction.
- At junctions, only interaction with trains running on the same line is considered.
- Signalling and block occupations are modelled in a simplified manner as minimum arrival- and departure headways at stations.
- A timetable needs to be provided that can be used as reference when estimating delays.
- Trains with scheduled stops meant for passenger exchange cannot depart ahead of schedule. If a stop is only due to a scheduled overtaking, trains can depart ahead of schedule.
- Trains running ahead of schedule have lower priority than trains running on time or late.
- Trains can always use running time allowance to catch up delays. It can be used even if the train is ahead of schedule, i.e. if a train is early it can use running time allowance to arrive even more early at the next station.

2.1 Dispatching algorithm

In real life operation, trains almost always suffer from disturbances. It is the job of the train dispatcher to resolve conflicts that occur when trains deviate from the planned timetable. In synchronous simulation of timetables with disturbances, the functionality of the train dispatcher needs to be modelled, Siefer (2008). This is the task of the dispatching algorithm. The algorithm in this paper solves potential conflicts by evaluating the impact of changing train order at stations, i.e. overtakings, on travel time. This is done starting with the first station on the line progressing from the first to the last train arriving at that station. When all trains are scheduled at one station, the algorithm proceeds to the next. Thus it is not completely the same as a time-based synchronous simulation where conflicts are solved in chronological order. Figure 3 shows a flowchart of the procedure.

The parameter a decides how many trains are included when different passing solutions are calculated. In dense traffic, it may not be enough to just consider pairwise changing the order of trains since the decision is likely to affect trains following close behind. Including more trains makes for better solutions but increase computing time and a typical number is around 3-6 trains. Given the number of trains and the number of tracks at the station, all possible passing combinations are calculated. If the number of tracks are equal to or greater than the number of trains a, the number of possible passing combinations is $a!$. For each combination, delays are estimated for all trains and the combination with the lowest total weighted delay is chosen. Even if a solution includes several trains, only the first departing train is fixed in the timetable before the algorithm

proceeds one train ahead in time. By considering several trains and only fixing the solution of the first, a rolling horizon approach is implemented that that makes for good solutions.

There is also a need to consider more than one station at a time when the best passing combination is found, Sahin (1999). This is necessary for two reasons. First, if the number of tracks available for overtakings is limited, trains may have to be passed at one station in order to make it possible to overtake another train at the next. The second reason is more specifically connected to how the dispatching algorithm works. If the speed differences of different train services are low, the additional delay for scheduling an overtaking may always be higher than keeping the same train order. Since the algorithm only looks at the delays locally, a slightly faster train may get trapped behind a slower train for several line sections. The fundamental limitation remains, but is reduced if the cost of possible solutions at several stations further downstream is considered.

In the implemented algorithm, delays at the next adjacent station are taken into account when finding the best solution at the current station. In practice, this is done by first calculating all possible passing solutions O_1 at current station m and their associated delays $C(p_i)$, Figure 3 and (1) - (3). For each solution in O_1, all possible solutions O_2 and their delays $C(q_j)$ at station $m+1$ are calculated, hence it exists many possible solutions O_2 for each solution in O_1. The total delay for a solution at station m is defined as the sum of the delay at m and the minimum delay of the corresponding available solutions at station $m+1$. When the best solution is found, departure times at station m and arrival times as $m+1$ is fixed for the first departing train. The part of the solution corresponding to O_2 is not stored, thus creating a spatial "look-ahead" similar to the temporal "look-ahead" created by considering several trains, but only storing the solution of the first. Delays of different trains can be assigned different weights, thus favouring solutions with low additional delays for trains with high weight.

As described above, the algorithm does not detect conflicts, but does just evaluate the consequences of changing train order, regardless if there are any conflicts or not. This is computationally inefficient, especially for low density timetables with few conflicts. For this reason, a functionality that identifies conflicts is implemented that in many cases reduce the number of solutions it is necessary to calculate. After calculating the first solution in O_1, which corresponds to not changing train order at station m, it is checked if it has a delay equal to zero. If the delay is zero and there exist a corresponding C_2 that is also zero, there is no conflict at station m. It is then unnecessary to calculate the rest of the solutions in O_1 and their corresponding O_2. This feature does not show in Figure 3.

$$s = \min(a, n(N_m)) \tag{1}$$

Let x be a passing order of s trains at a station

$$d_k(x) = t_k(x) - u_k \tag{2}$$

$$C(x) = \sum_{k=1}^{s} w_k(x) \cdot \max(0, d_k(x))^v \tag{3}$$

m: *station number*

m_{max}: *total number of stations*

n_{max}: *total number of trains at station m*

N_m: *the set of all non scheduled trains at station m ordered according to their arrival time*

