Simulation of rail traffic

Applications with timetable construction and delay modelling

HANS SIPILÄ

KTH Architecture and
the Built Environment

Licentiate Thesis in Transport Science
Stockholm, Sweden 2012

ROYAL INSTITUTE
OF TECHNOLOGY

Simulation of rail traffic

–

Applications with timetable construction and delay modelling

Hans Sipilä

Licentiate Thesis in Transport Science

Stockholm 2012

School of Architecture and the Built Environment
Department of Transport Science
Division of Transport and Logistics
SE–100 44 Stockholm

TRITA–TSC–LIC 12–003
ISBN 978–91–85539–87–1

Abstract

This thesis covers both applications where simulation is used on parts of the Swedish rail networks and running time calculations for future high-speed trains with top speed improvements on existing lines. Calculations are part of a sub-project within the Green Train research program (Gröna tåget). Higher speeds are possible with increased cant and cant deficiency in curves. Data for circular curve radii is used on existing lines combined with information on decided and on-going upgrades. Calculation of static speed profiles is made for a set of cant and cant deficiency values. Different train characteristics are used regarding top speed, starting acceleration and power to ton ratio. Running time calculations are made for these different train characteristics with the fictive speed profiles. In addition, different stopping patterns are applied.

Results are presented together with running times for two reference train types, one with carbody tilting and one without. It is clear that carbody tilting, allowing a higher cant deficiency, is important on many of the existing lines considering achieved running times. The benefit of tilting is marginal on newly built and future lines designed with large curve radii. However, on many of the existing lines the over 20 year old reference train with carbody tilting achieves shorter running times compared to a future train without tilt but with higher top speed. The work presented here has contributed with input to other projects and applications within the research program.

Simulation in RailSys is used to evaluate on-time performance for high-speed trains, between Stockholm and Göteborg in Sweden, and changes in timetable allowances and buffer times with respect to other trains. Results show that on-time performance can be improved with increased allowances or buffer times. In the case with increased buffers, other trains are pushed in the timetable with the intention of obtaining at least five minutes at critical places (e.g. conflicting train paths at stations) and as separation on line sections. On-time performance is evaluated both on aggregated (group) level and for trains individually. Some of the trains benefit significantly from the applied measures.

Prior to a simulation some of the delays have to be defined. This includes dwell extensions and entry delays, i.e. extended exchange times at stations and delayed origin station departures inside or at the network border. Evaluation of observed data give insight on the performance of a real network. However, separating primary (exogenous) and secondary (knock-on) delays is not straightforward. Typically the probabilities and levels of primary delays are defined as input, thus secondary delays are created in the simulations. Although some classification of delays exist in observed data, it is not sufficient without further assumptions and preparation.

A method for estimating primary running time extensions is presented and applied on a real timetable between Katrineholm and Hässleholm in Sweden. The approach consist of creating distributions based on deviations from scheduled running time. Since this represent total outcome, i.e. both primary and knock-on delays are included, the distributions are reduced by a certain percentage and applied in the simulations. Reduction is done in four steps, separately for passenger and freight trains. Root mean square error (RMSE) is used for comparing mean and standard deviation values between simulated and observed data.

Results show that a reasonably good fit can be obtained. Freight services show a higher variation than passenger train evaluation groups. Some explanations for this are difficulties in capturing the variations in train weights and speeds and absence of shunting operations in the model. In reality, freight trains can also frequently depart ahead of schedule and this effect is not captured in the simulations. However, focus is mostly on passenger trains and their on-time performance. If a good enough agreement and operational behaviour is achieved for them, a lower agreement for freight trains may be accepted.

Acknowledgements

Trafikverket (Swedish Transport Administration) and SJ AB provided funding for this research. I would like to thank my supervisors Bo-Lennart Nelldal and Oskar Fröidh for their support and advice. Thanks also to Magnus Wahlborg, a member of the reference group and contact person at Trafikverket responsible for the research program financing this thesis. In addition, I would like to thank the other members of the reference group: Armin Ruge, Kenneth Håkansson and Åke Lundberg (Trafikverket), Dan Olofsson, Mattias Mellberg and Britt-Marie Olsson (SJ AB), Johanna Törnquist Krasemann (BTH/LiTH) and Olov Lindfeldt (Vectura). Thanks to Anders Lindfeldt and Jennifer Warg for all the stimulating discussions and their valuable inputs. Finally, thanks to all colleagues and fellow PhD students at the division of Transport and Logistics (KTH) and to Δ.

Stockholm, May 2012

Hans Sipilä

List of publications

Reports

I. Sipilä, H., 2008. Körtidsberäkningar för Gröna tåget. (Running time calculations for the Green Train). In Swedish. Appendix B not included.

Papers

I. Sipilä, H., 2010. Simulation of modified timetables for high speed trains Stockholm–Göteborg. Presented at: First International Conference on Road and Rail Infrastructure, Opatija.

II. Sipilä, H., 2011. Calibration of simulation model on the Southern Main Line in Sweden. Presented at: Railway Engineering, 11:th International Conference and Exhibition, London.

Related publications not included in thesis

- Nelldal, B-L., Lindfeldt, O., Sipilä, H., Wolfmaier, J., 2008. Förbättrad punktlighet på X2000 – Analys med hjälp av simulering. Tech. Rep., KTH.
- Lindfeldt, O., Sipilä, H., 2009. Validation of a simulation model for mixed traffic on Swedish double-track railway line. Presented at: Railway Engineering, 10:th International Conference and Exhibition, London.
- Sipilä, H., 2010. Tidtabellsläggning med hjälp av simulering – Effekter av olika tillägg och marginaler på X2000-tågen Stockholm–Göteborg. Tech. Rep., KTH.
- Sipilä, H., Warg, J., 2012. Kapacitetsanalys av Södra stambanan. Tech. Rep., KTH.

Contents

1 Introduction

1.1 Background

Sweden is experiencing increasing capacity constraints across the rail network. Finding solutions that meet this growth becomes more and more important. In Sweden, both passenger and freight services have increased and planned investments in new rail infrastructure boost a future growth. More pressure is put on the existing rail network. Although it is a positive trend in an environmental sense, several problems with rail traffic are highlighted with this increased traffic load. More trains and increased capacity utilisation lead to higher probabilities for delay propagation in the system. This in turn has to do with operational characteristics, higher use of the infrastructure etc. It is obvious that in the long run new tracks and lines are needed, but there is a need for actions and strategies that can improve performance on the existing network in a shorter term.

In the past, the whole rail network (trains and infrastructure) was operated by the national railway company. Today functions are split among several operators, maintenance companies, an infrastructure agency, etc. Increased demand creates a need for more train traffic. Commuter, regional and long distance train systems increase, both in number of trains and area covered. Train stops on new stations and extended local and regional networks are demanded by municipalities etc. The main idea is beneficial but this also creates problems to the already congested rail network, especially on the main lines. Things to consider for improving capacity can also be related to train characteristics, e.g. longer trains, double-decker trains etc.

1.2 Objectives and scope

The main focus in this research project is to develop methods for using simulation in timetable planning. Today this work is mainly done by the Swedish Transport Administration (Trafikverket) in cooperation with train operators. Increasing demands for more train slots from operators, regions and municipalities require improved timetable evaluation methods. Even though the infrastructure puts restrictions on the possible timetables, there is almost an infinite number of timetable solutions for a large network (relevant ones). Different constraints reduce the degrees of freedom in timetable construction even further. Examples of these are:

- Regular intervals (periodic)
- Minimum buffer times (headways) between trains and allowances
- Speed mix
- Meets and overtakings
- Connections/transfers between trains (trip chaining)
- Demand for certain departure or arrival times

One possible approach in evaluating and varying some of the above factors is to use simulation software. RailSys, which is used in this thesis and at KTH, was developed by the railway group (IVE) at the University of Hannover and later at the University of Braunschweig. The commercial part is handled by Rail Management Consultants (RMCon). The software allows a microscopic description of the infrastructure in terms of links and nodes with tracks, points (turnouts), signals, speed configurations, etc.

The track model together with vehicle models give the basis for defining timetables. However, one of the most important factors in simulations of rail traffic and what differ simulation from general timetable planning is delay modelling. A realistic simulation model needs some input delay definitions to reflect real conditions. This stochastic feature is introduced in the model by different types of distributions. The synchronous simulations are later carried out for a predefined number of cycles.

Dispatching in RailSys is mainly based on relative priorities between train types. Delay threshold values can either increase or decrease the priority or leave it unchanged. Dispatching on priorities can be weighted to total expected delay when a conflict is recognised between two trains. Multiple conflicts are broken down to pairwise problems and handled by the dispatching function. The look-ahead time range for the dispatching function can be changed, depending on the setup these times can have a significant effect on outcome.

A general aim is also to develop and improve methods for using simulation in timetable planning. This can be both in a short term perspective and for strategic planning purposes. Some issues that can be of interest are for example evaluation of different timetable solutions and infrastructure modifications. It is important to understand what can and should be modelled and when simplifications can be made.

Running time calculations for varying train configurations regarding starting acceleration, power to ton ratio and top speed is described in section 4. Different sets of track cant and cant deficiency combined with radii for circular curves are used in calculating static speed profiles on existing and future lines. This work is part of the Green Train research program. The main objective in the program is to develop a concept proposal for a new, attractive high speed train adapted to Nordic conditions [8].

1.3 Limitations

In speed profile calculations with varying cant and cant deficiency only circular curve radii is considered. Transition curves before and after a circular curve can in some cases allow a lower speed than the circular part itself. However, this would require a more detailed analysis and was not within the scope of this study. When train traction characteristics are calculated, adhesion is not considered to affect the short-duration tractive power under normal conditions.

Delay distributions used in simulations are designed with the purpose of avoiding large values that in reality would result in dispatching measures which are difficult to model in the simulation software. A typical example is the case where trains use tracks normally intended for traffic in the opposite direction, this feature can only be modelled in some specific cases. In order to avoid so-called deadlocks, some limitations are also introduced at larger stations regarding possible paths between station entrance and exit signals. When it comes to the Swedish signal system and ATC not all features are modelled, improvements are possible with access to detailed data over interlocking and ATC information.

Shunting operation at stations is not modelled. This also applies for vehicle chaining and transfers between passenger trains. Although these characteristics can to some extent be considered in the simulations, they add more complexity and might not reflect actual conditions. Freight trains regularly depart ahead of schedule, in the simulations presented here freight trains are either on time or delayed. This is also considered when performance measures are compiled from observed and simulated data. A modelling approach capturing this property is under evaluation and comparative simulations are planned based on the model used on Southern Main Line.

1.4 Structure of the thesis

Section 2 gives an overview of some relevant studies regarding simulation of train traffic, timetable properties and possible evaluation parameters. Methodological issues and problems are discussed in section 3. Further details regarding methodology and results are presented in section 4–6 covering one report and two papers making up the second part of this thesis. Sections 7–9 cover conclusions, future work and a discussion of the contributions of this thesis. Fig. 1 gives an overview of the lines discussed further on in this thesis and station names and id:s used in most of the figures.

Figure 1: Map showing simulated lines (red) and evaluation stations. Parts of the connecting lines (blue) are included in the simulations.

2 Related research

Previous railway simulation studies consist, among others, of analysis on the Western Main Line between Stockholm and Göteborg. In [15] delays were categorised to operator, infrastructure and vehicle related events. The purpose was to reduce delays and compare that to an expected increase in on-time performance, with main focus on the high-speed trains. Results showed that even if the delays in all categories were reduced by 25–50%, on-time performance barely reached 90% at the end stations (trains with a maximum delay of five minutes).

An experimental design setup was used in [13] to both calibrate and validate simulation performance on the Western Main Line. In this study registration data and timetable represented the same period. A good fit is achieved comparing to operational outcome. Calibration and validation data sets were divided according to levels of exit delays, giving a high-low situation.

In [18] and [20] 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. Allowances and margins in railway scheduling is discussed in [19]. Recovery times can decrease or avoid delays for specific trains. These are obtained by adding allowances or supplements to the minimum running time. Buffer times are added to minimum headway between two trains and act as recovery times of the timetable. The aim with these properties is to increase the reliability of timetables.

Capacity on single track lines with additions of both freight and passenger trains is investigated in [24]. Simulation software Rail Traffic Controller (RTC) is used to evaluate a representative fictive route with varying train density. Simplifications are made to improve comparison of the effects of key variables regarding traffic composition and passenger train speed. Some randomisation of freight train departures are introduced. A homogeneous condition with a composition of 100% freight trains is defined as a base case, although the number of trains vary.

Comparing simulations where the total number of trains is constant show higher delays in the heterogeneous cases. The more passenger trains operated, the higher the variation in the delay of the freight trains. Higher passenger speeds will increase delays although the marginal increase in speed has a diminishing factor on the delays of freight trains when the network becomes saturated.

A similar study is presented in [4], in which a hypothetical single-track line is modelled. Combinations of three different types of freight trains and one passenger train with different percentages of each train type are simulated. One objective is to investigate what aspects of heterogeneity have the most pronounced impact on delay.

The principal metric used for comparison is average delay. In cases with only freight trains, heterogeneity in top speed, power to weight ratio and dispatching priority is varied separately for percentages of two selected freight train types. Results from simulations show that dispatching priority has a greater impact on delays than speed or power. Adding passenger trains gives higher delays compared to adding the same number of freight trains. This is due to both higher priorities and speeds.

A field study of the Red Line of the Massachusetts Bay Transportation Authority is presented together with a simulation model (SimMETRO) in [10]. 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.

Methodology for modelling a calibration process as a multi-variate optimisation problem and solving it with SPSA algorithm is presented in [11]. This methodology is applied in the simulation study discussed in [10] and results show that the calibration process improves the parameters and refines the input. The algorithm used is shown to be efficient.

Influencing factors on train punctuality and some applications of this is discussed in [16] which is partly based on earlier studies. Examples of some analysed punctuality factors are infrastructure capacity utilisation, cancellations, operational priority rules and temporary speed restrictions. Data regarding both delays and the influencing factors have been used with the aim to identify how the studied factors influenced the punctuality.

The study focused on commuter trains in the Oslo area and long distance trains on one line. Number of travellers, capacity utilisation and temporary speed restrictions are found to have a negative correlation to punctuality. Cancellations have a positive correlation. Several of the factors seem to have threshold values. Delays under these threshold values are absorbed during normal conditions, high delays remain in most cases until the final destination.

A simulation model for delay evaluation is presented in [9]. The model is validated on a regional railway in Taiwan. The main purpose is to evaluate the impacts of primary (first) and knock-on delays on the modelled timetable. Dispatching measures can reduce dwell and running times for delayed trains. Validation is done by comparing actual and estimated delays at selected stations. The case study shows that the simulation model can be used for evaluating knock-on delays reasonably well. However, the studied timetable is relatively homogeneous and there are no high-speed or freight trains.

Some topics regarding railway capacity and delay relationships is discussed in [14]. Capacity variation due to speeds, number of tracks etc. is illustrated with examples. Analytic, statistical and micro-simulation methods for delay analysis are presented and some of their advantages and drawbacks are listed.

Analytic methods can be useful at a strategic planning and design stage, however simplifying assumptions are often necessary. This may not be a problem considering that many future parameters are uncertain. Elements from queuing theory and optimisation are common in these approaches. Micro-simulation offers a useful approach in modelling interactions between trains, i.e. propagation of knock-on delays initiated by some exogenous event.

Defining input data, e.g. track layout and signalling, can take a significant amount of time. There is a risk of loosing resolution in general tendencies if only one specific setup is studied. Statistical methods using regression give an opportunity for establishing empirical relationships between capacity utilisation and knock-on delays for a given level of primary delays.

Research concerning simulation of rail networks is limited. Some of the material available present interesting approaches to simulation setup and evaluation techniques. Although national conditions differ, many of the basic questions and problems remain similar to conditions in Sweden. Finding more comprehensive ways of performing a calibration process is interesting, especially if simulation is to be used efficiently and provide reliable results in future timetable planning.

A couple of the references discuss the topic of adjusting and tuning a simulation setup with reference to observed data. Applications focused on modelling a limited line section with mostly one type of service could be interesting to use for a similar case in Sweden. However, evaluation of larger networks or lines usually requires modelling of different train categories with highly variable characteristics and this adds to the model complexity.

3 Methodology

3.1 Defining simulation infrastructure

Defining the rail network demands in many cases detailed data for track lengths, gradients, positions for signals, balises and turnouts, stopping points etc. Speed profiles need to be defined considering turnout speeds and other restrictions. This implies studying automatic train control information (ATC) and signal plans and blueprints. ATC is a train protection system for railways ensuring a safe operation of trains. As is normal in reality, infrastructure is separated into stations (nodes) and line sections (links) from one station to another.

Train routes within stations, from entry to exit signals, model possible ways trains can utilise and pass a station. If information is available concerning train route dependencies, e.g. a certain train route (movement) is not allowed while another route is set, this can in many cases be incorporated into the model. Separating stopping points for different train lengths and types adds to the level of detail.

The described infrastructure features do not necessarily take long time to implement if only a small network or a simple infrastructure layout is considered. However, defining large stations or junctions is relatively time-consuming. If this work is done in a qualitative way, there is usually only a need to apply some updates prior to a simulation. Depending on the objectives and scale of a project not all details need to be considered.

Over more than ten years the Swedish rail network has been coded into RailSys. This work has in most parts been carried out by KTH and later also by the Swedish Transport Administration (Trafikverket, previously Banverket). Today (year 2012) the majority of the network is included, but simulations has been carried out mostly on the main lines in the southern part of Sweden.

3.2 Vehicle models and timetables

Vehicle models (trains) are needed to carry out running time calculations, timetable evaluations and simulations. Examples of typical inputs are traction force diagram, weight, running resistance, top speed etc. While passenger train configurations are usually well known, freight trains show a high variance in configuration. This means that several freight train models are needed with varying lengths, weights, speeds and so on.

Timetables are constructed by assigning train types to different routes. If a predefined timetable exists, running time calculations give information on the amount of allowance available on different sections. Normally, in Sweden, a standard allowance of 3% is added to the minimum running time in order to model small variations in driver behaviour. This supplement is used throughout all the simulations and calculations described. Prior to a simulation, the amount of usable allowance in case of delays can be set.

One advantage with RailSys or other software/method that consider block occupation is the possibility of detecting additional conflicts that may not be clearly visible if block occupation is not considered.

3.3 Simulation of timetables

Defining a timetable demands that specific station routes are set for each train. If a real timetable is modelled, the routes can be established from track occupation plans. However, in a simulation (as in reality) there is a need for alternative routing through stations and these conditions need to be realistically defined. Dispatching parameter settings create time range options for different routing decisions. Depending on the network characteristics, traffic intensity and so on, these settings may need adjustments to avoid or attain certain events in the simulations.

Additional measures may be necessary if the number of deadlocks is too high, i.e. simulation cycles cannot be completed due to situations which the software is not able to resolve. These can consist of both real deadlocks which, if they occurred in reality, would require reversing train movements or situations where limitations in the routing module create deadlocks. These problems occur mostly on networks with single-track lines and junctions with complex routing. However, the number of deadlocks can usually be significantly reduced by making adjustments both in infrastructure and route settings. This means that the actual behaviour in station routing is not always possible to model.

Dispatching is mostly based on priorities assigned to different train types. These can be related to delays, i.e. train priority can increase or decrease if a specified delay limit is reached. A trade-off can be set between priorities and expected delay for an event where a dispatching decision is required. Trains can reduce their delays by using possible allowance. The percentage available is configured prior to a simulation. Values used in a previous simulation study done on the Swedish rail network are 30–40% for freight trains and 65–75% for passenger trains [13]. This setting becomes important for freight trains, since they show a high variation in configurations regarding lengths and weights.

3.4 Discussion of the methodological problems

Depending on the purpose of the simulations and on the network model (real, imaginary or a combination) the delay modelling can vary. In some cases it can be sufficient to use some self-defined distributions, e.g. a low-high level approach. This can give some insight on recovery performance, i.e. ability to reduce delays. For simulations based on existing networks historical delay data is commonly used. Three important delay types are usually considered prior to a simulation setup.

- Entry delays (initial): Used for trains entering at a network boundary or inside the network.

- Dwell/departure delays: Mostly used for stochastic behaviour in passenger exchange times, but also personnel issues or train malfunctions.

- Line delays: Extended running times due to infrastructure malfunctions, weather conditions, decreased vehicle performance, driver behaviour.

Entry delays are usually extracted out directly from historical delay data. This principle works fine at a network boundary. For trains created inside a network this practise is a little bit rough. However, this aspect has not been studied in detail. Dwell delays are difficult to define based on normal delay registration data, instead manual measurements carried out in previous projects are used [15]. Distinction is made according to type of train system and size of station with respect to passenger amount.

Some of the delays are related to extended running times between stations. The secondary (knock-on) part is created in the simulations as reactions to other disturbances, i.e. conflicts with other trains. Realistic scenarios also demand a primary part for these type of delays. The normal train registration data records deviations from scheduled timetables at every station. Delay increase of five minutes or more between two stations or at a station is usually cause reported. This means that separation in primary and secondary causes is possible. Unfortunately, there is not much information about smaller primary delays, 1–4 min. These have been shown to be important in earlier studies [13, 15]. Fig. 2 shows the proportion of cause reported primary delays related to total number of train runs on the Southern main line for a six month period.

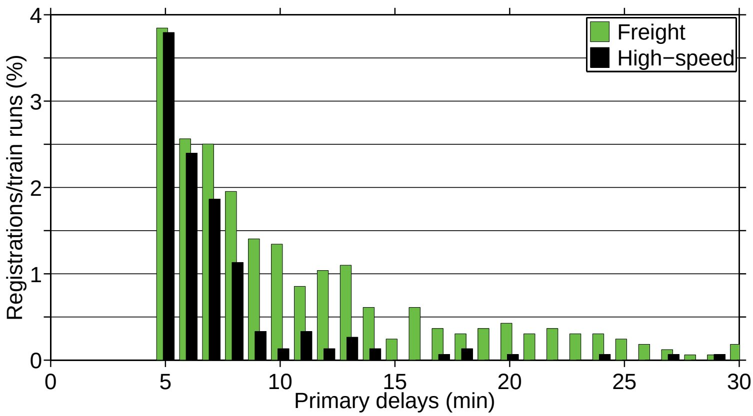


Figure 2: Primary cause reported delays for southbound trains on section Mjölby–Hässleholm (265 km).

One way of creating smaller primary delays prior to a simulation is to make some assumptions based on cause reported delays greater than or equal to five minutes and then make estimations for the 1–4 min interval. Another approach is to disregard all cause reports and construct delay distributions directly from the base data. Since this data includes both secondary and primary delays, the calculated distributions need adjustments in order to reflect a smaller part of the total delays.

3.5 Modelling running time extensions

If the actual primary delays to be modelled are small or trains traverse a short distance in the network, running time extensions may be excluded. It can then be sufficient to only model entry and dwell delays. However, in a large network with long distances the need for modelling running time extensions increases. Fig. 3 shows examples for two train groups on Southern Main Line from simulations without applying running time extensions for any train. In particular, high-speed trains do not experience similar delay development as is the case for observed values.

The significant delay increase in observed data on the last section was caused by a speed limit due to long term tunnel maintenance. In this period, no additional allowance was introduced in the timetable. The freight train group has, relatively speaking, smaller differences considering that freight operations are more difficult to model than passenger operations. This is discussed later.

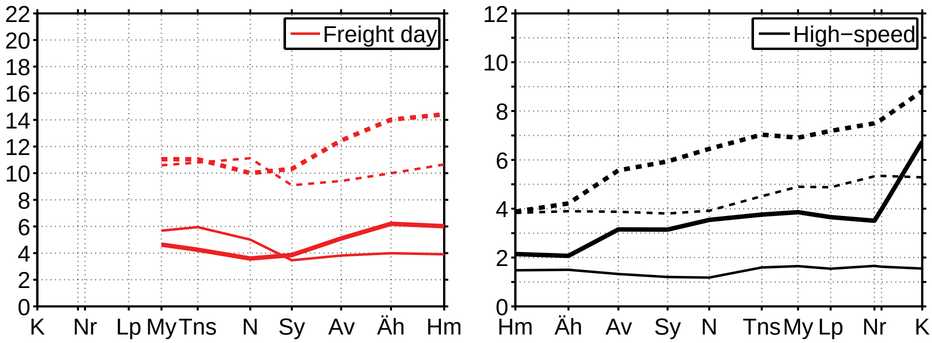


Figure 3: Simululated (thin) and observed values (bold), mean (solid) and standard deviation (dashed). No primary running time extensions are applied in this simulation.

Running time extensions have not been applied on every section between two stations. Instead the network is divided into subsections, where the boundary stations typically are junctions or stations with regular train turnarounds [13, 15, 21]. Distances for subsections used in simulations on the Southern Main Line are in the range of 35–50 km with 2–6 intermediate stations. Similar conditions apply also for the simulations performed on the Western Main Line.

Running time extensions have typically been applied in the middle of subsections. This assumption is checked by simulating cases where different approaches are used for positioning the extensions. Three additional cases are modelled and positions are set to first and last section combined with randomly pointing out a section (uniform distribution). Delay values are equal in every case.

Fig. 4 shows mean and standard deviation values for southbound high-speed and northbound daytime freight trains. The difference is marginal in the first case, although sections with high running time extensions are clearly distinguishable. The difference between delay development for the freight group is not large, in a relative sense. Levels for running time extensions are higher compared to the high-speed trains.

Due to a lower dispatching priority and speed freight trains use sidings frequently allowing faster passenger trains to pass. This procedure, if not pre-scheduled, results in knock-on delays and adds to the total delay development. Considering these four delay positioning cases, using randomisation seems to be a good alternative and most realistic considering that it models the stochastic behaviour of actual delay positions.

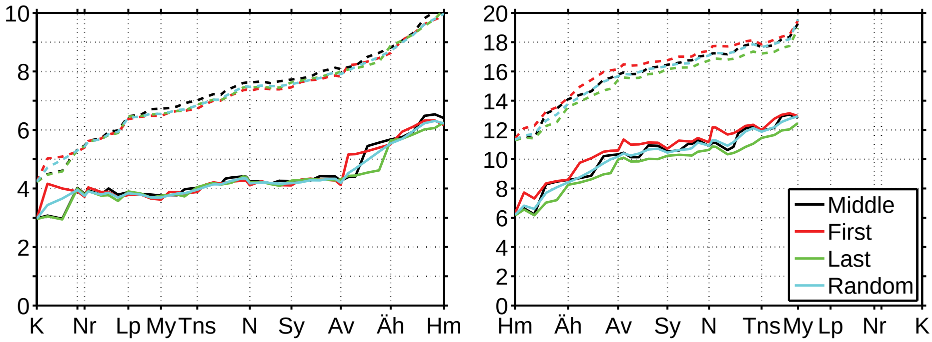


Figure 4: Simulated outcome with different schemes for positioning running time extensions on sections. Southbound high-speed trains (left) and northbound daytime freight trains (right). Mean (solid) and standard deviation (dashed).

4 Running time estimations for the Green Train

In the Green Train research program (in Swedish *Gröna tåget*) a new generation of high-speed trains are studied considering a system perspective. The objective has been to develop a concept proposal for a new, attractive high speed train adapted to Nordic conditions that is flexible for several different tasks on the railway and interoperable in the Scandinavian countries. This includes both technical, safety and environmental issues combined with market and economical analysis. Relations between vehicle configuration and infrastructure as well as operations and capacity are of special interest. Innovative solutions can improve both performance and economy which in turn contributes to better socio-economic conditions for rail traffic [8].

4.1 Background

Market and economy is a subproject within the Green Train program and deals with traffic patterns, running times and other topics which are of importance in creating an attractive product for travellers. This involves for example train interior design with comparisons between seat layouts, pricing and load factors. People's dispositions for choosing the train instead of other alternative means of travel are, in addition to comfort issues, also highly dependant on travel times and in a wider perspective on reliability, i.e. a high share of the trains should be on time, which is a quality issue.

