Quality on single-track railway lines with passenger traffic - Analytical model for evaluation of crossing stations and partial double-tracks

OLOV LINDFELDT

KTH Architecture and the Built Environment

Licentiate Thesis in Infrastructure
Stockholm, Sweden 2007
Quality on single-track railway lines with passenger traffic - Analytical model for evaluation of crossing stations and partial double-tracks

Olov Lindfeldt

Licentiate thesis
Division of Traffic and Logistics
Department of Transport and Economics
Stockholm 2007
© 2007 Olov Lindfeldt
olovl@infra.kth.se

TRITA-TEC-LIC 07-003

ISSN 1653-445X


Cover picture: Crossing at Tällberg (Sweden), February 2007. Note that the exit signal at the station limit is already clear, and so the locking time for the coming exit train path is minimized.
Abstract

Railway transportation is showing a substantial increase. Investments in new infrastructure, new fast and comfortable vehicles, and high frequency of service are important factors behind the increase.

Infrastructure configuration and timetable construction play important roles in the competitiveness of railway transportation. This is especially true on single-track lines where the travel times and other timetable related parameters are severely restricted by crossings (train meetings). The crossings also make the lines’ operation more sensitive to disturbances.

Since the major part of the Swedish railway network is single-track it is of great interest to examine the relationships between operation properties, such as travel times and reliability, and infrastructure configuration on single-track lines. The crossings are the core feature of single-track operation and this thesis focuses on the crossing time, i.e. the time loss that occurs in crossing situations.

A simplified analytical model, SAMFOST, has been developed to calculate the crossing time as a function of infrastructure configuration, vehicle properties, timetable and delays for two crossing trains. The effect of possible surrounding trains is not taken into account and all kinds of congestion effects are thus excluded from evaluation. SAMFOST has been successfully validated against the simulation tool RailSys, which shows that this type of simplified model is accurate in non-congested situations.

A great advantage of disregarding congested situations is that analysis is independent of timetable assumptions. The model also explicitly shows the effect of punctuality, which is of particular importance on single-track lines where the interdependencies between trains are strengthened by the crossings.

For the same reason, the timetable is severely constrained. Nonetheless, there is often a need for changes of the timetable (crossing pattern). The thesis proposes three simple measures of timetable flexibility, all based on assigned crossing time requirements. Together, these measures can be used to evaluate how infrastructure configuration, vehicle properties, punctuality etc affect possibilities to alter the timetable.

As an example of its application, SAMFOST has been used to evaluate the effect of shorter inter-station distance, partial double-track and combined crossing and passenger stop. These measures affect the operational properties quite differently.

More crossing stations result in a minor decrease in travel time (lower mean crossing time) but significantly higher reliability (lower crossing time variance). These effects are independent of punctuality, which is a valuable property.

A partial double-track results in shorter travel times and in some cases also higher reliability. Both effects are strongly dependent on punctuality and high punctuality is needed to achieve high effects.
A combined crossing and passenger stop results in a situation similar to that of a partial double-track. In this case it is important to point out that the assignment of time supplements in the timetable should be directly correlated to punctuality in order to achieve good operation.
Preface

This licentiate thesis is based on research performed between 2004 and 2007 at the Division of Traffic and Logistics at the Royal Institute of Technology (KTH).

The thesis consists of two parts; an introductory essay and three papers that form the basis of the thesis:


The results are also presented in the research report “Effekter av partiella dubbelsektrar och fler mötesstationer på enkeletrar”, Lindfeldt 2007 (in Swedish).
Acknowledgements

Banverket (the Swedish National Rail Administration) funded this research.

I would like to thank everyone who has been involved in this project. My supervisors Professor Bo-Lennart Nelldal and Professor Lars-Göran Mattsson, thank you for supervision, encouragement, support and advice. The members of the reference group, Magnus Wahlborg, Magdalena Grimm, Hans Dahlberg, Håkan Berell from Banverket (the Swedish National Rail Administration) and Sirpa Holmroos from SJ AB have given me valuable guidance and support.

Dr Piotr Lukaszewicz from KTH Railway Technology provided inspiring comments on my thesis at the final seminar and Dr Björn Kufver from Ferroplan supported me at the conferences in Prague and London. Anders Lindfeldt MSc, my youngest brother, showed me the usefulness of Matlab and helped me in several tricky Matlab-programming situations. Carl Grauers, train driver at Inlandsgods AB, provided valuable information about train driving, signalling etc.

Thanks also to my parents who introduced the railway to me in my early years. I also want to send a special thought to my father who died during the final phase of the thesis work. His skills in mathematics and science really inspired me to study at KTH, both for my master's and my doctor's degree.

Stockholm, November 2007

Olov Lindfeldt
# Contents

1 Introduction ........................................................................................................ 1  
1.1 BACKGROUND ........................................................................................................ 1  
1.2 OBJECTIVES ........................................................................................................... 2  
1.3 DELIMITATIONS ...................................................................................................... 3  

2 Related research .................................................................................................. 5  
2.1 SINGLE-TRACKS ...................................................................................................... 5  
2.2 KNOCK-ON DELAYS ............................................................................................... 6  
2.3 RESCHEDULING AND DELAY MANAGEMENT .......................................................... 7  
2.4 CONCLUDING REMARKS ....................................................................................... 8  

3 Methodology ......................................................................................................... 9  
3.1 MODELLING SINGLE-TRACK OPERATION ............................................................... 10  

4 Results ................................................................................................................ 13  
4.1 INFLUENCES OF STATION LENGTH (PARTIAL DOUBLE-TRACK) AND INTER-STATION DISTANCE ON DELAYS AND DELAY PROPAGATION ON SINGLE-TRACK LINES WITH REGIONAL RAIL TRAFFIC ............................................................... 13  
4.2 SAMFOST – A TIMETABLE-FREE WAY OF ANALYSING SINGLE-TRACK RAILWAY LINES ........................................................................................................ 16  
4.3 CROSSING TIMES ON SINGLE-TRACK RAILWAY LINES – DEPENDENCIES OF DIFFERENT INFRASTRUCTURE AND TRAFFIC FACTORS ..................................................... 23  

5 Discussion of the main contributions of the thesis ............................................ 31  
5.1 SIMPLE MODELS ..................................................................................................... 31  
5.2 PUNCTUALITY DEPENDENCIES ............................................................................ 31  
5.3 INFRASTRUCTURE AS A VARIABLE ....................................................................... 32  
5.4 TIMETABLE FREE ANALYSES AND TIMETABLE FLEXIBILITY .................................. 32  
5.5 TOOL FOR INFRASTRUCTURE AND TIMETABLE PLANNING .................................. 33  

6 Final remarks and future work ............................................................................. 37  
7 References ............................................................................................................. 39
1 Introduction

1.1 Background

Transportation is a large and important part of the economy and the need for transportation increases continuously. Road traffic is the major form of transportation. However, a combination of investments in railway infrastructure, new, fast and comfortable rail vehicles, and increased frequency of service has led to an extensive expansion of rail passenger traffic. The Mälar valley and the Sound region are good examples of geographical areas where travel by train has increased dramatically.

Along with increasing environmental awareness and a political desire to reduce emissions, the railway has a good opportunity to increase its market share and contribute to a sustainable society. Such an increase is strongly dependent on its competitiveness and so rail services need to be fast, frequent, comfortable and reliable.

These competitive factors, in turn, depend on technical properties in the railway system:

- Infrastructure layout and operational reliability.
- Vehicle design and operational reliability.
- Timetable.

In order to achieve fast, frequent and reliable services it is necessary to understand the relationships between infrastructure, vehicle, timetable and disturbances (lack of reliability). This thesis deals with some of these relationships on single-track lines.

Regarding speed, frequency and reliability of services, single-track lines exhibit special properties, most of them tightly connected to crossings. On single-track lines, with only ordinary crossing stations, each crossing implies longer running times. The crossings also imply decreased reliability since delays propagate between crossing trains. The limited crossing possibilities also constrain capacity and thereby also the frequency of services on single-track lines.

In order to increase the competitiveness of the railway transportation system, measures that limit these effects are of special interest. Examples of such measures are:

- Transformation of single-track into complete double-track.
- Partial transformation of single-track into double-track.
- Increased number (higher density) of crossing stations.
- Crossing combined with regular passenger stop.
• Greater punctuality.

A complete transformation of a single-track line into a double-track means a complete elimination of the problems related to crossings and results in a great leap in capacity as well as a decrease in travel time and disturbance sensitivity.

Since this type of investment is extensive, it is not always justifiable to go directly from a single-track to a complete double-track. In order to decrease the leaps and achieve a better match between investments and demand it is of great interest to consider less expensive measures that still, completely or partly, result in the desired effects. This is possible in some cases, though at the cost of different kinds of restrictions on the traffic. As traffic demand develops such restrictions will become more and more troublesome and at some point the complete measure is motivated.

Partial double-tracks, a greater number of crossing stations and combined crossing/passenger stop, all affect the characteristics of the railway line so that the travel times and/or disturbance sensitivity decreases. Capacity is also affected, but this change may be difficult to utilize without losing some of the travel time gained and/or the decrease in disturbance sensitivity. In order to understand how these measures alter the railway system and under what conditions they actually work as required, it is necessary to examine the operational properties of both the original system and the adjusted system.

The railway suffers from low punctuality due to frequent primary delays. A major reduction in these delays would help to make the operation of single-track lines more reliable and contribute to a decrease in average travel times. This measure is systematically examined in the thesis through the use of different levels of primary delays.

1.2 Objectives

This thesis has several objectives. The overall objective is to clarify the operational properties of single-track railway lines, i.e. how the traffic is affected by infrastructure configuration, timetable and disturbances. This includes examination of some important factors in infrastructure and vehicles as well as punctuality and how these factors interact.

Punctuality, in the form of primary delays, is a factor of particular importance in the operation of single-track lines. A special objective is therefore to show and analyse sensitivity to disturbances that occur on single-track lines.

A more specific objective is a further examination of three measures that can be used to decrease the negative effects of crossings. The work aims to show how time for crossing, and its variance, is affected by partial double-track, shorter inter-station distances and crossing combined with passenger stop. These measures are all strongly linked to the infrastructure. Therefore, unlike in many other studies, the infrastructure has to be treated as a variable. A secondary objective that follows from this is to show how the infrastructure can be treated practically as a variable by fictive model lines that do not exist in reality.
The three measures not only affect the properties at isolated points on the studied line. Properties along the whole line, or on continuous sections of line, are changed, thereby also changing the conditions for alternative timetables. In order to take this change into account, timetable flexibility also has to be studied.

It is beyond the scope of this thesis to give a strict definition of timetable flexibility. However, one important objective is to make a first attempt to give some sort of definition of this important concept, including punctuality to emphasise its importance for timetable construction.