$n(N_m)$: *the number of trains in N_m*

a: *the maximum number of trains considered when solving conflicts*

$C(x)$: *total weighted delay for passing solution x*

w_k: *delay weight for train k in passing order x*

d_k: *delay of train k in the passing order x*

v: *weight to penalize large delays*

N_m^s: *the set of the s first trains in N_m*

t_k: *arrival time at station $m+1$ ($m+2$) for train k in passing order x*

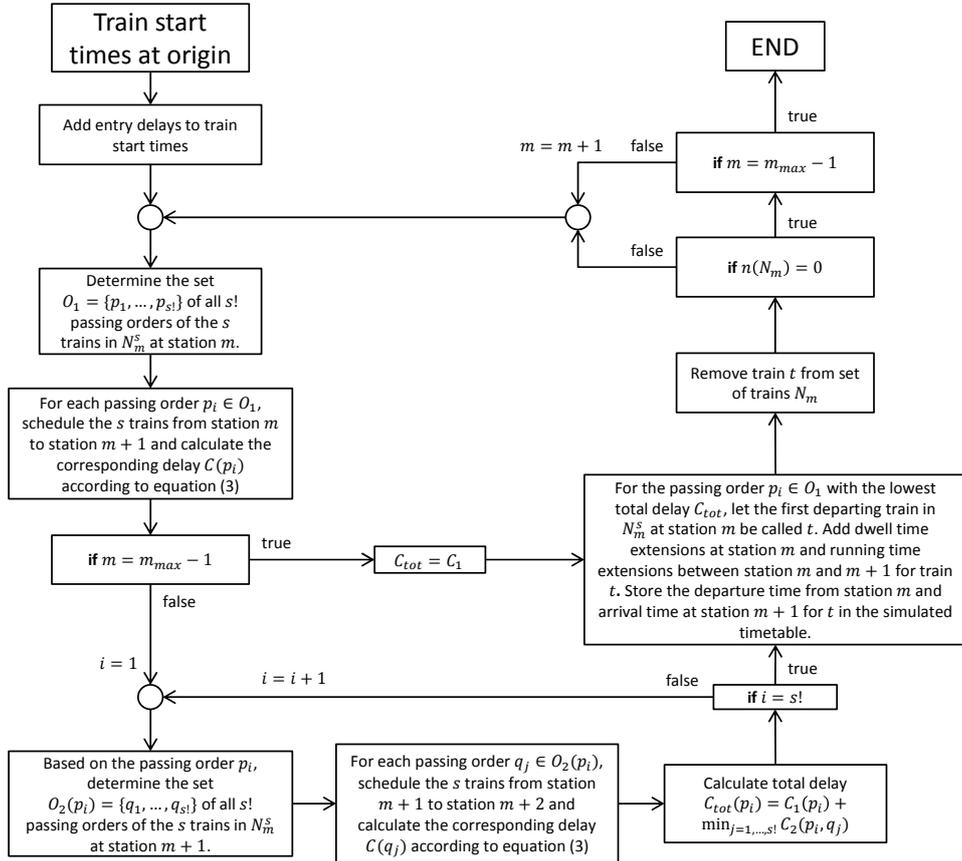u_k: *timetable arrival time at station $m+1$ ($m+2$) for train k*

Figure 3: Dispatching algorithm. Please note that there will only be *s*! possible passing orders if the number of tracks are equal to or greater than *s*.

2.2 Limitations

The developed simulation model has several limitations compared to a microscopic simulation tool. Many of them are due to the macroscopic modelling of the infrastructure and rolling stock. It is for example not possible to directly determine the effect of making block sections shorter or changing station design to allow for more parallel movements. However, many of these properties are implicitly taken into consideration in model parameters, and can therefore still be analysed to some extent.

Another limitation is that the dispatching algorithm cannot handle opposing traffic, hence it cannot be used to operate single tracks. The fundamental assumption that it is possible to solve all conflicts on one line section before proceeding to the next, is not valid if traffic comes from more than one direction.

Because the simulation is not purely time controlled, there is a limit to how large the stochastic primary delays can be without producing unrealistic model response. In reality, a large enough primary delay will affect decisions upstream, which is not possible in the model since all dispatching decisions for one line section is taken before moving on to the next. However, as long as a single running or dwell time extension is no longer than a few minutes, it should not influence the realism of the result. Another theoretical limitation is that the effects of an extreme bottleneck will not be transmitted upstream for the same reasons already mentioned.

3   Validation of simulation model

To determine how well the model can be used to simulate timetables with disturbances, a calibration and validation procedure is performed using the commercial simulation tool RailSys as reference. The advantage of using another validated simulation tool as reference, rather than real data, is that it allows for a more controlled environment and that it is easier to make many scenarios that can be used to validate the model response under different conditions. Figure 1 shows the validation procedure. The calibration procedure is the same but with a feedback loop via the "Calibrated parameters" box.
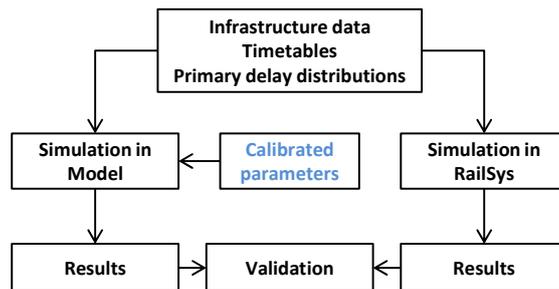

Figure 1: Validation workflow.