One key factor in attracting passengers is shorter running times. Fast trains have a comparative advantage over car and air travel on medium distances, i.e. 300–600 km. Travel times can be significantly lower than for cars but also compared to air travel if airport procedures and journeys to and from airports are considered.

High train speeds usually require new infrastructure, with mostly straight tracks and large curve radii. This makes it possible to separate high speed trains from those with lower speeds and improves the capacity situation. There is also a need to control new bottleneck situations which are likely to arise in some areas and in junctions where new and old lines connect. Another possibility is to increase speeds on existing lines. This can be achieved by using modern trains with track friendly dynamics and adjusting some track parameters. Updates in other infrastructure are also necessary at some locations.

As mentioned one of the main limitations in achieving higher speeds on existing lines is curve characteristics. Most of the main lines in Sweden have some longer sections with relatively straight track geometry. Adjusting cant in curves can improve static speed profiles for certain trains. One main focus in this subproject is therefore to analyse curve radii data on six dedicated lines and use different cant and cant deficiency set values to calculate new static speed profiles.

Other parameters affecting running times are top speed, acceleration and braking performance of the train itself. This relates to tractive effort, power output, train resistance etc. An obvious impact on total travelling times comes from stopping patterns, i.e. number of stopping stations. The main objective is to vary some of the characteristics affecting running times. In this study several new static speed profiles according to cant and cant deficiency parameters are considered. In addition different start acceleration and power output settings are evaluated. Calculations are done for several different routes and stopping patterns.

4.2 Calculation of maximal curve speed

The track geometry is a crucial factor for top speed and, by extension, running times. Sharp curves, i.e. small curve radii, imply limitations on maximum permitted speeds. Using trains with carbody tilting and bogies with good running characteristics can, to some extent, compensate for sharp curves. New speed profiles are based on existing line speeds, circular curve radius data and combinations of cant (D) and cant deficiency (I). Applied cant, also called super elevation, is the amount by which one running rail is raised above the other running rail. Cant deficiency is a measure of the resulting lateral acceleration (centrifugal force) in curves and is measured in the track plane.

Circular curves have characteristics which explain some of the steps used in the calculations. The curvature is constant, i.e. the radius is constant. For a constant speed circular curves give constant quasi-static acceleration, which justifies a constant cant value. Fig. 5 illustrates some of the definitions.

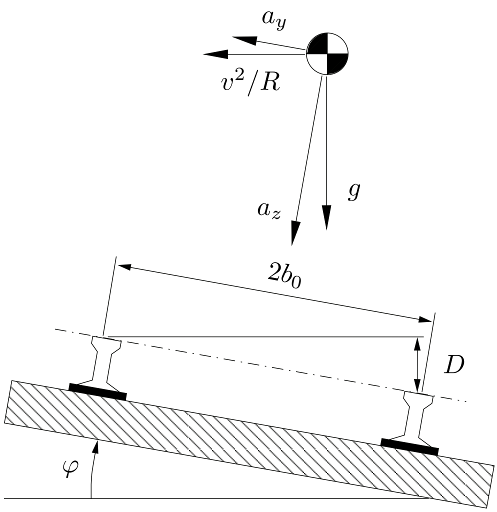


Figure 5: Horizontal (v^2/R) and vertical acceleration (g). Acceleration parallel (a_y) and perpendicular to track plane (a_z).

From fig. 5 follows that for a given speed (v), curve radius (R) and cant angle (φ) lateral acceleration (a_y) is calculated by eqn. 1. If φ is small the simplification in the second expression can be used.

$$a_y = \frac{v^2}{R} \cdot \cos\varphi - g \cdot \sin\varphi \approx \frac{v^2}{R} - g \cdot \frac{D}{2b_0} \tag{1}$$

A cant value that exactly balances the quasi-static lateral acceleration ($a_y = 0$) for a given speed and radius is called equilibrium cant (D_{EQ}) or theoretical cant. This can be estimated from

$$D_{EQ} \approx \frac{2b_0}{g} \cdot \frac{v^2}{R} = \left\{ v : \mathrm{m/s} \rightarrow \mathrm{km/h} \right\} = C\frac{v^2}{R} \tag{2}$$

Assuming standard gauge gives that $2b_0 = 1500$ mm. Note that in the right hand expression of eqn. 2 the radius is given in meters, speed in km/h and equilibrium cant in millimeters. This means that the constant $C \approx 11.8 \text{ mm} \cdot \text{m} \cdot \text{h}^2/\text{km}^2$. Cant deficiency ($I$) exists if the applied cant is lower than D_{EQ} and is a measure of the additional cant needed to obtain $D = D_{\mathrm{EQ}}$. Cant deficiency is proportional to remaining track plane acceleration (eqn. 3)

$$I = D_{\mathrm{EQ}} - D \tag{3}$$

If the cant deficiency is negative, i.e. $D > D_{\mathrm{EQ}}$, the difference is called cant excess (E). In this case there will be an unbalanced lateral force in the running plane and the resultant force will move towards the inner rail of the curve. This can give problems for slow and heavy trains since a lower speed generates a higher cant excess. Therefore a limit for maximal cant value at speed v_{f} is needed (eqn. 4).

$$D \leq E + C \cdot \frac{v_{\mathrm{f}}^2}{R} \tag{4}$$

Applied values in calculations are $E = 110$ mm and $v_{\mathrm{f}} = 90$ km/h [3]. In practise, this limitation has influence in curve radii from 1500 m and up. Additionally there should also be a limitation on cant deficiency with regard to cross-wind. In the speed profile calculations maximal cant deficiency is $I_{\mathrm{lim}} = 300$ mm up to 225 km/h (v_{lim}). Over this speed allowed cant deficiency decreases with 1 mm/km/h (k) [3]. Depending on considered alternative, the maximal cant deficiency can be lower than I_{lim}. Fig. 6 shows the principles for setting cant deficiency value for a specific curve radius.

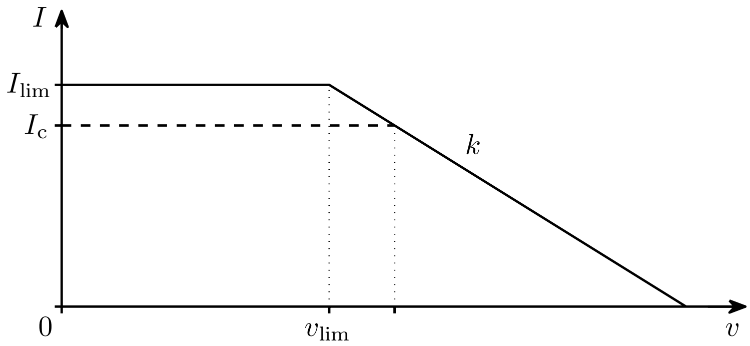


Figure 6: Allowed cant deficiency with regard to maximal case specific value, e.g. I_c, and to crosswind ($v > v_{\text{lim}}$).

Allowed cant deficiency is determined by solving for I_{cw} in eqn. 5 and taking the smallest value of I_{cw} and I_c. With this established the maximal curve speed is calculated with eqn. 6. Case specific cant value is used considering the limitation in eqn. 4.

$$I_{\text{lim}} + kv_{\text{lim}} - I_{\text{cw}} = \sqrt{\frac{k^2 R(D + I_{\text{cw}})}{C}} \tag{5}$$

$$v = \sqrt{\frac{R\,(D + \min\{I_c, I_{\text{cw}}\})}{C}} \tag{6}$$

Between a straight track section and circular curve or between two circular curves with different radius there is usually a need for transition curves. These are needed in order to achieve a gradual change of cant and curvature for comfort reasons. One characteristic is that the cant is changing as a function of the longitudinal position of the track and normally the mathematical form for a cant transition follows the mathematical form of the curvature which can be linear or non-linear. Transition curves have speed dependent limits for maximal rate of change of cant and cant deficiency [12]. In this study transition curves are omitted.

4.3 Speed profiles

Curve speed calculations are not done for all existing curves. The aim is to find sections where a constant speed can be used. Lines with many curves or variation in curve radii usually give profiles with frequent speed changes. On lines with large curve radii or a high share of sections with straight track, a high constant speed is used on longer distances. The benefit of raising an already high top speed with for example 10% over a short distance is usually small.

This principle is used when new profiles are designed. Line sectioning is mainly based on existing speed profiles. Where applicable sections are split in two or more parts. The curve with the smallest radius is limiting the maximal speed and this is used in the calculations. Speeds are not changed close to larger urban areas and on larger stations where other limitations can apply. Also other circumstances can limit train speeds, e.g. geotechnical.

Running time calculations assume that main tracks are used where possible which means that speed limitations from using diverging tracks in turnouts are avoided. However, at some positions on the analysed lines turnouts are passed in diverging positions and current speed restrictions are obeyed. Table 1 presents combinations of cant and cant deficiency used in the calculations [3]. The first two are existing profiles and considered as reference cases. Category B is for trains with no carbody tilt (P1) and category S for trains with carbody tilt (P2 and P3).

Table 1: Reference and computed speed profiles with cant and cant deficiency values (mm)

Profile	Cant	Cant def.	Information
P1	150	150	Reference category B trains
P2	150	245	Reference category S trains
P3	150	245	Carbody tilt
P4	160	165	No carbody tilt
P5	160	245	Carbody tilt
P6	160	275	Carbody tilt
P7	160	300	Carbody tilt

Speed profiles in Sweden have a base profile for category A trains. Almost all freight trains and some passenger trains belong to this category. Trains (vehicles) with track friendly running gear (radial steering bogies) are allowed to keep 10–15% overspeed relative to profile A on many lines. These are called category B trains. A maximal overspeed of 25–30% is allowed for track friendly trains with carbody tilt. Currently only X2 trains have this functionality in Sweden.

A vehicle is considered track friendly if it causes low maintenance costs on the track and on the vehicle itself. In particular this gives advantages on existing non-perfect tracks. Speeds are rounded down which means that actual overspeed normally is less than 10/30%. There are also other technical issues related to the track which can limit the overspeeds. In addition, many sections both with low and high speeds, have the same top speed for all trains due to other reasons than train category.

Maximal curve speed for some of the fictive profiles is illustrated in fig. 7. Profiles P5–P7 differ in the range of 0–15 km/h, while P4 get clearly lower speeds caused by the significantly lower cant deficiency value according to table 1. There is no difference between profiles P5–P7 in the higher speed range due to the crosswind limitation. Speeds are also rounded down to nearest five multiplicative.

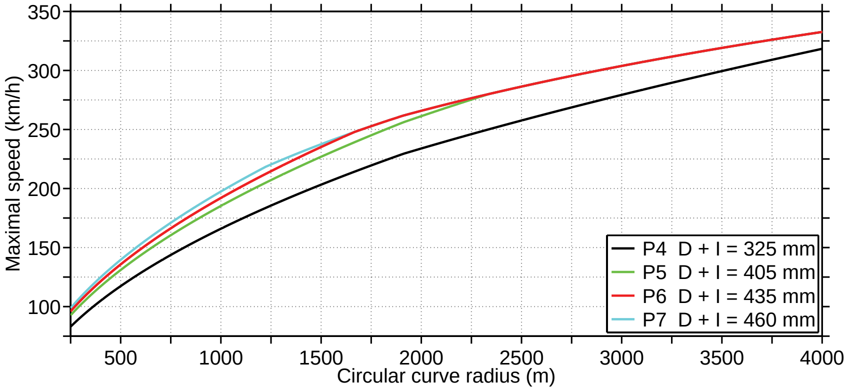


Figure 7: Maximal curve speed for different equilibrium cant values.

Fig. 8–10 show today's speed profiles for B-trains and S-trains (P1 and P2) and the calculated speed profiles without (P4) and with tilt (P6). Since the speed difference between profiles P5–P7 is small, P6 is chosen as a representative case for both speed diagrams and later in the results for running time calculations. Speed profile P6 is also judged to be most realistic among other things with regard to travel sickness.

Calculated speed profiles for Western Main Line use current speeds in the area close to Stockholm before the relatively newly built Grödinge Line with design speed 250 km/h for tilting trains (fig. 8). There are sections of varying lengths that allow higher speeds. In most cases curve radii is the limiting factor, at some stations speeds default to the current ones although radius data would allow higher speeds.

Noticeable higher speeds could be obtained between Töreboda and Skövde, over 300 km/h. This section has been used for high speed trials with a test train in the Green Train program. A movable bridge is the reason for the short speed limitation of 160 km/h at Töreboda.

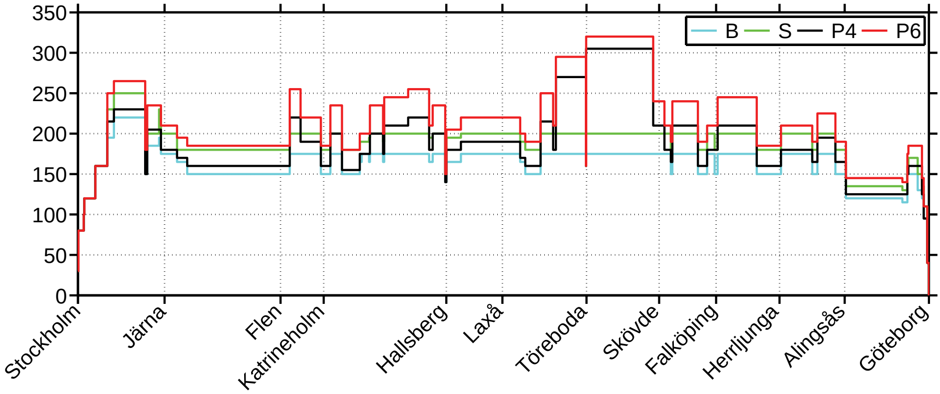


Figure 8: Possible speed profile Stockholm–Göteborg (Western Main Line), distance 455 km.

Profiles for Southern Main Line in fig. 9 are identical with fig 8 from Stockholm to Katrineholm. The section between Norrköping and Nässjö has many limiting curve radii in the range of 500–800 m, some even lower. The line is relatively straight between Nässjö and Eslöv, high-speed trains can run with at least 250 km/h. Some exceptions with speed reductions exist, mainly at stations. Existing speeds are used on the Danish side after Malmö.

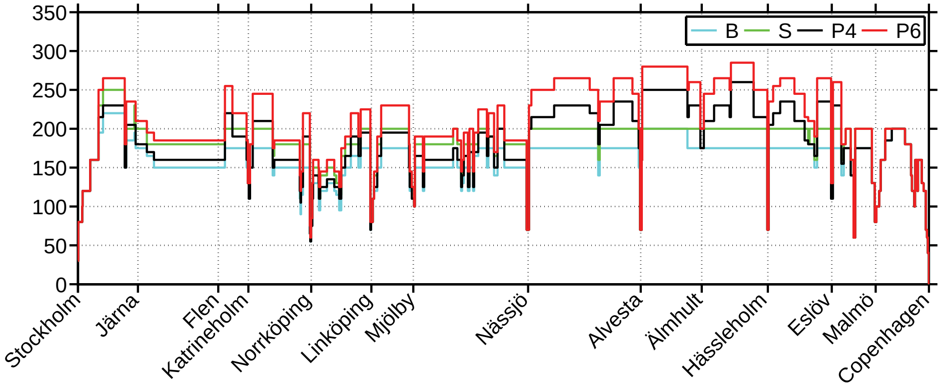


Figure 9: Possible speed profile Stockholm–Malmö (Southern Main Line) and Malmö–Copenhagen, distance 657 km.

Fig. 10 shows how the old winding section differs clearly from the converted or new sections, where speeds up to 300 km/h are possible with profile P6. Härnösand–Umeå has mostly radii around 2000–3000 m. It can be seen from fig. 10 that the split between new/upgraded and older sections is around 50%.

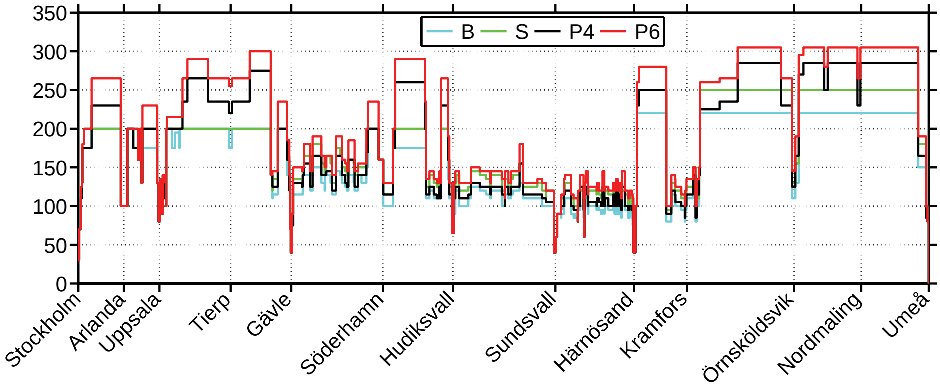


Figure 10: Possible speed profile Stockholm–Umeå (East Coast, Ådalen and Bothnia Line), distance 728 km.

4.4 Train characteristics

Performed running time calculations are made for trains with varying characteristics regarding maximum permitted speed (top speed), start acceleration and power to ton ratio. The aim is to analyse the impact on running times regarding these characteristics and stopping patterns for a number of different lines. Levels used are specified in [7].

Common for all trains are weight 360 tons, length 155 m and an addition for rotating masses with 5.6% (dynamic mass supplement). Braking characteristics are analysed as uniform deceleration of 0.6 m/s^2, which is so-called comfort braking. Deceleration is not linear in reality, but can be reasonable well approximated with a constant value. It is assumed that at least half of the axles are powered, meaning that the adhesive mass is at least 180 tonnes. Adhesion is not considered to affect the short-duration tractive power under normal conditions.

Traction force diagrams are defined for the modelled trains. Desired starting acceleration (a_s) determines the level of starting tractive effort. Available tractive power (P) is limited above a certain speed, therefore the tractive force (F) is reduced. Traction force diagrams are defined by eqn. 7 and 8 where train mass including dynamic mass supplement is used (m_e).

$$F = m_\mathrm{e} \cdot a_\mathrm{s} \tag{7}$$

$$P = F \cdot v \tag{8}$$

Aerodynamic properties of a train are important and, if well designed, can reduce required traction force, energy consumption and running times. Aspects to be considered are for example air drag, impulse resistance and aerodynamic loads caused by crosswinds and the slipstream along the train.

Aerodynamically induced forces are typically proportional to speed or the square of speed. The sum of mechanical and aerodynamic resistance ($D_{\mathrm{ma}\}}$) is approximated with eqn. 9 where $A = 2400$ N, $B = 60$ kg/s and $C = 6.5$ kg/m are vehicle constants [6].

$$D_{\mathrm{ma}} = A + Bv + Cv^2 \tag{9}$$

Gradients contribute to total resistance ($D = D_{\mathrm{ma}} + D_{\mathrm{g}}$) and will, depending on whether the train is facing a downhill ($D_{\mathrm{g}} < 0$) or uphill ($D_{\mathrm{g}} > 0$), reduce or improve acceleration characteristics. Available traction force is $F - D$ and the acceleration is calculated for a sufficiently small speed interval (eqn. 10).

$$a_i = \frac{(F_i + F_{i+1}) - (D_i + D_{i+1})}{2m_{\mathrm{e}}} \tag{10}$$

Traction force and acceleration diagrams for reference trains and some variants of the Green Train appear in fig. 11. The power assumed relates to the maximum short-duration power used for a limited time during full acceleration or electric braking. Assumed that power output is constant, a higher starting acceleration can be achieved by changing the gear ratio between motor and wheels. This would normally also give a lower top speed. However, in this study it is assumed that top speed levels are dependent on power output and running resistance, not on gear ratio settings.

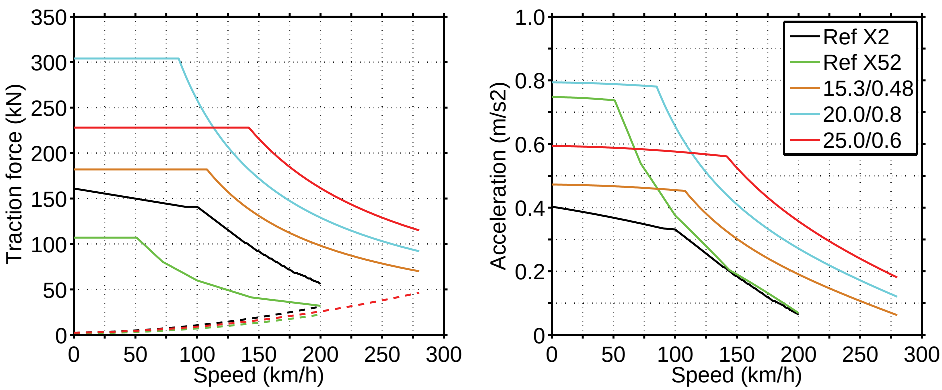


Figure 11: Traction force, total resistance with $D_{\mathrm{g}} = 0$ (dashed) and acceleration diagrams for reference trains and combinations of power output (kW/ton) and starting acceleration ($\mathrm{m/s^2}$).

Power requirements are usually defined by desired residual acceleration, i.e. acceleration level at top speed. The difference between obtained residual acceleration is clearly seen in fig. 11. Gradients affect the acceleration and can, depending on the distribution of negative and positive gradients, increase or decrease acceleration times. Trains which have a low residual acceleration can face problems in reaching or keeping top speed even at relatively moderate positive gradients. Fig. 12 shows acceleration times for selected trains on a flat track and with gradient. In this case a power to ton ratio of 15.3 kW/ton almost doubles the acceleration time for reaching 280 km/h with a 6‰ gradient compared to the flat track case.

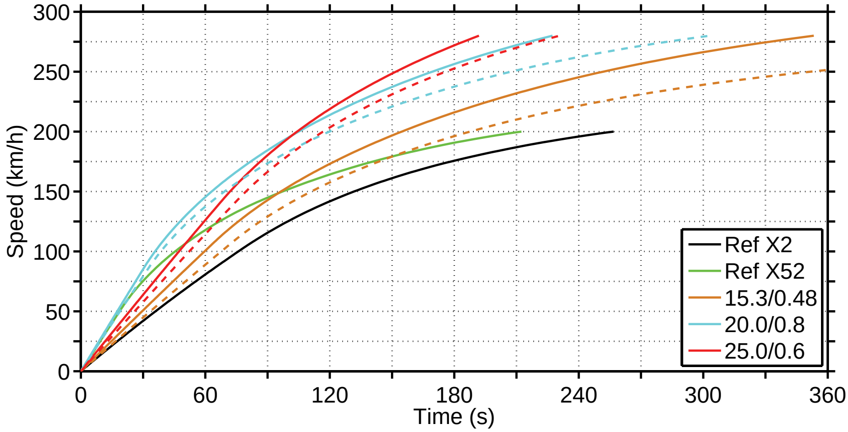


Figure 12: Acceleration time for reference trains and combinations of power output (kW/ton) and starting acceleration (m/s^2), flat track (solid) and gradient 6‰ (dashed).

4.5 Running time calculations

RailSys is used for running time calculations. For this purpose it is sufficient with a simplified infrastructure model, i.e. main tracks, stopping positions, signals and gradients. However, it is essential to set correct distances and speed profiles. Lines are modelled according to the situation year 2006–2007 including on-going construction projects of which some have been completed until year 2012. Vehicle models are defined according to the steps described in section 4.4 and assigned to the defined speed profiles.

Cases with different stopping patterns are defined for the studied relations. Generally most of them have one stopping pattern with few or no intermediate stops and one including several stops. Station dwell times include passenger exchange times and in some cases additional time for train meets on single-track sections. Patterns and dwell times are defined in [7].

Example timetables are presented in table 2–5 which show departure times for reference trains and for one with power output 20.0 kW/ton, starting acceleration 0.6 m/s^2, top speed 250/280 km/h and speed profile with 165/275 mm cant deficiency (P4 and P6 in table 1). Station values are rounded off to full minutes. Column *Time* shows unrounded total times including dwell times. Value $+T$ indicates added seconds. Note that all stops do not fit in table 5.

Table 2: Western Main Line, Stockholm–Göteborg, non-stop

Profile	Train	Time	+T	Stockholm	Göteborg
P1	X52	02:58:13	47	0.00	2.59
P2	X2	02:37:30	30	0.00	2.38
P4	250	02:40:53	7	0.00	2.41
P4	280	02:40:02	-2	0.00	2.40
P6	250	02:23:05	-5	0.00	2.23
P6	280	02:21:56	4	0.00	2.22

Table 3: Southern Main Line, Stockholm–Malmö–Copenhagen, 3 stops

Profile	Train	Time	+T	Stockholm	Hässleholm	Lund	Malmö	Copenhagen
P1	X52	04:32:49	131	0.00	3.36	4.05	4.15	4.35
P2	X2	04:09:31	89	0.00	3.14	3.41	3.51	4.11
P4	250	04:06:48	132	0.00	3.13	3.39	3.49	4.09
P4	280	04:06:41	139	0.00	3.13	3.39	3.49	4.09
P6	250	03:45:17	103	0.00	2.54	3.18	3.28	3.47
P6	280	03:43:10	110	0.00	2.52	3.16	3.26	3.45

Table 4: Eastern Link/Southern Main Line, Stockholm–Malmö–Copenhagen, 3 stops

Profile	Train	Time	+T	Stockholm	Hässleholm	Lund	Malmö	Copenhagen
P1	X52	04:03:04	116	0.00	3.06	3.35	3.45	4.05
P2	X2	03:51:24	36	0.00	2.55	3.22	3.32	3.52
P4	250	03:34:51	129	0.00	2.41	3.07	3.17	3.37
P4	280	03:31:26	154	0.00	2.38	3.04	3.14	3.34
P6	250	03:22:07	113	0.00	2.31	2.55	3.05	3.24
P6	280	03:16:30	90	0.00	2.25	2.49	2.59	3.18

Table 5: East Coast Line/Bothnia Line, Stockholm–Sundsvall–Umeå, 10 stops

Profile	Train	Time	+T	Stockholm	Uppsala	Gävle	Söderhamn	Sundsvall	Härnösand	Örnsköldsvik	Umeå
P1	X52	05:43:23	157	0.00	0.33	1.17	1.57	3.19	4.07	5.05	5.46
P2	X2	05:23:20	280	0.00	0.33	1.16	1.52	3.06	3.49	4.46	5.28
P4	250	05:13:10	230	0.00	0.31	1.10	1.48	3.04	3.49	4.41	5.17
P4	280	05:09:43	197	0.00	0.31	1.09	1.47	3.03	3.48	4.39	5.13
P6	250	04:50:14	286	0.00	0.30	1.07	1.41	2.51	3.31	4.20	4.55
P6	280	04:44:25	335	0.00	0.30	1.06	1.40	2.49	3.29	4.17	4.50

It is clear that an increase in top speed from 250 to 280 km/h give small time gains in relation to total time and distance in the examples. Profile P4 is aimed for non-tilting trains and in order to fully benefit from the higher top speed, curve radii must increase from around 2300 to 3100 m (fig. 7). This is mostly obtained on recently built lines. The situation is similar for trains with carbody tilting running on profile P6, an increase of top speed gives relatively small time gains in the presented cases. Changing from profile P4 to P6 (increasing allowed cant deficiency) gives the most significant improvements. The same effect is also seen for the reference trains comparing profile P1 and P2.

A better understanding of how running times vary regarding to train configurations and speed profiles is obtained by making calculations based on table 6. Depending on which line is studied 16 different power and acceleration settings are used for 2–3 levels of top speed and cant deficiency. Speed 320 km/h is only used on cases with future high-speed lines, for this speed 10.8 kW/ton is not useful. In these cases the highest power output is 30.0 kW/ton.