A methodological objective is to develop a simplified, mathematical traffic model and show an example of how such a model can be used. Simplifications require assumptions about traffic density, punctuality, dispatching rules, signalling systems, train movements etc. Given these assumptions the model helps to give important insights about system properties at a basic level. An important part of the work is to show under which conditions the model is accurate.

Using a mathematical model, fictive infrastructure configurations with a high degree of symmetries, as well as existing configurations with unique, non-symmetric properties, may be analysed. This type of comparison is also an aim, since analysing both an idealised case and a “real” case greatly increases understanding of the operational properties.

1.3 Delimitations

Only single-track railway lines, where crossings take place, are taken up. Nodes and all types of network and network effects are excluded. The analysed lines are assumed to be long (> 120 km) and without connections to other lines. Only passenger services are modelled and two different vehicle types are used: X50, a regional train with high acceleration and X2, a long distance train with lower acceleration. Both traffic directions have been operated by the same vehicle and mixed crossings are not analysed.

Capacity utilisation is considered to be moderate and the frequency of services does not exceed two trains/hour and direction.

All fictive model lines are constructed to obtain results that are easy to understand. Therefore, one standard station design is used and the inter-station distances are equal within each model line\(^1\). No gradients are modelled.

One type of signalling system is used: ERTMS, level 2. Continuous updating of driving permissions makes the trains behave in a logical and deterministic way. This delimitation eliminates the need to model ATC track antennas and their locations.

Punctuality is varied within the interval 100 – 350 seconds mean arrival (primary) delay. Better and poorer punctuality is not analysed. These limits are derived from

\(^1\) Partial double-tracks influence the flanking inter-station distances in a special way.
real delay distributions on the Swedish railway. In the analyses negative exponential distributions are principally used since they are easily adjusted for different punctuality levels.

Only situations without congestion effects are treated. This means that two following trains are always assumed to be more than approximately two inter-station distances apart, which implies that they are independent in crossing situations that appear with trains in the opposite direction. These delimitations in service frequency and punctuality make the probability of congestion very low (<1/200).

The modelled timetables contain no slack (time supplements) and so no recovery from delays is possible within the model. This delimitation is natural since the objective is to model crossings with disturbances. The need for supplements is rather a result of the analysis, since they are necessary for recovery from the delay propagation that occurs in crossing situations.

Driver behaviour is not modelled. All trains behave according to deterministic vehicle data and follow the same assumption regarding acceleration courses, use of maximum speed and deceleration courses.

Series of crossings are not modelled. However, this would be of great interest for further research. Such analyses need modelling of, or more detailed assumptions about, time supplements between the crossings.
2 Related research

The research in railway operation is extensive. Most modern research concerns double-tracked systems or entire networks of double-tracked lines and stations that connect them.

In many countries single-track lines are of minor interest since the traffic volumes correspond to double-track systems. For this reason the infrastructure related research on single-tracks is limited. Instead, most modern single-track literature is orientated towards timetable construction and reliability in existing infrastructure configurations.

Many studies deal with knock-on delays (delay propagation) and rescheduling. Although most of these studies are not applicable to single-tracks, they give a useful basic picture of the railway system, operation and possible modelling techniques.

Several types of analytical models are presented in the literature whereas simulation methods are more rarely discussed. Queuing theory and other types of statistical methods, where delay distributions are combined, are used to model interactions between delayed trains.

2.1 Single-tracks

Research on single-track systems generally concentrates on either infrastructure or timetable. This is a simplification since infrastructure and timetable are very closely interconnected. Analyses concerning just one of them therefore imply considerable assumptions as regards the other.

In [9] Petersen presents a simple model that can be used to describe the delay time as a function of the traffic intensity. The main timetable assumption, that makes the study general, is that departing times for trains are independent random variables that are uniformly distributed over the defined time period. Given this randomised timetable the time costs (delays) for crossings and overtakings are calculated. One conflict at a time is identified and resolved and so the trains are treated pair-wise. The most important advantage of models like this is that effects of changes in infrastructure parameters are easy to examine.

In [10] both timetable and line alignment are assumed to be known. Train performance is assumed to be deterministic and from this a method to find the best locations of crossing stations is presented. The track length at crossing stations is also considered and conditions for minimum crossing delay are examined. The method focuses on frequent small delays that can be managed by longer crossing stations (i.e. partial double-tracks) and time supplements. Longer delays are handled by secondary crossing stations. These may also be used for slower trains with low priority. The study concludes that single-tracks work quite well, as long as infrastructure and timetable are coherent and delays limited.

Higgins et al [6] describe a decomposition procedure that for a given cyclic timetable (day or week) for high-speed trains finds the numbers and positions of crossing
stations that minimise both the risk of delays and the delays caused by train conflicts. The timetable is specified only by information about earliest possible departure time of the trains. Each conflict is divided into two parts:

- Conflict delay: the part that is included in the timetable.
- Risk of delay: expected amount of delays due to unforeseen events.

The output from the model is both an optimal infrastructure and an optimal timetable. This combination gives a complete technical solution to the entire single-track problem. In most practical cases, however, there are other timetable-related constraints that ought to be included in the model.

In [5] Higgins et al develop their model further, resulting in a model that can be used both in dispatching and in long-term infrastructure planning. However, the model does not allow random delay events.

There are several examples of studies where the infrastructure design is fixed and the timetable is somehow constructed according to infrastructure constraints (and market demand). An early example of this is the mathematical treatment of two-way traffic on a single-track presented by Frank [4]. Using simplified models for train movements he calculates the capacity of a single-track line both for one-way traffic and for certain (fleet) systems of two-way traffic. The results are most applicable on freight or military transport systems but may also serve as a starting point for further studies.

Chen and Harker [3] present a sophisticated model for estimation of mean delays and delay variance for trains that operate on a single-track. In this model the inter-station distances are assumed to be even and the actual departure time of each train is randomised around a specified timetable time. The conflict resolution is handled through calculation of probabilities of conflict between every pair of trains. The study shows that shorter inter-station distances lead to lower mean delays and delay variances. The number of trains also influences the delays significantly.

### 2.2 Knock-on delays

Some modern research has also been carried out with an explicit focus on knock-on delays, or secondary delays as they are referred to in this thesis. Although these studies mostly concern stations and/or double-track lines, they clearly show the effects of delay propagation. Several studies in this area are timetable-independent, which makes them very useful in long-term planning and in other situations where the timetable is not known.

Carey [2] takes up different measures of reliability. He discusses the advantages and disadvantages of measures based on probabilities (i.e. observed delays) and measures not using probabilities. Almost all measures for prediction of reliability involve headways (time space between two consecutive trains) since longer headways generally reduce knock-on delays. Apparently there are several advantages of using simple measures that are not based on probabilities, although mathematical methods for more exact calculations are available.
In [8] Oetting proposes a model for calculating knock-on delays that appear in a system of serial infrastructure elements. This type of chaining is based on the fact that outgoing delay in one segment is equal to ingoing delay in the following section and that disturbances tend to spread in the opposite direction. This type of calculation technique, based on convolution, may be used for studies of infrastructure elements as well as studies of train-paths (timetables).

In [7] Huisman et al use queuing theory to analyze the dependencies and interactions between the individual components in a railway system. The operation is here defined by frequencies of service and no specific timetable is defined. Stations, junctions and sections (lines) are modelled according to their special properties. The model seems to be a good alternative to simulation and the result is, in some sense, mean values of all possible timetables that can be constructed from the frequencies that are given as input data. In the model occupation times as well as minimal headway times are assumed to follow negative exponential distributions. Therefore the calculated waiting times are slightly overestimated.

Other examples of literature concerning knock-on delays are given by Wendler [15], Carey [1] and Yuan et al [16].

2.3 Rescheduling and delay management

While the research on knock-on delays concerns small delays that occur with high frequency, the focus in rescheduling is on large delays and disturbances. In a rescheduling problem the aim is to restore the traffic to the timetable in such a way that the knock-on delays are minimised. The infrastructure is then a given constant as is the planned timetable. The literature on rescheduling is extensive and only a few examples are mentioned here.

In [13] and [14] Törnquist presents an optimisation approach to the rescheduling problem. A mathematical formula which allows an n-tracked network to be modelled is constructed. Alternative objective functions, such as total final delay and total cost associated with delays are used and four different rescheduling strategies are tested. The study shows that it is possible to find rescheduling solutions that limit the knock-on delays and/or cost associated with these delays. The most complete rescheduling strategy is sometimes too time-consuming for practical use. However, a more limited optimization strategy is often good enough.

A special field within the rescheduling research concerns connections between trains (most often passenger trains). One example is Schutter [12] who examines the possibilities to recover from delays by breaking connections. In the presented model the connections are represented by different kinds of synchronization constraints. In case of delay the so-called soft constraints may be broken, but at a cost that represents compensation activities and dissatisfaction for passengers. The method is feasible for real-time dispatching since the system uses a moving horizon in which the model is continuously updated.
2.4 Concluding remarks

Infrastructure configuration, timetable and delays are essential ingredients in railway operation. Modelling the operation of a railway system therefore requires assumptions of these parameters. Depending on these assumptions the analysis will fit into one (or more) of the boxes in the following matrix. Using this type of matrix it is possible to group the models presented in literature.

<table>
<thead>
<tr>
<th>Timetable</th>
<th>Infrastructure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant (existing)</td>
<td>Constant (existing) Reliability, rescheduling and delay management</td>
</tr>
<tr>
<td>Variable (no constraints)</td>
<td>Variable (no constraints) Do not apply</td>
</tr>
<tr>
<td>Known demand (frequencies)</td>
<td>Time table optimisation (knock-on delays) Combined infrastructure-and timetable optimisation</td>
</tr>
<tr>
<td>Variable (no constraints)</td>
<td>Examination of infrastructure potential General examination</td>
</tr>
</tbody>
</table>

When both timetable and infrastructure are constant (upper left) the main variable will be delays (punctuality) and different kinds of rescheduling and reliability models apply.

In cases where the timetable is a variable that is guided by some sort of constraints (e.g. demand requirements) we turn into the field of optimization, either of the timetable layout only, or of a combination of infrastructure and timetable. Knock-on delays play an important role in objective functions in this type of optimization.

If the timetable is completely unknown, and the aim is a general examination, rather than optimization, we are out for an investigation of infrastructure potential or a complete general examination. This situation sometimes occurs in long-term planning when both the timetable and the future development of demand are unknown.