3.1 Setup

In Lindfeldt (2011, 2013, 2014) extensive simulation experiments on a double track are performed. A sequential timetable generator is created and used to generate hundreds of timetables that are simulated in RailSys in a Monte Carlo simulation. Four factors are varied in the experiment: traffic density, timetable heterogeneity (mix of train types), primary delay distributions and distance between overtaking stations. Simulation results are analyzed in terms of exit delays, secondary delays and how allowances in the timetables are used to catch up delays. The methodology is developed in Lindfeldt (2011) and applied in Lindfeldt (2014) to determine the effect of freight trains skipping scheduled stops and departing ahead of schedule. In Lindfeldt (2013) data from the simulations is used to determine how several different measures of heterogeneity can be used to explain secondary delays.

The wide range of operational conditions represented in these studies make them ideal to reuse in the validation of the new model. Timetables from the previous studies are imported into the new model and the same stochastic distributions are used to model primary delays. Table 2 summarize the experimental setup used for calibration and validation.

The infrastructure models consist of one track operated in one direction, thus mimicking the operation of a double track with assumed independency of traffic in different directions. Overtaking stations are spaced equidistantly. Stations are placed at 20 km intervals in the first infrastructure model and at 40 km intervals in the second. The timetables are defined as cyclic timetables of up to three trains per cycle and using up to three different train types: high speed, intercity and freight trains. Taking cyclicity into account, this makes in total 14 unique combinations, in this paper referred to as timetable variants. Each timetable variant has a different mix of train types and therefore different degrees of heterogeneity, table 2. Each timetable variant contains several timetables with different traffic densities.

Perturbations include three different types of delays, entry delay, running time extension and dwell time extension. All three types are varied coherently for two levels and the distributions are based on empirical data from real operation in Sweden Nelldal (2008). Entry delays and running time extensions are modelled with empirical distributions and dwell time extensions as log-normal. For each of the in total 280 scenarios, a timetable with 35 cycles is constructed. Each timetable is simulated with 80 replications to achieve stable results. Table 1 shows characteristics of the different train types.

Table 1: Train type characteristics.

Train type	High-speed	Intercity	Freight
Vehicle (Swedish designation)	X50	X60	RC4, 1000 tons
Top speed [km/h]	200	160	100
Average speed [km/h]	168	125	95
Total running time [min]	72	90	127
Number of stops	4	6	2
Train Length [m]	54	107	400

Table 2: Experimental setup used for calibration and validation. A full factorial design with four factors and 280 scenarios.

Inter-station distance		Timetable variant (mix of train types)		Traffic density	Perturbation level
[km]	Variant number	HS: high speed, IC: intercity, FR: freight		Start headway, % of minimum	Low, high
	1	100% HS			
	2	100% IC			
	3	100% FR			
20	4	50% HS, 50% IC		116	
	5	50% HS, 50% FR			Low
	6	50% IC, 50% FR		138	
	7	67% HS, 33% IC			
	8	67% HS, 33% FR		171	
	9	33% HS, 67% IC			
40	10	33% HS, 33% IC, 33% FR		223	High
	11	33% HS, 33% FR, 33% IC			
	12	33% HS, 67% FR		322	
	13	67% IC, 33% FR			
	14	33% IC, 67% FR			

3.2 Calibration

Before the validation can take place, the model is calibrated against a subset of the available data. The calibration data consists of all traffic densities in timetable variants 1-6 for the 20 km infrastructure and high primary delays. The reason that timetable variants 1 - 6 are used is that they are the least complicated. Timetable variant 1-3 consist of only one train type each, and timetable variant 4-6 of two train types, 50 % of each. High primary delays make the effects of increased traffic density and timetable heterogeneity more evident, which makes it easier to calibrate the model.

The objective of the calibration is to minimize the difference in mean exit delay when the model outcome is compared to results from RailSys. It is done in an iterative process where results are evaluated by mean of visual inspection (diagrams like the one shown in Figure 4). Three parameters are calibrated individually for the three types of trains: how much of the available running time allowance that can be used to recover delays, minimum headway between trains on arrival and departure from stations and train type priority, Figure 2.

The reason that available running time allowance is included in the calibration is that it might be too easy to recover from delays in a model that does not model the signaling system and the dynamic behavior of train breaking and accelerating. The minimum headway is a parameter that is set explicitly in the model, while in RailSys it is given by the microscopic infrastructure model in combination with the detailed vehicle models. Train type priority is explicitly defined in both the model and in RailSys, but used in different ways in the dispatching algorithms.

The strategy for the calibration is as follows:

1. Adjust the amount of available running time allowance that can be used to recover from delays. Use the simulations with the lowest traffic density of the timetable variants that consist of only one type of trains (timetable variant 1-3). Using timetable variants with only one train type makes the results independent of priority. Using the lowest traffic density makes the result less dependent on minimum headway between trains.
2. Use all simulations of timetable variants 1-3 to determine the minimum headway between trains of the same type. A longer minimum headway will make the exit delay increase faster with higher traffic densities, i.e. it controls the slope of the line representing the mean exit delay shown in Figure 4. Compare with step 1 that defines "the intercept". Using timetable variants with only one train type makes the results independent of priority.
3. Finally, use all simulations of timetable variants 4-6 to adjust the time weights, priorities, assigned to delays of each train type. In fact, it is the pairwise ratios between the weights that are important, and in theory it is enough to determine two of the ratios to also know the third. This makes the problem of finding the right priorities overdetermined, and it may be necessary accept a compromise to find a solution.