Table 6: Train configurations (the number of combinations is 60 but depending on the case some of these are not considered), top speed 320 km/h and P = 30,0 kW/ton used on cases which include Eastern Link and/or Götaland High-speed Line

Parameter	Levels				
Top speed (km/h)	250	280	320		
Power output/mass P (kW/ton)	10.8	15.3	20.0	25.0	30.0
Starting acceleration a_{s} (m/s^2)	0.48	0.6	0.8	1.0	

Running time calculations are presented in diagrams corresponding to cases in table 2–5. Train configurations are shown for starting acceleration levels and power output groups. Running times include stops but no dwell times, i.e. acceleration and braking at stopping stations is modelled. A generally usable train should have performance that gives short running times on tracks of different character.

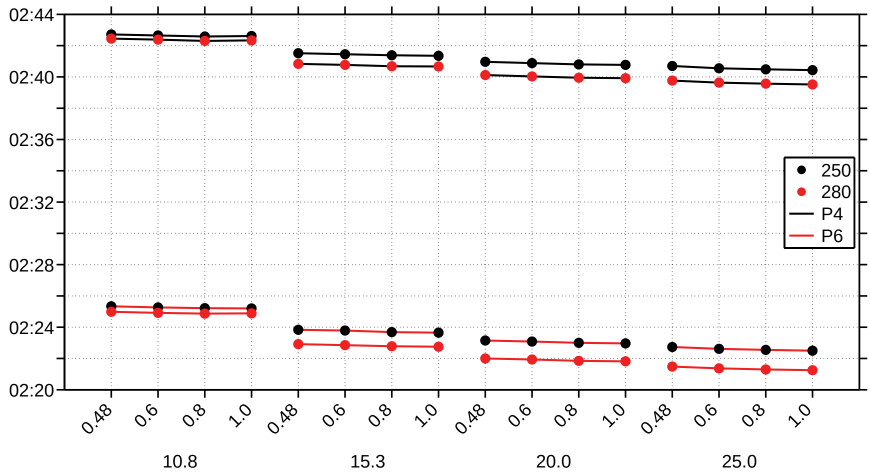


Figure 13: Western Main Line, Stockholm–Göteborg, non-stop. Tilting is important for reducing running times. Power output, starting acceleration and top speed have low elasticity, due to non-stop operation.

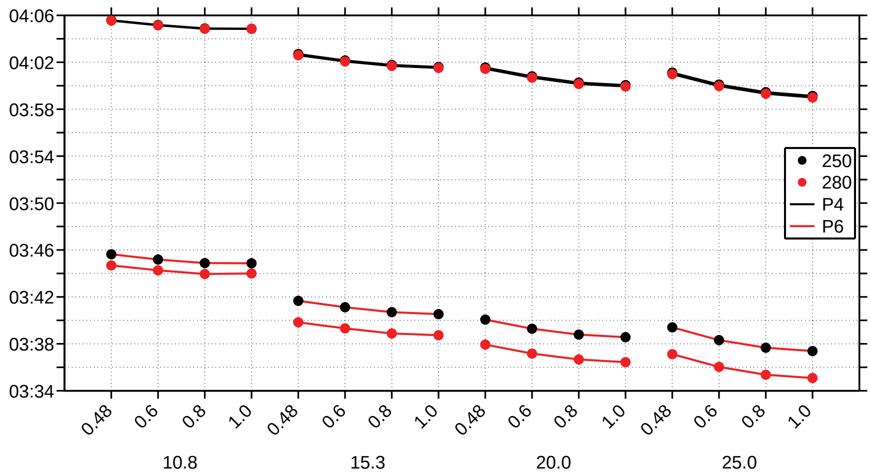


Figure 14: Southern Main Line, Stockholm–Malmö–Copenhagen, 3 stops. Tilting is important, starting acceleration has low elasticity. Almost no sections with higher speed than 250 km/h exist for profile P4 (fig. 9). Increasing power output from 10.8 kW/ton has a clear influence, especially for the higher speed level.

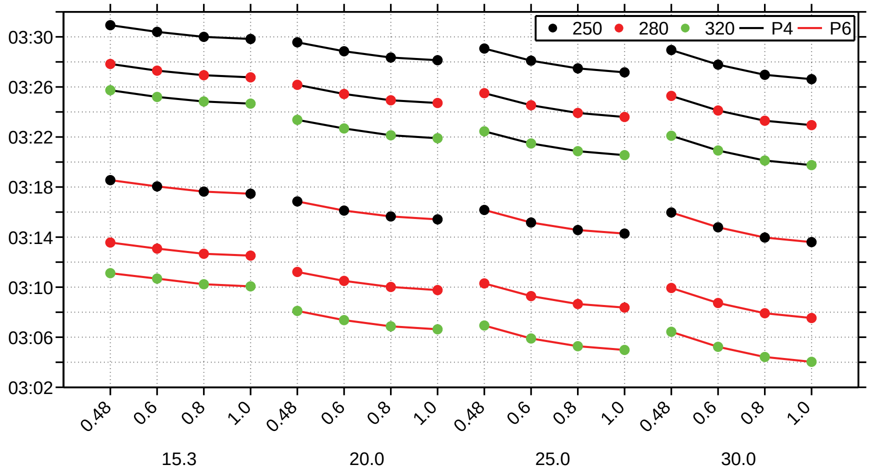


Figure 15: Eastern link/Southern Main Line, Stockholm–Malmö–Copenhagen, 3 stops. Relatively high elasticity for speeds and cant deficiency, low for power output and starting acceleration. Eastern Link allows high speeds for both tilting and non-tilting trains but since main part of the distance is on Southern Main Line, changing from profile P4 to P6 have a higher impact on running times than a speed increase.

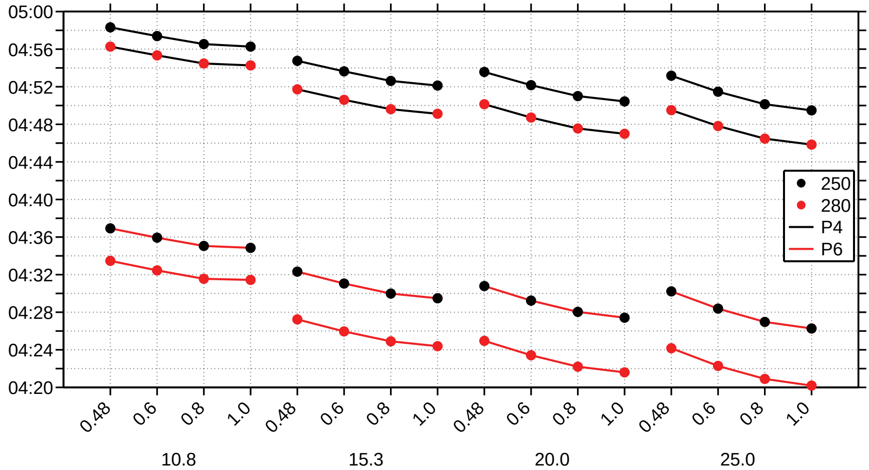


Figure 16: East Coast/Bothnia Line, Stockholm–Sundsvall–Umeå, 10 stops. Difference between cant deficiency levels is in the range 22–26 min, comparing trains with equal configurations aside from tilting. Power output in the range 15.3–25.0 kW/ton has relatively low elasticity. Some sections, especially on the northern part, enable speeds over 250 km/h resulting in time gains comparing the top speed levels.

4.6 Conclusions

The impact train characteristics have on running times varies depending on the speed profile. Lines with high continuous top speeds impose other requirements on trains than lines with shifting speeds. To make use of shorter sections with higher speeds a high power output is needed since accelerations in higher speed ranges are time consuming. Requirements on residual acceleration is a typical constraint on power output, especially for uphill gradients. This study assumes that an increasing starting acceleration, everything else being equal, does not limit top speed. Thus, acceleration time to full speed always decreases if starting acceleration is increased. Considering starting acceleration and power output, usually the first step from 0.48 to 0.6 m/s^2 and 10.8 to 15.3 kW/ton gives the most observable time gains.

According to fig. 7 top speed level is only marginally improved by increasing cant deficiency from 275 to 300 mm. The lower limit is also a normal value for express trains in Europe. Cant deficiency only affects comfort and the values are considered to give an acceptable centrifugal force in the train's passenger compartment given carbody tilting. On almost all of the investigated lines changing from non-tilting to tilting speed profile gives the highest time gain.

On newly built and future lines this has less effect, instead top speed levels show a high elasticity. The biggest advantage with tilting trains is observed on lines with small or medium curve radii, this advantage decreases on lines with large radii. Stockholm–Göteborg is a typical relation where trains without tilting functionality are not an option considering the running time differences between profile P4 and P6 or between the reference trains. Similar results are presented in [17].

As mentioned earlier, braking is not varied. However, increasing deceleration can give similar improvements as increasing starting acceleration and is useful on services with many stops and on lines with shifting speed profiles. Acceleration and braking performance are also important in real operations since they can be used to reduce delays. These characteristics can in general be kept on a moderate level not to decrease the passenger's comfort.

5 Timetable adjustments for high-speed trains

The introduction of tilting trains ($v_{\max} = 200$ km/h) in Sweden 1990, enabled significant running time improvements on section Stockholm–Göteborg and between Stockholm and Malmö, i.e. between Sweden's capital and the second and third largest cities. Further infrastructure improvements have contributed in shortening running times. However, for the last ten years demand for passenger train services has increased, the same holds for freight transport on rail. Several commuter train systems have expanded, both in frequency, distance and number of stops. The increased traffic also contributes to disturbance sensitivity. In this study simulation is used to evaluate how timetable adjustments for high-speed trains affect their on-time performance.

5.1 Background

The Western Main Line links Stockholm with Göteborg, the distance is 455 km. It is highly utilised on most sections. Commuter trains operate in both ends and several regional train systems use parts of the line. There are also InterCity-trains on the same relation, although some using a partly different route. Freight operations are significant, especially on the western part between Hallsberg and Göteborg. The speed mix or difference between train speeds is high on most parts. This limits capacity and adds to the disturbance sensitivity. In particular trains using higher speeds risk a relatively high delay increase when catching up slower trains or when affected by disturbances in general.

High-speed trains (X 2000) have 15–17 departures on a normal weekday. Some seasonal variations exist and regarding different weekdays. On-time performance, a generally used measure describing outcome in rail networks, for high-speed trains has dropped under the last years. The performance of train traffic is also much more focus today than before. This has lead to several studies, investigations and debates on how performance in the Swedish rail network can be improved.

5.2 Method

This study is based on an infrastructure model used in a previous study where the focus was to make cause specific reductions in delay distributions and evaluate the impact on on-time performance [15]. Specifically, the aim was to reduce delays caused by infrastructure, vehicle and operator related problems. Some rebuilds and additions are made in the model, including balises and speed settings. Delay input (distributions) is reused from this earlier study with additions from another simulation study done on the same line [13].

Parts of connecting lines are added in order to capture the potential conflicts occurring at some stations due to in- and outflow on the main line. The influence can be even more substantial considering that some are single-track lines. Whether this approach is necessary or not depends for example on the frequency of crossing conflicts and limitations on available station tracks. The time interval between outgoing and incoming trains is also modelled in a more realistic way.

The focus is mainly on adjusting timetables and evaluating outcome for high-speed trains. Applied delay distributions and dispatching parameters are not changed. High-speed train timetables between Stockholm and Göteborg have in general a 3% driver allowance based on the minimum running time. Node allowances are used to maintain a certain quality even if small disturbances occur, e.g. extended dwell times, variation of speed due to restrictive signal aspects, etc. These are usually applied in the range of 4 min per 200 km, resulting in 7–8 min between Stockholm and Göteborg. Rounding up to full minutes adds to the total allowance.

The modelled timetable is considered to represent a normal weekday, Thursday 29th January 2009, full traffic is assumed Passenger train timetables are usually similar comparing different weekdays but freight train operations can have significant variations. Three alternative simulation scenarios for high-speed trains are defined, aside from the reference case:

- Reference, scheduled timetable
- Decreased timetable allowance with 4 min
- Increased timetable allowance with 4 min
- Increased buffer times separating high-speed and other trains

Decreased and increased allowances are applied in the way that half of the timetable adjustments are made on section Stockholm–Hallsberg and the other half on section Hallsberg–Göteborg. Buffer times are increased so that high-speed and other trains both are separated by at least 5 min on lines and at stations with overlapping routes, this refers to timetable time. If block occupation time is compared the buffer time decreases. A buffer time of 4 min is allowed in situations where, for example, a freight train is scheduled to a side track followed by a high-speed train on a main track. Timetable for year 2009 had buffer times down to 3 min regarding high-speed trains at some stations. Buffer times on the bottle-neck south of Stockholm Central Station are not considered, these are 2–3 min.

Available data regarding real outcome is from August 2006 to January 2007. Simulation of the reference timetable is compared to this data, some adjustments are made to delay distributions and additional trains due to the extended network. Adjusting the timetable is made with the aim of minimising number and magnitude of changes for other trains. However, it is not possible to avoid disrupting regular interval passenger train timetables and in some cases completely reschedule freight trains.

5.3 Results

On-time performance diagrams are presented on aggregated level for all high-speed trains with respect to the different timetable cases. Fig. 17 shows on-time performance values for simulated reference timetable and real outcome. Construction work at Gnesta station caused significant speed restrictions in the period from where observed values are obtained. As a result, on-time performance decreased locally.

On-time performance increases close to the end station in direction Stockholm–Göteborg. This is due to allowances placed on this section, speeds are lower here (profile S in fig. 8). A similar allowance exists on the bottle-neck between Stockholm Central and Stockholm South Stations, a double-track section of around 2 km. There are also other differences between these two timetables, both for high-speed trains and other trains. However, the purpose of this comparison is to check the behavioural trend of the simulated on-time performance.

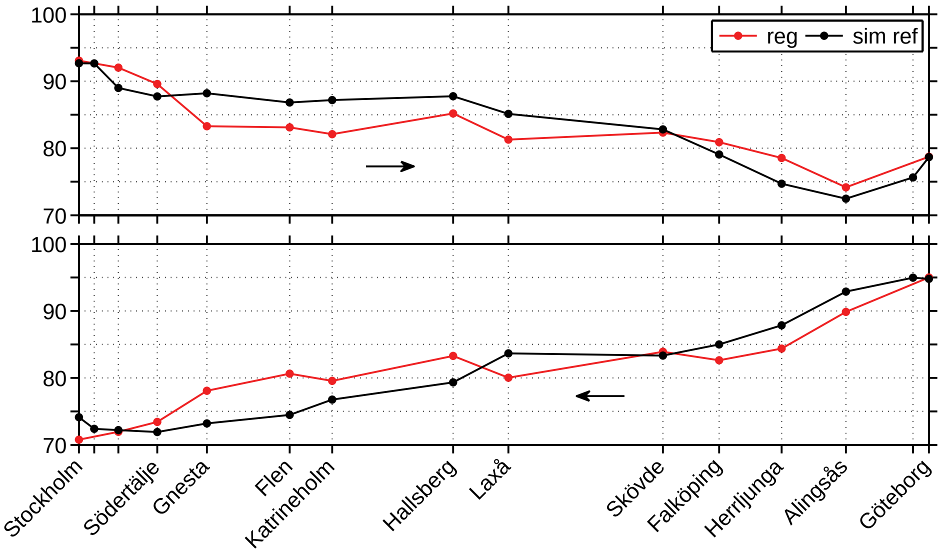


Figure 17: Observed (reg) and simulated (sim ref) on-time performance (%), 5-minute level, both directions.

Some sections are problematic, especially if compared to the relative distance. This can be explained by both dense traffic with limited overtaking possibilities and also by tight schedules with regard to minimum running time. Fig. 18 shows that on-time performance at end station decreases with approximately 10 percentage points in both directions with decreased allowances.

If allowances are increased improvements are higher for trains from Göteborg to Stockholm compared to the other direction, 10 and 5 percentage points. This difference in directions gets clearer comparing the case with increased buffer times. Improvements are also in this case considerably higher in northbound direction, for southbound trains the difference is marginal. One reason for this is that in this direction, more timetable changes are needed to satisfy the minimum buffer time requirement. Thus, there may be a higher potential for improvement.

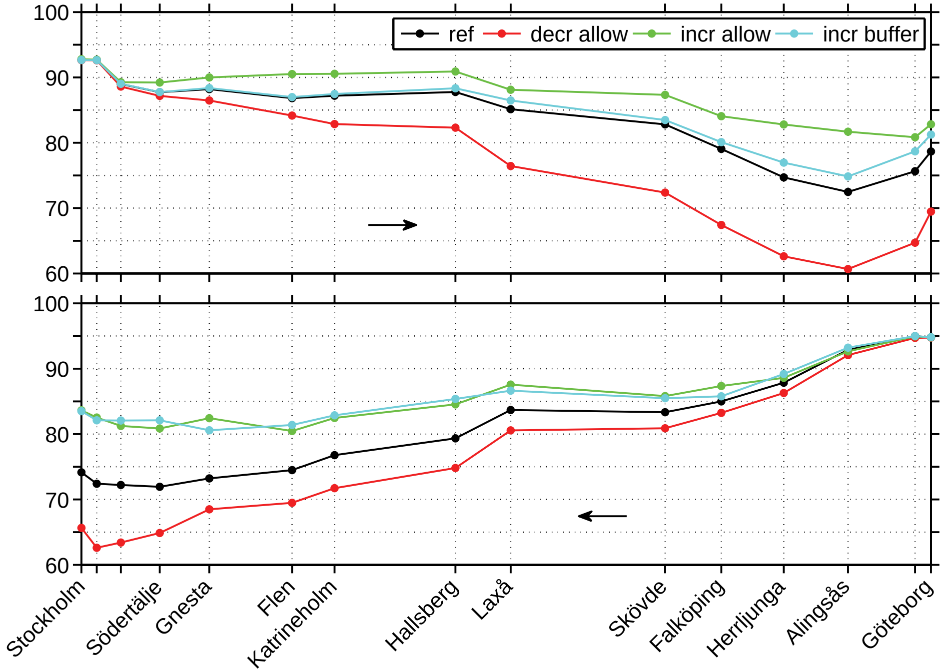


Figure 18: Simulated on-time performance for timetable cases (%), 5-minute level, both directions.

Simulated mean and standard deviation for arrival values are generally lower in Stockholm compared to Göteborg. A late high-speed train faces the risk of ending up behind a commuter train on the section Alingsås–Göteborg (45 km). These trains make several stops and overtaking possibilities are limited. In the simulations, no overtakings between passenger trains are allowed on this section. The situation for northbound trains close to Stockholm is not as sensitive compared to southbound trains close to Göteborg. The speed mix is lower and most high-speed and regional trains have the same number of stops.

Fig. 19 shows the difference in mean and standard deviation arrival values between reference and alternative timetables for each high-speed train. Positive values indicate improvements. Mean arrival values for the reference case are typically in the range of 2–5 min in Stockholm and 3–6 min in Göteborg. Corresponding standard deviations are in the range of 6–10 min in Stockholm and 10–13 min in Göteborg. Both directions have 3–4 trains that fall outside these intervals, southbound direction has the highest values. Fig. 19 can be compared with fig. 18, generally the case with increased allowance shows the best improvement for standard deviation, the difference for mean values is not as clear.

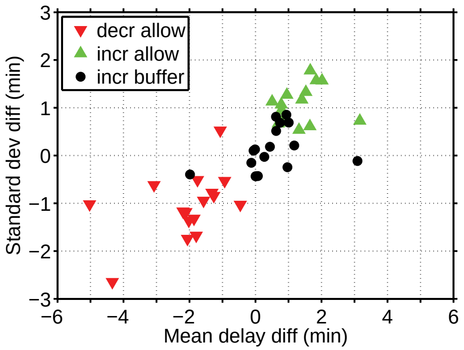


Figure 19: Difference in mean and standard deviation for arrival values between reference and alternative timetables, Göteborg (left) and Stockholm (right).

Changing from aggregated to individual level helps in finding specific trains and sections that may have a relatively big influence on aggregated on-time performance levels. Fig. 21 shows the situation for some trains, both reference and the case with increased buffer times. Two of the trains have significant drops in on-time performance between Gnesta and Södertälje. In these cases, time interval between passing high-speed train and departing commuter train is short. The common section is between Gnesta and Järna (17–18 km). For a late high-speed train there is a risk of getting behind the commuter train, instead of the other way around. There is no overtaking possibility in the simulations and it is also hard to accomplish in reality. Commuter trains make one intermediate stop on this section.

Fig. 20 shows outcome for one weekday in 2008, scheduled paths are shown as reference. In the first case a commuter train departure is pushed and the delayed high-speed train do not get an additional delay. The train sequence is changed in the second case, which causes a knock-on delay on the high-speed train. Two high-speed trains from different origin stations are scheduled successively (fleeting) and the first one is late in both cases, which may have resulted in a knock-on delay for the second high-speed train prior to this section.

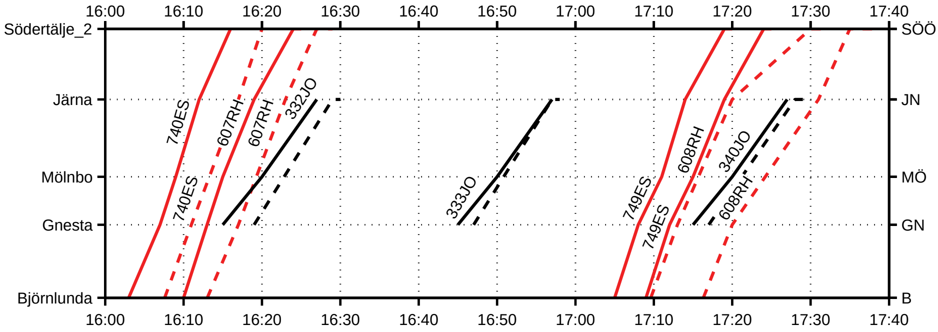


Figure 20: Real outcome with passing high-speed trains and departing commuter trains in Gnesta. Scheduled (solid) and observed values (dashed).

This scenario with small buffer times (3–4 min) at Gnesta occurs for six high-speed trains from Göteborg in the reference timetable. In the case with increased buffer times, the commuter trains in question are pushed. The effect of this is shown in fig. 21 and the level of on-time performance can be maintained to the end station without the reduction occurring in the reference timetable. The other exemplified trains show improvements in on-time performance as well. Considering all trains, the majority get improvements. Trains that already have a high on-time performance get only marginal or no changes, indicating that these trains already have a good slot regarding buffer times.

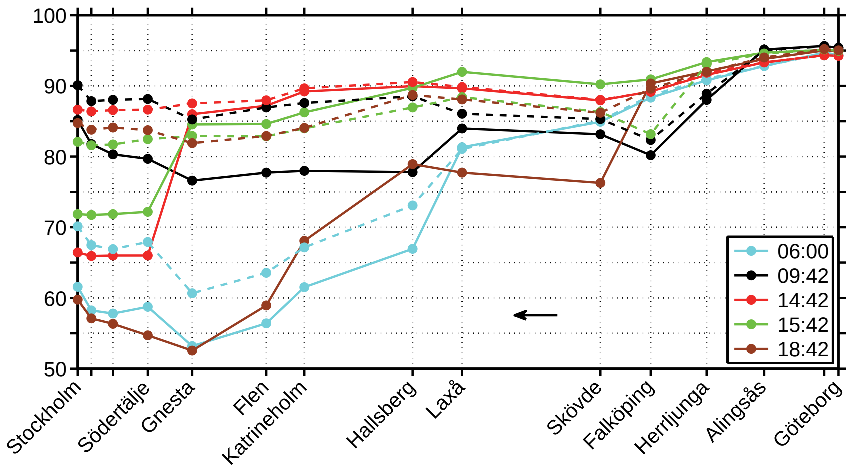


Figure 21: Simulated on-time performance (%) for selected northbound trains, reference timetable (solid) and case with increased buffer times (dashed), 5-minute level. Legend shows departure time from Göteborg.

A comparison is made by simulating hypothetical scenarios where some or all applied primary delays are removed (fig. 22). This gives an impression of how the different primary delay types affect punctuality. Simulations without entry or initial delays have the same offset at both start and end station. Studying on-time performance on one level only hides information, e.g. how values are distributed for punctual and unpunctual trains. In order to get a more comprehensive picture, on-time performance should be combined with distribution diagrams or mean delays.

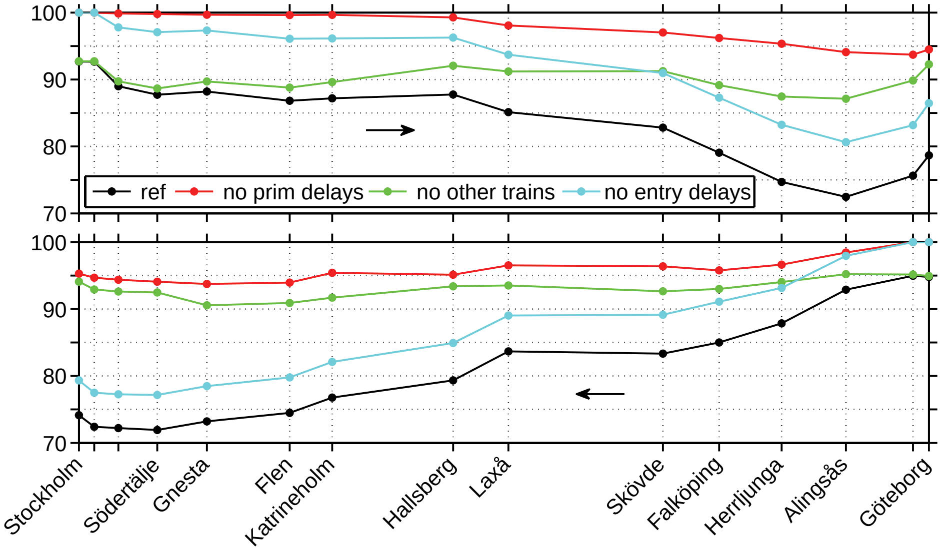


Figure 22: Simulated on-time performance (%) for reference timetable, case with no primary delays on high-speed trains, case with active primary delays but no other trains and case without entry delays, 5-minute level, both directions.

The high-speed trains seem to have enough allowance to handle their own primary delays if no knock-on effects from other trains are considered, i.e. on-time performance level is approximately the same at start and end station in both directions (fig. 22). The alternative with all primary delays inactive for high-speed trains, everything else equal to reference case, shows the level of knock-on delays caused by other trains. Since also entry delays are inactive, the trains are able to start in their planned slots and a majority of trains will also keep their slots almost for the entire route.

5.4 Conclusions

Simulations show that decreasing allowance with 4 minutes gives that on-time performance drops from 79 to 70% in Göteborg and 74 to 65% in Stockholm. The opposite scenario, with increased allowance, can improve on-time performance with 5–10 percentage points depending on direction. Better results are obtained on the aggregated level for northbound trains. However, on-time performance is better for southbound trains in the reference case.

If buffer times are increased between high-speed and other trains, on-time performance for arrival values in Göteborg improves from 79 to 83% and in Stockholm from 74 to 84%. Northbound direction shows a better improvement which can partly be explained by the directional difference in the amount of changes. Obtaining the defined buffer time limits required more changes for northbound trains in the reference timetable, thus implying that the relative improvement could be more pronounced.

Going from an aggregated to individual level shows that some trains suffer from large drops in on-time performance or increased mean delay on some sections. This can both be explained by locally tight situations with other trains, e.g. slower commuter trains coming in front of delayed high-speed trains, and also by the difference in real and modelled dispatching. Increasing buffer times between passing high-speed and departing commuter trains in Gnesta give clear improvements in the simulations. This strategy has also been adapted in reality and used in year 2010 timetable. Some aspects concerning on-time performance on this section and crossing train routes in Gnesta are discussed in [1].

Increased buffer times between high-speed and other trains is a strategy that could be used in gaining improved on-time performance. However, the space for other trains reduces meaning that some trains may not fit in the considered time space. If buffer times for other trains are squeezed making them more sensitive to delays with propagation effects this can eventually also affect the high-speed trains negatively. Further investigations of this can contribute in finding a realistic trade-off that can be used and give satisfactory results for the whole timetable.

