Simulation of rail traffic

Applications with timetable construction and delay modelling

HANS SIPILÄ

Licentiate Thesis in Transport Science
Stockholm, Sweden 2012
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.

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 consists of creating distributions based on deviations from scheduled running time. Since this represents 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 Wahlberg, 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 Lansberg (Trafikverket), Dan Olofsson, Mattias Mellberg and Brit-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 ∆.

Stockholm, May 2012

Hans Sipilä
List of publications

Reports


Papers


Related publications not included in thesis


List of publications

Reports


Papers


Related publications not included in thesis

## Contents

1 Introduction 1  
1.1 Background 1  
1.2 Objectives and scope 1  
1.3 Limitations 3  
1.4 Structure of the thesis 3  

2 Related research 5  

3 Methodology 9  
3.1 Defining simulation infrastructure 9  
3.2 Vehicle models and timetables 9  
3.3 Simulation of timetables 10  
3.4 Discussion of the methodological problems 10  
3.5 Modelling running time extensions 12  

4 Running time estimations for the Green Train 15  
4.1 Background 15  
4.2 Calculation of maximal curve speed 16  
4.3 Speed profiles 18  
4.4 Train characteristics 22  
4.5 Running time calculations 24  
4.6 Conclusions 29  

5 Timetable adjustments for high-speed trains 31  
5.1 Background 31  
5.2 Method 31  
5.3 Results 33  
5.4 Conclusions 38  

6 Estimating running time extensions 39  
6.1 Background 39  
6.2 Method 40  
6.3 Results 41  
6.4 Conclusions 46  

7 Summary and conclusions 47  

8 Contribution of the thesis 49  

9 Future work 51
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:
Nordic conditions [8].

to develop a concept proposal for a new, attractive high speed train adapted to
part of the Green Train research program. The main objective in the program is
used in calculating static speed profiles on existing and future lines. This work is
sets of track cant and cant deficiency combined with radii for circular curves are
celeration, power to ton ratio and top speed is described in section 4. Different
Running time calculations for varying train configurations regarding starting ac-
timation. Some issues that can be of interest are for example evaluation
time planning. This can be both in a short term perspective and for strategic
A general aim is also to develop and improve methods for using simulation in
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 Man-
agement 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 timeta-
bles. However, one of the most important factors in simulations of rail traffic and
what differ simulation from general timetable planning is delay modelling. A real-
istic simulation model needs some input delay definitions to reflect real conditions.
This stochastic feature is introduced in the model by different types of distribu-
tions. 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 ac-
celeration, 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 losing 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.

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.

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

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: Simulated (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.
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 dependent 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 competitive 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.

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 dependent 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 competitive 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 (f). 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.

\[ a_y = \frac{v^2}{R} \quad a_z = g \quad D = g h_0 \quad \phi \]

Figure 5: Horizontal (v²/R) and vertical acceleration (g). Acceleration parallel (a_y) and perpendicular to track plane (a_z).

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 (f). 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.

\[ a_y = \frac{v^2}{R} \quad a_z = g \quad D = g h_0 \quad \phi \]

Figure 5: Horizontal (v²/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 \( \phi \) lateral acceleration \( a_y \) is calculated by eqn. 1. If \( \phi \) is small the simplification in the second expression can be used.

\[
a_y = \frac{v^2}{R} \cos \phi - g \sin \phi \approx \frac{v^2}{R} - g \frac{D}{2b_0} \quad (1)
\]

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

\[
D_{EQ} = \frac{2b_0}{g} \frac{v^2}{R} = \left\{ v : m/s \rightarrow \text{km/h} \right\} = C \frac{v^2}{R} \quad (2)
\]

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 \( C \approx 11.8 \text{ mm} \cdot \text{m}^2/\text{km}^2 \).

Cant deficiency \( (I) \) exists if the applied cant is lower than \( D_{EQ} \) and is a measure of the additional cant needed to obtain \( D = D_{EQ} \). Cant deficiency is proportional to remaining track plane acceleration (eqn. 3)

\[ I = D_{EQ} - D \quad (3) \]

If the cant deficiency is negative, i.e. \( D > D_{EQ} \), 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 \( v_l \) is needed (eqn. 4).

\[ D \leq E + C \cdot \frac{v_l^2}{R} \quad (4) \]

Applied values in calculations are \( E = 110 \) mm and \( v_l = 90 \) km/h [3]. In practice, 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 crosswind. In the speed profile calculations maximal cant deficiency is \( I_{max} = 300 \) mm up to 225 km/h \( (v_{max}) \). 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 \( I_{max} \). Fig. 6 shows the principles for setting cant deficiency value for a specific curve radius.