Table 3: Results of calibration.

	Freight	Inter City	High Speed
Usable running time allowance [%]	50	83	83
Minimum headway [s]	171/133	99/77	117/91
Priority (weight) [-]	1	10	50

In table 3, usable running time allowance refers to how much of the scheduled running time allowance that can be used to recover from delays in the model. In RailSys the running time allowance is defined as 6 % of the running time, and 100 % can be used to reduce delays as soon as a train is late. Calibration results show that the freight trains can use significantly less running time allowance than the passenger trains. It seems plausible that freight trains with low acceleration and braking performance have a harder time utilizing the running time allowance, compared to modern high performance passenger trains. However, the validation will show that another fundamental difference between the two models also affects the result.

In a conflict-free timetable the scheduled headway must by definition be greater than the minimum headway for all block sections. On double tracks with heterogeneous traffic, it is important to consider minimum headway at the first and the last block section between two stations, i.e. at departure and arrival at stations. Headway at departure determines how fast a train can depart after being overtaken by another train. Headway at arrival determines how close the overtaking train can be to the overtaken train.

The timetables used for calibration and validation in this study have the same minimum scheduled headway for all train types, 180 s on arrival and 140 s at departure from stations. The minimum headway, however, depends on many parameters such as, signaling, train braking performance, train speed, train length, etc. It is therefore expected that the calibration shows that the freight train have a significantly longer minimum headway than the passenger trains. It should be noted that these headways are estimated

between two trains of the same type, while in reality it depends on the combination of the train types of the preceding and the following train.

The calibration showed large difference in train type priority. This is not surprising since in the dispatching in RailSys is configured to mainly apply strict priority in this study, rather than minimize the total delay. It is also worth mentioning that priorities in RailSys is static in this setup, i.e. does not change when a train is late or early. Figure 4 shows how the calibrated model fits the reference. The greatest differences between the two models occur at the highest traffic densities, where the model does not show as large increase in mean and standard deviation of the exit delay as RailSys, further discussed in the next section.

3.3 Validation

The mean and standard deviation of the exit delay is used in the validation. Aside from the data set used for calibration, four other data sets are available for validation. The first consist of the remaining timetable variants of the setup with 20 km between overtaking stations and high primary delays that are not used for calibration. Compared to the calibration data set, the remaining timetable variants have a different heterogeneity. The second data set is based on the same infrastructure model but with lower primary delays. In the third and fourth data set, the infrastructure model is changed to one with 40 km between overtaking possibilities and primary delays are varied from high to low, respectively. No additional calibration is performed after the one described in the previous section is finished.

The data set used for calibration and the first set used for validation is studied in detail in Figure 4. It shows mean and standard deviation of the exit delays from both models. The fit is not perfect but still good and both mean and standard deviation are at the same levels and show approximately the same reaction to increased traffic density in both models. For some timetable variants, the mean and standard deviation does not increase as much with traffic density in the model as it does in RailSys. This behaviour was observed already in the calibration phase and is due that some trains get extremely high secondary delays in RailSys. However, it should be noted that the traffic is extreme at the highest traffic densities with overtakings scheduled at every station.