On-time performance is checked on an aggregated level for other passenger trains and freight trains on the same evaluation stations as for high-speed trains. The amount of trains varies from station to station since no chaining is done. Results of this show only marginal differences for passenger trains, on-time performance for freight trains have the same behaviour on most stations. Improvements are achieved in the case of increased buffer times at some locations. A more detailed description of the results from this work is presented in [21].

6 Estimating running time extensions

One of the difficulties in preparing input to simulations is to design realistic running time extensions, i.e. primary delays that are applied on trains between stations. The importance of using this type of disturbances is emphasized in previous simulation studies in Sweden [13, 15]. If only dwell extensions and initial/entry delays are modelled the delay increase (punctuality decrease) along a studied line typically does not reflect the historical data it is compared with. This depends of course on other aspects as well, e.g. traffic density, available allowance etc. In this section a method for estimating primary running time extensions from historical data is presented.

6.1 Background

Run time extensions are used for modelling variations in travel times from one station to another. These extensions can have several causes, e.g. decreased acceleration performance caused by weather related or vehicle problems, line speed restrictions, signalling failures etc. Freight train configurations not matching their planned schedule, e.g. higher train weight and lower top speed, can also cause delays. Knock-on delays are not causing all variation between observed and scheduled running times.

Delay statistics can be divided into two parts. General statistics show deviation relative a planned schedule, mostly at stations, and have a resolution of one minute. This data does not indicate on delay causes. The other part consists of delays with cause reports. These are based on limits where a train's running time between two stations is compared to the scheduled running time. A cause report is added, if a delay increase exceeds or is equal to a specified limit. This data can later be sorted into primary and secondary (knock-on) events.

In an ideal situation, the cause reported data would capture all delay events (small or large) and the construction of distributions for running time extensions would be easy. However, the time limit used means that smaller disturbances are not reported. This also means that delays which grow over several stations but do not reach up to the time limit between registration points does not necessarily get reported, although the final delay may be large. Cause reported data available in this study is on a five minute level, i.e. delay increase has to be at least five minutes in order to get a cause report.

In this study, cause reported delays are not used, instead delay distributions are estimated from regular statistics based on train registrations. The idea is to change the initially obtained distributions, since these contain a mix of both primary and knock-on delays and apply this in a simulation model on parts of the Southern main line. The evaluated part consists of the section Katrineholm–Hässleholm, although the simulated network is extended and includes some of the connecting lines.

6.2 Method

The data used for estimating running time extensions is based on general delay statistics (registration data) from January to June 2008. A regular Thursday in January 2010 is chosen as timetable. Additionally trains from the previous and following night are defined making the simulated timetable 32 hours in total. Connecting lines are included to the first station with the aim of capturing crossing effects on the main line and possible time differences between outbound and inbound train movements.

The main line is divided into sections of 35–50 km in such a way that most trains using a certain section traverse from station A to F or F to A (fig. 23). Delay statistics is received for junctions and turnaround stations defining the sections and network boundary stations. For example, in fig. 23 there is no data available for stations B to E. Most trains are identifiable with this approach, although a few trains cannot be entirely tracked.

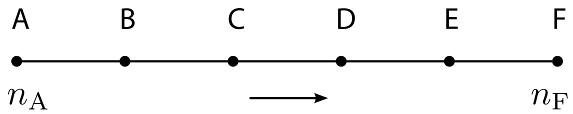


Figure 23: Definition of line sectioning with time registrations n_{A} and n_{F}.

Saturdays, Sundays and holidays are filtered out from the data. Due to a major disruption on one of the sections, causing significant delays, seven regular weekdays are excluded. Data is sorted on stations and additional processing is needed to identify paths for individual trains. Only data for trains (train numbers) running on minimum 20% of the total evaluated days are used for making distributions.

For all qualified trains the difference between n_{A} and n_{F} is considered. Positive values indicate delays and negative values mean that trains are ahead of their schedule. However, this is not modelled in the simulations and therefore negative registration values defaults to zeros. Four different cases are possible for the difference value (t) and they can be described by

$$n_{\mathrm{A}} \geq 0 \;\rightarrow\; t = \begin{cases} n_{\mathrm{F}} - n_{\mathrm{A}} & \text{if } \; n_{\mathrm{A}} \leq n_{\mathrm{F}} \\ 0 & \text{if } \; n_{\mathrm{A}} > n_{\mathrm{F}} \end{cases} \qquad n_{\mathrm{A}} < 0 \;\rightarrow\; t = \max\{0, n_{\mathrm{F}}\}$$

Base distributions with t-values are compiled for different train groups, sections and directions. Fig. 24 gives an example on observed and adapted running time deviations. In this case pure deviations are considered and not the conditions used for trains ahead of schedule discussed above. Fig. 24 clearly illustrates the varying freight train performance. Passenger train groups are mostly well defined but freight trains show significant variation in running times and train configurations.

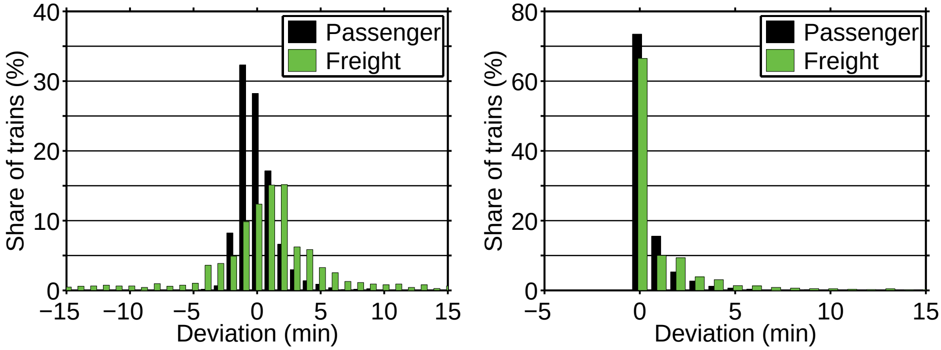


Figure 24: Deviation from scheduled running time (min). Southbound trains Mjölby–Tranås. Positive values indicate delay increase. Observed distribution (left) and partly adapted distribution for simulations (right).

Freight train grouping is based on running distance and further split into day and night groups. Distributions are reduced by keeping a certain percentage of values for every minute level, i.e. the remaining number of values are regarded as primary running time extensions. Number of registrations are kept constant meaning that removed registrations with an actual delay are considered as zero values.

Reductions are applied in four levels (20, 40, 60 and 80%) and varied separately for freight and passenger trains, resulting in 16 simulations. The results are later compared to real outcome and evaluated by using root mean square error (RMSE) for mean and standard deviation values.

Dwell and entry delays are handled as described earlier. Freight trains are matched on train numbers or time and grouped according to mean and standard deviation values in the simulations. Scripts are needed to simplify construction of input delay files used by the simulation software. An additional benefit is that random seed can be controlled meaning that other applied non-varied delays are identical in the simulations.

6.3 Results

Delay cases are evaluated by comparing simulated and real outcome and using RMSE as measure of performance (MOP). Comparison is made for six representative train groups at stations for which real delay data is available. Every group uses at least three measurement stations. Fig. 25 shows results for the applied combinations of running time extensions in southbound direction. Freight train groups have in many cases generally higher values than passenger trains, which can partly be explained by greater differences between planned schedules and actual operation. The character of the originally defined distributions are such that reductions give a higher impact compared to passenger trains.

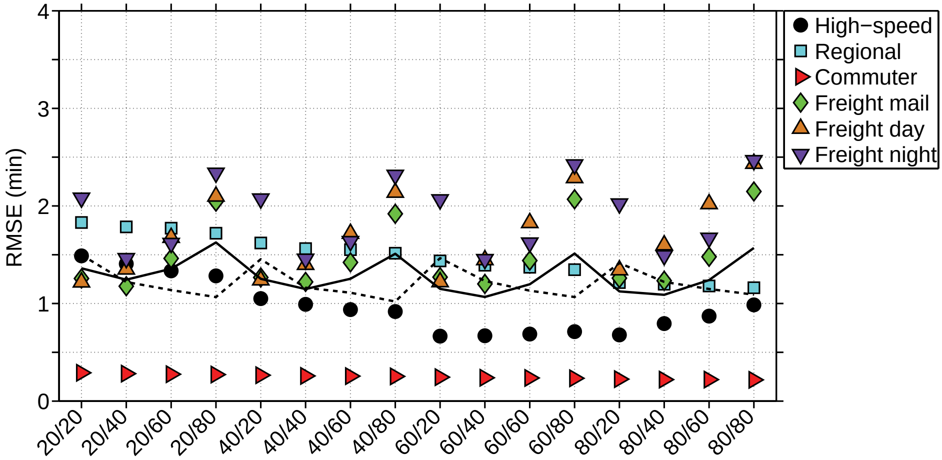


Figure 25: RMSE points for mean values. Solid line shows mean value for RMSE mean and dashed line for RMSE standard deviation. Case definition on x-axis, passenger/freight.

The commuter train group (Norrköping–Linköping–Mjölby) has almost no variation comparing the different cases. Probability getting a primary running time extension is low and varying reduction levels have little influence. This can also be an indication that allowances make up possible delays efficiently. All of the passenger train groups have little variation when reduction level is changed for freight trains. These groups are in most cases given a higher dispatching priority than freight trains.

RMSE values for northbound direction (fig. 26) show a higher degree of variance for freight trains compared to the other direction, again this is due to unreduced distributions with high probability for delays. Passenger train variation is relatively small. The reductions used are always same in both directions meaning that a recommended case should simultaneously give acceptable results in both directions. Alternatively, if one assumes a low level of interaction between directions, two different cases can be chosen. However, this should be simulated and checked.

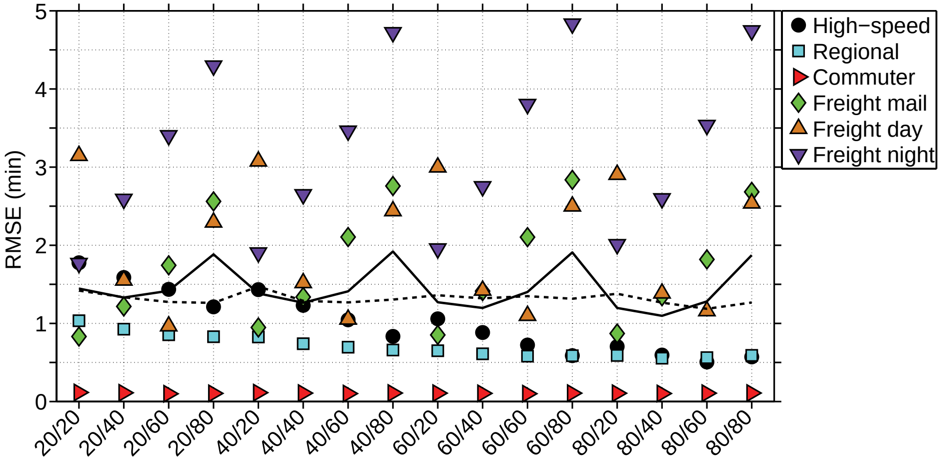


Figure 26: RMSE points for mean values. Solid line shows mean value for RMSE mean and dashed line for RMSE standard deviation. Case definition on x-axis, passenger/freight.

In fig. 25 freight trains have the lowest RMSE values if original distributions are reduced to 40%. Best fit for passenger trains is on the 60% level, especially high-speed trains are assigned more weight than the other groups. However, the freight train groups have high spread and changes from one delay level to another do not correlate well, i.e. there is no single delay level where all freight groups have minimum values simultaneously.

Fig. 25 and 26 gives only a goodness-of-fit measure and do not indicate on actual delays. Plotting delay development throughout the studied line is descriptive and possible problem areas can be identified. Fig. 27–29 show mean and standard deviation values for selected delay cases. It is clearly observable that agreement for freight trains is worse than for passenger trains. There is usually no good simultaneous agreement between day and night freight train groups and between mean and standard deviation values. In this sense northbound direction is worse.

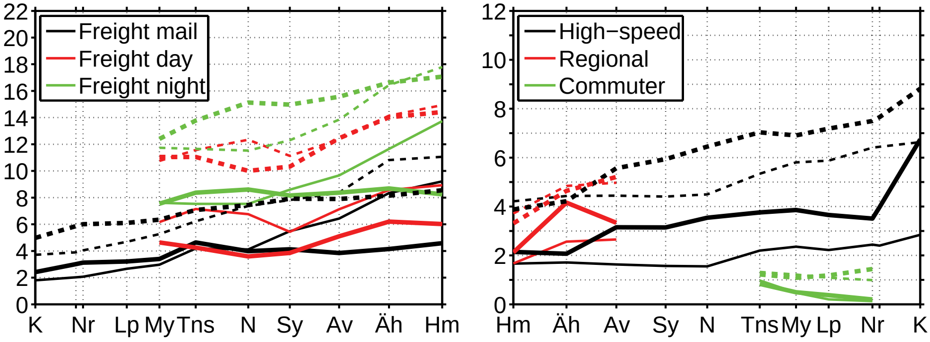


Figure 27: Simulated (thin) and observed values (bold), mean (solid) and standard deviation (dashed). Southbound freight trains 80/80 and northbound passenger trains 20/20.

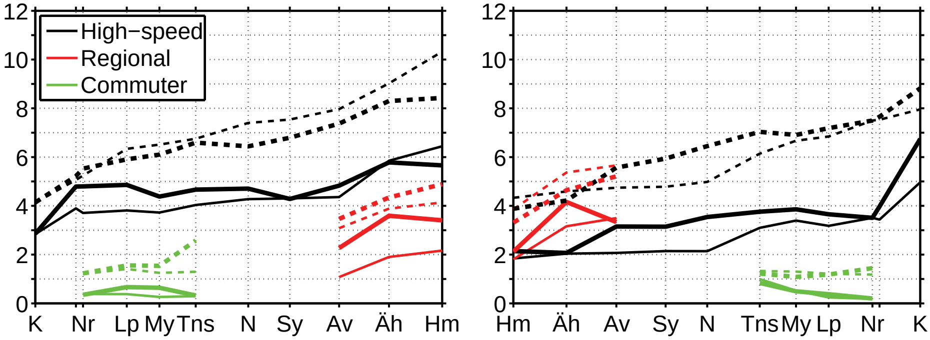


Figure 28: Simulated (thin) and observed values (bold), mean (solid) and standard deviation (dashed). Passenger trains 60/40.

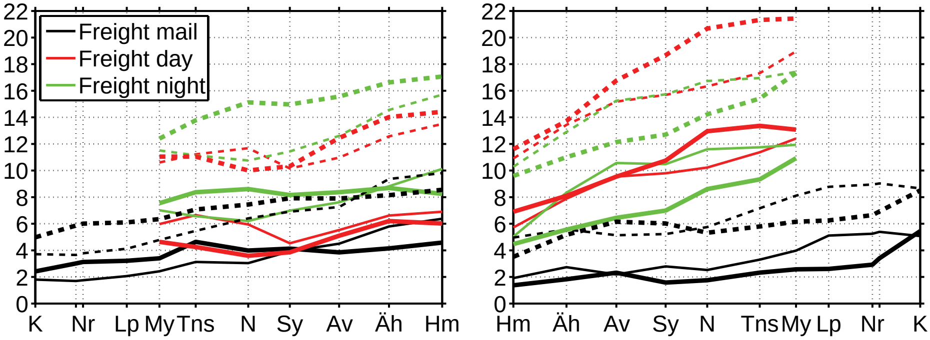


Figure 29: Simulated (thin) and observed values (bold), mean (solid) and standard deviation (dashed). Freight trains 60/40.

Comparisons between observed and simulated data can also be made by checking exit delays given an entry delay interval. This can for example give information on the possibility of reducing an entry delay and give some insight on the breakpoint, i.e. if an entry delay is in this range or higher the possibility of significantly reducing this delay before exit is small.

Table 7 shows the number of registrations used for the intervals in fig. 30. Simulation values represent case 60/40 according to fig. 25 and 26. In the lower start delay intervals, many of the trains are able to remain in the same interval on exit, alternatively reduce their delay completely. The higher start delay intervals show a larger spread on exit, the relatively small number of registrations adds to this.

Considering freight trains, they are both prone to delay increase and in the same time have the possibility of reducing delays considerably. The first issue is due to their dispatching priority compared to passenger trains, this holds especially in the simulations. Many of the so-called timetable technical stops for freight trains can be cancelled or at least reduced in time due to the operative scenario. This makes it possible to catch up on delays.

Table 7: Number of observations in start delay intervals used in fig. 30

	Southbound high-speed trains					Northbound freight trains				
Delay	0–2	2–4	4–6	6–8	8–10	0–12	12–24	24–36	36–48	48–60
Reg	715	290	143	81	46	1150	125	40	33	22
Sim	1368	522	245	145	97	4071	449	89	72	55

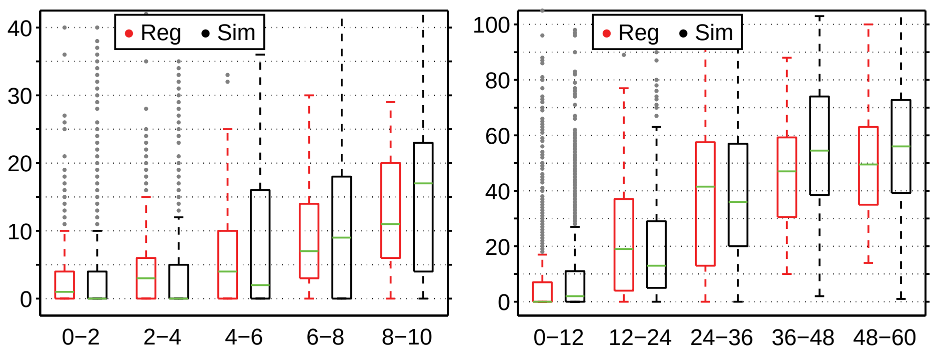


Figure 30: Observed (Reg) and simulated (Sim) outcome in minutes for trains depending on start delay intervals (x-axis). Southbound high-speed trains Katrineholm–Hässleholm (left) and northbound freight trains Hässleholm–Mjölby (right).

6.4 Conclusions

Results show that a reasonably good fit is possible to obtain using deviations from scheduled running time as the basis for estimating primary running time extensions. The subdivision used is relatively rough, a finer breakdown can give a better estimation. However, it must be considered that the scheduled timetable is rounded to full minutes and if the distance considered is small this can contribute with a relatively large error. Using data for every station is valuable also for analysis of real outcome and this will add knowledge that is useful in setting up simulations.

A general problem with the observed data (registrations) is that arrival times are only reported according to scheduled stops. This means that stops made on other stations cannot be identified in a straightforward way. Comparing running times and checking for changes in train sequence is needed in these cases. Track circuit data can also provide a solution to this, but processing this for many stations on a large network can be challenging. Since passenger trains rarely make this type of unscheduled stops this mainly concerns freight trains.

Passenger train operations show a better fit than freight trains. Difficulties in modelling freight operations are for example length and weight variations, top speed limitations and stops including shunting. It should also be noted that the simulated timetable and observed data stems from different time periods. Although many of the train slots are almost equal there are differences and these mostly concern the freight trains. Using observed data from a longer time period and comparing outcome from different timetable periods is recommended.

7 Summary and conclusions

Running time calculations for concept trains show that significant time gains can be achieved by increasing cant and allowing a higher cant deficiency. Although the calculated speed profiles do not consider the limitations imposed by transition curves, a good estimation of possible running times is obtained. No significant infrastructure improvements are assumed, except already decided, in the developed speed profiles. A number of today's restrictions are also included although some of these should be eliminated.

Examples of other infrastructure related issues to consider are design of the overhead wire, maximal speed passing a platform, signalling, vertical curves and ballast pick-up. Current ATC in Sweden is designed for speeds up to 270 km/h, this limitation disappears when ETCS is introduced. Vertical radii are seldom less than 10000 m on the main lines and this is well within the limits for the speeds considered in this work. Future trains must be able to cope with minimum standards regarding track geometry quality.

Results show that carbody tilting is essential in achieving the possible benefits on older lines with moderate curve radii. However, this is not the case for newly built and future lines with high curve radii. Emphasis should instead be put on top speed and other related prerequisites. The higher the proportion of high-speed line the less the benefit of tilting and the greater the benefit from higher top speed. Low power to ton ratio, e.g. 10.8 instead of 20 kW/ton, has a clear negative influence on running times. This effect is mostly pronounced in cases with station stops and increases with the number of stops. Increased power to ton ratio is also important in achieving more efficient and powerful regenerative braking that saves energy.

The influence of an increased rate of acceleration and deceleration above 0.6 m/s^2 has modest influence on running times in most cases. For regional services with frequent stops it may however be useful with higher acceleration and deceleration values, e.g. 0.8 m/s^2. In real operations track capacity issues will influence the possible combinations of traffic and train types. The speed difference between the trains limit the available capacity, thus it is of interest to increase average speeds also for other trains or separate high-speed trains from other slower trains. This can be done both by increasing the number of tracks, e.g. upgrade double-track to four tracks, or dedicate lines for high-speed trains only. Obtaining full benefits out of high-speed trains demands construction of new railway lines.

Simulations of timetable modifications indicate quite clearly the impact on the aggregated on-time performance for high-speed trains. How and where on the line changes are applied will probably influence the outcome, however this has not been checked and is not within the scope of this study. The most beneficial case for on-time performance at end stations could be to add all additional allowance close to the end stations. This would however be more of a quick fix or a change aimed for the public timetable.

Many of the problems for high-speed and other trains are initiated by small delays which in some cases eventually implies missed train slots. This in turn leads to further delays due to slower trains in front of faster trains. Minimizing these events usually calls for better on-time performance along the line or at some critical places and not just at the end. Thus, allowances may be applied more strategically regarding individual trains and positions. However, extending running (journey) can also lower the attractiveness of train travel. On the other hand, improving on-time performance and reliability is a key factor in attracting passengers.

Increasing buffer times between trains complies partly with the same purpose as increasing or redistributing allowances. the aim is to ensure that the scheduled train order is kept, i.e. a train remains in its slot even with a small or moderate delay typically in the range of 3–5 min. Sufficient buffer times will, more or less, improve on-time performance for all trains. Assuming that high-speed trains with small delays are prioritized in dispatching decisions, other trains get pushed and delayed. If buffer times can absorb more of these events, delay propagation will reduce. The drawback is obvious, buffer times consume capacity and possible train slots. Again it comes down to a trade-off between acceptable delay levels and desired traffic.

Considering the use of real observations as delay input results show that a quite good fit with real outcome can be obtained, even if only data from a selection of stations is used. More extensive analysis is possible if all train registration data for all stations is included. Overtaking situations and ad-hoc solutions can be identified, which can improve simulation data input.

This method can also be tested on other lines and networks to see if some general conclusions can be drawn. It could for example mean finding a relationship between primary delay distributions, running distance, number of stations and traffic intensity. Since one of the main ideas with timetable simulations is to test modified or completely new timetables, a set of general delay distributions could be helpful.

A limitation in the observed data is the fact that a train is recorded according to its planned schedule, which means that arrival time stamps are only created at stations where the train has a scheduled stop. If an overtaking is carried out ad-hoc, which happens frequently, this event will show up as an extended running time from the previous station. When this is observed in the data, the cause of this can be an actual running time extension (primary or knock-on) or that the train is passed by another train. Deciding which of the two cases apply can be done by comparing the train sequence on the succeeding stations.

8 Contribution of the thesis

The work done on running time calculations with future speed profiles give insight in the relation between train characteristics, stopping patterns and future speed profiles. Results of this kind are important in making prognoses and estimations of possible market effects. Some of the calculated speed profiles combined with other infrastructure related data have been used for evaluating eco-driving strategies and use of regenerative electric brakes with regard to travel time and energy consumption. Parts of the results from the work presented here are used in other reports within the Green Train research program [2, 8, 23]. The simulation model for Southern Main Line has also been used to evaluate possible timetables for the Green Train with an increased speed from 200 to 250 km/h [22].

Simulations on the Western Main Line give an example on an application in timetable planning. Although assumptions are made considering possibilities to make adjustments for other trains, the simulations give indications on the effects of increased and reduced allowances for high-speed trains. Results from the case with increased buffer times highlight the importance of adequate time distances between trains and at critical places.

Evaluating trains individually and not only on aggregated level give information on which trains have problems considering on-time performance and affect the aggregated value negatively. A specific example with tightly scheduled buffer times between passing high-speed trains and departing commuter trains illustrates the benefits of increased buffer, i.e. improved on-time performance.

Using observed data is a normal approach in finding appropriate levels of disturbances in a calibration process. The different type of perturbations used are important and improve the stochastic behaviour, i.e. the model becomes more realistic. Since running time extensions cannot be identified in a straightforward way from observed data, estimations are needed. The approach used in simulations on the Southern main line show that a reasonably good fit is possible to obtain by using distributions based on running time deviations and reduce them with the purpose of separating primary and knock-on delays.

In order to see if this method can be used more generally, other lines or networks should be investigated. More complete observed data, i.e. all stations and possibly from a longer time period, is preferred and in particular important considering single-track lines. The aim is to find adequate measures for delay modelling by that contribute to future use of simulations in the timetable planning process.

9 Future work

Using simulation in timetable planning requires a good understanding in delay modelling, dispatching measures etc. Valuable insight can also be apprehended by studying delay data for line or network. This gives information and can highlight some key factors that should be considered in analysing the current scenario, designing future timetables, preparing simulations and so on. Normally there is no real need to simulate a historical timetable, however for calibration and validation purposes this is of interest since delay data and simulated timetable conforms in time.

Strategic planning does not necessarily imply choosing between timetables, it can also include evaluating different infrastructure solutions. Many cases consist of comparing alternatives and studying relative differences, there is however still a need to model delays in a realistic way. The calibration process can also be done more efficiently by using an optimization approach and simultaneously allow more parameters to be varied. There is also a need to understand the impact of different dispatching measures in the software, e.g. priorities, time windows for decisions etc. This will hopefully give insight in the level of compliance between real and simulated dispatching.

Simulation of single-track lines impose a challenge, both in evaluation of observed data and in dispatching. It is however necessary to study this since there are long and heavily utilized single-track sections in Sweden. Although much focus is put on double-track lines, many of the issues of shortcomings in capacity concern single-track lines as well. Modelling this type of train operation is challenging, especially with a mix of trains and unbalance in the flow regarding direction. For example, investigating train meets on single-track with three trains by adding a third station track turned out to be difficult in a modelling perspective [5].

Freight trains departing ahead of schedule is also an interesting area considering simulation studies. Although many investigations focus on passenger trains it may still be of interest to model freight services more precisely. This can also lead to higher attractiveness regarding studies where the main focus is on freight trains and their performance. Preliminary work on this is done on parts of the model included in the Southern Main Line simulations. The method used will be further developed and compared with results from the work presented here.

References

[1] Anand, N., and Anayi, M. Improving punctuality of train traffic on Western main line of Swedish railway network. In *Proceedings of AMSE 2009 Rail Transportation Division Fall Conference* (Fort Worth, USA, 2009).

[2] Andersson, E. Green train – Concept proposal for a Scandinavian high-speed train – Final report – Part B. Tech. rep., KTH Railway Group, 12–02, 2012.

[3] Andersson, E., and Persson, R. Fall för gångtidsberäkningar i Gröna tåget – Förslag. Memorandum, KTH, November 2006.

[4] Dingler, M., Lai, Y.-C., and Barkan, C. Impact of train type heterogeneity on single-track railway capacity. In *Journal of the Transportation Research Board.* (March 2009).

[5] Eveby, S. Samspelseffekter vid kapacitetseffekter – Exempel från Dalabanan. Master's thesis, TSC–MT 11–002, KTH, 2011.