The idea of this thesis is to provide some general knowledge about operative properties of single-track lines. Infrastructure and timetable as well as delays thus appear as variables and the examination becomes rather descriptive. To achieve this it is necessary to disregard congested situations. Congestion appears for special combinations of infrastructure, timetable and delay patterns and is, as the literature already states, of special interest. However, congestion is an undesirable and troublesome state that appears when systems are overloaded. In order to understand the effects of congestion it is necessary to start with more simple non-congested situations.

Even if the infrastructure and timetable are treated as variables they have to be limited in some way. A single-track implies several constraints on the timetable. Therefore the proposed “timetable-free” analysis is more natural on single-track systems that have inherent limitations.
3 Methodology

Railways are complex technical systems. They are often considered to be static, stiff and inflexible. As long as only constant, non-dynamic parts, such as the infrastructure, are considered, the system is quite easy to understand. Reality is different. The variance in different parameters makes the system difficult. Some examples:

- **The timetable** creates a well defined structure. However, the capacity is utilised differently every day since the actual timetable varies from one day to the other due to delays, extra trains and cancelled trains.

- **The available capacity**, which is a very important condition for the timetable, varies over time. Failures, construction work, accidents and delays all make the available capacity vary over time.

- **Vehicle properties**, relative to those assumed during timetable construction, vary. Important examples are freight trains whose train mass often differs from the timetabled train mass, change of vehicle types without corresponding change of the timetable, partial vehicle failures, weather conditions that affect adhesion etc.

- **The railway system is used and operated by humans. Human behaviour** varies naturally from one time to another. Train crews, dispatchers and passengers all contribute to this variance.

All these variances make the railway system complex and very interesting to analyse. To find general relations in this noise of superposed variances is not easy. Moreover, strong interactions between factors can be expected, that make the task of analysis even more difficult.

There is no obvious choice of method of analysis. Two alternative methods appear:

1. Analysis of a specific situation in detail.
2. General analysis given simplifying assumptions and delimitations.

The first method results in deep knowledge of a specific traffic situation. By delimiting the examination to a specific situation it is possible to use a commercial railway operation simulation tool. This type of simulation requires detailed information about infrastructure, timetable, dispatching and disturbances, which in fact implies a great many hidden assumptions.

A simulation can be compared to a spot test. Most parameters have to be set by hand since the available simulation programs for railway operation do not handle automatic parameter variation. Therefore, only a few variants may be analysed, which in practice requires some kind of simplifying assumptions!
The program codes of the available simulation programs are not open. Therefore, it is in many cases difficult to interpret the results when it is not obvious how the programs work.

The second method means that only quite simple traffic situations are analysed. Given simplifying assumptions, mathematical models become possible. These may be performed to handle several variables simultaneously.

This thesis aims to capture fundamental knowledge that is general rather than specific. For this reason it is natural to start with simple cases, seek simple relations and learn about important principles. After that it is possible to increase the complexity and perform detailed studies.

Simplifying assumptions and delimitations are therefore accepted at this basic level and a mathematical model that is adjusted to the actual issues is a natural method. Compared to simulation a more analytical model has several advantages:

- The researcher controls all assumptions and parameters. Modelling of dispatching and vehicle movements are two important examples.

- The model may be adjusted in detail so that a systematic examination of important variables is possible. Infrastructure configuration and punctuality are examples.

- The model may be constructed without specific assumptions about timetables, which means that the infrastructure can be more unbiased examined.

- The model may be adjusted for calculation of new concept variables. One important example is timetable flexibility.

- The results become easier to grasp. This is particularly true if independence assumptions are used.

A simple mathematical model is well suited for systematic analyses of a great many simple fictive cases. A central part of this project has therefore been the construction of a mathematical model that gives general results for simple operative cases.

### 3.1 Modelling single-track operation

The operation of single-track lines is well suited for analytical modelling since the traffic is well-defined by the crossing stations. The crossing situations, where trains travelling in opposite directions meet, are probably the most important part of single-track operation since they cause time losses and delay propagation. These time losses are hereafter referred to as crossing time.

In order to examine the influence of crossings a model, named SAMFOST, has been constructed. The model stands on two fundamental assumptions:
1. Two crossing trains are independent before crossing.

2. Different crossing situations are independent of each other. This implies that following trains are always far enough apart not to interfere with each other when they cross a train in the opposite direction.

These assumptions have high validity in non-congested situations, i.e. when the combination of infrastructure configuration, timetabled frequency of services and punctuality is such that following trains only occasionally come so close that independence is broken.

SAMFOST performs a stepwise analysis where each step means that a new set of parameters have to be assigned:

Step 1: The combination of infrastructure and vehicle data together with passenger stop data, give a timetable-free characteristic of the line.

Step 2: The addition of delay data and timetabled crossing point gives the full distribution of the crossing time.

![Figure 3-1 The two model steps in SAMFOST. Bold boxes are output data from the model that may be analysed further.](image)

The most important advantage of the two fundamental assumptions is that the first model step becomes independent of further assumptions about timetable and punctuality. This makes it possible to define the so-called crossing time function, which is a timetable- and punctuality-free description of the infrastructure properties. The crossing time function shows the time needed to perform a crossing as a function of the theoretical crossing point that would be realised on a double-tracked line.

In the second step the punctuality is taken into account and hereby the model clearly shows how punctuality influences the operative result. Using this stepwise modelling
a deeper understanding of the combination of infrastructure design, timetable and punctuality becomes possible.

In order to check the validity of the model a crossing time function was derived from data from the simulation of a reference case. The differences between SAMFOST and the simulated results (RailSys) turned out to be very small.

After choosing mathematical modelling as method of analysis some questions about the methodology remains, i.e. how the developed model shall be used in order to answer the specific questions that is part of the project objectives.

Unlike most previous research, the infrastructure is here treated as a variable. In order to let the infrastructure vary, the exact configuration of existing railway lines has to be abandoned. A natural way to do this is to examine fictive railway lines with extensive symmetries.

In these infrastructure designs important parameters such as station design and inter-station distances follow defined standards. The symmetries make the results easier to interpret. Within each infrastructure variant the inter-station distances are equal, all stations are identical, the same vehicle and punctuality are used for both directions etc.

To complete the examination of the three specific objective measures a factorial experiment is performed. Varying more variables, in this case six, simultaneously helps to increase the generality of the results. Moreover such an experiment reveals hidden interaction effects that connect the variables to each other.
4 Results

4.1 Influences of station length (partial double-track) and inter-station distance on delays and delay propagation on single-track lines with regional rail traffic

This paper shows the advantages of simple analytical models. The crossing time function is introduced and merely by looking at the function the characteristics of a single-track line become clear.

The crossing time is defined as the extra time needed to perform a crossing on a single-track compared to a double-track where crossings do not imply any extra time consumption. The crossing time function shows how the crossing time varies along the line with local minima at the locations of crossing stations and local maxima in-between them. Figure 4-1 shows an example.

![Crossing time function for a line with nine, equally spaced, crossing stations.](image)

Without knowing anything about the timetable the crossing time function tells us how the crossing possibilities vary along the studied line. Figure 4-1 shows the crossing time function for a completely symmetric line with 9 identical crossing stations (denoted B1, B2, F1-6 and C), equal inter-station distances and no passenger stops.

Each crossing station results in a time interval with low crossing time. This is the lowest achievable crossing time given the actual combination of infrastructure design and vehicle performance. Compared to a double-track, that has a constant zero crossing time function, the increase in travel time, that has to be paid for a crossing, is rather high.
The crossing time function can be used to estimate the time supplement that has to be added in the timetable for different locations of the timetable crossing point. However, this estimation will only be correct as long as the trains arrive with a zero delay difference, i.e. when they arrive in the planned time relation to each other.

As soon as the delay difference is non-zero the realised crossing time will take another, often higher, value than the planned crossing time. The actual crossing time will thus consist of a deterministic part that is given by infrastructure and vehicle parameters and a stochastic part that depends on the actual delay difference as well as infrastructure and vehicle factors. The latter part varies with the delay difference and may be referred to as delay propagation.

The second part of the paper deals with the effect of partial double-tracks and decreased inter-station distances, respectively. As shown in figure 4-1 some conclusions may be drawn without knowledge about the location of the timetable crossing point or the punctuality. By introducing a distribution for the delay difference the complete crossing situation for a given timetable crossing point may be illustrated and examined. Figure 4-2 shows an example.

Figure 4-2 Crossing time function for different inter-station distances and probability density function for delay difference (dashed).

Figure 4-2 shows an example of a probability density function and crossing time functions for three different inter-station distances: 15 km (bold solid), 9 km (thin dashed) and 3 km (thin solid). The combination of a crossing time function and a probability density function makes the situation clear and intuitive. As indicated, a decrease in inter-station distance results in a higher frequency and lower maximum crossing time. The minimum crossing time is not affected since all stations have exactly the same features!
Combining different inter-station distances with three levels of arrival punctuality gives the following results:

- Shorter inter-station distances have a general effect. The effect on mean crossing times is small but rather independent of the arrival punctuality. This is also true for the crossing time variance. Altogether this means that shorter inter-station distance is a suitable measure when robustness is more important than travel times.

Please note that this feature is a consequence of the assumption that all stations are identical in design. If a passenger stop is introduced at one or some of the stations this symmetry is lost and the system properties are dramatically changed. This is treated in the following papers.

The other measure that is examined in the paper is partial double-tracks (increased station length).

![Figure 4-3 Crossing time function for a line with partial double-track and probability density function for delay difference (dashed).](image)

In the example shown in figure 4-3 the crossing time function corresponds to a line with a 13 km partial double-track and four ordinary crossing stations on each side. The location of the probability density function shows the position of the timetable crossing point, i.e. the point having the highest probability density.

When an ordinary crossing station is replaced by a partial double-track, the lowest crossing time decreases. The resulting lowest level only depends on the speed restriction at the entrance and exit points. Within this time interval no signal interference between the crossing trains occurs, which is always the case in the
corresponding situation on an ordinary crossing station. An extension of the double-track also results in a wider time interval having this lowest crossing time.

The effect of a partial double-track is thus two-fold: the crossing time is locally reduced and this reduction has a greater extension compared to an ordinary crossing station. Both features are important when the line is operated and the actual delay difference varies stochastically according to some distribution.

Combining the two functions it is possible to calculate the crossing time distribution for a given timetable crossing point. For different infrastructure designs, giving different crossing time functions, and/or different punctuality levels, resulting in different density functions, it is possible to compare the effect of different parameters.

In the paper different lengths of the partial double-track are combined with three punctuality levels and the result may be summarised thus:

- Partial double-tracks have only a local effect on the crossing time function. Therefore the effect of a partial double-track is highly dependent on the arrival punctuality of the trains. Low punctuality means that the mean crossing time does not decrease as much as for higher punctuality. Even more important, however, is the fact that the crossing time variance is substantially higher when arrival punctuality is low.