The reason for the somewhat irrational behaviour of the intercity trains, e.g. in timetable variant 7, is the timing of overtakings between the intercity trains and high-speed trains. Poor timing means that intercity trains are scheduled to wait at the station for a longer time before it is overtaken, which in turn reduce exit delay. Finally, it is interesting to observe the difference when the model is restricted to look only one station ahead when deciding on the best dispatching solution. Delays increase somewhat for high priority trains and decrease for low priority trains, especially at high density timetables. This effect is expected and means that the dispatching algorithm is less efficient if it only looks one station ahead instead of two when calculating the best dispatching solution.
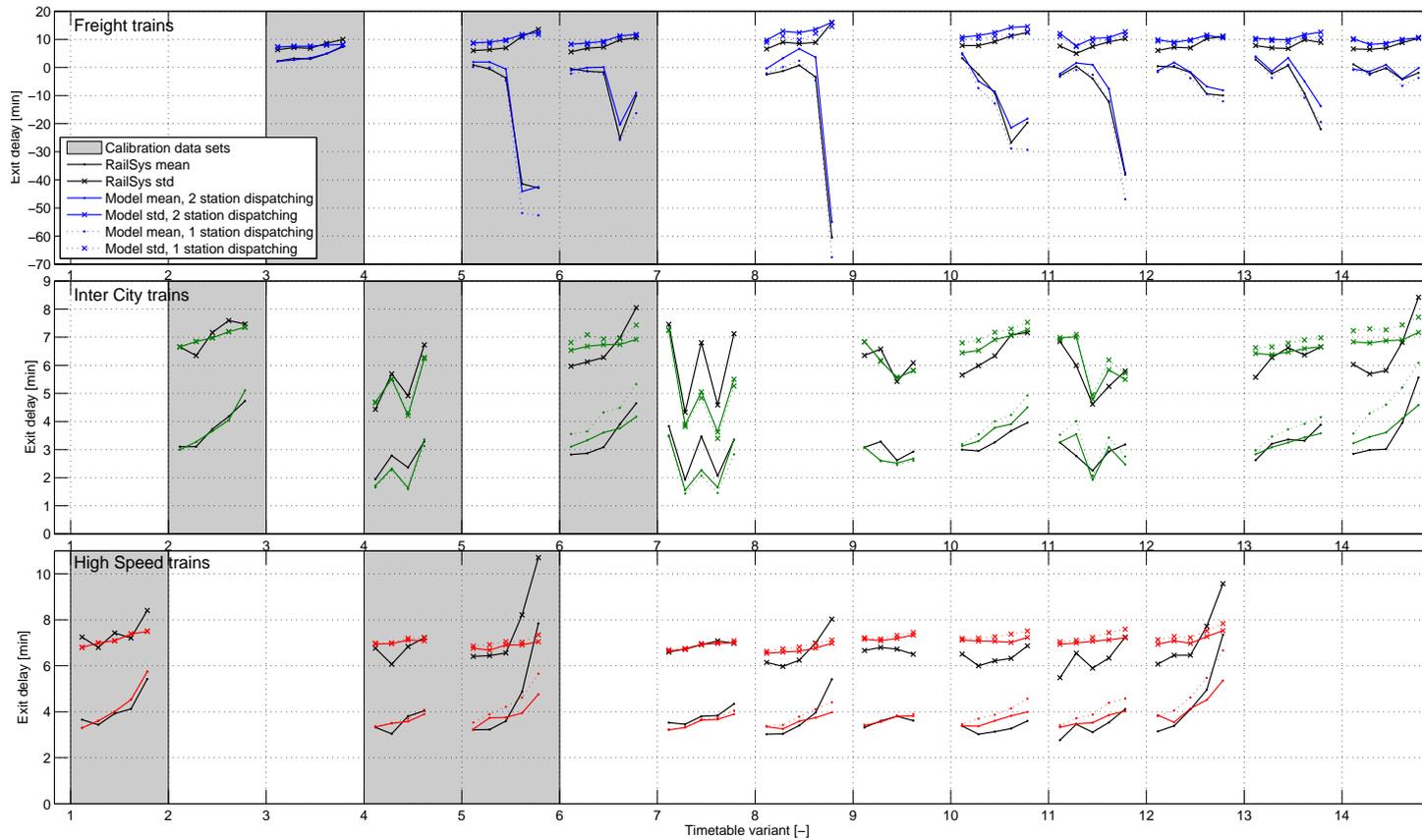
Figure 4: Setup with 20 km infrastructure, high primary delays. The x-axis shows the 14 different timetable variants in the setup (different mixes of train types, see table 2). Each timetable variant has been simulated at four or five different traffic densities (increasing to the right within each variant). Legend is the same for all three subplots, just different line colors.
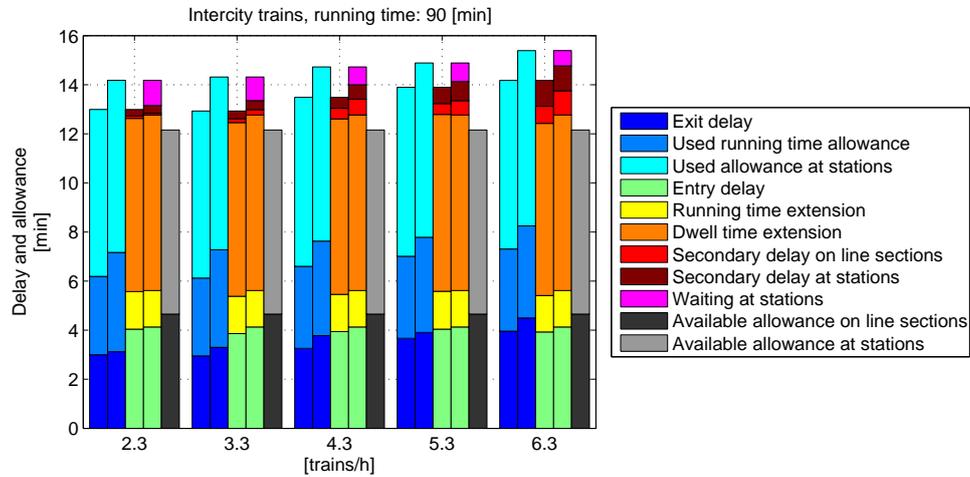
Figure 5: Bar graph of simulation results for IC trains for all five timetables in timetable variant 10, 20 km infrastructure, high primary delays. Results are shown pairwise with RailSys results on the left and model results on the right bars..

Figure 5 shows detailed data about the performance of the IC trains in timetable variant 10 in the first validation data set. All five timetables in the timetable variant are shown (traffic density increasing to the right). It is possible to read out how trains are affected, on average, by entry delays, primary delays, secondary delays and timetable allowances as well as the resulting exit delay. Results of the two models are shown next to each other for comparison (RailSys to the left).

Figures of this type is useful in the validation process as it reveals more information about how the models work compared to just looking at the exit delay. It is for example easy to confirm that the same amount of primary delays is applied in both models and in all timetables. Generated secondary delays are similar in both models and grow with increasing traffic density. For each model, the sum of exit delay and used allowances, blue bars, is the same as the sum of primary delays, secondary delays and time spent waiting at stations. This shows that the calculation of secondary delays is consistent with the calculation of used allowance.