[6] Fröidh, O. (Ed.) Gröna tåget – Framtida tågprestanda och bangeometri – Förstudie. Memorandum, KTH, November 2005.

[7] Fröidh, O. Gångtidsberäkningar i Gröna tåget. Memorandum, KTH, November 2006.

[8] Fröidh, O. Green train – Basis for a Scandinavian high-speed train concept – Final report – Part A. Tech. rep., KTH Railway Group, 12–01, 2012.

[9] Hwang, C.-C., and Liu, J.-R. A simulation model for estimating knock-on delay of Taiwan regional railway. *Journal of the Eastern Asia Society for Transportation Studies 8* (2010).

[10] Koutsopoulos, H., and Wang, Z. Simulation of urban rail operations: application framework. *Transportation Research Record: Journal of the Transportation Research Board No. 2006* (2007), 84–91.

[11] Koutsopoulos, H., and Wang, Z. Calibration of urban rail simulation models: A methodology using spsa algorithm. In *Proceedings of the 2011 Winter Simulation Conference* (2011).

[12] Kufver, B. *Optimisation of horizontal alignment for railways – Procedures involvong evaluation of dynamic vehicle response.* PhD thesis, TRITA–FKT Report 2000:47, KTH, 2000.

[13] Lindfeldt, O., and Sipilä, H. Validation of a simulation model for mixed traffic on Swedish double-track railway line. In *Proceedings of Railway Engineering 2009, 10th International Conference and Exhibition, London, UK* (2009), M. Forde, Ed., Engineering Technics Press.

[14] Mattsson, L.-G. *Critical Infrastructure: Reliability and Vulnerability.* Springer-Verlag, 2007, ch. 7, pp. 129–150.

[15] Nelldal, B.-L., Lindfeldt, O., Sipilä, H., and Wolfmaier, J. Förbättrad punktlighet på X2000 – Analys med hjälp av simulering. Tech. Rep. TRITA–TEC–RR 08–001, KTH, 2008.

[16] Olsson, N., and Haugland, H. Influencing factors on train punctuality – results from some norwegian studies. *Transport Policy 11* (2004), 387–397.

[17] Persson, R. Tilting trains – benefits and motion sickness. *Journal of Rail and Rapid Transit 224* (2010).

[18] Radtke, A., and Hauptmann, D. Automated planning of timetables in large railway networks using microscopic data basis and railway simulation techniques. In *Computers in Railways IX* (2004), WIT Press.

[19] Rudolph, R. Allowances and margins in railway scheduling. In *Proceedings of the World Congress on Railway Research* (Edinburgh, Scotland, 2003), pp. 230–238.

[20] Siefer, T. *Railway Timetable and Traffic.* Eurailpress, 2008, ch. 9, pp. 155–169.

[21] Sipilä, H. Tidtabellsläggning med hjälp av simulering – effekter av olika tillägg och marginaler på X2000-tågen Stockholm–Göteborg. Tech. Rep. TRITA–TEC–RR 09–007, KTH, 2010.

[22] Sipilä, H., and Warg, J. Kapacitetsanalys av Södra stambanan – Effekter av ökad trafik och ökad hastighet från 200 till 250 km/h. Tech. rep., KTH Railway Group, 1203, 2012.

[23] Sjöholm, M. Eco-driving and use of regenerative electric braked for the Green Train – The effect on travel time, energy consumption and brake wear. Tech. rep., KTH Railway Group, 1101, 2011.

[24] Sogin, S., Barkan, C., and Saat, M. Simulating the effects of higher speed passenger train in single track freight networks. In *Proceedings of the 2011 Winter Simulation Conference* (2011).

Report I

Körtidsberäkningar för Gröna tåget

–

Analys av tågkonfigurationer

Arbetsrapport

Hans Sipilä

KTH Arkitektur och samhällsbyggnad
Avdelningen för Trafik och logistik
KTH, 100 44 Stockholm
www.kth.se

Publikation 0802
ISBN 978-91-7415-133-6

Sammanfattning

Gröna Tåget är ett forskningsprojekt där en ny generations snabba tåg i Sverige studeras i ett systemperspektiv som omfattar marknad och ekonomi samt teknik. I delprojektet Marknad och trafik beräknas bland annat körtider för olika typtåg där parametrar som acceleration, effekt, retardation samt största tillåtna hastighet varieras inom vissa intervall. Körtider beräknas för ett urval linjer med varierande uppehållsmönster. Med utgångspunkt i Stockholm tas körtider och tidtabeller fram för Västra och Södra stambanan, Svealandsbanan samt kombinationen Ostkust-, Ådals- och Botniabanan. Beräkningar görs även för Västkustbanan och den i framtiden planerade Götalandsbanan inklusive Ostlänken.

Olika värden på rälsförhöjning och rälsförhöjningsbrist används för att skapa nya statiska hastighetsprofiler utifrån befintlig data om radier för cirkulärkurvor. Hänsyn har inte tagits till den begränsning i kurvhastighet som otillräckliga övergångskurvor kan medföra. Körtider beräknas sedan för typtågen på de olika hastighetsprofilerna. I ett senare skede kan resultaten användas till att bestämma intervall (minsta-största värde) för tekniskt, ekonomiskt och komfortmässigt möjliga tågprestanda som funktion av bangeometrin.

I resultaten och i analysen av körtidsberäkningarna blir oftast den största körtidsskillnaden mellan tåg med olika effekt, 10,8 och 15,3 kW/ton. Körtidsvinsten som fås av en effekthöjning avtar med ökande effekt. Nivån på tågets startacceleration har en tydlig påverkan på körtiderna om tåget gör flera uppehåll. De typtåg som har lägst effekt inom respektive sth-nivå kan anses som effektsvaga, residualaccelerationen blir låg.

På vissa banavsnitt varierar den beräknade statiska hastighetsprofilen kraftigt. I dessa fall kan det bli svårt att utnyttja banhastighetshöjningar om det enbart gäller en kortare sträcka. I framtiden kan eventullt några betydande hastighetsnedsättningar kring vissa stationer byggas bort. Trängseleffekter uppkommer ofta i verklig trafik där snabbtåg kombineras med långsammare tåg, vilket ytterligare kan påverka valet av bland annat effekt- och accelerationsnivå. Tåg med hög effekt och bra prestanda innebär även en högre belastning på kraftförsörjningen.

I rapporten presenteras grafiskt de olika hastighetsprofilerna på respektive bana. Ett avsnitt behandlar totala körtider för olika konfigurationer. Dessutom har tidtabeller tagits fram för respektive trafikeringsvariant. Som jämförelse används referenstågen X2 och X52.

Innehåll

1 Inledning

Forskningsprojektet om Gröna tåget (GT) handlar om att studera en ny generation snabba tåg i Sverige i ett systemperspektiv. Systemperspektivet omfattar såväl teknik som marknad och ekonomi, säkerhet och miljö, och att betrakta sambanden mellan fordon och infrastruktur, trafikering och kapacitetsutnyttjande. Genom innovativa lösningar går det att förbättra ekonomi och prestanda till exempel genom att utveckla ”low-cost high-speed” och förbättra samhällsekonomin för tågtrafiken i ett helhetsperspektiv [1].

1.1 Bakgrund

Ett delprojekt i Gröna tåget är Marknad och trafik. I delprojektet finns ett behov av att definiera teknisk standard i de viktigaste parametrarna som påverkar körtiderna. Olika varianter på typtåg definieras där parametrar varieras inom vissa intervall vad beträffar acceleration, effekt, retardation samt största tillåtna hastighet (sth). Körtider beräknas för ett urval linjer (typlinjer) med varierande uppehållsmönster. Dessa ska användas som underlag för marknadsanalyser för att kunna avgöra nyttorna med förbättrade prestanda. Som referens i körtidsberäkningarna används fordon X2 (X2000 tågsätt 6-vagnar) samt X52 (Regina). Utifrån data om radier för cirkulärkurvor skapas nya hastighetsprofiler där rälsförhöjning och rälsförhöjningsbrist varieras. I alla genererade hastighetsprofiler tas hänsyn till den aktuella linjens ovillkorliga fasta hastighetsnedsättningar (det vill säga de som inte beror på kurvgeometrin) enligt linjebok eller liknande data.

1.2 Syfte

Syftet är att göra körtidsberäkningar för de olika typ- och referenstågen i simuleringsverktyget RailSys. Beräkningarna sker för sex grundbanor med kombinationer [2]. Normalt beräknas två olika trafikeringsmönster på respektive bana. I bilaga B presenteras en del av resultaten i form av enkla tidtabeller där varje typtåg körs på de olika hastighetsprofilerna. I ett senare skede kan resultaten användas till att bestämma intervall för tekniskt, ekonomiskt och komfortmässigt möjliga tågprestanda som funktion av bangeometrin.

2 Infrastruktur och hastighetsprofiler

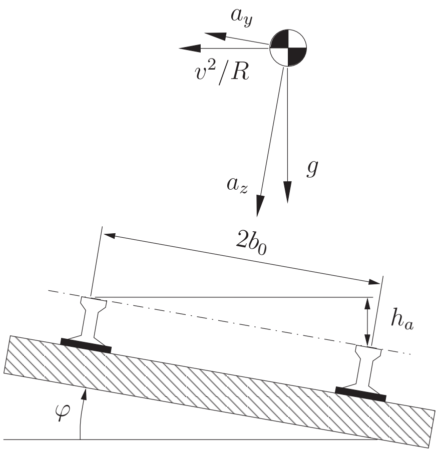
De beräknade hastighetsprofilerna beror i hög grad av befintlig infrastruktur, särskilt kurvor i horisontalplanet påverkar största tillåtna hastighet längs en bana. Beräkning av nya statiska hastighetsprofiler görs genom att använda olika kombinationer av värden på rälsförhöjning (h_a) och rälsförhöjningbrist (h_b) med hänsyn tagen till tillåtet rälsförhöjningöverskott (h_o). Hastighetsprofiler beräknas enligt tabell 1.

Tabell 1: Hastighetsprofiler med maximal rälsförhöjning och rälsförhöjningsbrist (mm)

Profil (P)	h_a	h_b	Information
1	150	150	Referens GT kategori B enligt linjebok
2	150	245	Referens GT kategori S enligt linjebok
3	150	245	Beräknad hastighetsprofil
4	160	165	Beräknad hastighetsprofil GT-B+
5	160	245	Beräknad hastighetsprofil GT-S
6	160	275	Beräknad hastighetsprofil GT-S+
7	160	300	Beräknad hastighetsprofil GT-S++

2.1 Rälsförhöjning och rälsförhöjningsbrist

Rälsförhöjning (h_a) är en anordnad förhöjning av ytterrälen i en kurva relativt innerrälen. Denna lutning minskar effekten av centrifugalkrafter (sidoaccelerationer) som uppträder vid körning i kurva [3]. Rälsförhöjningen h_a definieras enligt figur 1. Mätbasen för rälsförhöjningen ($2b_0$) är avståndet mellan de nominella hjul-räl kontaktpunkterna.



Figur 1: Horisontal (v^2/R) och vertikal acceleration (g). Acceleration parallell (a_y) och vinkelrät spårplanet (a_z).

Accelerationen a_y kan i termer av centrifugal- och tyngdacceleration samt med rälsförhöjningsvinkeln φ skrivas som

$$a_y = \frac{v^2}{R} \cdot \cos\varphi - g \cdot \sin\varphi \tag{1}$$

För tillräckligt små φ gäller att $\cos\varphi \approx 1$ vilket tillsammans med $\sin\varphi = h_a/2b_0$ (figur 1) ger att spårplansaccelerationen kan uttryckas som

$$a_y \approx \frac{v^2}{R} - g \cdot \frac{h_a}{2b_0} \tag{2}$$

Den rälsförhöjning som vid en given fordonshastighet och kurvradie ger spårplansaccelerationen $a_y = 0$ kallas teoretisk rälsförhöjning (h_t). För normalspårvidd gäller att $2b_0 = 1500$ mm vilket tillsammans med $g = 9{,}81$ m/s^2 och hastigheten v i km/h ger

$$h_t \approx \frac{2b_0}{g} \cdot \frac{v^2}{R} \approx \Big\{v : \mathrm{m/s} \to \mathrm{km/h}\Big\} \approx 11{,}8 \cdot \frac{v^2}{R} \tag{3}$$

Rälsförhöjningsbrist (h_b) innebär att den verkliga rälsförhöjningen är mindre än den teoretiska och är ett mått på den ytterligare rälsförhöjning som behövs för att uppnå $h_a = h_t$. Rälsförhöjningsbrist är proportionelig mot kvarvarande sidoacceleration och kan skrivas som

$$h_b = h_t - h_a \tag{4}$$

Om rälsförhöjningsbristen är negativ ($h_a > h_t$), finns ett rälsförhöjningsöverskott (h_o). Det är den överflödiga rälsförhöjning som behöver tas bort för att uppnå $h_a = h_t$. Rälsförhöjningen i en cirkulärkurva med radie R skall begränsas med avseende på tillåtet rälsförhöjningsöverskott för långsamma godståg med sth = V_g

$$h_a \leq h_o + 11{,}8 \cdot \frac{V_g^2}{R} \tag{5}$$

I ekvation 5 kan $h_o = 110$ mm samt $V_g = 90$ km/h ansättas [4]. Begränsningarna kommer i praktiken in i kurvradier från cirka 1500 m och uppåt. Maximal rälsförhöjningsbrist bör dessutom begränsas med avseende på sidvind. I de framtagna hastighetsprofilerna gäller att $h_b = 300$ mm tillåts upp till 225 km/h, varefter tillåten h_b sjunker med 1 mm/km/h. Villkoret för h_b blir då

$$525 - h_b = \sqrt{\frac{R \cdot (h_a + h_b)}{11{,}8}} \tag{6}$$

Om den rälsförhöjningsbrist som finns i tabell 1 betecknas h_{b1} och den som ges av ekvation 6 med h_{b2}, blir rälsförhöjningsbristen som används för vidare beräkningar $h_b = \min\{h_{b1}, h_{b2}\}$. Med hänsyn tagen till begränsningarna på h_a samt h_b från ekvation 5 och 6 beräknas maximala hastigheten $V_{\max}$ på ett godtyckligt banavsnitt med dimensionerande radie R_{dim} ur

$$V_{\max} = \sqrt{\frac{R_{\mathrm{dim}} \cdot (h_a + h_b)}{11{,}8}} \tag{7}$$

I de flesta fall gäller att $h_a = 0$ på rakspår (raklinje). Mellan raklinje och cirkulärkurva finns normalt en övergångskurva i vilken spårets krökning ($1/R$) ändras linjärt med längdmätningen i horisontalplanet (s). På samma sätt anordnas övergångskurvor mellan cirkulärkurvor med olika radier.

En ändring i rälsförhöjningen uppnås i en rälsförhöjningsramp där h_a normalt ändras linjärt med s. Krav finns på maximala värden för ramplutning (dh_a/ds), rampstigningshastighet (dh_a/dt) samt rälsförhöjningsbristens ändringshastighet (dh_b/dt) [5]. På grund av dessa krav kan den hastighet som en cirkulärkurva medger enligt ekvation 7 begränsas. I de hastighetsprofiler som beräknas används enbart data för cirkulärkurvor.

2.2 Beräkning av hastighetsprofiler för typsträckor

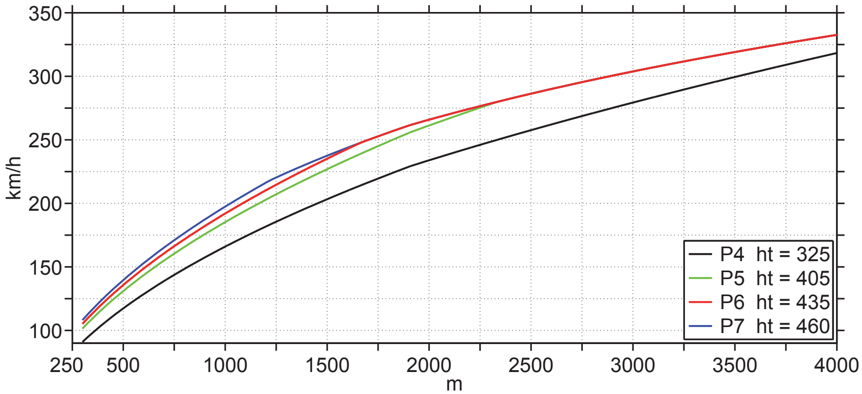
Hastighetsprofiler för typsträckor beräknas genom att i ett första läge dela in banan i delsträckor baserat på variationer i största tillåtna hastighet (sth) enligt linjeboken. Dimensionerande kurvradie (cirkulärkurva) bestäms därefter enligt BIS (Ban Informations System) för varje delsträcka. Om antalet cirkulärkurvor är n, ges den dimensionerande kurvradien R_{dim} av

$$R_{\text{dim}} = \min\{R_1, R_2, \ldots R_n\} \tag{8}$$

På dubbelspårssträckor bestäms normalt R_{dim} gemensamt för båda spåren förutsatt att grundhastigheterna är lika. På längre banavsnitt med samma sth kan variationer i R motivera en uppdelning i mindre delsträckor för att uppnå bättre hastighetsdifferentiering. Ibland begränsas R_{dim} på ett kortare banavsnitt och sätts lika med R_{dim} från föregående eller efterföljande sträcka, även om nuvarande sth kortvarigt höjts med 10–20 km/h. Detta genomförs i de fall där R_{dim} för ett relativt sett kort banavsnitt är märkbart högre än R_{dim} för de omgivande avsnitten.

Om en speciell orsak anges till en viss sth i linjeboken, till exempel geoteknisk, används samma sth även för GT hastighetsprofiler. Principen tillämpas även i närheten av och inom större stationer. Angiven sth kan även kortvarigt begränsas av växelpassager. Om ingen R_{dim} är definierad på ett banavsnitt följer GT hastighetsprofilerna 3 samt 5–7 linjebokens sth för S-tåg, medan GT hastighetsprofil 4 använder linjebokens sth för B-tåg (normalt 10 % hastighetsöverskridande). GT hastighetsprofiler rundas av nedåt till närmaste femtal, oavsett värdet på V_{max} som erhålls ur ekvation 7.

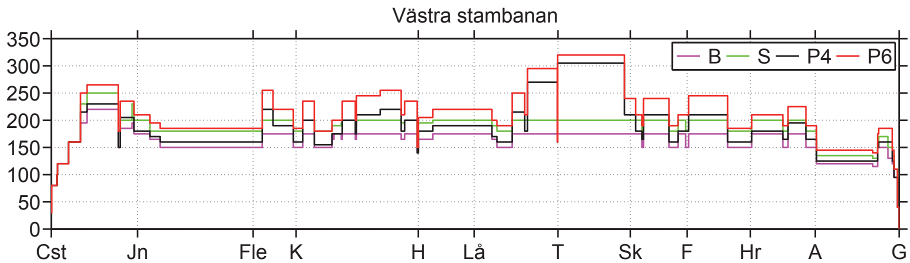
Profil 5 ligger normalt 5–10 km/h lägre än profil 6 medan skillnaden mellan profil 6 och 7 normalt är 5 km/h. För stora dimensionerande kurvradier ($R_{\text{dim}} > 2247$ m), innebär begränsningen på tillåten rälsförhöjningsbrist vid höga hastigheter att profil 5 konvergerar mot samma V_{max} som 6 och 7. I de beräknade hastighetsprofilerna avrundas V_{max} nedåt. Profil 5–7 har därför ofta samma hastighet om $R_{\text{dim}} > 2066$ m. Figur 2 visar beroende mellan hastighet och kurvradie för olika h_t beräknade enligt ekvation 5–7.



Figur 2: $V_{\max}$ vid olika kurvradier R.

2.3 Västra stambanan

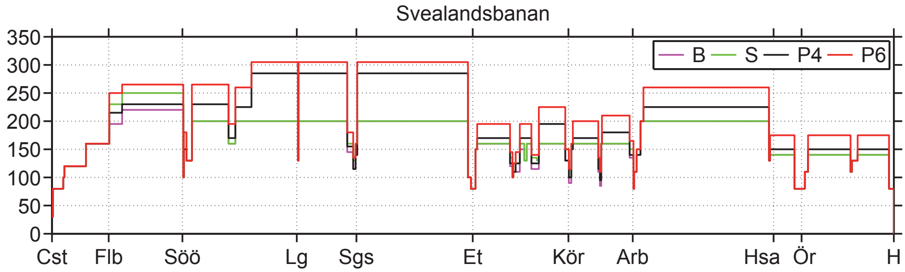
Linjebokens sth används fram till Flemingsberg, därefter definieras R_{dim} för beräkning av hastighetsprofiler fram till Olskroken med undantag för Södertälje syd övre och Hallsberg. På grund av bron över Göta kanal vid Töreboda tvingas tågen kortvarigt sänka hastigheten till 160, vilket förlänger körtiden något. De högsta hastigheterna uppnås mellan Gårdsjö och Skövde, framförallt efter Töreboda. Hastighetsprofilerna varierar tydligt i sth mellan Flen och Hallsberg samt mellan Skövde och Göteborg. GT hastighetsprofilernas karaktär följer som väntat de idag gällande hastigheterna (figur 3). I hastighetsdiagrammen visas dels dagens sth för B- och S-tåg samt GT hastighetsprofil 4 och 6 enligt tabell 1.



Figur 3: Hastighetsprofiler för Västra stambanan (Stockholm–Göteborg).

2.4 Svealandsbanan

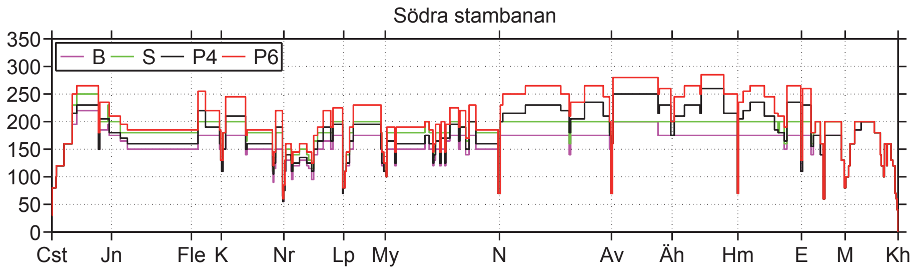
Hastighetsprofilerna är identiska med figur 3 fram till Södertälje syd övre där linjebokens sth följs. För tåg mot Eskilstuna finns en 130-växel vid Läggesta där det partiella dubbelspåret upphör. Motsvarande växel för tåg mot Stockholm finns vid Ryssjöbrink (om tågen går i normal vänstertrafik). Linjebokens sth används för GT hastighetsprofiler vid Strängnäs, Eskilstuna, Rekarne, delvis vid Kungsör samt mellan Arboga och Jädersbruk. Samma förutsättningar gäller även Hovsta (130-växel), kring Örebro, delvis mellan Mosås och Kumla, samt vid Hallsberg. De högsta hastigheterna erhålls på den relativt nybyggda sträckan mellan Södertälje och Eskilstuna samt mellan Jädersbruk och Hovsta. Hastighetsprofilerna (figur 4) varierar betydligt mellan Eskilstuna och Arboga.



Figur 4: Hastighetsprofiler för Svealandsbanan (Stockholm–Örebro–Hallsberg).

2.5 Södra stambanan

Mellan Stockholm och Katrineholm är hastighetsprofilerna beräknade för Västra stambanan (figur 3). Vid Katrineholm används linjebokens hastigheter. Därefter är R_{dim} definierad fram till Lund, med undantag för stationerna Norrköping, Linköping, Mjölby, Nässjö, Alvesta, Hässleholm och Eslöv. Efter Lund används linjebokens hastigheter kombinerade med hastighetsprofiler för Citytunneln. Karaktären på banan medför täta variationer i sth (V_{max}) mellan Katrineholm och Nässjö. Från Öresundsbron till Köpenhamn används hastigheter enligt baninformation från Banedanmark [6]. Efter Nässjö och vidare söderut uppnås relativt höga hastigheter, dock finns kortare sträckor där R_{dim} tydligt begränsar hastigheten (figur 5).

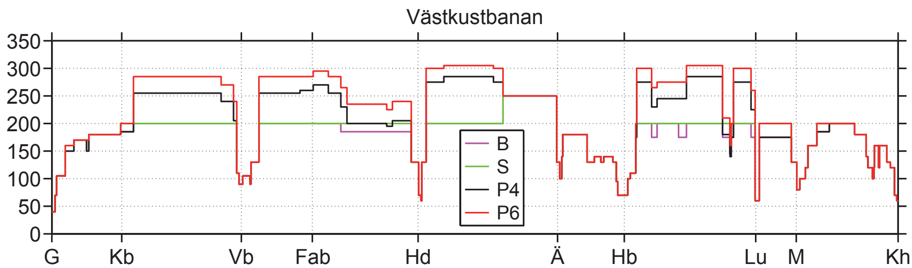


Figur 5: Hastighetsprofiler för Södra stambanan (Stockholm–Katrineholm–Malmö) och Malmö–Köpenhamn via Citytunneln.

2.6 Västkustbanan

GT hastighetsprofiler följer linjeboken till Kungsbacka. Anledningen till att hastigheter för GT inte beräknas på sträckan är geotekniska förhållanden (problem med lera) och begränsad kapacitet. Efter Kungsbacka beräknas hastigheter som beror av R_{dim} till Varberg. En ny dubbelspårig sträckning planeras mellan Varberg (ombyggd station) och Hamra. Eftersom byggstart ligger framåt i tiden (planeras till år 2010–2015) används befintlig enkelspårig bana för körtidsberäkningar på sträckan. Mellan Torebo och Heberg finns ett nyligen färdigställt dubbelspår och en ny station i utkanten av Falkenberg. Fram till Halmstad används R_{dim} mellan Kungsbacka och Varberg samt på sträckan Hamra–Biskopstorp. Efter Halmstad beräknas hastigheter fram till Hallandsåsen varefter $V_{\mathrm{max}} = 250$ används till Ängelholm.

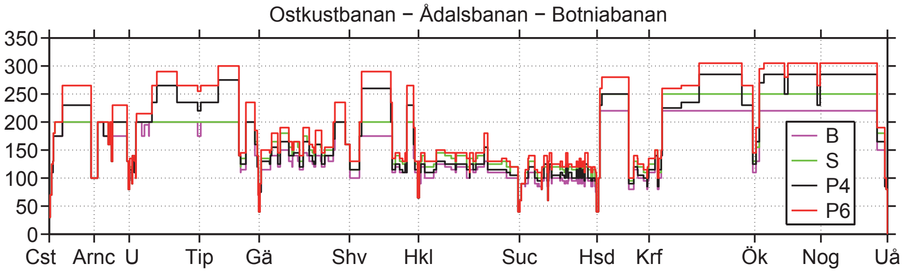
På sträckan Ängelholm–Helsingborg används sth i linjeboken även för GT hastighetsprofiler. Anledningen till detta är variationer i banans relativt sett låga sth. Mellan Helsingborg och Lund definieras R_{dim} med undantag för passagen genom Kävlinge. Från Lund och vidare till Köpenhamn är hastigheterna samma som i figur 5. Hastighetsprofiler för Västkustbanan visas i figur 6.



Figur 6: Hastighetsprofiler för Västkustbanan (Göteborg–Lund) och Lund–Malmö–Köpenhamn via Citytunneln.

