This is the accepted version of a paper presented at *Railway Engineering 2011. London, UK. 28th June 2011 - 30th June 2011*.

Citation for the original published paper:

Sipilä, H. (2011)
Calibration of Simulation Model on the Southern Main Line in Sweden.
In:

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

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-76754
Calibration of a simulation model on the Southern Main Line in Sweden

Hans Sipilä
KTH Royal Institute of Technology
Department of Transport Science
Teknikringen 72
100 44 Stockholm, Sweden
hans.sipila@abe.kth.se

Keywords: Railway operation, simulation, train delay

Abstract
Suitable analysis methods are needed for evaluation of future timetable scenarios, both in short term operational planning and for strategic planning with a longer time horizon. One method is to use simulation software which makes it possible to model large networks. The Swedish Transport Administration (Trafikverket) is in a process where the aim is to start using simulation software RailSys as a tool for timetable planning. This will at first be applied for long term strategic planning with the possibility to also use it in operational planning further on.

The main focus in this paper is to estimate primary run time extensions from registered data. Ideally these should only represent primary causes, e.g. decreased vehicle performance, variation in driver behaviour or infrastructure malfunctions. These extensions are important in order to make simulations more realistic.

Different reduction levels of registered data are tested in order to estimate primary run time extensions. Registered data used are absolute values without distinction between primary and secondary causes. Calibration simulations are done on the Southern main line in Sweden where the mix of high and low speed trains is substantial.

1 Introduction
Considering the increase in rail traffic volumes, finding solutions that meet this growth becomes more and more important. Sweden is no exception, both passenger and freight traffic has increased even though freight volumes are impacted by economic fluctuations. Simulation is a helpful evaluation tool in estimating rail traffic performance. The effect of different infrastructure layouts, timetables, vehicle types etc. can be studied and give useful insight for decision makers.

In this project RailSys is used for evaluating delay level variations between Katrineholm and Hässleholm, which is a subsection of the Southern main line in Sweden (figure 1). In RailSys infrastructure and train runs are modelled on a microscopic level with points, signals, gradients, speed profiles, dispatching decisions etc. Stochastic properties are introduced with delays affecting trains at predefined positions (primary/exogenous), which are further passed on as knock-on delays (secondary/reactive) to other trains. Preparation of complex infrastructure layouts is time consuming, but once this is done usually only small measures are needed for keeping it up to date. Most of the Swedish national rail network is coded in RailSys, jointly by KTH and the Swedish Transport Administration over a ten year period.
1.1 Timetable properties

Important factors which influence overall performance in a railway system are, e.g., recovery times in timetable and speed variations for different types of trains. Heterogeneous timetables have a high mix of slow and fast trains, which reduces overall capacity. This is the case for most of the main lines in Sweden. Timetables where speed variation is low are referred to as homogeneous.

Recovery times are used in timetables to compensate for disturbances and avoid passing on delays to other trains. Run time and dwell supplements are used to reduce influence of driver behaviour, weather conditions, extended passenger exchange times etc. Scheduled buffer times between trains also affect system robustness, in other words the ability to recover from disturbances. These factors are mostly intended to handle small deviations, (Rudolph, 2003). When larger disturbances occur, these will usually propagate to several trains, in particular on main lines with high traffic loads.
1.2 Delay definitions
Delays are mostly divided into two categories. Primary delays occur when trains are faced with disturbances that are not caused (at least not directly) by other trains, typically infrastructure and vehicle problems. Secondary delays, on the other hand, are caused by other trains. This type of delay is also referred to as a reactive or knock-on delay. Three types of preconfigured delays are used in the simulations:

- Entry delays
- Dwell time extensions
- Run time extensions

Entry delays influence trains entering the network at boundary stations. This also applies for trains created inside the network. Dwell time extensions model variations in passenger exchange times and other disturbances occurring at stations. Run time extensions are used on line sections between stations, a common event is when trains must pass a red signal, e.g. due to problems with track circuits. Weather conditions or vehicle problems can also force trains to run with reduced speed.

2 Related research
Previous railway simulation studies at KTH consist, among others, of studies on the Western main line between Stockholm and Gothenburg. In Nelldal (2008) delays were divided according to operator, infrastructure and vehicle error events. The purpose was to reduce delays and compare that to an expected punctuality increase, with main focus on the high-speed trains. Results showed that even if these three categories were reduced by 25–50%, punctuality barely reached 90% at the end stations (trains at most five minutes late).