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

\[
a_y = \frac{v^2}{R} \cos \phi - g \sin \phi \approx \frac{v^2}{R} - g \frac{D}{2b_0} \quad (1)
\]

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

\[
D_{EQ} = \frac{2b_0}{g} \frac{v^2}{R} = \left\{ v : m/s \rightarrow \text{km/h} \right\} = C \frac{v^2}{R} \quad (2)
\]

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 \( C \approx 11.8 \text{ mm} \cdot \text{m}^2/\text{km}^2 \).

Cant deficiency \( (I) \) exists if the applied cant is lower than \( D_{EQ} \) and is a measure of the additional cant needed to obtain \( D = D_{EQ} \). Cant deficiency is proportional to remaining track plane acceleration (eqn. 3)

\[ I = D_{EQ} - D \quad (3) \]

If the cant deficiency is negative, i.e. \( D > D_{EQ} \), 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 \( v_l \) is needed (eqn. 4).

\[ D \leq E + C \cdot \frac{v_l^2}{R} \quad (4) \]

Applied values in calculations are \( E = 110 \) mm and \( v_l = 90 \) km/h [3]. In practice, 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 crosswind. In the speed profile calculations maximal cant deficiency is \( I_{max} = 300 \) mm up to 225 km/h \( (v_{max}) \). 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 \( I_{max} \). Fig. 6 shows the principles for setting cant deficiency value for a specific curve radius.
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)

<table>
<thead>
<tr>
<th>Profile</th>
<th>Cant</th>
<th>Cant def.</th>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>150</td>
<td>150</td>
<td>Reference category B trains</td>
</tr>
<tr>
<td>P2</td>
<td>150</td>
<td>245</td>
<td>Reference category S trains</td>
</tr>
<tr>
<td>P3</td>
<td>150</td>
<td>245</td>
<td>Carbody tilt</td>
</tr>
<tr>
<td>P4</td>
<td>160</td>
<td>165</td>
<td>No carbody tilt</td>
</tr>
<tr>
<td>P5</td>
<td>160</td>
<td>245</td>
<td>Carbody tilt</td>
</tr>
<tr>
<td>P6</td>
<td>160</td>
<td>275</td>
<td>Carbody tilt</td>
</tr>
<tr>
<td>P7</td>
<td>160</td>
<td>300</td>
<td>Carbody tilt</td>
</tr>
</tbody>
</table>

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.

Fig. 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 Grodingle 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.
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.

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%.
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 eqns. 7 and 8 where train mass including dynamic mass supplement is used ($m_e$).

\[ F = m_e \cdot a_s \]  \( (7) \)
\[ P = F \cdot v \]  \( (8) \)

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 ($D_{\text{max}}$) is approximated with eqn. 9 where $A = 2400 \, N$, $B = 60 \, kg/s$ and $C = 6.5 \, kg/m$ are vehicle constants [6].

$$D_{\text{max}} = A + Bv + Cv^2$$

(9)

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

$$a_i = \frac{(F_i + F_{i+1}) - (D_i + D_{i+1})}{2m_e}$$

(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.

![Traction force and acceleration diagrams for reference trains and some variants of the Green Train](image)

Figure 11: Traction force, total resistance with $D_k = 0$ (dashed) and acceleration diagrams for reference trains and combinations of power output (kW/ton) and starting acceleration (m/s²).

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 high speeds. The sum of mechanical and aerodynamic resistance ($D_{\text{max}}$) is approximated with eqn. 9 where $A = 2400 \, N$, $B = 60 \, kg/s$ and $C = 6.5 \, kg/m$ are vehicle constants [6].

$$D_{\text{max}} = A + Bv + Cv^2$$

(9)

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

$$a_i = \frac{(F_i + F_{i+1}) - (D_i + D_{i+1})}{2m_e}$$

(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.

![Traction force and acceleration diagrams for reference trains and some variants of the Green Train](image)

Figure 11: Traction force, total resistance with $D_k = 0$ (dashed) and acceleration diagrams for reference trains and combinations of power output (kW/ton) and starting acceleration (m/s²).