2.7 Ostkustbanan, Ådalsbanan och Botniabanan

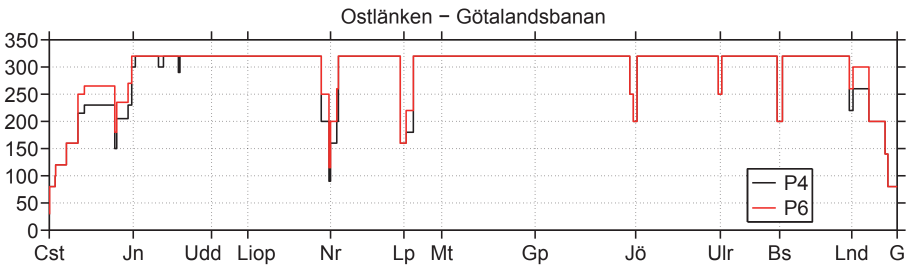
Linjebokens sth används fram till Kummelby samt mellan Arlanda och Knivsta. R_{dim} används mellan Kummelby och Arlanda samt mellan Knivsta och Uppsala. Norr om Uppsala och fram till Gävle beräknas GT hastighetsprofiler till största delen med R_{dim}, med undantag för Uppsala–Samnan och ett kortare banavsnitt vid Skutskär. På sträckan Gävle–Sundsvall används linjebokens sth bland annat vid passage av Söderhamn, Iggesund, Hudiksvall och Gnarp. Kurvradierna varierar kraftigt på sträckan. Från Sundsvall och vidare norrut används bandata för upprustade/nybyggda Ådalsbanan och Botniabanan med avseende på avstånd och kurvradier. Ådalsbanan har idag täta variationer i sth framförallt på grund av begränsande kurvradier. Beräknade V_{max} blir därför låg. Efter Härnösand får banan delvis ny sträckning vilket möjliggör högre hastigheter. Nybyggda Botniabanan byggs med R_{dim} som till största delen ligger kring 2000–3000 m (figur 7).



Figur 7: Hastighetsprofiler för Ostkustbanan, Ådalsbanan och Botniabanan (Stockholm–Umeå).

2.8 Ostlänken och Götalandsbanan

Hastigheter överensstämmer med tidigare exempel mellan Stockholm och Järna. På Ostlänken och Götalandsbanan har inga ytterligare hastighetsprofiler beräknats, sth ligger till största delen på 320 km/h (figur 8). Platser med relativt sett stora hastighetssänkningar är Norrköping, Linköping, Jönköping, Ulricehamn och Borås. I närheten av Göteborg sjunker sth gradvis. Skillnader i hastighet för B- respektive S-tåg förekommer endast marginellt. En parallell bana över Nyköping planeras med anslutningspunkter vid Uddeby och Lisstorp. Skillnad i gångsträcka för tåg som kör via Nyköping jämfört med huvudbanan är ungefär +2500 m, hastigheten varierar mellan 120–200 km/h. Kopplingspunkter mellan Götalandsbanan och Södra stambanan är planerade till Mantorp i närheten av Mjölby och Gripenberg söder om Tranås.



Figur 8: Hastighetsprofiler för Ostlänken och Götalandsbanan (Stockholm–Göteborg).

3 Prestanda för typtåg

Som en del i analysen av Gröna tåget har tio typtåg med delvis olika egenskaper valts ut för närmare analys. Främst varieras hastighet, accelerationsprestanda och effekt (tabell 2). Syftet är att avgöra hur de olika parametrarna påverkar körtiderna på olika banor med varierande antal uppehåll. Tidtabeller för typtågen finns i bilaga B.

Tabell 2: Tågtyper som används för körtidsberäkningar

	Sth (km/h)			Effekt/massa (kW/ton)					Acc. a_s (m/s^2)			Ret. (m/s^2)	
Tåg	250	280	320	10,8	15,3	20,0	25,0	30,0	0,48	0,60	1,00	0,60	0,80
101	•			•					•			•	
102	•				•					•		•	
103	•					•				•		•	
104	•					•					•	•	
105	•					•					•		•
111		•			•				•			•	
112		•				•				•		•	
113		•					•				•	•	
121			•			•				•		•	
122			•				•			•		•	
123			•					•		•		•	

Gemensamt för typtågen är vikten 360 ton, längden 155 m och ett dynamiskt masstillskott på 5,6 %. Retardationen antas vara konstant under hela inbromsningsförloppet, vilket är en approximation. Utifrån data från bland annat tabell 2 definieras dragkraftskurvor för typtågen. Minst hälften av axlarna har drivning och vid normala förhållanden antas uttagbar dragkraft endast marginellt påverkas av adhesion. Dragkraften (F) är konstant upp till en viss hastighet (v_a) och bestäms av

$$F = m_e \cdot a_s \tag{9}$$

I ekvation 9 är m_e tågets massa inklusive dynamiskt masstillkott, startacceleration (a_s) enligt tabell 2. För den avtagande delen av dragkraftskurvan gäller med konstant effekt (P)

$$P = F \cdot v \tag{10}$$

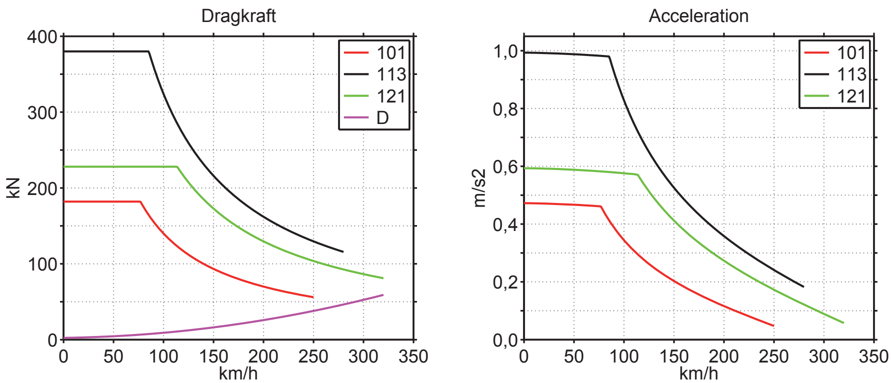
Ett aerodynamiskt väl utformat tågsätt är nödvändigt för att minimera dragkraftsbehov, energiförbrukning och körtider. Summan av mekaniskt och aerodynamiskt motstånd (D_{MA}) kan approximeras med ekvation 11, där A, B och C är fordonsberoende konstanter.

$$D_{MA} = A + Bv + Cv^2 \tag{11}$$

För ett aerodynamiskt väl utformat tåg (6 vagnar) antas $A = 2400$ N, $B = 60$ kg/s och $C = 6{,}5$ kg/m. De föreslagna värdena gäller ett 3,3 m brett fordon som väger 60 ton per vagn med små boggikjolar och med normal sidvind beaktad [1]. Stigningsmotståndet (D_S) i lutningar bidrar till det totala motståndet ($D = D_{MA} + D_S$). Beroende på om tåget befinner sig i ett uppförs- eller nedförslut gäller att $D_S > 0$ eller $D_S < 0$. Den tillgängliga dragkraften blir $F - D$ och accelerationen för ett tillräckligt litet hastighetsintervall i bestäms av

$$a_i = \frac{\overline{F}_i - \overline{D}_i}{m_e} \tag{12}$$

I figur 9 visas exempel på dragkrafts- och accelerationskurvor. Hastigheten v_a där kurvorna ändrar karaktär varierar beroende på vilket a_s och P som används.



Figur 9: Dragkrafts- och accelerationskurvor för några av typtågen i tabell 2 ($D_s = 0$).

4 Körtidsberäkningar

4.1 RailSys

Beräkning av körtider genomförs i RailSys, där infrastrukturmodeller av de aktuella banorna byggts upp. I RailSys definieras noder och länkar. En nod kan till exempel vara en signal, växel eller hastighetstavla och har ett bestämt kilometertal (position). En länk har egenskaper såsom längd, hastighet och lutning. I RailSys kan järnvägsnät byggas upp med olika detaljgrad beroende på användningsområde samt tillgång och kvalitet på indata. Banornas lutningar påverkar gångtiderna och används av RailSys för att beräkna stidningsmotståndet. Lutningar är inlagda på Västra stambanan, Svealandsbanan (inklusive förlängningen Valskog–Hallsberg), Ådalsbanan och Botniabanan. Lutningar finns även på Ostlänken och Götalandsbanan, undantaget bibanan via Nyköping.

I fordonsdelen definieras dragenheter (lok) som kan kompletteras med vagnar när själva tåget skapas. Ett komplett tåg kan även definieras i dragenhetsmodulen där bland annat sth, längd, vikt, adhesionsvikt, dragkraftskurva, tillskott för dynamisk massa, motstånd och andra parametrar bestäms. RailSys beräknar utifrån givna data en accelerationskurva för aktuellt tåg, retardationen är konstant. I tågmodulen anges bland annat vilken hastighetsprofil tåget skall följa samt vilka tillägg man vill använda. I samtliga beräkningar används ett tidstillägg som ger en körtidsförlängning med 3 % (förarmarginal).

4.2 Trafikeringsvarianter

Körtider för typtågen beräknas på samtliga hastighetsprofiler beskrivna i tabell 1. Olika trafikeringsvarianter (tabell 3) används för att bland annat ge en uppfattning om hur uppehåll påverkar gångtiderna för de olika typtågen. Uppehållstid på stationer inkluderar resandeutbyte, och även en schablontid för tågmöte vid enkelspår i vissa fall (variant OKB). Trafikplatssignaturer finns i bilaga A.

Tabell 3: Trafikeringsvarianter med uppehållstider vid stationer

Variant	Uppehållstid (min)																
	Cst	Flb	Söö	K	H	Sk	F	Hr	A	G							
VSB–0	•	0	0	0	0	0	0	0	0	•							
VSB–8	•	1	1	2	2	2	2	2	1	•							
	Cst	Flb	Nkv	Lg	Sgs	Et	Kör	Arb	Ör	Öb	Kla	H	Sk	G			
SB+VSB–3	•	0	0	0	0	2	0	0	2	0	0	0	2	•			
SB–11	•	1	1	1	1	2	1	1	2	1	1	•	–	–			
	Cst	Flb	Söö	K	Nr	Lp	My	N	Av	Hm	Lu	M	Hie	Cph	Kh		
SSB–3	•	0	0	0	0	0	0	0	0	2	2	2	0	0	•		
SSB–13	•	1	1	2	2	2	2	2	2	2	2	2	1	1	•		
	G	Mdn	Kb	Vb	Fab	Hd	Laov	Bån	Ä	Hb	Lkö	Lu	M	Hie	Cph	Kh	
VKB–3	•	0	0	0	0	2	0	0	0	0	0	2	2	0	0	•	
VKB–14	•	1	1	2	2	2	1	1	2	2	2	2	2	1	1	•	
	Cst	Arnc	U	Gä	Shv	Hkl	Suc	Hsd	Krf	Ök	Uåö	Uå					
OKB–10	•	1	2	2	2	2	5	2	2	2	1	•					
	Gä	Lse	Shv	Id	Hkl	Gnp	Suc	Trå	Hsd	Krf	Ök	Ökn	Hum	Nog	Höf	Uåö	Uå
OKB–15	•	1	2	1	2	1	5	1	2	3	2	1	1	3	1	1	•
	Cst	Flb	Söö	Nk	Nr	Lp	My	N	Av	Hm	Lu	M	Hie	Cph	Kh		
OL+SSB–3	•	0	0	–	0	0	0	0	0	2	2	2	0	0	•		
OL+SSB–13	•	1	1	2	2	2	2	2	2	2	2	2	1	1	•		
OL+GB+ SSB–3	•	0	0	–	0	0	–	0	0	2	2	2	0	0	•		
	Cst	Flb	Söö	Nk	Nr	Lp	Jö	Ulr	Bs	Lnd	G						
OL+GB–0	•	0	0	–	0	0	0	0	0	0	•						
OL+GB–9	•	1	1	2	2	2	2	1	1	1	•						

I tabellerna som finns i bilaga B avrundas ankomsttiderna vid stationerna till jämna minuter. Den avrundade tiden används sedan för vidare beräkning fram till nästa uppehåll där ny avrundning görs. Samtliga tider i tabellerna är avgångstider med undantag för sista tiden som är ankomsttid till slutstation. Om passagetid utan uppehåll angetts är tiden avrundad enbart i tabellen och ej i beräkningen. Om tiden betecknas med $(m : s)$ där m är minuttal och s sekundtal, sker avrundning enligt nedan [7]

$$(m : s) = \begin{cases} m & \text{om } s \leq 6 \\ m + 1 & \text{om } s > 6 \end{cases} \tag{13}$$

I tabellerna anges även körtid utan avrundning (Tid) samt tillägget i sekunder som uppkommer på grund av ekvation 13 (+T). Vid jämförelser mellan körtider för olika typtåg eller hastighetsprofiler används främst tiden utan avrundning eftersom skillnaden i tillägg varierar.

4.3 Körtidsberäkningar för olika tågkonfigurationer

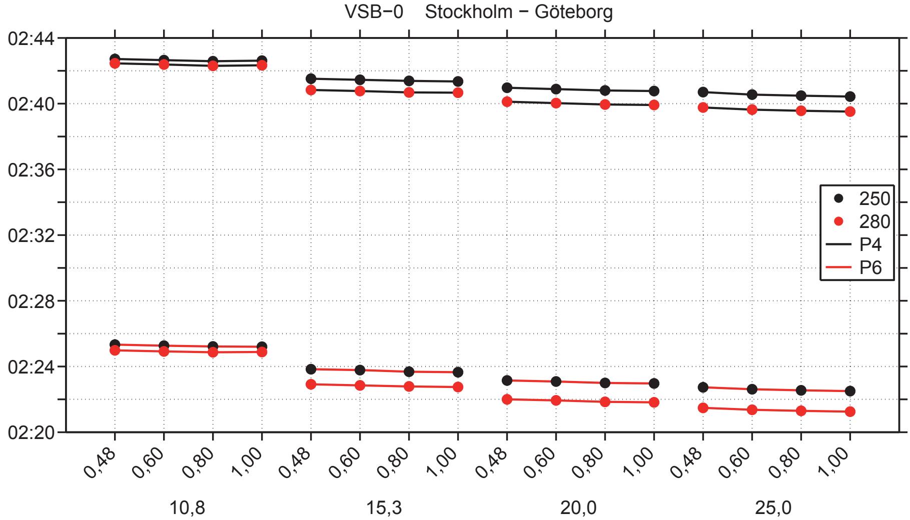
Tågens prestanda påverkar körtiderna i varierande grad. Hur mycket en höjning eller sänkning av till exempel effekten påverkar tiden beror i sin tur av banans hastighetsprofil, lutningar och antal uppehåll. I en analys av hur sth, effekt och startacceleration påverkar körtiderna på trafikeringsvarianterna i tabell 3 används sexton tågkonfigurationer per sth-nivå. Några av dessa sammanfaller med typtågen i tabell 2.

Effekten och startaccelerationen varieras i fyra nivåer för sth 250 och 280. I varianter som inkluderar Ostlänken/Götalandsbanan beräknas även sth 320 och effekt 30,0 kW/ton. Däremot används inte 10,8 kW/ton eftersom effekten inte räcker för att uppnå hastigheten 320 km/h. Bromsprestandan (0,60 m/s^2) varieras inte i analysen. Tabell 4 visar de beräknade konfigurationerna. Skillnaden i körtid för tåg med retardation 0,60 och 0,80 m/s^2 framgår i bilaga B, jämför typtåg 104 och 105.

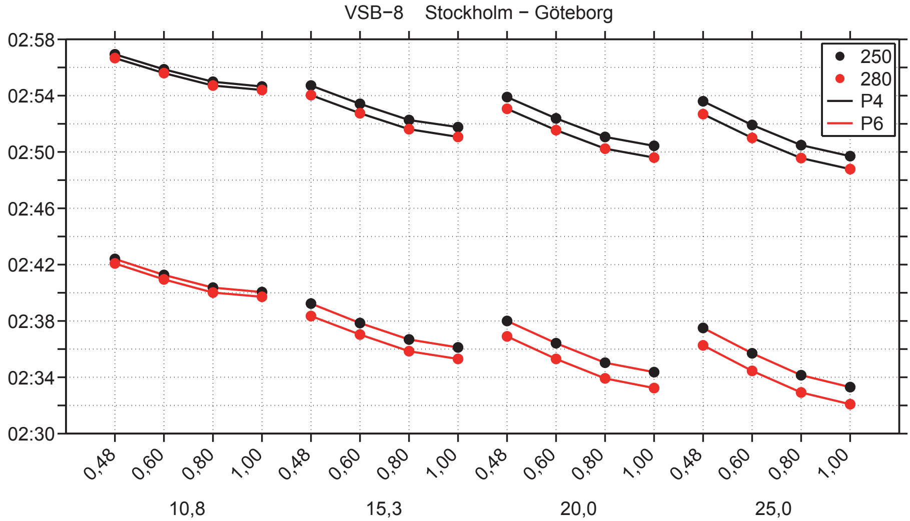
Tabell 4: Beräknade tågkonfigurationer. Sth 320 och P = 30,0 kW/ton används på varianter som inkluderar Ostlänken/Götalandsbanan

Parameter	Nivå				
Största tillåtna hastighet sth (km/h)	250	280	320		
Effekt/massa P (kW/ton)	10,8	15,3	20,0	25,0	30,0
Startacceleration a_s (m/s^2)	0,48	0,60	0,80	1,00	

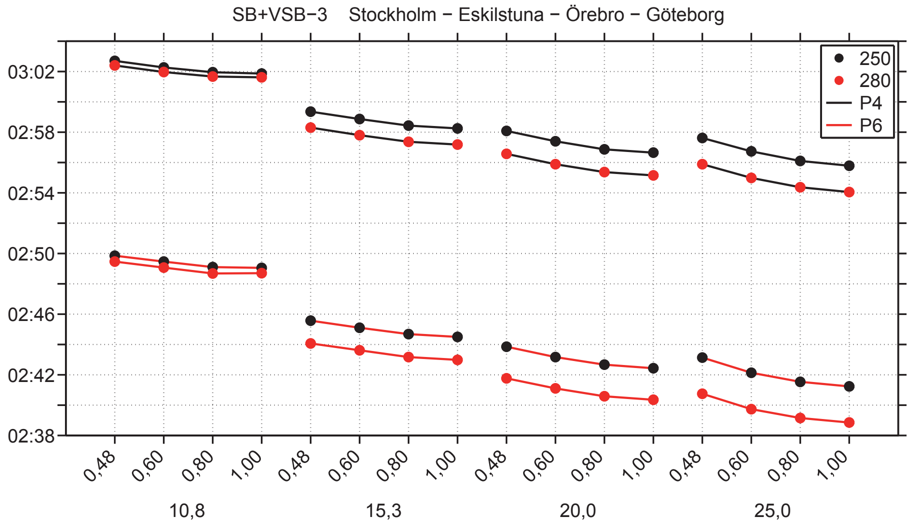
Resultaten av körtidsberäkningarna presenteras i diagram för respektive trafikeringsvariant. Uppehållstiderna är i samtliga fall noll sekunder. Tågkonfigurationerna finns på x-axeln och körtiderna på y-axeln. Hastighetsprofiler som visas är profil 4 (svart linje) och profil 6 (röd linje). Sth-nivåer anges med punkter (ifyllda cirklar), svart för 250, röd för 280 och grön för 320.



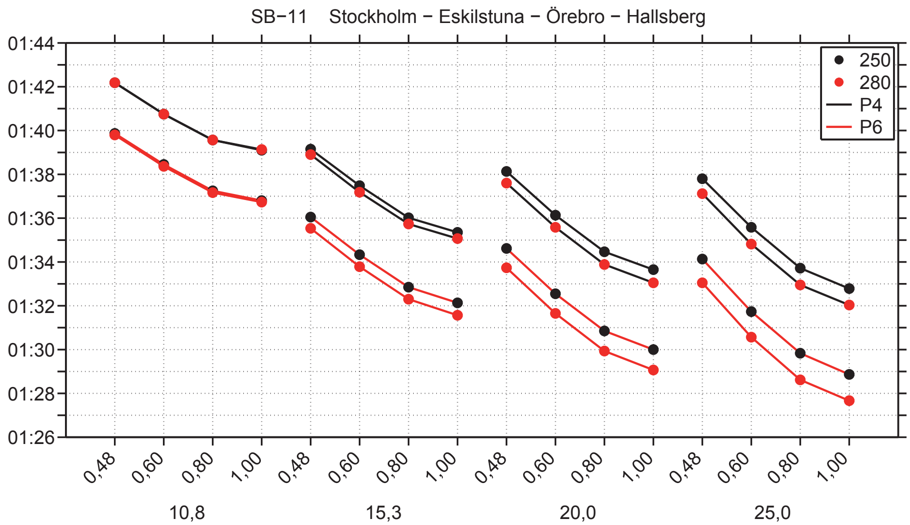
Figur 10: Eftersom tågen inte gör några uppehåll är tidsskillnaderna mellan olika a_s och P små och ligger i intervallet 0–4 minuter. En ökning från sth 250 till 280 ger endast 1–2 minuters tidsvinst. Banavsnitt som tillåter hastigheter över 250 är få.



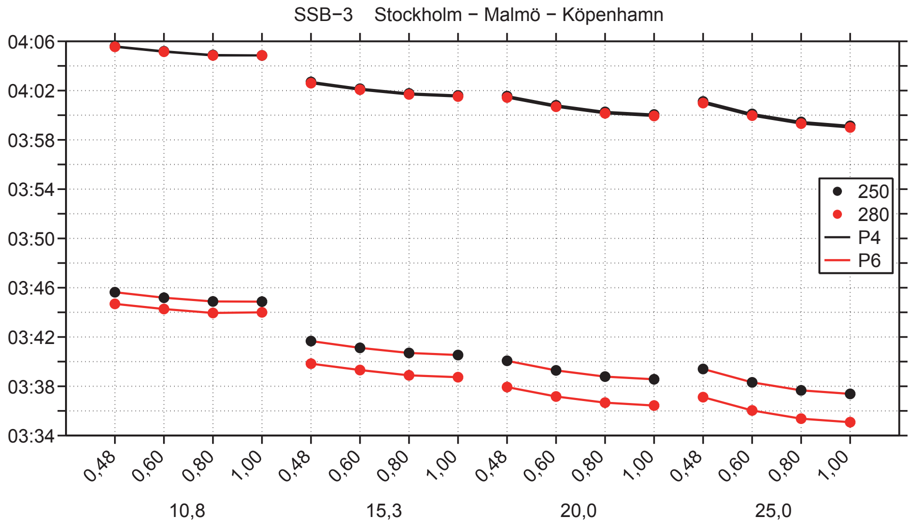
Figur 11: Skillnader mellan olika nivåer på a_s och P är tydligare än i föregående fall (VSB–0). En höjning av a_s ger 2–4 minuter tidsförkortning beroende på vilken effektnivå som avses. Effektsteget 10,8 till 15,3 påverkar körtiden mest. Effektnivåerna överlappar varandra för tåg med $P \geq 15{,}3$ kW/ton. En ökning av P ger inte automatiskt en körtidsförkortning om inte samtidigt a_s bibehålls eller höjs.



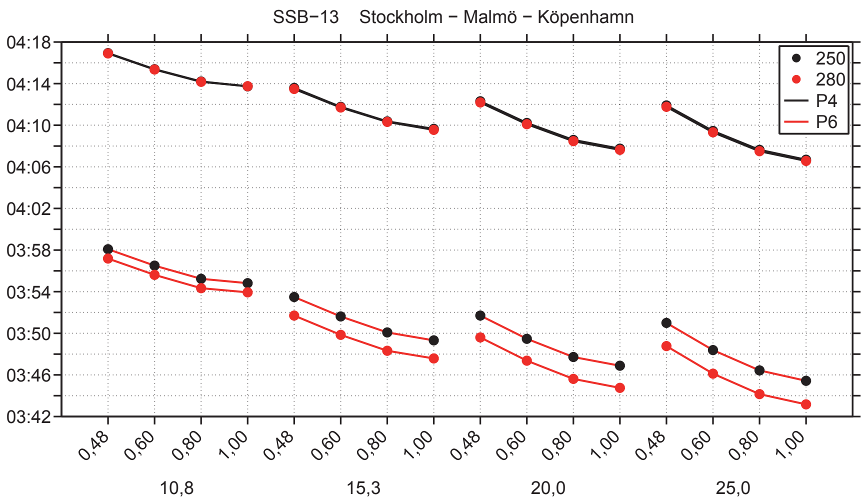
Figur 12: Nivån på a_s har ingen stor påverkan på körtiderna på grund av få stationsuppehåll. De banavsnitt på Svealandsbanan som tillåter hastigheter kring 280–300 för profil 4 och 6 medför att tågen med $P = 10{,}8$ får betydande accelerationstider enligt figur 25. Skillnaden i körtid mellan sth 250 och 280 blir tydligare med ökande P.



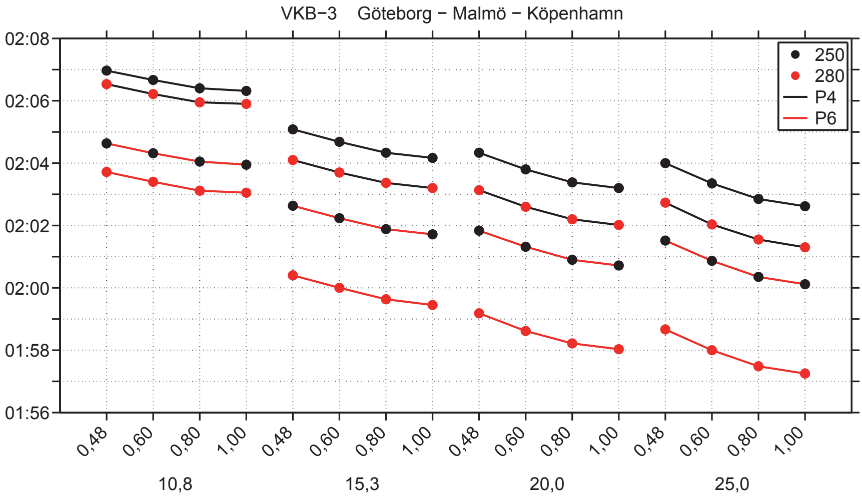
Figur 13: Stationsuppehållen bidrar till att skillnaden i körtid för tåg med sth 250 och 280 är liten. Däremot har a_s en tydlig påverkan, effektnivåerna överlappar varandra. Den körtidsförkortning som erhålls av en höjning i effekt avtar tydligt med ökande effekt. En ökning av a_s ger däremot ungefär lika mycket för varje effektnivå.



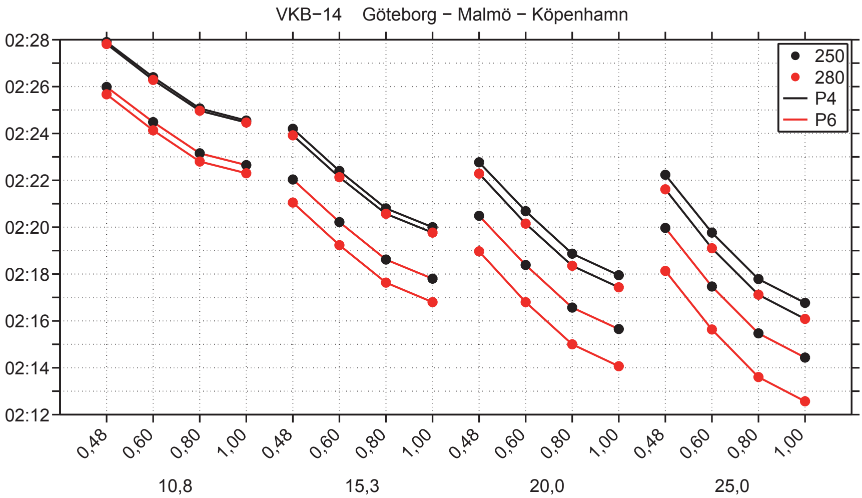
Figur 14: Södra stambanans hastighetsprofil har få sträckor som tillåter hastigheter över 250, skillnaderna mellan sth-nivåerna är därför marginella. Om tågen körs efter profil 4 finns ingen skillnad alls mellan olika sth. Hastighetsprofil 6 ger endast cirka 1–2 minuters skillnad mellan 250- och 280-tåg. Däremot är skillnaden mellan profil 4 och 6 tydlig. Eftersom endast tre uppehåll genomförs påverkar valet av a_s körtiden med endast 1–2 minuter. Precis som i tidigare fall ger effektökningen från 10,8 till 15,3 störst tidsförkortning.