An experimental design setup was used in Lindfeldt (2009) to both calibrate and validate simulation performance on the Western main line. A good fit was achieved comparing with operational outcome. Calibration and validation data sets were divided according to levels of exit delays. This gave a high-low situation. In this study registration data and timetable represented the same period.

Some of the knowledge and results obtained in Nelldal (2008) and Lindfeldt (2009) were further used in a project were the aim was to analyse how smaller timetable adjustments affect punctuality for high-speed trains (Sipilä, 2010). Variation of timetable allowances and buffer times indicated for example that punctuality increased if buffer times between high-speed trains and other trains were designed with at least 4–5 minutes. At critical stations, where delayed high-speed trains risk to get behind slower trains, increased buffer times from 3 to 5 minutes gave positive effects.

In Radtke (2004) and Siefer (2008) different simulation methods are described, e.g. synchronous and asynchronous modelling. Differences between microscopic and macroscopic approaches are also explained. Furthermore, the basic functionalities of RailSys are presented.

A field study of the Red Line of the Massachusetts Bay Transportation Authority is presented together with a simulation model (SimMETRO) in Koutsopoulos (2006). Different control strategies to improve the operating efficiency are tested and compared. Evaluation methodology and measures of performance parameters are discussed and the importance of making both a calibration and a validation process is emphasized.
3 Method
Implementation of primary run time extensions has been shown to have a significant influence on simulation results, i.e. they make the model more realistic (Lindfeldt, 2009). It is however not evident which level and intensity should be used for these delays. Statistics obtained from registered delay data provide a good overall picture, evaluation of delay development characterized by mean delays or punctuality are often used as measures of performance in a rail network. This absolute data is not telling anything about delay causes, although general conclusions can be made.

3.1 Timetable
Simulated timetable represent a normal weekday, Thursday January 29 in year 2010. All trains pre-scheduled in the national timetable are created. Most passenger trains belong to patterns with regular frequencies. Freight trains are split into four groups, long distance day and night along with short distance freight services. Mail freight trains run with significantly higher top speeds (160 km/h) than regular freight trains and are treated separately.

Passenger train vehicle types are usually well known and relatively easy to model in RailSys. Freight train configurations are based on assumptions and checked with respect to scheduled run times and available allowances in RailSys. Mostly a long and heavy train type is tried first and if necessary changed to a type with better performance (less weight and/or higher top speed). Most freight trains in Sweden run with maximum speed 80–120 km/h. Normal freight top speed on the Southern main line is 100 km/h.

There are no big differences between timetables from year 2008 and 2010 concerning the studied area. A majority of the passenger trains run in the same time slot, some trains have minor adjustments of 1–2 minutes. InterCity trains have more departures compared to 2008. Freight trains show more variance, although many trains do not deviate much if equal train numbers are compared. For trains with greater differences, the scheduled stops are usually different.

3.2 Delay data
Trains running on Swedish national network are registered at every station as deviations relative to the scheduled timetable. These registrations are usually triggered in the signalling system which also keeps track of train numbers. Depending on the trigger point position, automatic corrections are added so that the measurements represent the station middle point. Values are also truncated to full minutes. This means for example, that if the actual registration has a deviation of 2 minutes and 59 seconds, this appears as 2 minutes in the statistics.

To this basic delay data cause reports are added, this is however only done for delays which have a growth of at least five minutes between two registration points. Several cause report codes exist and they are classified into primary and secondary events. Small delays, either they occur at stations or on line sections, have no cause report assignments. Small growing delays, which eventually can become large, are therefore not easily traceable. This implies a problem when it comes to run and dwell time extension modelling. Dwell time delay distributions are reused from previous projects (Nelldal, 2008), which in turn are based on manual measurements carried out at different stations. Figure 2 shows cause reported delays on the Southern main line for freight and high-speed passenger trains (X2000).
Figure 2: Primary cause reported delays for selected southbound trains on section Mjölby–Hässleholm (265 km). Weekdays, January to June 2008.