- The speed restriction at entrance and exit points delimits the effect of partial double-tracks. A less restrictive speed would result in higher crossing time variance but the decrease in mean crossing time would be significant.

Some of these features change when a passenger stop is introduced on the partial double-track. This is treated in the following papers.

4.2 SAMFOST – a Timetable-free Way of Analysing Single-track Railway Lines

This paper deals with the conditions for independence that are assumed in the SAMFOST model, validation of the model against the simulation tool RailSys and some important applications such as combined crossing and passenger stop. Finally, an existing Swedish railway line is used to exemplify the usefulness of the model.

Two important independence assumptions open for a stepwise systematic approach like SAMFOST. The crossing time is determined by a great many factors such as infrastructure, vehicle, timetable crossing point, driver behaviour, delays, surrounding trains etc. In SAMFOST the infrastructure, vehicles and driver behaviour are regarded as deterministic, the delays are regarded as stochastic and surrounding trains are excluded from the model.

Disregarding trains other than the two crossing ones and assuming deterministic behaviour of other factors means that the crossing time only depends on the arrival delay difference of the two crossing trains. This makes it possible to calculate a crossing time function for each combination of infrastructure, vehicle and pattern of
passenger stops. In a case with several interacting trains a given delay difference would result in different crossing times depending on the positions and delays of surrounding trains.

It is therefore only possible to disregard congestion effects as long as situations where more than two trains interfere are rare. How often congested situations occur depends on infrastructure design, vehicle parameters and punctuality. For a given combination of infrastructure, vehicle and stopping pattern, it is therefore possible to calculate a congestion-free capacity for different levels of punctuality. In this thesis a situation where interference between following trains occurs less often than 1/200 is regarded as congestion-free. This level is somewhat conservative and so all types of congestion effects may be ignored. In practical use a higher value may be accepted.

![Figure 4-4: Congestion-free capacity for three different combinations of infrastructure, vehicle and passenger stop pattern.](image)

Figure 4-4 shows how the congestion-free capacity falls when punctuality deteriorates and that the combination of infrastructure design, vehicle parameters and passenger stop pattern is of great importance. The upper curve represents a situation with short inter-station distances, high acceleration vehicles and no passenger stop while the lower curve represents longer inter-station distances, lower acceleration vehicles and passenger stop at the timetabled crossing station. In all cases congestion is accepted in 0.5 % of the crossing situations. SAMFOST can thus be applied with high validity for all points under each curve.

This diagram of congestion-free capacity is an important result in itself since it gives an idea of the correlation between capacity and punctuality. When a line is operated under non-congestion conditions the traffic is likely to be stable and controllable.
Non-congested operation should be an aim in timetabling in order to avoid severe punctuality problems in the operation.

Given the two assumptions of independence the SAMFOST model has been validated through comparison with the simulation tool RailSys [11]. Random samples of deceleration courses, acceleration courses and combinations of both were taken and run in both models. Only small differences, less than one second per run, occurred.

The dispatching and signalling functions were also validated. Two important dispatching situations occur on single-track lines: choice of crossing station and track disposition at the chosen station. When a crossing station has been chosen and track disposition decided, signalling functions are crucial for the trains’ interaction during the crossing course.

The validation shows that SAMFOST chooses crossing station properly. Unfortunately, it was not possible to validate the track disposition since RailSys does not model this in a proper way. The actual crossing course, including the important signalling functions, was validated with good results. In a large sample of delay differences (runs), the difference in crossing time between the models never exceeded 1.5 seconds. A similar validation was performed to test the effects of a combined crossing and passenger stop, which is a more complicated dispatching situation. This validation also turned out well.

Altogether the validation shows that SAMFOST models crossing situations very similarly to RailSys, as long as the fundamental assumptions about independence hold.

Having validated the model and analyzed its delimitations it can be used for real analyses. One situation of special importance is the combination of crossing and passenger stop.

If a crossing is planned at a station where the trains have a regular passenger stop the time for deceleration and acceleration, as well as most of the waiting time, is useful time that does not become crossing time. Therefore, the combination of crossing and passenger stop implies time efficient crossings. This can be seen in the crossing time function in figure 4-5.
Most often at Swedish stations the platforms are located at one end of the station which results in a slightly asymmetric crossing time function. Referring to the crossing time function (solid line) in figure 4-5, the effects of a passenger stop may be summarised as follows.

- Time efficient crossings are made possible due to the stop. A time interval occurs, whose length depends on the timetabled dwell time. Speed restrictions at the farther end of the station imply a small crossing time.

- The time interval within which the stop station is used becomes wider than would be the case without a stop.

- The maxima surrounding the stop station become higher due to accelerations and decelerations that increase the run time on the single-track sections surrounding the stop station.

A passenger stop means that the amplitude of the crossing time function increases significantly. This means that the variance and sensitivity of punctuality also increase compared to a line without any passenger stop.

Figure 4-5 Crossing time function for a line with 9 crossing stations (solid) and passenger stop at mid-station. Different locations of timetable crossing point imply different positions of the probability density function for the delay difference (dashed).
Figure 4-6 Mean crossing time (solid) and standard deviation (dashed) as a function of inter-station distance. Three levels of punctuality.

Figure 4-6 shows that the mean crossing time (solid curves) is highly dependent on punctuality. The uppermost curve corresponds to low punctuality, \text{Exp}(350\text{s})-distributed arrivals, whereas the lowermost curve corresponds to higher punctuality, \text{Exp}(150\text{s}). This dependence is a result of the described asymmetry that implies that the timetable crossing station cannot be replaced by any other station without greatly increasing the crossing time.

Although the level of the crossing time is significantly lower compared to a case without any passenger stop, the punctuality dependence implies that it is difficult to find a feasible timetable since the amount of supplements is correlated to punctuality.

The standard deviation (dashed curves) shows only a slightly stronger dependence on punctuality compared to the case without any passenger stop. However, the level of the deviation is much higher, making it more difficult to construct a robust timetable.

The analyses described so far assume that the timetable crossing point is located symmetrically on a crossing station. Different kinds of timetable dependencies often make it impossible to obtain the optimal timetable crossing point for every crossing. It is therefore of interest to examine different locations of the timetable crossing point. This means that the probability density function is located at different locations along the crossing time function, as exemplified in figure 4-5.

The timetable crossing point is here defined as the theoretical point where the crossing would take place at arrival delay difference equal to zero if the line was double-track. Assuming high punctuality, i.e. \text{Exp}(150\text{s})-distributed arrival delays in both directions, and letting the density function shift stepwise along the crossing time function, the mean crossing time and its standard deviation for each location...
result in figure 4-7. This figure shows results for seven different inter-station distances, i.e. seven different crossing time functions.

![Figure 4-7 Mean crossing time (solid) and standard deviation (dashed) as a function of the location of timetable crossing point for different inter-station distances.](image)

Figure 4-7 clearly shows the differences between infrastructure designs (curves) as well as differences between locations (horizontal axis). Uppermost curves represent 15 km inter-station distance while lowermost curves represent 3 km inter-station distance (step length 2 km/curve). The shifting inter-station distance is reflected in different frequency and different crossing time levels.

Every minimum corresponds to a crossing station. At the middle crossing station all trains have a passenger stop which results in an absolute minimum for the mean crossing time.

Looking at the standard deviation it becomes clear that a shorter inter-station distance gives lower crossing time variance while a passenger stop results in higher crossing time variance.

This type of mean crossing time functions and standard deviation curves is also an important methodological result of this research since the method displays the impact of punctuality on crossing time and timetable construction.

SAMFOST has also been used to analyse an existing single-track railway line, the Svealand line from Södertälje to Eskilstuna southwest of Stockholm in central Sweden. This analysis shows the effect of real asymmetries with unequal inter-station distances and different punctuality for different traffic directions.
Figure 4-8 Crossing time function (solid thin). Mean (solid bold) and standard deviation (dashed) for crossing time. Two different levels of arrival punctuality are shown.

Figure 4-8 shows the characteristics of the Svealand line. The solid thin line is the crossing time function revealing four crossing stations, of which one is combined with a passenger stop (Nkv) and three are not, one partial double-track and two flanking double-tracks.

The figure also shows mean crossing time functions and standard deviation curves for two different levels of punctuality: existing punctuality and increased punctuality. The existing punctuality results in a mean function that is quite insensitive to changes in the crossing time function, which results in a rather low amplitude and small differences between alternative timetable crossing points.

If punctuality is improved the mean crossing time function follows the crossing time function much better and the amplitude is much higher. This follows the fact that the mean time function converges towards the crossing time function when punctuality increases. When punctuality is absolute and no delays occur, the two functions coincide.

Another important result of this analysis is that the timetable crossing points on double-tracks near the border point to a single-track need special attention. This is the fact to the left of Eskilstuna (Et) and to the right of Södertälje (Söö) where double-track sections help to decrease, but not completely eliminate, the crossing time. The crossing time variance in particular is high on these sections.
4.3 Crossing Times on Single-track Railway Lines – Dependencies of Different Infrastructure and Traffic Factors

In the third paper the concept of timetable flexibility is defined and exemplified. SAMFOST is then used to examine the impact of six different infrastructure and traffic factors on the crossing time at the optimal timetable crossing point as well as on the timetable flexibility.

The mean crossing time function and standard deviation curves that were derived in the second paper may be used to define timetable flexibility. When a new railway line is projected or a new timetable is to be constructed it is of great importance to assemble knowledge about potential crossing points’ features. One way of doing this is to calculate the mean crossing time function and standard deviation curve and use these to classify and compare alternative timetable crossing points.

The paper presents a first attempt to choose accepted timetable crossing points out of all possible points along the studied line. Two requirements for timetable crossing points are natural: as low a mean crossing time as possible and as low a crossing time variance as possible. In this first, and very simple, test only the mean crossing time function is used.

Figure 4-9 Mean crossing time function (solid) and standard deviation curve (dashed) for a line with five crossing stations. A tolerance line (horizontal at approx. 200 s) marks the accepted timetable crossing points located below it.

A tolerance level is defined 5 seconds above the highest located (local) minimum of the mean crossing time function. All timetable crossing points that have a lower mean crossing time than this tolerance level are regarded as accepted timetable crossing points. Other kinds of choice criteria, also including the standard deviation...
might be a good alternative, but for this first study only the simplest rule has been applied.

Figure 4.9 shows an example. The mean crossing time function (solid) originates from a line with five crossing stations. Trains stop at the mid-station which makes the mean function fall whereas the standard deviation curve rises. Following the simple rule every timetable crossing point that has a mean crossing time that is lower than the horizontal tolerance line is regarded as accepted.

The choice rule works very well as long as the mean crossing time function does not vary irregularly. From practical timetabling it is known that all crossing stations are accepted timetable crossing points. It is therefore reasonable to accept all local minima, although a more restrictive rule would only accept points near to the absolute minimum.