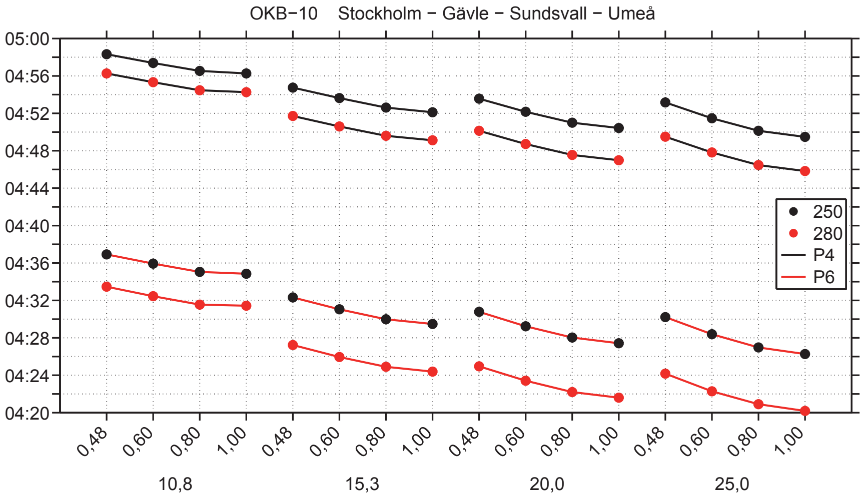
Figur 15: Liknande mönster som i figur 14. Många av uppehållen sker vid stationer där omgivande sth är relativt låg idag. I de fallen ger inte en inbromsning respektive acceleration lika mycket tidstillägg jämfört med om tåget passerar utan att stanna.



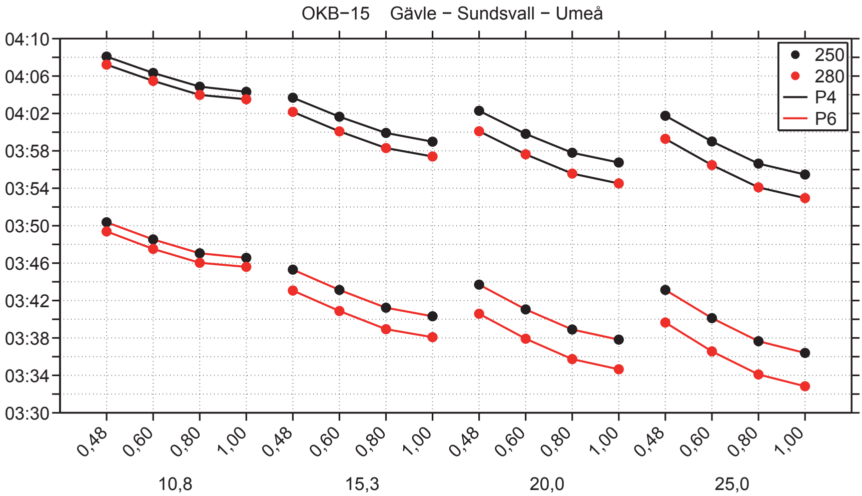
Figur 16: På Västkustbanan finns några längre sträckor som tillåter hastigheter på 280 eller mer. Skillnaden mellan sth 250 och 280 blir som mest 3 minuter för tåg med $P = 25{,}0$. En ökning av effekten har relativt liten påverkan på körtiderna. Tydligast är den i steget 10,8 till 15,3 med sth 280. På grund av endast tre stationsuppehåll har a_s ingen större påverkan.



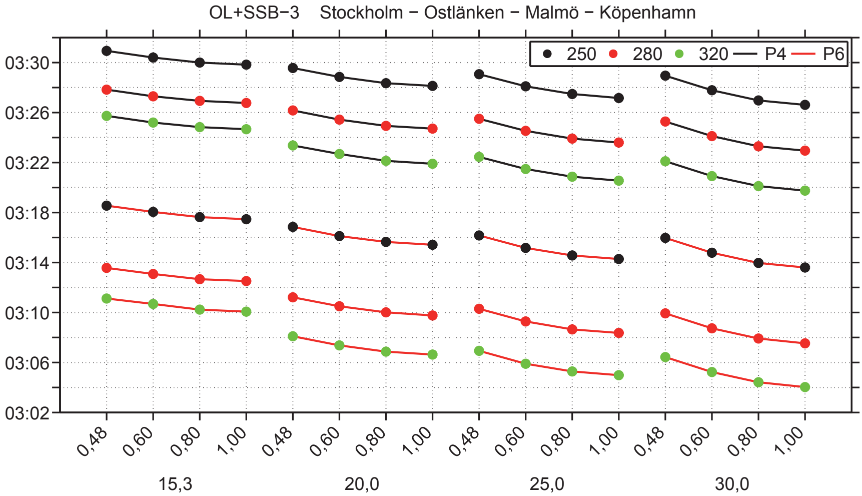
Figur 17: Tidsskillnaden mellan olika a_s varierar mellan 4–5 minuter. Skillnaderna i sth-nivå är små. En ökning av P ger mellan 1–5 minuter beroende på vilket effektsteg som avses samt nivån på a_s.



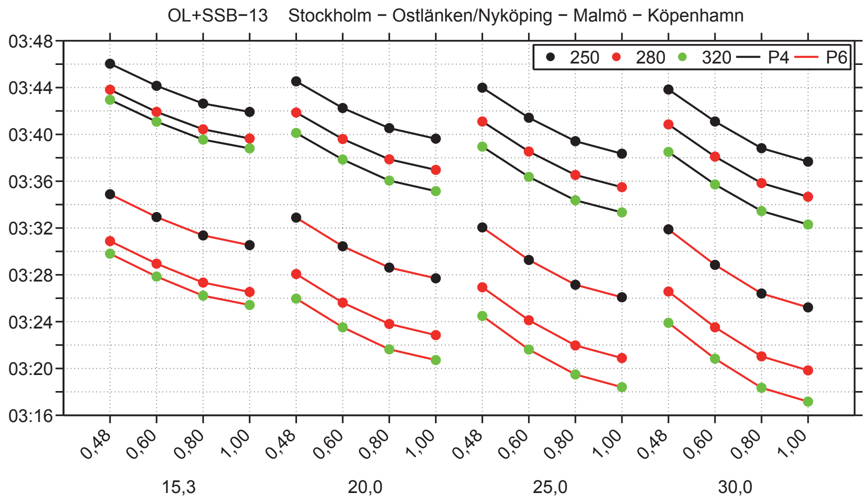
Figur 18: Hastighetsprofilen mellan Stockholm och Umeå har några kortare sträckor som tillåter hastigheter på 280 fram till Gävle. Skillnaden i körtid för tåg med sth 250 och 280 (som mest 6 minuter) får anses liten med tanke på banans längd. Om tågen kan köras efter hastighetsprofil 6 istället för 4 ger det ungefär 20 minuter kortare körtid. Liksom i tidigare fall ger inte effekthöjningarna några större tidsvinster i stegintervallet 15,3–25,0 givet att a_s är lika. Skillnaden mellan svagaste ($P = 10{,}8$ / $a_s = 0{,}48$) och starkaste tåg ($P = 25{,}0$ / $a_s = 1{,}00$) för sth 280 är 13 minuter.



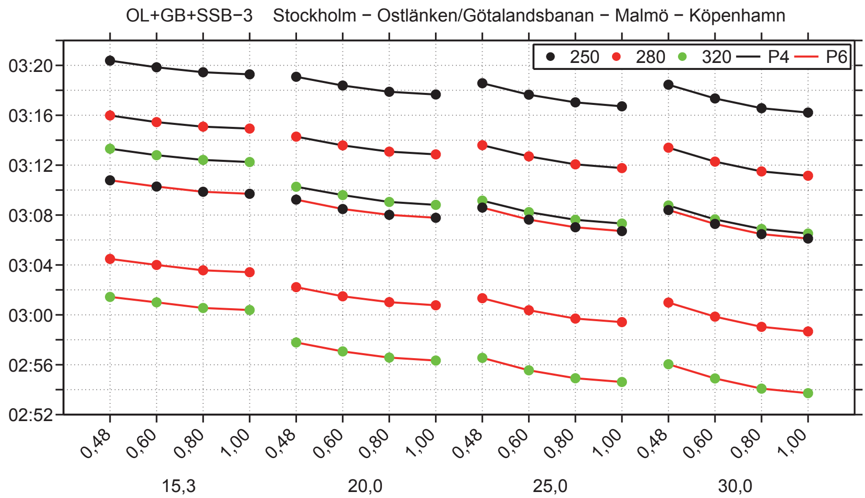
Figur 19: Liknande mönster som i föregående fall. Skillnaderna är inte lika tydliga på grund av den kortare sträckan. Antalet uppehåll bidrar också till att skillnaden i körtid för tåg med sth 250 respektive 280 blir marginella. Körtidsskillnaden mellan tåg med lägsta och högsta effekt och startacceleration är 16 minuter.



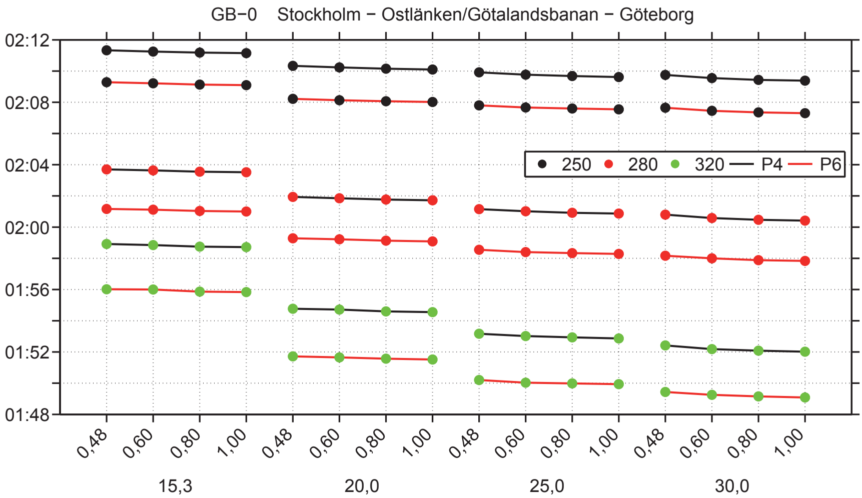
Figur 20: Hastigheten 320 tillåts på Ostlänken mellan Järna och Linköping. Hastighetssänkningar finns vid Norrköping och Linköping. Skillnaden för tåg med sth 250 och 280 varierar mellan 3–6 minuter. Vid en jämförelse mellan sth 280 och 320 krymper skillnaden till 2–4 minuter. För 320-tåg är största körtidsskillnaden med avseende på effektnivå 8 minuter. Antalet stationsuppehåll gör att a_s inte påverkar tiderna nämnvärt. Eftersom en stor del av sträckan inte tillåter hastigheter över 250, ger hastighetsprofil 4 längre tider än profil 6, även vid en jämförelse för tåg med sth 250 mot 320.



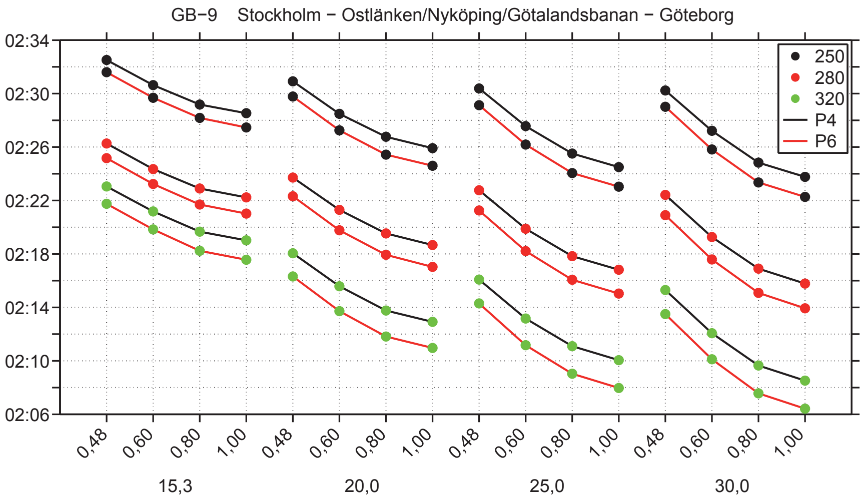
Figur 21: Skillnaderna mellan olika sth är något mindre än i föregående fall på grund av att uppehållen är fler. För tåg med $P = 30{,}0$ och $a_s = 1{,}00$ är skillnaden störst med avseende på effekt, 9 minuter. Tågen med $P = 15{,}3$ har en accelerationstid på nästan 750 sekunder för att komma upp i hastigheten 320 (figur 25), vilket ger en tydlig effekt på körtiderna.



Figur 22: Jämfört med variant OL–SSB–3 går tågen på Götalandsbanan till Gripenberg innan övergång sker till nuvarande Södra stambanan, vilket förkortar körtiderna med cirka 10 minuter. Tåg med sth 320 som körs enligt hastighetsprofil 4 får nästan samma körtider som tåg med sth 250 som körs enligt profil 6.



Figur 23: Inga uppehåll mellan Stockholm och Göteborg medför att a_s inte har någon påverkan alls. Nivån på effekten har också liten påverkan, skillnaden för tåg med sth 320 för effektsteget 15,3 till 20,0 är tydligast. Någon större skillnad mellan profil 4 och 6 finns inte.



Figur 24: Tågen körs via bibanan genom Nyköping. Skillnaderna mellan hastighetsprofilerna är marginella av samma anledning som i föregående exempel. Höjning av a_s ger mellan 1–3 minuter i förkortad körtid. Största skillnaden i effektnivå är cirka 11 minuter. Körtiden varierar mellan 3–8 minuter för tåg med olika sth.

5 Slutsatser

Vilka körtider som erhålls påverkas både av tågprestanda och banstandard. Vilken tågprestanda som är lämplig varierar beroende på vilken typ av bana tåget främst skall trafikera. En bana med hög kontinuerlig sth ställer delvis andra krav på fordonet än en bana med stora variationer i sth. Antal stationsuppehåll påverkar körtiderna olika mycket beroende på vilken prestanda tåget har.

5.1 Hastighetsprofiler

Ett tåg med bra gångegenskaper i spåret kan tillsammans med andra egenskaper innebära att högre hastigheter kan användas på befintliga banor. För att erhålla avsevärda höjningar i sth krävs dock oftast åtgärder i banunderbyggnad och att begränsande kurvradier åtgärdas. Dessutom ställs krav på anpassning av övriga tekniska system, till exempel kontaktledning och signalsystem.

Kortare banavsnitt som har högre sth än hastighetssteget innan och efter kan vara svåra att utnyttja. Om skillnaden i sth är för stor i förhållande till längden på det banavsnitt som har högre sth kan accelerations- och retardationstiderna bli relativt långa. Detta gäller främst vid höga sth och beror till stor del av tågets effekt. En hög kontinuerlig sth medför kortare körtider, men om hastighetsprofilen bryts med stora hastighetssänkningar tillkommer betydande körtidsförlängningar, även om sänkningarna ligger på kortare sträckor.

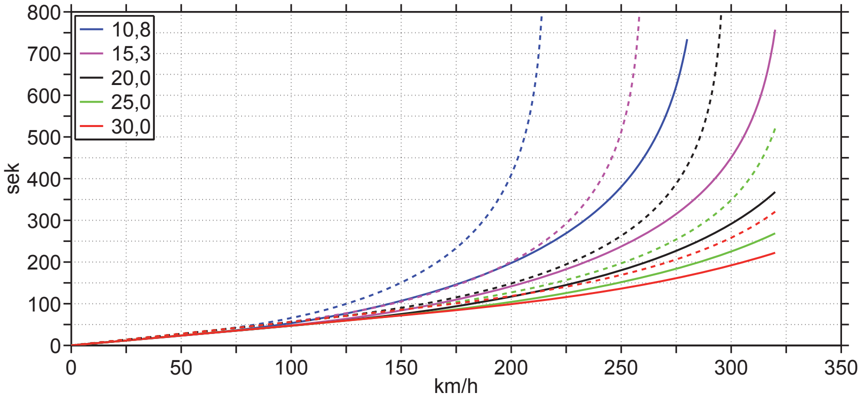
I hastighetsprofilerna för de olika banorna förekommer ibland stora hastighetssänkningar vid passage av stationer. Oftast begränsas sth av kurvor, växlar, plattformar och ibland av signalsystem. Däremot kan orsakerna som begränsar sth i förhållande till omgivande linjesträckor vid en del mindre och medelstora stationer i vissa fall åtgärdas och bidra till kortare körtider i framtiden.

Diagrammen som visar hastighetsprofilerna innehåller dagens B- och S-profil samt två beräknade hastighetsprofiler, P4 och P6 (tabell 1). Vid höga kurvradier konvergerar hastigheterna för de beräknade profilerna (undantaget P4) som en följd av sidvindsbegränsningen som införs i ekvation 6. Skillnaderna i sth är små för profil 5–7. I bilaga B finns körtider beräknade på samtliga hastighetsprofiler.

5.2 Tågprestanda

Effektbehovet per massa ökar med stigande hastighet och krav på residualaccelerationen i höga hastigheter är ofta dimensionerande för effekten. I ett lägre hastighetsintervall har a_s en tydligare påverkan på accelerationstiden. När den möjliga dragkraften är lika med tågets gångmotstånd kan tåget inte längre accelerera, det kan endast hålla konstant hastighet (jämviktshastighet). Tåg med effekt 10,8 kW/ton har svag accelerationsprestanda vid högre hastigheter och en ökning från 275 till 280 km/h tar drygt 100 sekunder. Samma förhållande gäller för tåg med effekt 15,3 kW/ton men på grund av högre effekt kan tåget i teorin uppnå 320 km/h.

Figur 25 visar hur effekten påverkar tiden under ett accelerationsförlopp. Gångmotstånd och acceleration är beräknade enligt ekvation 11 och 12 ($a_s = 0{,}60$). Tiderna förlängs om tågen accelererar i uppförslutning ($D_S > 0$). För tåg med låg effekt kan redan en liten lutning innebära att tåget inte uppnår sin sth förrän lutningen minskar eller försvinner helt. Accelerationstiderna ökar betydligt vid en uppförslutning på 10 ‰, vilket figur 25 visar. De tre tågen med lägst effekt klarar inte av att uppnå sina respektive sth. Vid nedförslutningar ($D_S < 0$) kan accelerationstiderna upp till en given hastighet analogt minskas.



Figur 25: Accelerationstider ($v_{\text{start}} = 0$) för olika effektnivåer utan lutning (heldragen linje) och med en uppförslutning på 10 ‰ (streckad linje).

I trafikeringsvarianter med många uppehåll har startaccelerationen en tydlig påverkan på körtiderna. Vid få stopp ger en måttlig höjning av a_s endast marginell påverkan. Effekten på körtiderna avtar med ökande a_s, en höjning från 0,60 till 0,80 m/s^2 ger en större körtidsförkortning jämfört med en höjning från 0,80 till 1,00 m/s^2. I de beräknade körtiderna ger oftast den första accelerationshöjningen från 0,48 till 0,60 m/s^2 den största körtidsförkortningen jämfört med efterföljande steg.

En höjning av effekten (P) har liknande påverkan på körtiderna som a_s. Största körtidsförkortningen uppnås i första effektsteget 10,8 till 15,3 kW/ton. Skillnaderna mellan olika a_s blir tydligare med ökande effekt. För banor med stor variation i hastighet ger en högre effekt att hastighetsprofilen kan utnyttjas bättre, speciellt gäller detta i de högre hastighetsintervallen. På en hastighetsprofil med liten variation har effekten mindre påverkan på körtiderna. Kraftiga och långa uppförslutningar ställer också krav på tågets effekt. I verklig trafik tvingas tågen ibland till att sänka hastigheten på grund av restriktiva signalbesked som beror av andra tåg. De situationerna bidrar också till kravet på tillräcklig startacceleration, effekt och retardationsförmåga.

En ytterligare faktor som påverkar körtiderna är tågens sth. Hur stora tidsvinster som kan uppnås beror till stor del av hastighetsprofilen och effektnivån. Tillräcklig effekt är nödvändig för att inte accelerationerna skall ta alltför lång tid. I en del fall är skillnaden mellan tåg som har sth 250 jämfört med sth 280 liten vilket, indikerar att hastighetsprofilen har få eller inga sträckor där hastigheter över 250 km/h tillåts. Kortare sträckor med högre hastighet har oftast marginell påverkan eftersom tid åtgår för acceleration och retardation.

Referenser

[1] O. Fröidh, E. Andersson, P. Hurtig, B. Kufver, R. Persson. Gröna Tåget - Framtida tågprestanda och bangeometri - Förstudie. PM 2005-11-25.

[2] O. Fröidh. Gångtidsberäkningar i Gröna Tåget. PM 2006-11-24.

[3] E. Andersson, M. Berg. Järnvägssystem och spårfordon. Institutionen för Farkostteknik, KTH, 2001.

[4] E. Andersson, R. Persson. Fall för gångtidsberäkningar i Gröna tåget - Förslag. PM 2006-11-15.

[5] M. Lindahl. Track geometry for high-speed railways - A literature survey and simulation of dynamic vehicle response. Institutionen för Farkostteknik, TRITA-FKT 2001:54, KTH, 2001.

[6] Banedanmark `www.bane.dk`

[7] Banverket Trafik. Riktlinjer för tidtabellskonstruktion på statens spåranläggningar. Föreskrift TF 601, 2000.

A Trafikplatssignaturer

I hastighetsdiagrammen och i beskrivningen av trafikeringsvarianter (tabell 3) används Banverkets trafikplatssignaturer för stationer inklusive hållplatser/hållställen (tabell A1). GB och OL anger att platsen ligger på Götalandsbanan respektive Ostlänken. CT betyder Citytunneln i Malmö.

Tabell A1: Trafikplatssignaturer

A	Alingsås	Hsa	Hovsta	My	Mjölby
Arb	Arboga	Hsd	Härnösand	N	Nässjö C
Arnc	Arlanda C	Hum	Husum	Nk	Nyköping C (OL)
Av	Alvesta	Höf	Hörnefors	Nkv	Nykvarn
Bs	Borås C (GB)	Id	Iggesund	Nog	Nordmaling
Bån	Båstad N	Jn	Järna	Nr	Norrköping C
Cph	Kastrup (flygpl.)	Jö	Jönköping (GB)	Sgs	Strängnäs
Cst	Stockholm C	K	Katrineholm C	Shv	Söderhamn V
E	Eslöv	Kb	Kungsbacka	Sk	Skövde C
Et	Eskilstuna C	Kh	København H	Suc	Sundsvall C
F	Falköping C	Kla	Kumla	Söö	Södertälje syd
Fab	Falkenberg	Krf	Kramfors	Tip	Tierp
Flb	Flemingsberg	Kör	Kungsör	Trå	Timrå
Fle	Flen	Laov	Laholm V	U	Uppsala C
G	Göteborg C	Lg	Läggesta	Udd	Uddeby (OL)
Gnp	Gnarp	Liop	Lisstorp (OL)	Ulr	Ulricehamn (GB)
Gp	Gripenberg (GB)	Lkö	Landskrona Ö	Uå	Umeå C
Gä	Gävle C	Lnd	Landvetter (GB)	Uåö	Umeå Ö
Hb	Helsingborg C	Lp	Linköping C	Vb	Varberg
Hd	Halmstad C	Lse	Ljusne	Ä	Ängelholm
Hkl	Hudiksvall	Lu	Lund	Äh	Älmhult
Hie	Hyllie (CT)	Lå	Laxå	Öb	Örebro S
Hm	Hässleholm	M	Malmö C (CT)	Ök	Örnsköldsvik
H	Hallsberg	Mdn	Mölndals nedre	Ökn	Örnsköldsvik N
Hr	Herrljunga	Mt	Mantorp (GB)	Ör	Örebro C

B Tidtabeller

Körtider från RailSys används för att skapa tidtabeller för referens- och typtågen (tabell 2) för de olika trafikeringsvarianterna. Tiderna (klockslagen) är beräknade enligt ekvation 13 i avsnitt 4.2.

Bilaga B ej bifogad i avhandlingen

Paper I

17–18 May 2010, Opatija, Croatia
First International Conference on Road and Rail Infrastructure

Simulation of modified timetables for high speed trains Stockholm – Göteborg

H. Sipilä
Division of Traffic and Logistics, Department of Transport and Economics, Royal Institute of Technology (KTH), Stockholm, Sweden

Abstract

In this research project, involving KTH, Swedish Rail Administration and train operating company SJ, timetable simulation is used on different areas in the Swedish rail network. The objective is to develop methods and strategies which can be useful for both long and short term timetable planning in the future.

This study presents some of the result from simulations done on the Western main in Sweden. Traffic is mixed with significant differences in average speeds. This makes the high speed passenger trains (X2000) sensitive for small delays and disturbances caused by other, much slower trains.

Simulation software RailSys was used to evaluate how increased and decreased time supplements for high speed trains affect punctuality. Also buffer times were increased between X2000 and other trains. Delay distributions from previous projects were implemented in order to model entry delays together with dwell and run time extensions.

Results show that increased buffer times can have a significant effect on punctuality. Some of the studied trains have a situation with dense traffic and high occurrence of overtakings. This clearly increases average delays and contributes to a lower punctuality.

Keywords: train traffic, timetable, simulation

1 Introduction

The Western main line is a highly utilized double track line between Swedens's two biggest cities, Stockholm and Göteborg (Fig. 1). High speed trains (X2000), commuter train systems together with several regional and InterCity systems share sections of the approximately 450 km long line. Freight traffic is also dense, especially between Hallsberg (H) and Göteborg. The Southern main line connects at Katrineholm (K) which gives a significant inflow and outflow of trains.

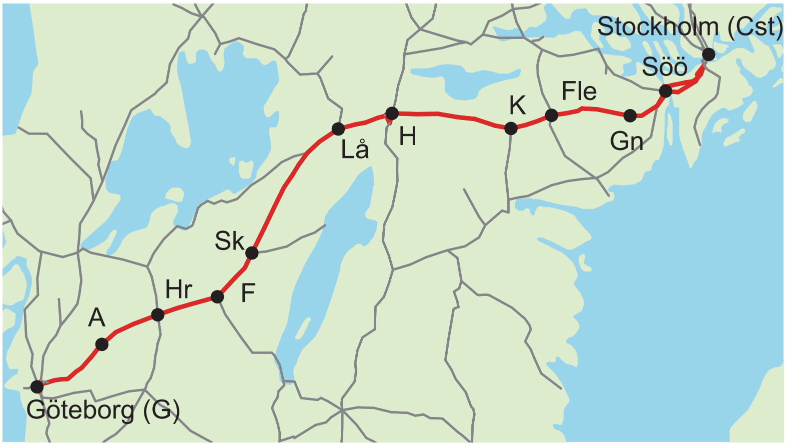


Figure 1. Western main line, Stockholm – Göteborg, with signatures for intermediate stations used in punctuality diagrams.

1.1 Background

Train slot demand has increased for both passenger and freight trains which makes the capacity stretched. The mixture of trains with different speeds contributes to delay propagation. High speed trains are sensitive to this, due to relatively large speed reductions which can result from disturbances. Punctuality is a widely used reliability measure. It calculates the percentage of trains arriving or departing within a certain number of minutes from scheduled time. In this study a 5-minute margin is used. Punctuality for high speed trains (X2000) between Stockholm and Göteborg drops with 15–20 percentage points depending on the direction. In an ongoing research project, methods and strategies for simulation based timetable planning are investigated. The project is a cooperation between KTH, Swedish Rail Administration and Sweden's largest passenger train operating company SJ.