One way smaller run time extensions can be modelled, is to make some assumptions from cause reported delays and apply these to lower levels. The relation between primary and secondary causes can give a hint on what levels to use for extensions below five minutes. Furthermore it is not evident if a reported primary delay has passed on knock-on delays to other trains, partly since separating station and line incidents is difficult. Cause reports can of course be filtered to avoid influence from incidents which are difficult to model in simulations.

In theory every train have a defined schedule. A majority of trains are regular and run with intervals from every day to once a week. Temporary adjustments in timetables means in general that train numbers are changed. This complicates the preparation of data to simulations, since it is difficult to distinguish these variant numbers from ad-hoc trains.

As mentioned earlier, trains are registered at every station along their route. However, these registrations are based on the timetable. Trains with scheduled stops have both arrival and departure times, this gives two registrations and makes it possible to separate between station and line events. If no stop is scheduled, only departure values are registered. This is not so much of a problem when dealing with passenger train statistics, since they mostly make station stops according to schedule. Freight trains however show significant day to day variations relative their scheduled timetable. Running ahead of schedule can be as common as running behind, deviations are high. Hence, the operational timetable is unique for every day. Freight trains have mostly three different types of scheduled stops or a combination of these.

- Timetable technical, overtaking or crossing situation
- Shunting, adding/removing cars
- Driver relief station

Depending on the real time conditions timetable technical stops are frequently cancelled. On the other hand a lot of unscheduled stops are carried out. Analysis of recorded delay data would be easier if registrations were made for both arrival and departure times. Some of the unscheduled stops can be found by comparing scheduled run times relative to adjacent stations. As a drawback, this gives problems in differentiating between run time extensions (exogenous or knock-on) and unscheduled stops.
3.3 Preparation of data

RailSys offers the possibility of using either negative exponential or empirical distributions for generating preconfigured delays. Other types of distributions can be used if they are made discrete. In order to achieve more flexibility and make input variation easier, delay files (perturbed timetables) are created in Matlab and then read by RailSys prior to a simulation. This makes it possible to split up simulations to subsets of cycles and use different delay settings in the same total simulation.

3.3.1 Run time extensions

The main objective in this project is to create run time extensions based on train registration data as simulation input, without considering any additional cause reports. Junctions and turnaround stations are used as representative locations for calculation of empirical run time distributions. The studied network is divided into sections according to figure 1. This approach means that both primary and knock-on delays are included. Distributions are then reduced to different levels in the simulations. Figure 3 illustrates the principle for calculating run time extensions. The vast majority of trains that run on a typical section will pass both station A and F. Section lengths are between 35–50 km.

![Figure 3: Train run from station A to F with registrations \( n_A \) and \( n_F \).](image)

For all trains that have passed an evaluation section, a relative time value \( t \) is calculated. Early trains \((n < 0)\) are handled as if they were on time, late trains have \( n > 0 \). Four different combinations of registrations are identified. Distributions are compiled according to these cases. Only trains with delays on arrival at station F contribute with positive values other than zero to the distributions. The principles are shown below.

\[
\begin{align*}
n_A \geq 0 \quad &\Rightarrow \quad t = \begin{cases} 
n_F - n_A & \text{if } n_A \leq n_F \\
0 & \text{if } n_A > n_F 
\end{cases} \\
n_A < 0 \quad &\Rightarrow \quad t = \max\{0, n_F\}
\end{align*}
\]

For the simulations scenarios, run time distributions compiled from registrations, are reduced in order to cut down the impact of knock-on delays. It is assumed that total delays on studied sections are higher than primary delays, which in part are unknown. The opposite could occur on long sections with unusual large timetable allowances, since most trains would be able to make up for smaller primary delays. Distributions are reduced by keeping a certain percentage \((R)\) of registrations for every minute level \((m)\). The total number of registrations is kept equal between original and reduced distributions. This gives that reduced registrations are given zero values. If the number of registrations for original distributions are denoted \( N \) and corresponding reduced registrations by \( n \), the reduction process can be written as

\[
\begin{align*}
m = 0 \quad &\Rightarrow \quad n_0 = N_0 + \sum_{m=1}^{m_{\text{max}}} (1 - R) \cdot N_m \\
m > 0 \quad &\Rightarrow \quad n_m = R \cdot N_m
\end{align*}
\]

6
Four different levels are used for run time distributions and applied on two main groups, passenger and freight trains (table 1). Original distributions are compiled for seven passenger train groups and four freight groups in each direction. Train numbers are matched as far as possible, remaining trains are grouped according to time and route. This is mostly an issue for freight trains.