This reasoning gives that the simple rule of choice may be applied in sections “A” (outside outermost dashed vertical lines) and “C” in figure 4.9, whereas sections marked “B” still remain to be decided on. This could be done by also introducing the standard deviation. However, to keep this first attempt to define timetable flexibility as simple as possible, the simple rule has also been applied in “B-sections”. This will probably overestimate the effect of a passenger stop or other properties that decrease the crossing time locally.

All accepted timetable crossing points contribute to the timetable flexibility and the group of accepted points may be subject to different kinds of evaluation to describe the flexibility. When doing this analysis, it is very important to keep in mind that punctuality affects the mean crossing time function. The paper briefly discusses possibilities to make the effects of punctuality strong or weak.

In the paper three different measures for timetable flexibility are defined. All of them are quite natural when a mean crossing time function is studied:

- Share of accepted timetable crossing points.
- Spread in position (horizontal axis), i.e. position variance for accepted timetable crossing points.
- Spread in crossing time (vertical axis), i.e. crossing time variance for accepted timetable crossing points.

Together these measures tell us a great deal about the possibility to find different timetable solutions with accepted crossing properties. The first two measures are fundamental whereas the third is a sort of check for the imperfect rule of choice.

In the second part of the paper a factorial experiment is presented. A model like SAMFOST is useful for multivariate analyses since the combined effect of several variables may be examined systematically, thereby making it possible to find and examine the effect of not only single variables, but also groups of variables that cause different kinds of interaction effects.
One simple way to perform a multivariate analysis is a factorial experiment at two levels. In such an experiment a number of factors are chosen. Each factor may take one of two values, denoted high and low level. The “experiment” is then run for all combinations, \(2^n\) in total, and for each run the values of the response variables of interest are saved. By evaluating the saved values, conclusions may be drawn about both main and interaction effects.

Six factors were chosen, forming a \(2^6\)-experiment, and levels were assigned as follows.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Description</th>
<th>Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Track length at station / on double-track where timetable crossing point is located</td>
<td>Low (–): 0.75 km, High (+): 10 km</td>
</tr>
<tr>
<td>B</td>
<td>Inter-station distance</td>
<td>Low (–): 15 km, High (+): 7.5 km</td>
</tr>
<tr>
<td>C</td>
<td>Passenger stop</td>
<td>Low (–): No stop, High (+): 60 s</td>
</tr>
<tr>
<td>D</td>
<td>Speed restriction at points at station where timetable crossing point is located</td>
<td>Low (–): 100 km/h, High (+): 160 km/h</td>
</tr>
<tr>
<td>E</td>
<td>Vehicle type</td>
<td>Low (–): X2, High (+): X50</td>
</tr>
<tr>
<td>F</td>
<td>Arrival punctuality (mean arrival delay)</td>
<td>Low (–): 200 s, High (+): 100 s</td>
</tr>
</tbody>
</table>

Four factors, A, B, C and F have been examined earlier in different ways. To widen the examination two more factors were added. According to previous results speed restriction at end points on partial double-track (D) might be a variable of interest. To test the effects of lower acceleration the vehicle type (E) was also varied.

The levels were chosen to be reasonable and then the experiment was run with respect to five different response variables:

- **Centred timetable**: mean and standard deviation for crossing time.

- **Timetable flexibility**: simple flexibility, position spread and crossing time spread for accepted timetable crossing points.

After 64 runs the six main effects and 58 interaction effects were estimated for each response variable.

The results for a centred timetable, i.e. timetable crossing point fixed to the mid-station (or partial double-track), are shown in figure 4-10 and 4-11. Each main effect is here defined as the difference between the mean value over experiments with the factor on its high level (32 experiments) and the mean value over experiments with the factor on its low level (32 experiments).

In a similar way each two-way interaction effect is calculated as a difference between the effect of changing the first factor from low to high level when the second factor
is low and the effect of changing the first factor from low to high level when the second factor is high.

In the figures all effects are shown by absolute value and each bar represents the effect of a change from low to high level. The sign above each bar indicates whether the effect decreases (-) or increases (+) the response variable.

Figure 4-10 Main (dark bars) and interaction (light bars) effects on mean crossing time.

Figure 4-11 Main (dark bars) and interaction (light bars) effects on standard deviation of crossing time.

Figure 4-10 shows that passenger stop (C) and partial double-track (A) are the most important factors for the mean crossing time. However, a strong interaction between these factors (AC) tells us that the effects are not additive. The lowest main effect is that of inter-station distance (B) that affects the mean crossing time least.

An important interaction occurs between partial double-track (A) and speed restriction at points (D). In this case the AD-interaction is almost as high as the D-effect itself. This is explained by the fact that a high speed at the points is only useful when combined with a partial double-track. With an ordinary crossing station, having a track length of only 750 m, higher speed than 100 km/h gives only a minor effect.
The results from these factorial experiments may also be presented in tables including the signs of each effect. The following tables, omitted in the paper but shown here in order to clarify the results, show the six main effects and the 15 two-way interaction effects. Please note that all effects are related to the overall mean value presented in the first column denoted “µ”. This mean value corresponds to a mean level of all factors. Due to linearity assumptions of the model all tabled effects take half the value compared to figures 4-9 and 4-10 showing the entire difference between low and high level.

<table>
<thead>
<tr>
<th></th>
<th>μ</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean [s]</strong></td>
<td>103</td>
<td>-30</td>
<td>-7</td>
<td>-44</td>
<td>-11</td>
<td>-11</td>
<td>-14</td>
</tr>
<tr>
<td><strong>Std [s]</strong></td>
<td>53</td>
<td>-11</td>
<td>8</td>
<td>5</td>
<td>-3</td>
<td>-11</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>AB</th>
<th>AC</th>
<th>AD</th>
<th>AE</th>
<th>AF</th>
<th>BC</th>
<th>BD</th>
<th>BE</th>
<th>BF</th>
<th>CD</th>
<th>CE</th>
<th>CF</th>
<th>DE</th>
<th>DF</th>
<th>EF</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean [s]</strong></td>
<td>5</td>
<td>19</td>
<td>-10</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>8</td>
<td>-4</td>
<td>1</td>
<td>-2</td>
<td>1</td>
</tr>
<tr>
<td><strong>Std [s]</strong></td>
<td>4</td>
<td>-14</td>
<td>5</td>
<td>-1</td>
<td>-3</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>-2</td>
<td>-1</td>
<td>-4</td>
<td>0</td>
<td>-1</td>
<td>1</td>
</tr>
</tbody>
</table>

Knowing the level of each factor the expected mean crossing time can be calculated as a sum according to the following equation. Please note that the sign of each factor must be included. Each parenthesis shall therefore be replaced by either +1 or -1.

\[
\theta_{\text{hine}} = \mu(\pm)A(\pm)B(\pm)C(\pm)D(\pm)E(\pm)F
\]

\[(\pm)(\pm)AB(\pm)(\pm)AC(\pm)(\pm)AD(...(\pm)(\pm)EF\]

The indices \(i, j, k, l, m, n\) denote the level signs of the six factors, i.e. + or -, \(\mu\) denotes the overall mean value taken over all 64 runs and capital letters A-F, AB-EF denote estimations of effects that can be read in the tables above. In order to emphasise that all values are estimations, each effect letter is marked with a horizontal line on the top. The following example shows how the formula works.

Assume that we have a line with a partial double-track, an inter-station distance of 15 km, a passenger stop, 160 km/h at the points of the partial double-track, X2-vehicles and punctuality at 200 s. This means that the signs will be + - + - - -. If only main and two-way interaction effects that have an absolute effect ≥ 5 seconds are considered, the resulting mean crossing time can be estimated as:

\[
\theta_{\text{hine}} = \mu + A- B+ C+ D- E- F- AB+ AC+ AD- CE = \\
= 103 - 30 + 7 - 44 - 11 + 11 + 14 - 5 + 19 - 10 - 8 = 46 \text{ s}
\]

This example clearly shows how a partial double-track and a passenger stop decrease the mean crossing time, whereas a longer inter-station distance, a vehicle with weak acceleration, and low punctuality increase the mean crossing time. Also note how the strong AC-interaction reduces the effect, “increasing” the crossing time by 19 seconds since A- and C effects are not additive.
If the standard deviation is analysed with the same method a somewhat different result appears, see figure 4-11. First, the inter-station distance (B) turns out to be the most important single factor which means that a short inter-station distance results in a low crossing time variance. However, the combination of passenger stop (C) and partial double-track (A) is very important for the variance. When a passenger stop is introduced at an ordinary crossing station it results in a substantial increase in crossing time variance. This increase can be eliminated if the passenger stop is combined with a partial double-track. This fact is reflected in a strong interaction between these factors.

Figures 4-12 and 4-13 show how the six factors affect timetable flexibility. All effects are shown by absolute value and each bar represents the effect of a change from low to high level. The sign above each bar indicates whether the effect decreases (-) or increases (+) the response variable.

Figure 4-12 Main (dark bars) and interaction (light bars) effects on simple flexibility.

Figure 4-13 Main (dark bars) and interaction (light bars) effects on position spread of accepted timetable crossing points.
The inter-station distance (B) is the factor that has the greatest impact on the simple flexibility, i.e. the share of the line that has accepted mean crossing time. This is natural since a decrease in inter-station distance is a general measure that works along the whole line and alters the mean crossing time function entirely.

Also, a partial double-track increases the simple flexibility, but in this case the new timetable crossing points make a time interval along the double-track. Please observe the strong interaction effect AB that means that these effects are not additive.

Looking at the second ratio of timetable flexibility, it is clear that a shorter inter-station distance increases the spread of timetable crossing points whereas the effect of a partial double-track is very low.

It is remarkable that punctuality in fact greatly affects timetable flexibility. Given the definitions of the two flexibility ratios a higher punctuality actually means less timetable flexibility, which means that it becomes more difficult to find a feasible timetable when punctuality is high.

Also the flexibility ratios may be calculated with the sign formula presented above using the tabled results below.