In RailSys, running time allowance can only be used to catch up delays if a train is late. In the model it can be used no matter if the train is late, on time, or early. This is a fundamental difference between the two models and is the reason why significantly more running time allowance is used by the model compared to RailSys, Figure 5. For the same reason passenger trains can arrive early at a scheduled stop in the model but not in RailSys. Trains arriving early at a stop must wait for the scheduled departure time, indicated in magenta. This difference has an impact on the calibration of usable running time allowance and is the main reason why the calibration showed that freight trains should only be allowed to use 50% of their running time allowance.
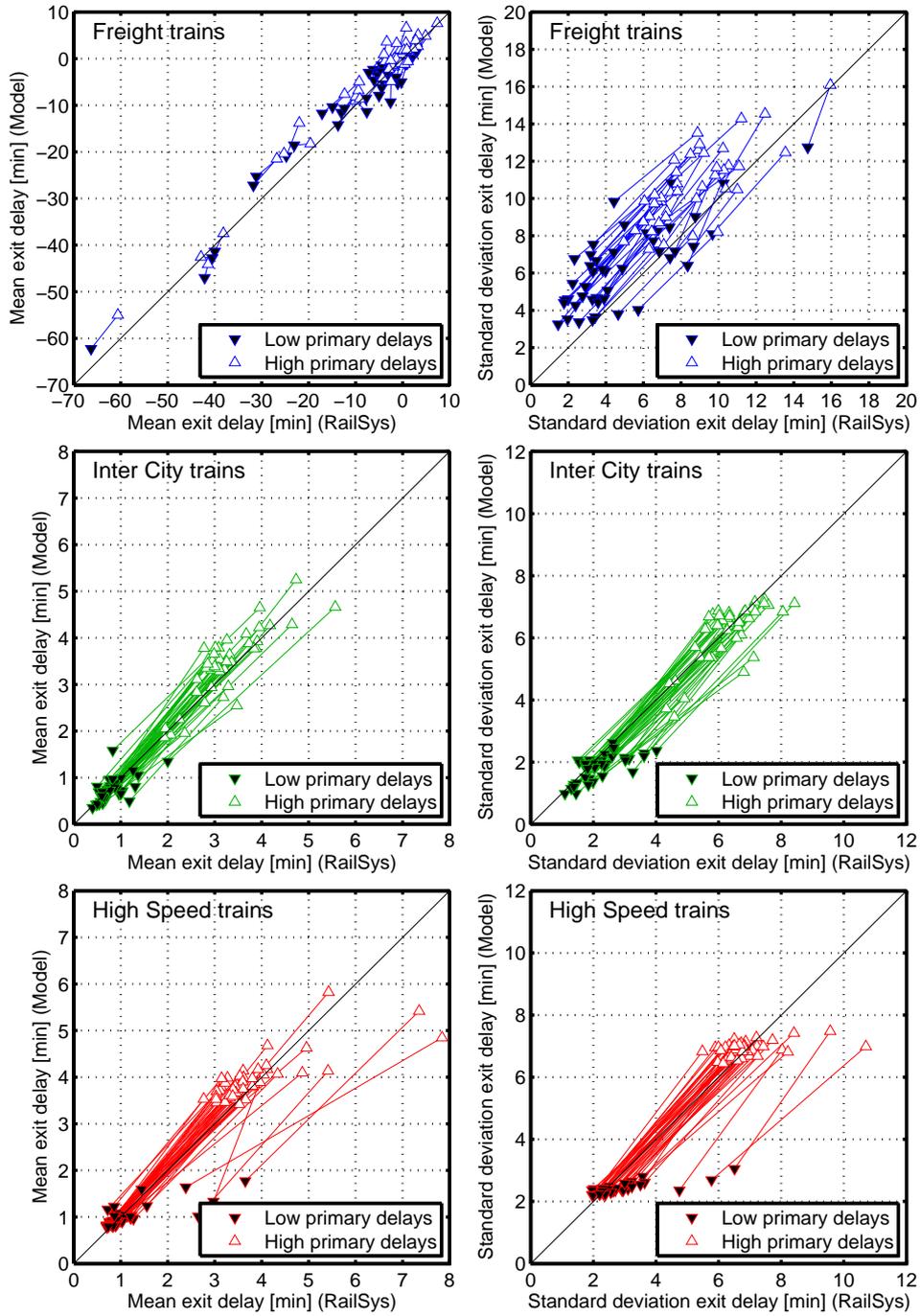
14

Figure 6: Scatter plots showing the reaction in mean (left) and standard deviation (right) of the exit delay when primary delays are decreased using the 20 km infrastructure variant.

Figure 6 shows scatterplots of the exit delay for the different scenarios based on the 20 km infrastructure variant. RailSys values are on the x-axis and model values on the y-axis. Each marker represents one timetable variant at one traffic density level. The same scenarios but with different levels of primary delays are connected with a line which makes it possible to compare the response to the change in applied primary delays of the two models. In the plot, passenger trains arriving ahead of schedule has been considered as they arrive on time. The model shows a larger (12-13 %) decrease in mean exit delay than RailSys when primary delays are reduced. The same applies for the standard deviation. For freight trains however, the model response in mean delay is on average 158% larger than RailSys. The large difference is due to that the running time allowance is available for early trains on the model, while it is not in RailSys, as already discussed.