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 high speeds.
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/s²), flat track (solid) and gradient 6‰ (dashed).](image)

### 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/s², 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>
<thead>
<tr>
<th>Time</th>
<th>Value +T</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note that all stops do not fit in table 5.
Table 2: Western Main Line, Stockholm–Göteborg, non-stop

<table>
<thead>
<tr>
<th>Profile</th>
<th>Train</th>
<th>Time</th>
<th>+T</th>
<th>Stockholm</th>
<th>Göteborg</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>X52</td>
<td>02:58:13</td>
<td>47</td>
<td>0.00</td>
<td>2.59</td>
</tr>
<tr>
<td>P2</td>
<td>X2</td>
<td>02:37:30</td>
<td>30</td>
<td>0.00</td>
<td>2.38</td>
</tr>
<tr>
<td>P4</td>
<td>250</td>
<td>02:40:53</td>
<td>5</td>
<td>0.00</td>
<td>2.41</td>
</tr>
<tr>
<td>P4</td>
<td>280</td>
<td>02:40:02</td>
<td>-2</td>
<td>0.00</td>
<td>2.40</td>
</tr>
<tr>
<td>P6</td>
<td>250</td>
<td>02:25:55</td>
<td>-5</td>
<td>0.00</td>
<td>2.23</td>
</tr>
<tr>
<td>P6</td>
<td>280</td>
<td>02:21:56</td>
<td>4</td>
<td>0.00</td>
<td>2.22</td>
</tr>
</tbody>
</table>

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

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

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

<table>
<thead>
<tr>
<th>Profile</th>
<th>Train</th>
<th>Time</th>
<th>+T</th>
<th>Stockholm</th>
<th>Hässleholm</th>
<th>Lund</th>
<th>Malmö</th>
<th>Copenhagen</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>X52</td>
<td>04:03:04</td>
<td>116</td>
<td>0.00</td>
<td>3.06</td>
<td>3.35</td>
<td>3.45</td>
<td>4.05</td>
</tr>
<tr>
<td>P2</td>
<td>X2</td>
<td>03:51:24</td>
<td>36</td>
<td>0.00</td>
<td>2.55</td>
<td>3.22</td>
<td>3.32</td>
<td>3.52</td>
</tr>
<tr>
<td>P4</td>
<td>250</td>
<td>03:34:51</td>
<td>230</td>
<td>0.00</td>
<td>2.41</td>
<td>3.07</td>
<td>3.17</td>
<td>3.37</td>
</tr>
<tr>
<td>P4</td>
<td>280</td>
<td>03:31:26</td>
<td>234</td>
<td>0.00</td>
<td>2.38</td>
<td>2.94</td>
<td>3.14</td>
<td>3.34</td>
</tr>
<tr>
<td>P6</td>
<td>250</td>
<td>03:22:07</td>
<td>113</td>
<td>0.00</td>
<td>2.31</td>
<td>2.55</td>
<td>3.05</td>
<td>3.24</td>
</tr>
<tr>
<td>P6</td>
<td>280</td>
<td>03:16:30</td>
<td>90</td>
<td>0.00</td>
<td>2.25</td>
<td>2.49</td>
<td>2.59</td>
<td>3.18</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Profile</th>
<th>Train</th>
<th>Time</th>
<th>+T</th>
<th>Stockholm</th>
<th>Uppsala</th>
<th>Gävle</th>
<th>Söderhamn</th>
<th>Sundsvall</th>
<th>Härnösand</th>
<th>Örnsköldsvik</th>
<th>Umeå</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>X52</td>
<td>05:43:23</td>
<td>157</td>
<td>0.00</td>
<td>1.17</td>
<td>1.57</td>
<td>3.19</td>
<td>3.47</td>
<td>4.07</td>
<td>5.05</td>
<td>5.46</td>
</tr>
<tr>
<td>P2</td>
<td>X2</td>
<td>05:23:20</td>
<td>280</td>
<td>0.00</td>
<td>1.16</td>
<td>1.52</td>
<td>3.06</td>
<td>3.49</td>
<td>4.46</td>
<td>5.28</td>
<td></td>
</tr>
<tr>
<td>P4</td>
<td>250</td>
<td>05:13:10</td>
<td>230</td>
<td>0.00</td>
<td>1.10</td>
<td>1.48</td>
<td>3.04</td>
<td>3.49</td>
<td>4.41</td>
<td>5.17</td>
<td></td>
</tr>
<tr>
<td>P4</td>
<td>280</td>
<td>05:09:41</td>
<td>207</td>
<td>0.00</td>
<td>1.10</td>
<td>1.47</td>
<td>3.01</td>
<td>3.49</td>
<td>4.37</td>
<td>5.13</td>
<td></td>
</tr>
<tr>
<td>P6</td>
<td>250</td>
<td>04:51:14</td>
<td>286</td>
<td>0.00</td>
<td>1.07</td>
<td>1.41</td>
<td>2.51</td>
<td>3.51</td>
<td>4.20</td>
<td>4.55</td>
<td></td>
</tr>
<tr>
<td>P6</td>
<td>280</td>
<td>04:44:25</td>
<td>335</td>
<td>0.00</td>
<td>1.06</td>
<td>1.40</td>
<td>2.49</td>
<td>3.29</td>
<td>4.17</td>
<td>4.56</td>
<td></td>
</tr>
</tbody>
</table>