Table 1: Percentage parameters ($R$) for simulations, passenger (P) and freight trains (F)

<table>
<thead>
<tr>
<th>Train category</th>
<th>Simulation cases</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>P</td>
<td>20</td>
</tr>
<tr>
<td>F</td>
<td>20</td>
</tr>
</tbody>
</table>

3.3.2 Entry delays

If passenger trains are ahead of schedule it is mostly in the range of a couple of minutes, depending on which station is considered. Since trains cannot be created ahead of their schedule in RailSys, early registrations are defined as zeros (on schedule). Figure 4 shows a typical scenario for station departure values where most passenger trains have a scheduled stop. Freight trains show a high spread and a significant proportion departs before scheduled time. Entry delay distributions are compiled from registered departure values and applied to corresponding train groups and stations (figure 4). Entry delays for freight trains are clustered into subgroups, according to mean, standard deviation and frequency for each train number.

![Figure 4: Departure distribution for northbound trains at Hässleholm. Trains ahead of schedule have values < 0 (left). Train entry grouping for southbound freight trains at Skänninge, network boundary station north of Mjölby (right).](image)

4 Results

Simulation results are analysed in two ways. At first, selected train groups are evaluated with respect to correlations between varied delay levels and other groups. This can show how sensitive these groups are to overall changes in the setup. Second part deals with comparing real and simulated outcome, also here a selection of train groups is analysed. This is useful in preparing for example timetable and/or infrastructure scenario simulations.
4.1 Variation in a group and interaction with other groups

An investigation of how sensitive selected train groups are for primary delay level variations in their own group compared to influences from another group is done by checking exit delay variations. High-speed passenger trains are compared with freight trains in day group. These are chosen since both groups run for a relatively long distance on the simulated network and during daytime. In figure 5 group internal variations are represented by moving from curve to curve along the y-axis. Changes imposed by the comparison group is shown on the x-axis.

![Figure 5: Exit delay variation for high-speed passenger trains and freight trains in day group, mean (solid) and standard deviation (dashed curves).](image)

In this case, train groups are affected more by internal changes than variations in the other group. Northbound direction shows high exit delay variation for the freight group, which indicates that their original run time distributions are relatively heavy. Delay levels are varied simultaneously for all freight train groups, which means that for example the short distance freight group can cause disturbances, especially since both groups have similar dispatching priority parameters. This is also a factor in the interference between passenger and freight trains in general. In the performed simulations, passenger trains generally have a higher dispatching priority than freight trains. If the train group with higher priority also shows significant original run time delays, these trains could have much more influence on trains with lower priority. In this case however, the cross influence is small.

4.2 Simulations compared with real outcome

In most cases it is of interest to compare a simulation model to its real counterpart. For railway operations this can be done by comparing combinations of delays, punctuality or some other measures. In order to capture the delay situation, averages and deviation values are used. Data used for creating simulation inputs are shown as reference in delay plots. Visually checking simulation results give a good first impression if the fit is good or bad. However, a more general method is preferable to get a better overview when many cases are compared. For measure of performance (MOP), the frequently used root mean square error (RMSE) is calculated for mean and standard deviation. In the expression below simulated and real data is denoted by $X_i$ and $Y_i$.

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (X_i - Y_i)^2}$$
Representative groups are chosen to highlight differences between various train categories. The purpose is to get a general idea on adequate settings from table 1 to use for further simulations, for example testing new timetables. Three cases are chosen, two extremes and one in between. Figure 6 shows mean delay RMSE values for evaluated train groups.

![Figure 6: Root mean square error for mean delays.](image)

It is obvious that the commuter train system show a good agreement, almost independent of applied run time extensions. Freight train groups have higher RMSE values. Regarding total agreement, it can be seen that case 10 should be chosen of the three. Figure 6 gives only a goodness-of-fit indication, it does not reveal anything about the actual delay levels. Plotting delay development throughout the studied line is descriptive, since sections with significant increase or decrease in delay can be identified.