<table>
<thead>
<tr>
<th></th>
<th>µ</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple flex.</td>
<td>79%</td>
<td>11%</td>
<td>17%</td>
<td>8%</td>
<td>0%</td>
<td>-3%</td>
<td>-6%</td>
</tr>
<tr>
<td>Position std</td>
<td>477</td>
<td>-5</td>
<td>38</td>
<td>12</td>
<td>-1</td>
<td>-8</td>
<td>-13</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AB</th>
<th>AC</th>
<th>AD</th>
<th>AE</th>
<th>AF</th>
<th>BC</th>
<th>BD</th>
<th>BE</th>
<th>BF</th>
<th>CD</th>
<th>CE</th>
<th>CF</th>
<th>DE</th>
<th>DF</th>
<th>EF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple flex.</td>
<td>-9%</td>
<td>-3%</td>
<td>0%</td>
<td>1%</td>
<td>2%</td>
<td>-6%</td>
<td>0%</td>
<td>1%</td>
<td>0%</td>
<td>-2%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Position std</td>
<td>2</td>
<td>13</td>
<td>1</td>
<td>0</td>
<td>-4</td>
<td>-18</td>
<td>0</td>
<td>3</td>
<td>6</td>
<td>1</td>
<td>-7</td>
<td>-6</td>
<td>0</td>
<td>-1</td>
</tr>
</tbody>
</table>

The factorial experiment shows that several combinations must be examined if some sort of “general” view is to be obtained. For many combinations the results from the experiment cases confirm previous conclusions and results. One of the clearest conclusions seems to be that shorter inter-station distances are a simple measure that acts on the crossing time variance and timetable flexibility, but not on the mean crossing time. A partial double-track, on the other hand, is a much more complex measure (several interaction effects indicate this) whose effects are more difficult to explain. Also, a passenger stop turned out to be a complex measure with a strong negative effect on the increase in crossing time variance. Irrespective of this, the important decrease in mean crossing time is a valuable property, in particular when punctuality is high.
5 Discussion of the main contributions of the thesis

The contributions of this thesis may be divided into five areas. A simplified method of analysis has been introduced that gives clear, well-structured results, the importance of punctuality is emphasised, and the infrastructure is treated as a variable which makes it possible to gain knowledge about alternative configurations. All analyses are, in principle, free from timetable assumptions and a concept of timetable flexibility is presented. The proposed model can be useful in infrastructure planning as well as timetable construction.

5.1 Simple models

Despite a well-defined infrastructure configuration and timetable, railway operation is complex. Different kinds of variances contribute to this complexity. Most of the methods of analysis at hand require assumptions about these variances.

An important contribution of this work is that a few fundamental assumptions open for simple models. These models are simple to use and the results are understandable and easy to display. By using these models the results become more general and systematic.

Its ease of use, together with the systematic analysis the model allows, is an important advantage in infrastructure planning where the results may be used to design more detailed studies with other techniques. In timetabling the simple models may contribute when standards and rules-of-thumb are to be decided.

It could be argued that the fundamental assumptions make the model too rough. This is probably true in cases where good predictions of future demand and operation exist. However, this is in fact seldom the case. Instead, reality shows that assumptions used to make predictions during the construction of railway systems often differ from the actual traffic that is operated when the line is in use. It is often of great importance to know somewhat more about the system properties.

This fact indicates that a more general knowledge of the operation conditions is valuable. By using less detailed overview perspective it is possible to study how the system would react if operation conditions change.

5.2 Punctuality dependencies

Almost the whole Swedish railway network suffers from low punctuality. On several lines the punctuality problems delimit the capacity since the number of secondary delays accepted is limited. Still, punctuality is usually used only as a follow-up measure. Although the secondary delays predominate, punctuality is seldom systematically used as feed-back in timetable construction. Even less often is the infrastructure design evaluated with respect to actual punctuality.

In this work punctuality is one of the most important parameters. An important contribution is that the impact of punctuality is shown explicitly. Together, the crossing time function and the mean crossing time function clearly show that
punctuality matters. The crossing time function shows the situation with full punctuality where the effect of infrastructure configuration is clearly seen, while the mean crossing time function, for a low punctuality case, shows a situation that is apparently non-sensitive to the infrastructure configuration.

So the effects of infrastructure measures are in fact highly dependent on punctuality. This again shows the need for a systematic variation of important parameters when railway operation is analysed.

The work also shows examples of the impact of punctuality on capacity. The calculation of congestion-free capacity, where punctuality is combined with infrastructure parameters, clearly states that capacity is extremely dependent on punctuality, as long as the occurrence of congested situations has to be limited. For the Swedish railway this is an important contribution since lack of systematic feedback and knowledge of punctuality-capacity relations mean that several lines are over-utilised compared to the required punctuality levels.

5.3 Infrastructure as a variable

In most other studies the infrastructure is treated as a constant, or a variable that can take on only a few predefined values. This thesis shows examples of how the infrastructure, in a very systematic way, can be treated as a variable. This means that fictive model lines are examined, instead of existing infrastructure designs that are highly asymmetric and unique.

This is an important contribution in itself, since almost all other studies focus on a given infrastructure. In order to learn more about railway operation one has to allow the infrastructure to vary within a rather wide area. In some sense this fact means that a new dimension of the analysis is accepted.

One result of this is that not only one length of partial double-track or inter-station distance is examined. It is therefore possible to draw conclusions about marginal effects of double-track extensions and station distance shortenings. It is also possible to make a clear definition of partial double-track, display patterns that result from station properties that are repeated at every station etc.

The general knowledge that is gained by analysing symmetric cases is very useful when existing lines are to be improved. Also in this case it is useful to treat the infrastructure as a variable that is changed systematically starting from the existing configuration. Results from symmetric cases here tell us what kind of improvements that are of interest to examine.

5.4 Timetable free analyses and timetable flexibility

One of the most important contributions of this thesis is the idea of timetable-free analyses. In many cases railway operation analyses are based on some kind of timetable assumption that implies that only one, or a few, operation variants (timetables) are tested.
However, the timetable develops much faster than the infrastructure and sooner or later the system is operated differently than it was originally planned for. There are several Swedish examples of this and so timetable-free methods would be an important complement to existing methods of operation analysis.

By evaluating several hypothetic timetables the analysis becomes much more general and the knowledge of the system deeper. The train separation that occurs on single-track lines makes it possible to perform this kind of analyses.

A natural development of the timetable-free concept is the evaluation of timetable flexibility. This idea ought to be an important contribution, in particular when asymmetries from a real line are included in the analysis. It is useful to analyse timetable flexibility both during timetable construction and in infrastructure planning. In the latter case it is a way to check sensitivity to timetable changes relative the dimensioning “planning timetable”.

The main contribution regarding timetable flexibility is the idea and conceptual thoughts rather than the proposed definitions that need to be developed further. Some kind of flexibility measure is probably required when infrastructure measures, such as partial double-tracks and decreased inter-station distances are to be completely analysed.

5.5 Tool for infrastructure and timetable planning

SAMFOST is a valuable tool in infrastructure planning as well as timetable construction. This is exemplified by the evaluation of different inter-station distances and lengths of partial double-track. A fundamental result from these evaluations is that the operation of line sections with only ordinary crossing stations and no passenger stops are insensitive to punctuality. It is, however, difficult to achieve large decreases in crossing time by introducing only crossing stations. Instead, this measure mainly affects the crossing time variance. Shortened inter-station distances therefore mean increased reliability rather than increased time efficiency in crossings.

Partial double-tracks, on the other hand, result in shorter crossing times. This, however, is achieved at the cost of increased sensitivity to punctuality. Partial double-track is therefore an appropriate measure in some situations but not generally. The same holds for a combined crossing and passenger stop. In this case it is important that the time supplement added in the timetable is adjusted to the level of punctuality.

These results are important to the Swedish Rail Administration, who are planning for new single-track lines and upgrades of existing lines. The clear results of partial double-tracks and passenger stops tied to crossings also confirm the experience of sensitivity to punctuality that is characteristic for such lines.

The following figures show a real example where SAMFOST has already been used to evaluate alternative improvements of a Swedish line, the Svealand line southwest of Stockholm.
Figure 5-1 shows the characteristics of the existing Svealand line. The line is constructed for 60-minute traffic which implies a crossing every 30 minutes. In order to make 30-minute traffic with time efficient crossings, the infrastructure has to be improved.

Two different improvement strategies are shown in the following figures. In figure 5-2 the existing partial double-track is extended and a new partial double-track is introduced 15 minutes run time (one direction) away from the first one. This strategy maintains the main features of the line with a highly varying crossing time function.

An alternative improvement is shown in figure 5-3. In this case the existing partial double-track is extended to the right so that it reaches the existing double-track system. On the rest of the line three new crossing stations are introduced. This results in a less fluctuating crossing time function.
Figure 5-2 Improvement strategy: New and extended partial double-track. Crossing time function (solid thin), mean crossing time (solid bold) and standard deviation for crossing time (dashed).

Figure 5-3 Improvement strategy: New crossing stations and “eliminated” partial double-track. Crossing time function (solid thin), mean crossing time (solid bold) and standard deviation for crossing time (dashed).
6 Final remarks and future work

This thesis deals with several fundamental properties of single-track railways. The SAMFOST model is a first step to describe these properties. Having examined only simple operational cases it is natural to go on and model also more complex situations.

An obvious case of special interest is the effect of series of crossings that interfere. To model such cases it is natural to consider time supplements and their effects on operation. Not only the amount of supplement but also the distribution along the line is important.

No internal sources of disturbance are modelled in SAMFOST, i.e. the trains are only given initial delays. One way to make the model more realistic is to include delays that occur during the modelled train courses such as dwell time extensions etc. It is also possible to include driver behaviour in the model, which would make the trains behave more stochastically. This factor may also represent shifting adhesion situations etc.

Another situation, which is closely related to series of crossings, is operation during congestion. When more than two trains affect a crossing a second order of crossing time, which is caused indirectly by the position of other trains, is added. Analysing such situations would probably give valuable knowledge about capacity of single-track lines.

This thesis concerns only homogenous passenger traffic. In reality, single-tracks are most often operated with mixed traffic. A natural continuation would therefore be to examine cases with mixed traffic, freight trains etc.

All these extensions mean that more variables are included. This in turn indicates that a more systematic, multivariate analysis may be useful. Therefore, an important part of the future work is to develop methods to examine and describe the effect of several variables.

Another possible continuation is a further development of the timetable flexibility concept. This could be done either on the basis of the existing SAMFOST model, or on a further developed model that includes the features discussed above. Further knowledge of timetable flexibility may be very valuable in the work of timetable construction as well as in the infrastructure planning.

The work has explicitly shown the importance of punctuality. Since punctuality is a great problem in railway operation, deeper studies of how punctuality affects timetable flexibility, capacity and infrastructure design are valuable in the whole railway sector.

Two principles of infrastructure design are examined in this thesis. The analysis shows that there are several other infrastructure strategies that could be of interest. Therefore, a natural continuation of the work is to examine more infrastructure configurations. One example of interest might be shorter inter-station distances locally around stations with passenger stops. Such a configuration would help to
reduce the mean crossing time as well as the standard deviation for crossings planned to the passenger stop station.