25
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

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top speed (km/h)</td>
<td>250 280 320</td>
</tr>
<tr>
<td>Power output/mass $P^*$ (kW/ton)</td>
<td>10.8 15.3 20.0 25.0 30.0</td>
</tr>
<tr>
<td>Starting acceleration $a_s$ (m/s$^2$)</td>
<td>0.48 0.6 0.8 1.0</td>
</tr>
</tbody>
</table>

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.

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 ($v_{\text{max}} = 200$ 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.

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.

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

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.

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.

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

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.

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.

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.

\[
\begin{array}{cccccc}
A & B & C & D & E & F \\
\eta_A & & & & \eta_F
\end{array}
\]

Figure 23: Definition of line sectioning with time registrations \(\eta_A\) and \(\eta_F\).

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 \(\eta_A\) and \(\eta_F\) 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

\[
\begin{cases}
\eta_F - \eta_A & \text{if } \eta_A \leq \eta_F \\
0 & \text{if } \eta_A > \eta_F
\end{cases}
\]

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.

\[
\begin{array}{cccccc}
A & B & C & D & E & F \\
\eta_A & & & & \eta_F
\end{array}
\]

Figure 23: Definition of line sectioning with time registrations \(\eta_A\) and \(\eta_F\).
Fig. 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.
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.
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 give 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.

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

<table>
<thead>
<tr>
<th>Delay (x-axis)</th>
<th>Southbound high-speed trains</th>
<th>Northbound freight trains</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Reg</td>
<td>Sim</td>
</tr>
<tr>
<td>0-2</td>
<td>715</td>
<td>1368</td>
</tr>
<tr>
<td>2-4</td>
<td>143</td>
<td>245</td>
</tr>
<tr>
<td>4-6</td>
<td>46</td>
<td>97</td>
</tr>
<tr>
<td>6-8</td>
<td>1150</td>
<td>4071</td>
</tr>
<tr>
<td>8-10</td>
<td>40</td>
<td>89</td>
</tr>
<tr>
<td>0-12</td>
<td>72</td>
<td>55</td>
</tr>
<tr>
<td>12-24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24-36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36-48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>48-60</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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).

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

<table>
<thead>
<tr>
<th>Delay (x-axis)</th>
<th>Southbound high-speed trains</th>
<th>Northbound freight trains</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Reg</td>
<td>Sim</td>
</tr>
<tr>
<td>0-2</td>
<td>715</td>
<td>1368</td>
</tr>
<tr>
<td>2-4</td>
<td>143</td>
<td>245</td>
</tr>
<tr>
<td>4-6</td>
<td>46</td>
<td>97</td>
</tr>
<tr>
<td>6-8</td>
<td>1150</td>
<td>4071</td>
</tr>
<tr>
<td>8-10</td>
<td>40</td>
<td>89</td>
</tr>
<tr>
<td>0-12</td>
<td>55</td>
<td>55</td>
</tr>
<tr>
<td>12-24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24-36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36-48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>48-60</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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/s² 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/s². 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.

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/s² 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/s². 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 imply 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.

Many of the problems for high-speed and other trains are initiated by small delays which in some cases eventually imply 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