The corresponding numbers for the 40 km case is for passenger trains 3-6 % and for freight trains 127% respectively. Some outliers are also visible where the model underestimates exit delay of especially high speed trains. Most of these outliers correspond to high traffic density scenarios.

A complete summary of the validation is shown in Figure 7 and table 4. The figure shows the mean difference between the two models for calibration and validation data sets. It is clear that the difference between the two models is greatest for freight trains where the model overestimates the mean and standard deviation of the exit delay. The passenger trains show a much better fit of the mean exit delay, with the possible exception of the high speed trains in validation data set 3 and 4. Regarding the standard deviation, a slight tendency to underestimation is visible.

Figure 7 shows if delays are consequently over or under estimated. However, positive and negative errors may cancel each other out and give a low mean error. For this reason the absolute difference as well as the normalized absolute difference can be found in table 4. Even if the freight trains show the largest absolute errors, the corresponding relative error is small ($< 8\%$), due to the great range of their exit delay.
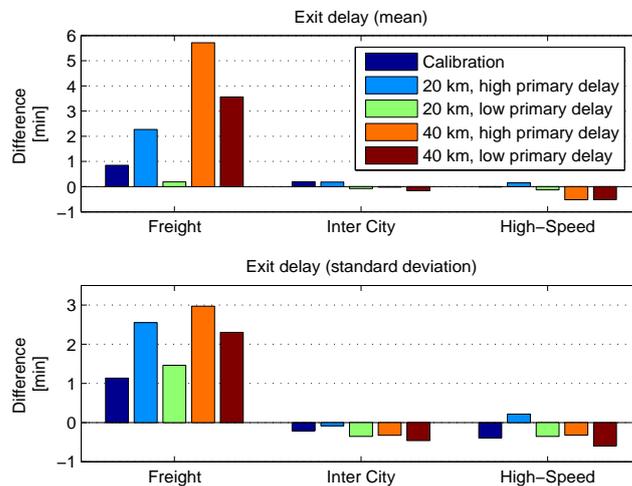

Figure 7: Calibration and validation results, mean difference between the model and RailSys. Positive values indicate that delays are overestimated.

Table 4: Summary of the calibration and validation of the model against RailSys. Values in parentheses are normalized by the range of the data set.

	Absolute difference in mean exit delay [min, (%)]			Absolute difference in standard deviation of the exit delay [min, (%)]		
	Freight	Inter City	High Speed	Freight	Inter City	High Speed
Calibration	1.4 (2.7)	0.33 (12)	0.54 (11)	1.6 (3.2)	0.45 (16)	0.65 (14)
20 km, high delays	2.7 (4.1)	0.39 (11)	0.43 (9.4)	2.6 (4.0)	0.54 (15)	0.55 (12)
20 km, low delays	2.4 (3.5)	0.16 (11)	0.24 (8.2)	1.9 (2.9)	0.43 (28)	0.48 (16)
40 km, high delays	5.7 (7.9)	0.41 (12)	0.84 (9.4)	3.0 (4.2)	0.58 (17)	0.75 (8.4)
40 km, low delays	4.9 (6.6)	0.23 (16)	0.53 (6.3)	2.3 (3.2)	0.48 (33)	0.69 (8.2)

Looking at the final results of the validation, a natural question to ask is if the calibration was successful or could perhaps been done in a different way to get better results. For passenger trains, the results are satisfactory and it is probably hard to achieve much better results. For freight trains however, results show that delays are overestimated already in the calibration data set. Detailed analysis, like the one showed in Figure 5, reveals that freight trains seem to suffer from too much secondary delay at stations, rather than using too little allowance. This is at the same time as passenger trains receive less secondary line delays than in RailSys, which is the reason they do not react as strongly to increased traffic density in many of the timetable variants that include freight trains. The reason that freight trains suffer from too much delays is that the model reduce train priority for early trains. Indeed, it might be a bit drastic to set the time weight of an early train to zero. Instead it could be more appropriate with a significant reduction in priority in order to favor better solutions.

The final conclusion of the validation is that the model works well when compared to the established and validated micro simulation software RailSys. It responds well to change in traffic density, timetable heterogeneity, primary delays and inter-station distance. In the cases with the largest discrepancies, the main reason is found to be how the two models handle trains running ahead of schedule with respect to train priority and availability of running time allowance.

4   Conclusions

The model presented in this paper simulates timetables using a Monte Carlo approach. Infrastructure and rolling stock is modelled at a macroscopic level while timetables and perturbations have a more detailed representation. The dispatching algorithm applies a rolling horizon approach to solve conflicts and can consider several trains when estimating the impact of different dispatching decisions. It also takes into account the consequence of each decision on the possibility of making good decisions at the next station.

The model shows good results when validated against the micro simulation software RailSys. Model response is satisfactory when parameters such as traffic density, heterogeneity, primary delays and inter-station distance are changed. Some differences between the two models become apparent in the validation process and concerns how early trains are modelled. In the model trains ahead of schedule are allowed to use available running time allowance to run faster than planned, while this is not the case in RailSys. In this paper it mostly affects the results of freight trains.