The infrastructure measures examined in this thesis affect the operation properties quite differently. Construction costs also differ. Therefore it is of interest to compare different measures. A complete comparison requires evaluation of public economy. This includes traffic factors such as travel times, frequency of service, punctuality as well as infrastructure factors such as construction costs and operation costs. This type of comparison is a natural continuation of the evaluation presented in this thesis.

All these possible fields of further studies aim to increase knowledge of a complex technical system that is surprisingly difficult to model. Nevertheless, more knowledge of railway operation is very important if the railway is to contribute to the development of society.
7 References


**Paper I**

Olov Lindfeldt

“Influences of station length and inter-station distance on delays and delay propagation on single-track lines with regional rail traffic”

Influences of station length and inter-station distance on delays and delay propagation on single-track lines with regional rail traffic

O. Lindfeldt
Division of Traffic and Logistics, Department of Transport and Economics, Royal Institute of Technology, Stockholm, Sweden

Abstract
Train services on single-track lines suffer from time losses due to crossings, imposed by the bidirectional traffic. The time losses, in this paper denoted crossing time, are caused by constraints in the infrastructure and delay propagation, which give a stochastic contribution that varies from one crossing situation to another.

Two examples of infrastructure improvements that decrease the crossing time are examined: partial double-track at the location of timetabled crossing and decreased inter-station distances. A mathematical model is used to evaluate these improvements.

Partial double-tracks seem to be very efficient when traffic intensity and delay variances are moderate. Shortened inter-station distances give less effect but are less sensitive to delay variance and give valuable additional line capacity.

The used model assumes independence between crossing trains, which imposes a moderate capacity utilisation. In more congested situations simulation methods are needed to make more complex crossing patterns possible.

Keywords: single-track, delay propagation, partial double-track, inter-station distance
1 Introduction

On single-track lines train services often have lower average speed and less punctuality than on double-track lines. This is a result of the bidirectional traffic that calls for crossings. These crossings take place at crossing stations with two or more parallel tracks. In many cases the crossing itself leads to time losses, compared to a crossing on a double-track line. The crossing time consists of three parts:

- A static lowest time that is given by the infrastructure. This is a minimal (theoretical) crossing time that occurs also when the crossing trains arrive at the station in an optimal way.
- A planned time that is due to infrastructure imperfections, such as unequal inter-station distances, irregular traffic or other types of planned time supplements.
- A stochastic time due to arrival delays, varying from one crossing situation to the other.

These three factors are not additive. High values on the planned part of the crossing time can overtake the static part and make the local infrastructure less constraining. A planned crossing time also decreases the impact of arrival delays. One important case is when a crossing is combined with an ordinary stop for passenger exchange. In such situations influences of infrastructure constraints are often minimised and the impact of arrival delays lessened.

Assuming perfect conditions in surrounding infrastructure, no time supplements and no stops for passenger exchange, the lowest crossing time and the stochastic contribution can be analysed systematically.

Regional train services have some important demand features that are hard to combine with constraints imposed by single-track lines:

- Short travel times. Regional train services must reach a high average speed.
- High frequency. At least in the morning and afternoon rush hour a regular and frequent service is required.
- High punctuality. People travelling to and from their places of work demand high punctuality.

Trying to meet these requirements, crossing restrictions are a great problem. This paper describes two examples of improvements that decrease crossing time: partial double-tracks and shorter inter-station distances. Other measurements, such as higher speed in points etc, will be left to future research.
2 Method

2.1 Model

In order to analyse crossing phenomena on single-track lines, a mathematical model has been constructed. In the model it is possible to vary several parameters, hence different kinds of analyses can be performed. The most important variables are: inter-station distances, track lengths for each station, speed restrictions at points, interlocking time for each station, vehicle retardation and acceleration data and distributions of arrival delays (primary delays).

One advantage of this model is that one or more variables can be varied systematically. As will be shown later, characteristics of different infrastructure designs can be evaluated very easily.

In order to make the model as simple and transparent as possible, some idealising assumptions were made:

- Signalling system: ERTMS level 2, e.g. continuous update of driving permissions.
- Crossing trains are independent before crossing. This assumption requires a moderate capacity utilisation. At higher levels, interferences between trains will cause further restrictions on crossing. The model serves to give a first idea about characteristics of single-track lines. A very interesting analysis of highly congested cases could be performed by means of a simulation at a later stage, in order to find the limitations of this simple model.
- Everything that happens outside the model is taken into account by distributions of arrival delays. Each train only crosses one other train in the modelled line.
- No time supplements in timetable. One output of the analysis is the need for (and effect) of time supplements.
- Dispatching follows the very simple rule of minimising the sum of added delay for crossing trains. This assumption also requires a moderate capacity utilisation.
- No gradients.

Given the infrastructure model, the vehicle model and the idealising assumptions delay characteristics for a line can be calculated. Fig. 2-1 shows the crossing time function, i.e. the crossing time as a function of difference in arrival delay.
Figure 2-1 Crossing time function showing the crossing time for different values of arrival delay difference. Each crossing station is used during a time interval between two adjacent maxima (vertical dotted lines). Station signatures B2-B1 shown. C denotes timetabled crossing station, F flanking stations and B border stations.

For every five seconds of delay difference the model has solved a crossing conflict. This is done by first detecting the conflict and alternative crossing solutions. Second, a crossing course, involving train movements, is performed for each station alternative. Finally, the station giving the lowest crossing time is chosen to give the crossing time value for the current delay difference.

Note that the reference level, crossing time $= 0$, represents a situation where both trains run on the line without any time losses due to crossing, which corresponds to a double-track line. In fig. 2-2 the tested line is shown schematically.

The length of the model, 120+20 km, is chosen statistically so that at least 99 % of all crossings will take place within the model, given that both distributions of arrival delays follow a specific worst-case distribution with very high variance, taken from the Swedish railway’s delay data.

In solving crossing conflicts, the model has to choose the best station for each conflict. Therefore, the border stations, B1 and B2 in fig. 2-2, must be treated carefully, since the model does not know anything about stations beyond these.

So far no attention has been paid to the fact that the difference in arrival delay follows a distribution. The outcome of different differences of arrival delays will thus be given different weights according to a distribution. This distribution can be calculated through convolution of the two crossing trains’ arrival delay distributions:
\[ f_Z(z) = \int f_X(\tau) f_Y(\tau - \tau) d\tau. \] 

(1)

In eqn (1) \( \tau \) denotes arrival delay for the down train, \( t \) denotes arrival delay for the up train and \( z \) denotes difference in arrival delay. \( f_X \) and \( f_Y \) are the corresponding probability density functions and \( f_Z(z) \) the resulting probability density function. Eqn (1) requires that the delays of two crossing trains are independent.

As a first step, negative exponential distributions have been used. Real delay statistics show that expected values on single-track lines in Sweden are 150-350 seconds. A negative exponential distribution, given the same expected value, shows a lower variance than empirical distributions.

![Figure 2-2 Example of line model with uniform inter-station distance 15 km and track length 750 m (Swedish standard).](image1)

![Figure 2-3 Three levels of probability density functions for difference in arrival delay. Standard deviation 210, 350 and 475 s.](image2)

Two equal negative exponential distributions \( (f_X = f_Y) \) give a Laplace distribution when convoluted, eqn (2). In fig. 2-3 this distribution is shown for three values on the scale parameter \( \lambda \). Corresponding expected values for \( f_X \) and \( f_Y \) were 150, 250 and
350 s respectively. Note that the same distribution has been used for down- and up trains, making the curves symmetrical.

\[ f_z(z) = \frac{\lambda}{2} e^{-\lambda|z|}. \]  

(2)

A very interesting question is how sensitive the outcome of different infrastructure improvements is to the shape of this distribution. Some measures will turn out to be quite dependent, while others will not.

### 2.2 Choice of variables

Several improvements are possible in order to obtain less time consuming crossings and lower disturbance sensitivity. Two interesting alternatives are:

- **Partial double-track** located where the crossings are planned to take place. When station tracks are longer than a certain length, crossings without any interference between trains are possible and the station becomes a partial double-track. To make this kind of improvement cost-efficient, some kind of regularity with a basic interval timetable is required so that utilisation is high. This timetable also has to be stable for several years. A severe drawback for a partial double-tracks is that it is a local improvement whose performance differs from surrounding ordinary stations. Choosing another timetable, whose crossing pattern differs from the optimal one, results in lower infrastructure utilisation.

- **Decreased inter-station distances.** If inter-station distances are decreased the maximum crossing time will be lower, but the minimum crossing time will not change. An important advantage of decreased inter-station distances is that they increase the overall capacity of the line. This means that the investment can be used either to run more trains, maintaining existing punctuality, or to run the same number of trains as before the improvement but with higher punctuality.

To evaluate the effect of a partial double-track located at the timetabled crossing, the track length of station C in figs. 2-1 and 2-2 was increased stepwise by 500 m from 750 m to 28,250 m. Inter-station distances F2-C and C-F1 thus decreased by 250 m each time. For moderate increases (≤ 15 km) this method seems to be of practical interest. The shortened inter-station distances on each side of station C are then part of the total improvement. For longer extensions the question is whether to adjust flanking inter-station distances for reasons of capacity or economy.

The other variable of special interest, inter-station distance, requires a different approach. In this case all stations, C included, were of same length: 750 m. An interval of interesting inter-station distances was defined as 3-15 km. Between these limits calculations were performed with a step of 500 m. As already mentioned, a minimum model length of 120+20 km was set up using delay statistics. Given this minimum, the number of modelled stations increased with decreasing inter-station
distances. Moreover, the number of stations had to be odd (symmetry reasons) and so this theoretical number, 120/inter-station distance, must be rounded upwards to the nearest odd integer.

These adjustments result in different total model lengths for different inter-station distances, all of them \( \geq 140 \) km. The varying model length could make comparisons more difficult, but since the probability of crossing outside the middle-120 km model part is negligible, this will not cause any real problems.

3 Results

3.1 Increased track length of timetabled crossing station

Fig. 3-1 shows how the crossing time depends on the difference in arrival delay. The upper, solid, curve represents a reference case with 15 km inter-station distances and 750 m track length for all stations (shown in fig. 2-2). Note that each station, located between two adjacent maxima, has a centred interval with low crossing time of about 140 s. At these intervals both trains arrive at the station so simultaneously that neither has to brake all the way to stop. On the slopes up towards a maximum, where the crossing is moved to the adjacent station, one train has to stop and wait, which gives linearity. On the slopes, the stochastic part of the crossing time is predominant, whereas in the minimum areas infrastructure constraints such as speed restrictions in points and signal interference, due to short station tracks, are the foremost causes of crossing time.