Figure 7–9 present some case combinations used in figure 6. Situation with high and low delay levels in figure 7 show that, especially development for mean values have high deviations relative observed data. Passenger train scenario with low levels show acceptable values, although the simulation fails to model the observed delay increase for northbound high-speed trains between Norrköping and Katrineholm. This increase is explained by a speed reduction due to tunnel renovation work, that started in the beginning of 2008. Both directions were affected and at first no compensation was used in the timetables. Trains with high scheduled speeds are especially sensitive to large speed reductions.

The other passenger train groups have acceptable results. In the following figures mean delays have solid lines and standard deviation dashed, observed values are bold. A majority of the freight trains run between Mjölby and Hässleholm (and further on to Malmö). Although some trains in these groups also run on section Katrineholm–Mjölby, this number is significantly smaller. For station codes, see figure 1. Simulation case 10, which have lower RMSE values than case 1 and 16 is shown in figure 8 and 9. They indicate better agreement than in figure 7, at least for mean values. Northbound freight trains in night group deviate clearly from observed values. Difference in standard deviation is significant for some of the freight train groups. The evaluated commuter train group had good delay statistics, which is also the case for year 2010.
Figure 7: Simulated (thin) and observed delays (bold curves) in minutes. Southbound freight trains, 80/80 case 16 (left). Northbound passenger trains, 20/20 case 1 (right). Mean values are solid lines and standard deviations are dashed.

Figure 8: Simulated (thin) and observed (bold) delays in minutes. Passenger trains, 40/60 case 10, both directions. Mean values solid, standard deviations dashed.

Figure 9: Simulated (thin) and observed (bold) delays in minutes. Freight trains, 40/60 case 10, both directions. Mean values solid, standard deviations dashed.
Differences between start values can partly be explained by dispatching decisions done in RailSys when trains are initialized. Compiled entry delay distributions are based on actual departure values, where the real dispatching decision is already made. In the simulations this decision is added to the entry delays, which means that trains with higher priority influence other trains before train runs start. In some cases passenger transfer connections, that are not modelled in these simulations, can influence results.

Freight train modelling is difficult due to several timetable and operational characteristics. Trains ahead of schedule are not modelled, although this can be achieved to some degree once trains already are running in the network. Running two trains in parallel, i.e. same direction on two tracks, is also hard to achieve without causing other unwanted situations in the simulations. This practice is not uncommon, especially during night time, and timetables are also planned with these solutions.

Some trains stop at designated stations for shunting, which imposes a problem in estimating dwell times. One way of handling this is by analysing specific data concerning arrivals and departures for these stations. As a first approximation, ordinary delay statistics can give some insight. Freight trains with shunting stops can be modelled by splitting up the train runs and apply initial delays to model variations in departures. However, this might not be preferable if station track (side track) occupation is of importance. Shunting movements on stations are not practical to handle in large network simulations and the effects on results will probably not be very transparent.

Train priority thresholds also play an important role and can, depending on the situation, have a significant impact on simulation results. In these simulations, trains with different start priorities rarely fall under another train category. Their value can however drop under other trains in the same category. These parameters show a dynamic behaviour in real operations and make them therefore tricky to model.

5 Conclusions

Although some operational characteristics cannot be modelled, simulation offers a good tool for evaluation rail network performance. Compared to for example analytical methods, large areas with dense traffic can be analysed. Micro simulation also models interaction between train runs at stations and on line sections. Quality of the results depends strongly on input data handling, e.g. definition of infrastructure, vehicle modelling etc.

A key factor is to design realistic delay conditions. Different approaches can be used in this process. Simulation results show that good agreement can be obtained by using train registrations, without considering cause reported data. Differences in timetables, although small, can cause some of the larger deviations between simulated and observed values. Other influencing factors, such as dispatching parameters and freight train performance assumptions, also play a role in the outcome.

Registration data for all stations in a network give more possibilities in evaluating changes in train order sequences, i.e. overtakings. These estimations can increase the precision in run time distributions, which in a reduced state are used for stochastic extensions. In this paper, only two levels are used simultaneously. Some improvements can probably be obtained by using varying levels on different sections. This increases the number of simulation cases needed for finding optimal settings.

No validation simulations are performed at this stage. These are important to carry out if reliable conclusions are to be made. Future work should include both calibration and validation studies of networks with different characteristics. Even if the total number of trains can be much lower than in this project, single-track sections imply more dispatching problems than usually observed on double-track lines. Extensive simulations involving single-track lines are challenging, both regarding parameter setup and result interpretation.
6 References