A calibration strategy is developed that systematically uses timetables of different

traffic densities and train mixes to calibrate parameters such as amount of usable running time allowance, minimum headway and train priority. It works well for passenger trains, but is less successful for freight trains because of the differences between the two models already mentioned. Applying the same methodology to calibrate the model against a real railway line using data from real operation, may however prove difficult because it is not easy to find operating conditions that differ as much as in the fictive scenarios used in this calibration. However, one general conclusion that can be drawn from the calibration is the importance of modelling different headways for different train types to get correct simulation results. In a model that uses a microscopic representation of the infrastructure this comes automatically from the signal block occupation, but not in a model with macroscopic infrastructure representation.

Even if the infrastructure models used in this study are quite simple, no complex junctions or varying signalling configurations, it is still interesting to observe that a macroscopic model can perform so well when compared to a microscopic model. Hence, the final conclusion is that the simulation model can be used to evaluate capacity of double track railway lines.

The objective with the developed simulation model is that it will be included in larger model for capacity evaluation. With only a few modifications, the dispatching algorithm presented in this paper can be used to schedule timetables at the same time as it is fast enough to easily generate hundreds or thousands of timetables for real life sized scenarios. Using many timetables in the capacity analysis makes it less timetable dependent, which can for example be desirable when a future timetable unknown.

For long term capacity planning, microscopic simulation may represent a level of detail that is not necessary. The macroscopic infrastructure model and the possibility to generate timetables save a lot of time when the simulations are prepared, compared to if a microscopic simulation tool is used. Instead, this time can be spent on improving the analysis in other aspects, e.g. including more scenarios.

The capacity model will make it possible to consider market demand on both timetable and operation when capacity is analysed, e.g. demand for regularity of passenger services, short travel times and punctuality. It can also be used to improve studies based on socio-economics, e.g. defining strategies for making more efficient timetables or in analyses of infrastructure investments where investments costs are weighed against benefits in terms of reduced travel times and delays.

5 References

Abril M. et al., 2007. An assessment of railway capacity. Transportation Research Part E.

Fischetti M., et al., 2009. Fast approaches to improve the robustness of a railway timetable, Transportation Science.

Landex, A., 2008. "Methods to estimate railway capacity and passenger delays", PhD Thesis, Technical University of Denmark.

Lindfeldt, A., 2011. "Investigating the impact of timetable properties on delay propagation on a double track line using extensive simulation", presented at Railway Engineering 11: th International Conference and Exhibition, London.

Lindfeldt, A., Sipilä, H. 2014. "Simulation of freight train operations with departures ahead of schedule", in proceedings Computers in Railways 2014

Lindfeldt, A., 2013. "Heterogeneity measures and secondary delays on a simulated double-track" In proceedings of 5:th International Seminar on Railway Operations Modelling and Analysis, Copenhagen, Denmark.

Lindfeldt, O., 2009. "Analysis of capacity on double-track railway lines". Transport Planning and Technology.

Lindfeldt, O., 2010. "Impacts of infrastructure, timetable and perturbations in operation of double-track railway lines with mixed traffic". 12th World Conference on Transportation Research 2010, Lisbon.

Lindner, T., 2011. "Applicability of the analytical UIC Code 406 compression method for evaluating line and station capacity", Journal of Rail Transport and Planning & Management, p 49-57.

Meng, L., et al., 2013. "Assessing the absolute traffic carrying capacity: a train timetabling approach", In proceedings of 5:th International Seminar on Railway Operations Modelling and Analysis, Copenhagen, Denmark.

Nelldal, B-L., et al., 2008. "Förbättrad punktlighet på X2000 – analys med hjälp av simulering, (Improved punctuality for X2000 – a simulation approach). KTH, 2008. (In Swedish)

Radtke, A. et al., 2001. "Handling of railway operation problems with RailSys". Proc. of the 5th World Congress on Rail Research. Cologne.

Sahin, I., 1999. "Railway traffic control and train scheduling based on inter-train conflict management", *Transportation Research,* B 33, pp. 511-534.

Salido, M., et al., 2012. "Robustness for a single railway line: Analytical and simulation methods" Expert Systems with Applications.

Siefer, T.: Simulation. In: Hansen, I. A.; Pachl, J. (editors): Railway timetable and traffic, pp 155–169. Eurailpress, 2008.

Sipilä, H., 2010. "Simulation of modified timetables for high speed trains Stockholm-Göteborg" Proceeding of First International Conference of Road and Rail Infrastructure

Sipilä, H., 2014. "Evaluation of single track timetables using simulation", in Proceedings of the 2014 Joint Rail Conference.

Sogin, S., et al., 2011. "Simulating the effects of higher speed passenger trains in single track freight networks", Proceedings of the 2011 Winter Simulation Conference

Trafikverket, 2014. "Samhällsekonomiska principer och kalkylvärden för transportsektorn: ASEK 5.1" Borlänge. In English: Swedish Transport Administration, 2014. Socio-economic principles and calculation values for the transport sector: ASEK 5.1.

UIC Code 406 1st edition, Capacity, International Union of Railways, 2004.