1.2 Aim

The research project aims in investigating how simulation of networks with disturbed traffic can be used to improve timetable planning. This study focuses on how timetable changes, e.g. increased or decreased supplements, affect punctuality for a specific train group and what actions can be taken to improve performance. Simulation software RailSys has been used at KTH in several research projects and capacity studies. In 2005 Swedish Rail Administration decided to implement RailSys as a tool for capacity studies and extend it to timetable planning in the future. This paper presents some of the results from a more comprehensive research report [1].

2 Method

Synchronous simulation models, where all events happen in the same order as in reality, offers a powerful tool in evaluating and analyzing rail network performance. This means that dependencies between train runs, connections and vehicle circulation can be included in the simulation [2]. It is also possible to perform changes in infrastructure, timetable, dispatching strategies and evaluate different setups. This can give further input to both short and long term planning. In the performed simulations only the timetable is varied. Infrastructure, vehicle models, delay distributions and other settings are kept unchanged.

2.1 Infrastructure model

The infrastructure model consists of tracks, stations, signalling system, speed profiles etc. Track and signalling plans used by Swedish Rail Administration provide detailed data to the infrastructure model. Some connecting lines are included in the network to make inflows and outflows more realistic. With this approach, influence of crossing train runs is included in the simulations. This is important, especially if the traffic is frequent.

2.2 Modelling of primary delays

To reflect real operation different type of disturbances are assigned to the trains. Entry delays, dwell time and run time extensions affect trains at relevant locations in the network. An important part is therefore to construct realistic delay distributions which can be used in RailSys. A previous project focused on reducing disturbance levels caused by infrastructure and vehicle malfunctions combined with operator related errors. In that study delay distributions were designed from real operational data based on year 2006 [3].

Entry delays are applied at stations where trains are created in the simulation model. These delays are quite easy to obtain from delay statistics. The whole set of trains is divided into smaller groups which gives more precise disturbance levels.

Dwell time extensions model variable boarding and alighting time for passengers at stations. Data for these delay distributions were compiled from manual measurements (performed by SJ) of passenger exchange times. Representative measured stations were chosen to model large, medium and small stations from an exchange time view. Dwell time extensions are difficult to estimate from registered delay statistics [4].

An important part of the simulation setup is to apply realistic run time extensions which potentially affect trains network wide. Since these extensions are meant to represent primary delays, registered data needs to be filtered by available cause-reported delays. This may be difficult to analyze since cause reports are made manually by dispatchers and do not cover run time extensions less than 6 minutes on single inter-station distances.

A slow growing delay, which for example could be the result of decreased vehicle performance, can lack a cause report even if the delay eventually becomes large. The network was divided into subsections, where larger stations or junctions are used as border points. Run time extensions are applied roughly in the middle of these subsections.

An ideal situation is to use distributions for run time extensions which represent only primary delays. Knock-on delays are created in the simulations and can later be evaluated. When delay variance is high, which is common for freight trains, it may be difficult to construct subgroups of trains and to adjust the delay distribution levels acting on them [3, 5]. High primary delays, originating from major infrastructure or vehicle breakdowns in the registered data, are avoided by truncating the distributions at some levels. These events usually require advanced dispatching measures, e.g. rerouting through other line sections.

2.3 Timetable

Passenger train configurations are usually well known. However, this does not apply for freight trains which mostly have a large variance in both length and weight. Also day to day variation for specific train numbers can be significant. Most of the freight trains were modelled with weight 1000 tons, length 400 m and with maximum speed 90–100 km/h. Average length and weight values for freight trains running on weekdays in October 2008 between Hallsberg and Göteborg were 410 m and 880 tons [6]. Usage of running time allowances to compensate for delays was limited to a certain percentage of the excess time and applied differently to passenger and freight trains [5].

Recovery times can decrease or avoid delays for specific trains. These are obtained by adding allowances or supplements to the minimum running time. Buffer times are added to minimum headway between two trains and act as recovery times of the timetable [7]. Most of the timetable changes made on other trains are minor adjustments. However, some trains (mostly freight) are completely rescheduled. Simulated timetables are:

- Reference timetable for year 2009 (T09)
- Decreased timetable supplements with 4 minutes for X2000
- Increased timetable supplements with 4 minutes for X2000
- Increased buffer times between X2000 and other trains

The simulated reference timetable was created according to scheduled train plan valid for Thursday 29th January 2009, which was a normal weekday and not linked to any public holiday. Some preplanned freight trains were cancelled, this was compensated with additional scheduled extra trains. X2000 trains are, with a few exceptions, running every hour with departures between 5 and 21, most of them make four intermediate stops. The stopping pattern shows some variance regarding to which stations are used, however all trains have a scheduled stop in Skövde (Sk). In addition non-stop trains operate during morning and afternoon peak hours.

3 Results

First approach was to analyze the X2000 trains on aggregated level where results are calculated for the whole group in each direction. It is also of interest to study individual trains. This requires that the number of simulated cycles is sufficient in regard to output variance when different sets of random numbers for generated primary delays are used. Performed simulations were done with 1600 cycles, which gave an acceptable variance in output data.

3.1 Aggregated level

The simulated reference timetable shows that some sections are problematic in both directions, especially regarding to relative distance (Fig. 2). This is partly explained by dense traffic on some sections and also by tightly scheduled supplements or allowances. X2000 trains have usually some extra timetable supplements when approaching Göteborg and Stockholm, which may reduce delays. Whether this feature can be used or not depends on the traffic situation further ahead.

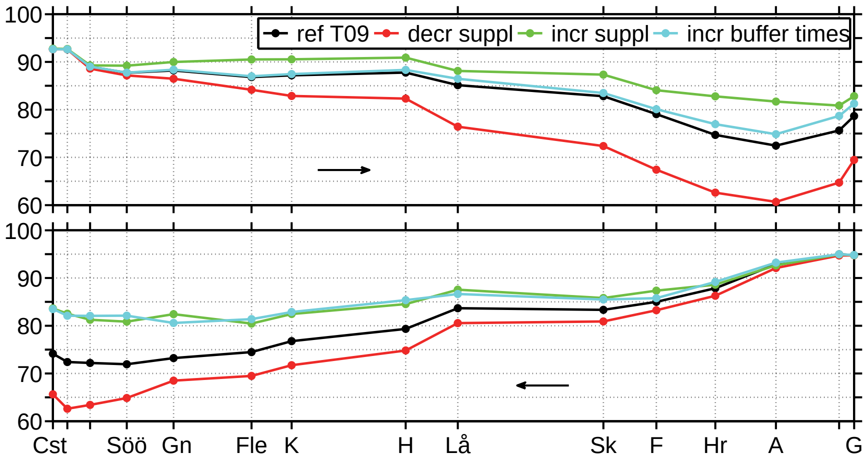


Figure 2. Simulated punctuality (%) on 5-minute level for X2000 trains. Reference timetable T09 with three alternatives. Relative scale on station axis.

If timetable supplements are decreased arrival punctuality at end stations drops with 10 percentage points. Approximately the opposite occurs for increased supplements, trains bound for Stockholm have a larger improvement (Fig. 2). A considerable number of timetable changes was performed in order to adjust other trains when supplements were changed for X2000. As mentioned before primary delays are fixed, meaning that all trains experience the same frequency and level of primary disturbances in each simulated variant.

In the third variant buffer times between X2000 and other potentially conflicting trains were increased to 5 minutes. This holds both for line sections and for crossing train movements at stations, except on the Stockholm bottleneck. Simulated punctuality improved significantly for trains bound for Stockholm. Results in the other direction showed only marginal improvements (Fig. 2). The amount of changes applied in the timetable was also less in this direction, therefore potential improvement was smaller.

3.2 Individual level

Punctuality gives only limited information about train delays. The large number of small delays is not fully described by the quite rough punctuality measure. A more comprehensive picture can be obtained if means and standard deviations are studied [4]. It can also be valuable to break up aggregated groups and present each train individually. This gives insight in whether trains in a group have equal or spread performance. Poor performance for one or a few trains could imply low quality train slots with risk for high disturbances, timetable and running time mismatch or problems with dispatching measures in the simulations.

Simulated mean arrival delays and standard deviations are lower for trains arriving in Stockholm compared to Göteborg (Fig. 3). One probable explanation is that if a late train approaching Alingsås ends up behind a slower train, significant delay increase will occur. Commuter trains make 10 intermediate stops on this section, overtaking possibilities are limited. Trains approaching Stockholm from the Western main line have smaller differences in scheduled average speeds. Compared to the Göteborg area, changed sequence of trains have a lighter impact on delay development.

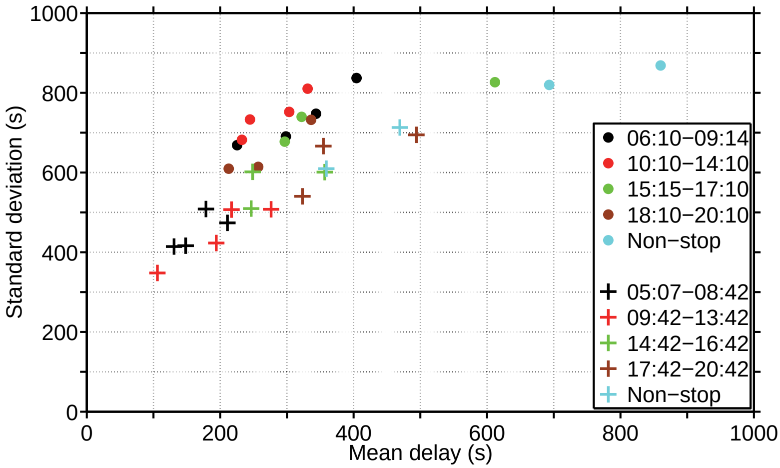


Figure 3. Arrival values in Göteborg (•) and Stockholm (+). Departure time intervals from start station in legend. Timetable T09.

Non-stop trains have generally higher mean and standard deviation values. Main causes are tight train slots, smaller supplements and lack of ability to reduce delays on scheduled stops. The variant with decreased supplements gives larger spreading in mean and standard deviation values. Increased supplements and buffer times have less spreading compared to timetable T09 in Fig. 3. However, Stockholm has smaller values in average compared to Göteborg in all variants.

An example of improved punctuality with increased buffer times for selected trains is shown in Fig. 4. Some trains with initial high punctuality performance obtain only marginal or no improvements. An obvious reason is that these trains already have robust train slots where the probability for small disturbances caused by other trains is low. Trains with low punctuality levels have tight buffer times on several sections or stations with other trains.

Frequent scheduled overtakings of freight trains is also a disturbance source. Some trains benefit from postponed commuter train departures, thus locally creating more robust solutions. Altered overtaking stations and time adjustments for freight trains can also create better conditions.

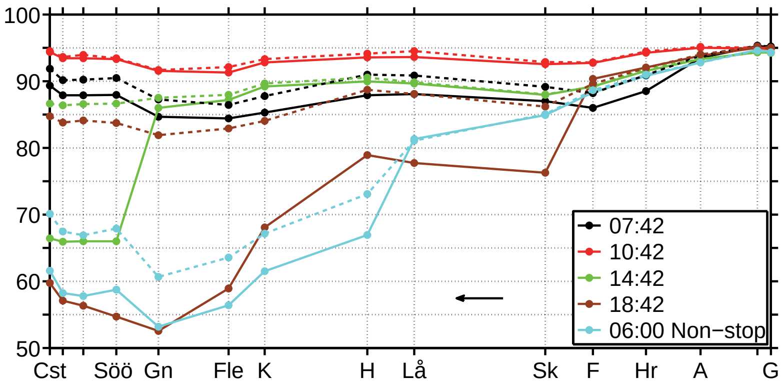


Figure 4. Simulated punctuality (%) for selected trains to Stockholm, timetable T09 (solid) and increased buffer times (dashed).

4 Conclusions

In one sense the situation for X2000 trains is better in the Stockholm area compared to Göteborg. When disturbances occur and scheduled train sequences are changed, the differences in average speeds are lower in Stockholm. The traffic is of course intensive, especially during peak hours. A substantial improvement was observed for specific trains when scheduled buffer times to departing commuter trains was increased from 2–3 to 5 minutes in Gnesta. Commuter train turnarounds at this station are also discussed in [8]. Freight train levels in Stockholm are quite small compared to Göteborg region.

On the other side of the Western main line mixture of trains with different speeds is significant. Freight traffic is intensive and even though stations with overtaking possibilities exist, initial delays are likely to increase. X2000 trains leaving Göteborg during late afternoon and evening have a higher probability for decreased punctuality compared to morning departures. In particular northbound freight train departures from Göteborg are dense during afternoon hours. Freight trains ahead of schedule are not uncommon in Sweden, which further increase the variance in real operations. This is however difficult to model, early departures are therefore assumed to be on time in delay statistics used in RailSys.

Punctuality effects on other trains in the simulated variants were small. However, this was only checked on fairly rough aggregated levels. This means that most of the trains on which these evaluations were based on, were hardly affected by the timetable changes. This kind of analysis could be performed in more detail, with emphasis on trains scheduled close in time to X2000.

High variances for arrival delays in Fig. 3 indicate that general timetable supplements are difficult to apply. Instead individual trains with poor performance could be studied to find solutions. Simulations with common random numbers show that deviation between mean delays for individual trains is approximately $\pm 2\%$ in average and $\pm 6\%$ as most. A complete calibration and validation process was not conducted in this study, instead input was used from earlier projects performed at KTH [3, 5].

References

[1] Sipilä, H.: Tidtabellsläggning med hjälp av simulering – Effekter av olika tillägg och marginaler på X2000-tågen Stockholm – Göteborg (Simulation based timetable costruction – Effects of allowances and supplements on X2000-trains Stockholm - Göteborg). KTH, 2009. (In Swedish)

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

[3] Nelldal, B-L.; Lindfeldt, O.; Sipilä, H.; Wolfmaier, J.: Förbättrad punktlighet på X2000 – analys med hjälp av simulering (Improved punctuality for X2000 – a simulation approach). KTH, 2008. (In Swedish)

[4] Yuan, J.: Stochastic modelling of train delays and delay propagations in stations. PhD thesis, Delft University of Technology, 2006.

[5] Lindfeldt, O.; Sipilä, H.: Validation of a simulation model for mixed traffic on a swedish double-track railway line. Railway Engineering – 2009, 10th International Conference, London, UK, 2009.

[6] Nelldal, B-L.; Lindfeldt, A.; Lindfeldt, O.: Kapacitetsanalys av järnvägsnätet i Sverige – Delrapport 1–3 (Capacity analysis of the Swedish rail network). KTH, 2009. (In Swedish)

[7] Rudolph, R.: Allowances and margins in railway scheduling. Proceedings of the World Congress on Railway Research, pp 230–238. Edinburgh, Scotland, 2003.

[8] Anand, N.; Anayi, M.: Improving punctuality of train traffic on Western main line of Swedish railway network: Simulation approach. Proceedings of ASME 2009 Rail Transportation Division Fall Conference, Forth Worth, USA, 2009.

Paper II

CALIBRATION OF SIMULATION MODEL ON THE SOUTHERN MAIN LINE IN SWEDEN

Hans Sipilä
Royal Institute of Technology (KTH)
Division of Transport and Logistics
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 in-frastructure 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.

Figure 1: Southern main line map, evaluated section shown in red. Parts of the blue sections are included in the simulated network. Distance Katrineholm–Hässleholm is 400 km.
(Map: Josef Andersson, KTH)

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 prescheduled 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).

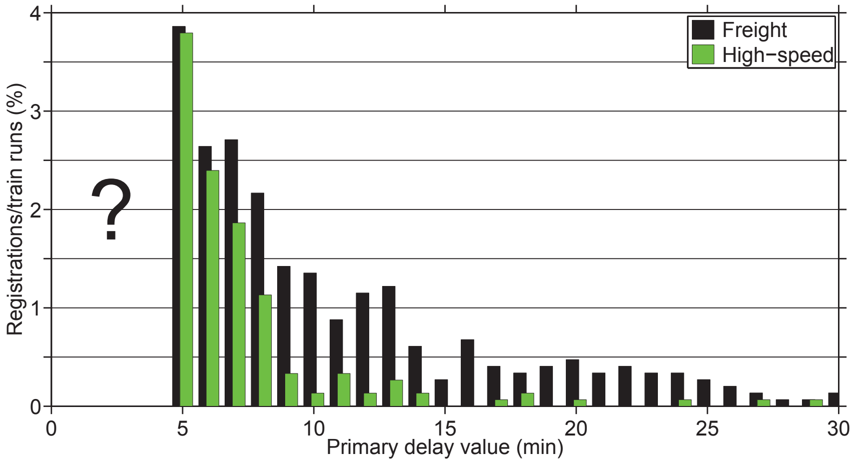


Figure 2: Primary cause reported delays for selected southbound trains on section Mjölby–Hässleholm (265 km). Weekdays, January to June 2008.

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.

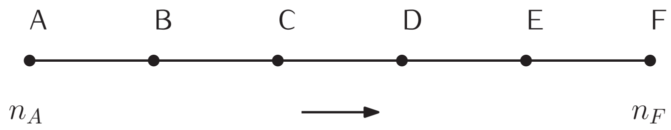


Figure 3: Train run from station A to F with registrations n_A and n_F.

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.

$$n_A \geq 0 \quad \Longrightarrow \quad t = \begin{cases} n_F - n_A & \text{if } \; n_A \leq n_F \\ 0 & \text{if } \; n_A > n_F \end{cases}$$

$$n_A < 0 \quad \Longrightarrow \quad t = \max\{0, n_F\}$$

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

$$m = 0 \quad \Longrightarrow \quad n_0 = N_0 + \sum_{m=1}^{m_{\max}} (1 - R) \cdot N_m$$

$$m > 0 \quad \Longrightarrow \quad n_m = R \cdot N_m$$

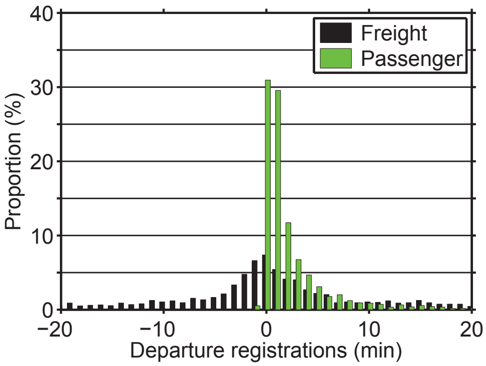
Four different levels are used for run time distributions and applied on two main groups, passenger and freight trains (table 1). Original distributions are compiled for seven passenger train groups and four freight groups in each direction. Train numbers are matched as far as possible, remaining trains 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)

Train category	Simulation cases 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
P	20	20	20	20	40	40	40	40	60	60	60	60	80	80	80	80
F	20	40	60	80	20	40	60	80	20	40	60	80	20	40	60	80

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.



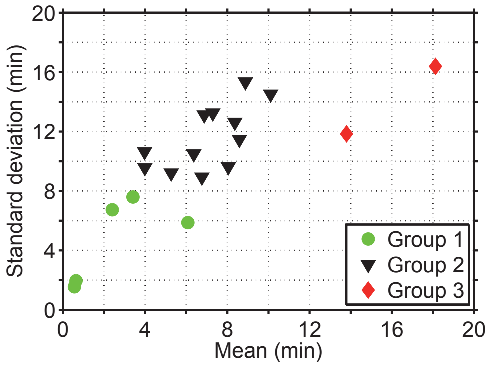


Figure 4: Departure distribution for northbound trains at Hässleholm. Trains ahead of schedule have values < 0 (left). Train entry grouping for southbound freight trains at Skänninge, network boundary station north of Mjölby (right).

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.

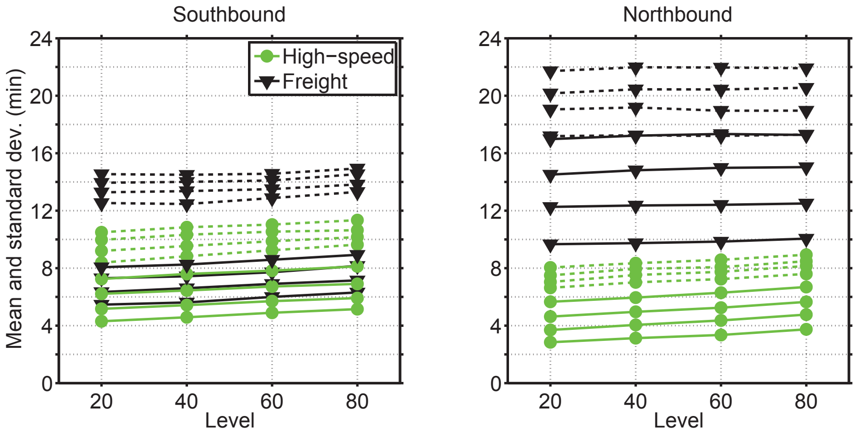


Figure 5: Exit delay variation for high-speed passenger trains and freight trains in day group, mean (solid) and standard deviation (dashed curves).

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 and Y.

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - Y_i)^2}$$

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.

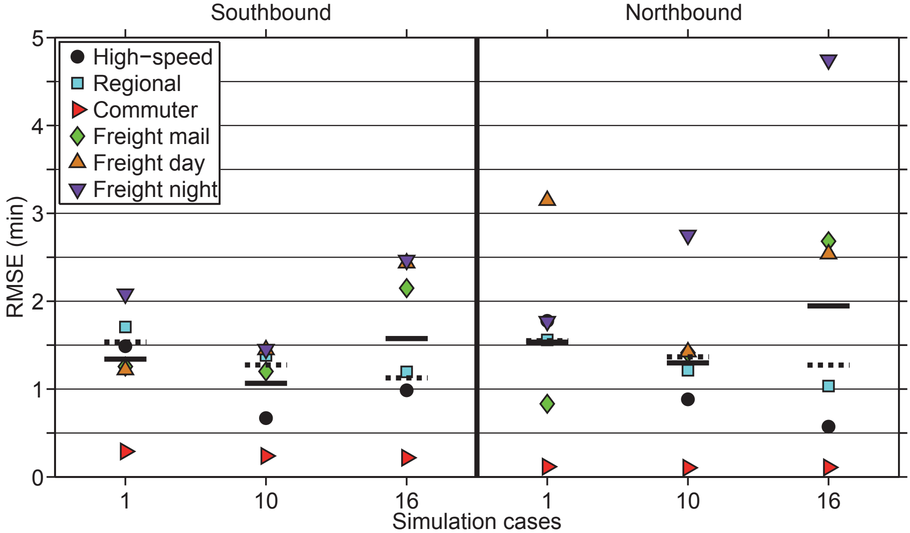


Figure 6: Root mean square error for mean delays. Average values of RMSE denoted by short lines, solid for mean and dashed for standard deviation.

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.

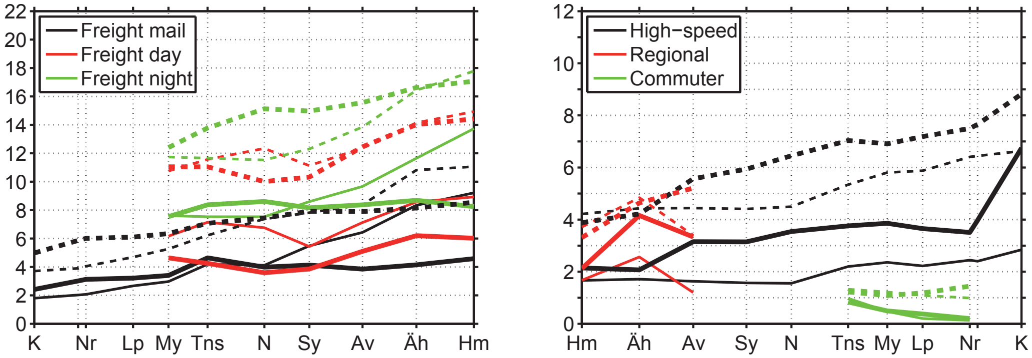


Figure 7: Simulated (thin) and observed delays (bold curves). Southbound freight trains, 80/80 case 16 (left). Northbound passenger trains, 20/20 case 1 (right). Mean values are solid lines and standard deviations are dashed.

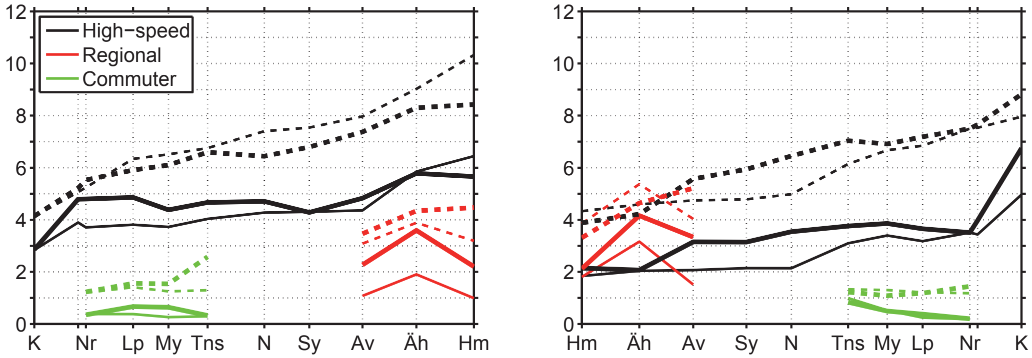


Figure 8: Simulated and observed delays. Passenger trains, 40/60 case 10.

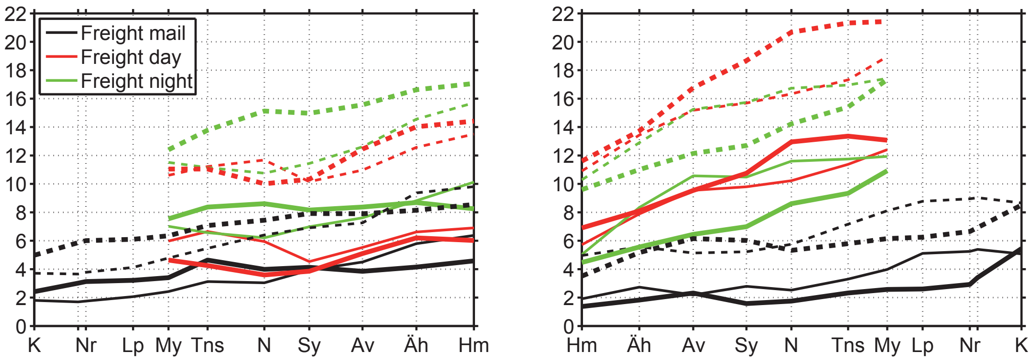


Figure 9: Simulated and observed delays. Freight trains, 40/60 case 10.

Differences between start values can partly be explained by dispatching decisions done by 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.

Koutsopoulos, H.; Wang, Z.: Simulation of urban rail operations: Application framework. Proceedings of the Transportation Research Board. 2006.

Lindfeldt, O.; Sipilä, H.: Validation of a simulation model for mixed traffic on a swedish double-track railway line. Railway Engineering – 2009, 10th International Conference, London, UK, 2009.

Nelldal, B-L.; Lindfeldt, O.; Sipilä, H.; Wolfmaier, J.: Förbättrad punktlighet på X2000 – analys med hjälp av simulering (Improved punctuality for X2000 – a simulation approach). KTH, 2008. (In Swedish)

Radtke, A.; Hauptmann D.: Automated planning of timetables in large railway networks using microscopic data basis and railway simulation techniques. Computers in Railways IX. WIT Press, 2004.

Rudolph, R.: Allowances and margins in railway scheduling. Proceedings of the World Congress on Railway Research, pp 230–238. Edinburgh, Scotland, 2003.

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

Sipilä, H.: Simulation of modified timetables for high speed trains Stockholm–Göteborg. In: Lakušić, S. (editor): Proceedings of the First International Conference on Road and Rail Infrastructure (CETRA 2010), pp 575–581. Opatija, Croatia, 2010.

TRITA-TSC-LIC 12-003
ISBN 978-91-85539-87-1