If track length increases, the curve for crossing time will shift in the time interval covered by station C. This is shown in fig. 3-1. Equidistance between two thin solid curves is 4 km, so that the upper one represents a track length of 4,750 m and the lower 24,750 m. Due to all symmetries in the model parameters the result is quite easy to understand:

- Maxima surrounding the partial double-track are lower. This is a result of shorter inter-station distances on each side. Note that thin solid curves coincide with the reference curve out to the left, and right, of intersection at the new maxima. This means that an extension only affects a specific (local) time interval.
- Curves corresponding to partial double-tracks get a new, lower and wider minimum (90 s). This minimum crossing time depends only on infrastructure constraints, e.g. speed restrictions at entrance and exit points for one of the crossing trains. Within a time interval crossing can take place without signal interference between the trains. The length of this interval is equal to the width of the minimum area, which depends on the track length.
Fig. 3-1 also shows the probability density function for difference in arrival delay (standard deviation 210 s), dashed line reaching 0-level at ± 1 000 s. Note that a track length of 16 750 m covers a large part of the area below the density curve (between vertical lines). The effect of an extension is thus double. First, the minimum level becomes lower, leaving only static infrastructure factors. Second, the interval within which the minimum value occurs becomes wider. A larger proportion of crossings get the minimum crossing time. These facts result in a substantial decrease in average crossing for the first kilometres of extension.

### 3.2 Decreased inter-station distances

The result of different inter-station distances is shown in fig. 3-2. The upper, solid curve still represents a reference case where all inter-station distances are set to 15 km and track length to 750 m. A reduction to 9 km inter-station distances gives a second curve, dotted; with maximum crossing time about 225 s. A further reduction to 3 km gives a curve that oscillates around 140 s, which is difficult to distinguish in the figure.

The effect of decreasing inter-station distances is lower maxima and an increasing share that lies at minimum levels (140 s in this case). In fact, the decrease in maxima is linear from 15 km to about 5 km. In this interval, stochastic delay propagation plays an important role in building up the crossing time, although the lowest crossing time level, depending only on infrastructure and vehicle parameters, is important in situations where trains arrive simultaneously to a station. The stochastic delay propagation is proportional to inter-station distance, which will be shown in fig. 3-4.
At shorter distances, 3-4 km, the crossing time part that is due to infrastructure constraints dominates and covers stochastic effects. Stochastic effects can thus not be seen in fig. 3-2 for inter-station distance 3 km.

![Crossing time functions](image)

**Figure 3-2** Crossing time functions. Reference case with inter-station distance 15 km (bold solid). Inter-station distance 9 km (thin dashed), 3 km (thin solid) and probability density function for arrival delay difference (dashed curve).

### 3.3 Average crossing time and standard deviation

The average crossing time delay is given by eqn 3:

\[
\bar{t}_c = \int_{-\infty}^{\infty} f_z(z) * g(z) dz.
\]  

(3)

Here \( z \) denotes difference in arrival delay, \( f_z \) is the probability density function and \( g \) is the crossing time function shown in figs. 3-1 and 3-2. Curves for average crossing time are shown in fig. 3-3 and 3-4. The results of three different \( f_z \)-functions, standard deviation 210, 350 and 475 s, are shown.
Figure 3-3 Average crossing time (solid) and crossing time standard deviations (dashed) for partial double-track. Results from three different distributions (fig. 2-3) for difference in arrival delay are shown.

Figure 3-4 Average crossing time (solid) and standard deviations (dashed) for partial double-track. Results from three different distributions (fig. 2-3) for difference in arrival delay are shown.
In fig. 3-3 and 3-4 solid curves represent average crossing time and dashed curves standard deviations for the crossing time. In fig. 3-3, showing results for partial double-track, the average crossing time falls from a maximal level of about 200 s, consisting of the minimum crossing time and substantial stochastic delay propagation, towards the lowest crossing time (here 90 s) where delay propagation is almost negligible. Note the effect of lower arrival punctuality that makes these curves start at a higher level and fall more slowly.

For the inter-station distance case the minimal crossing time depends on the fact that these “short” stations always cause crossing time due to signal interference between crossing trains. These curves (fig. 3-4) therefore fall from a maximal level towards this static lower limit determined by infrastructure performance. Note that a change in arrival punctuality influences neither the average crossing time nor the standard deviation significantly.

### 4 Conclusions

Presented results show some principal differences between the two alternative improvements.

Partial double-tracks are most efficient when variance in arrival delay difference is low. A lower standard deviation of arrival delay difference than 210 s, lowest curves in fig 6, and thereby higher precision, would make this improvement even more efficient. An even greater effect could be obtained with higher speeds at entrance and exit points on the partial double-track, resulting in a lower minimum level of crossing time.

Shorter inter-station distances, in particular below 10 km, seem to give a result that is more independent of arrival punctuality. Both these results provide a moderate capacity utilisation, so that alternative crossing stations are available in most crossing situations.

Partial double-tracks only have a local effect. Rather high minimum levels of standard deviations reflect this. Crossing situations that fall outside the double-track still result in high crossing times due to long inter-station distances. In case of shorter inter-station distances standard deviations fall faster and to a much lower minimum level.

High arrival punctuality is an important requirement to obtain independence between crossing trains. It is therefore very interesting to find out at which variance level, and/or capacity utilisation level, this simple model starts to give serious under-estimations of crossing time. Such a check could be done in a simulation tool, which makes interactions between several crossings possible through active dispatching. A qualified guess is that the simple model works quite well if traffic intensity is lower than three trains an hour in each direction, provided that variance in arrival delay is moderate.
The studied cases are quite idealised in themselves. Further work will therefore be done to examine asymmetric cases: between traffic directions, in the infrastructure etc. Partial double-tracks are constructed for one (or a few) timetable structure(s). This is a very important (negative) property that can be examined by systematically changing the planned crossing position.

A very important situation is when a (planned) crossing is combined with a regular stop for passenger exchange, which gives a lower crossing time. An analysis of such cases can indicate how timetables should be constructed.

Hopefully this work will contribute to a deeper understanding of the features of single-track railway lines and the possibilities to operate fast services with high punctuality on such lines. The complete work will be presented in a licentiate thesis in 2007.

5 Acknowledgements

This project has been financially supported by the Swedish National Rail Administration.

References

[1] Barter, W., Forecasting robustness of timetables, Comprail 2004
Paper II

Olov Lindfeldt

“SAMFOS - a Timetable-free Way of Analysing Single-track Railway Lines”

SAMFOST - a Timetable-free Way of Analysing
Single-track Railway Lines

O. Lindfeldt
Royal Institute of Technology, Transport and Economics,
Teknikringen 72, 100 44 Stockholm, Sweden, e-mail: olov1@infra.kth.se

Abstract

Operation of single-track railway lines is characterised by the stochastic crossing time. Important properties such as capacity, run times and overall punctuality are strongly dependent on the crossing time. In non-congested situations the crossing time can be modelled analytically. SAMFOST is a model that shows in a transparent way how the crossing time varies along a railway line. Including arrival delays, the mean crossing time (and standard deviation) can be calculated for different timetable crossing points. An important contribution is that the dependence between arrival delays and crossing time is shown explicitly.

SAMFOST has been successfully validated against the accepted simulation tool RailSys. The validation shows that SAMFOST has an accepted accuracy as long as the crossing time depends only on the two crossing trains (non-congested situations).

Using SAMFOST, different types of measures to decrease the crossing time (and its variance) have been analyzed. For example, a combined crossing and passenger stop gives a considerable decrease in the mean crossing time, but at the same time the variance increases. A combined crossing and passenger stop thus becomes more dependent on punctuality than a crossing at a station without passenger stop.

Keywords: single-track, railway operation, delay

1 Introduction

The railway is a complex traffic system. Factors such as infrastructure design, timetable and punctuality interact in a complicated way. The ongoing increase in utilisation of the railway system calls for a better understanding of its traffic properties. In order to display and describe some of the fundamental characteristics of single-track railway lines, a Simplified Analytical Model For Single-tracks, SAMFOST, has been developed.

This paper will give a short background, followed by a description of the problem and the model and its validation against an existing simulation model. At the end, two examples of applications will be given. The first shows how fictive model lines can be used to generate general results about changes in infrastructure and timetable etc. The second example shows how an existing, highly asymmetric, railway line is characterised when analysed with SAMFOST.

About 80% of Swedish railway lines are single-track. Single-track lines have lower capacity, lower average speeds and higher disturbance sensitivity than double-track lines. All these differences are due to train crossings where the two traffic directions
become interdependent. Some measures can be taken in order to minimise the influence of crossings: partial double-tracks, more crossing stations or combined crossings and passenger stops are three important examples.

A crossing is characterised by the crossing time, i.e. the time that is needed to carry out the crossing. The crossing time is thus the extra time that has to be added to the run time because of the crossing and it depends on factors such as infrastructure design, timetabled crossing point, vehicle parameters, driver behaviour, punctuality etc. The last two factors are more or less stochastic which also makes the crossing time stochastic. The crossing time will therefore have a distribution. This distribution directly influences traffic properties such as capacity, running times, secondary delays, timetable flexibility etc.

The occurrence of crossings implies time separation of the services, which in turn opens up for powerful mathematical models to be constructed for single-track lines. SAMFOST is a simple and transparent mathematical model that can be used to increase understanding of single-track railway traffic. It handles infrastructure parameters as well as timetable and punctuality. In SAMFOST, the analyses are performed in steps and the first two are independent of the timetable crossing point.

A major advantage of mathematical models is that several variables can be handled systematically at the same time. Interactions between different variables can be examined. A theoretical alternative to using a mathematical model like SAMFOST is to perform a large number of simulations. In practice this is not possible since today’s simulation tools are not suited to automatic variable change and so this method will be very time consuming. It can therefore be argued that mathematical models can be very useful in many cases.

2 Related research

The research on railway operation, including scheduling, re-scheduling (disturbance management) and infrastructure planning is extensive. Scheduling and re-scheduling seem to be a better-covered area than infrastructure configuration and its influences on traffic operation and operation possibilities. This is in a way natural since the infrastructure is much more static and harder to change. Nevertheless, knowledge of infrastructure properties is essential when robust timetables are to be constructed.

Higgins et al [3] put forward an optimization model to determine the required number and location of crossing stations. The crossing time is decomposed into “conflict delay” and “risk of delays” which corresponds to the lowest crossing time at a station and the stochastic delay propagation part of the total crossing time respectively. The model has several similarities to SAMFOST but both the train runs and the delays are modelled in less detail.

Petersen [9] presents an analytical model for mean delays due to crossings and overtakings as a function of traffic density. All departures are random, the model handles the trains pair-wise and the delays due to interferences are calculated roughly. Higgins et al [4] propose a model to optimise the timetable on a single-track, disregarding initial delays. Chen and Harker [2] propose a somewhat more developed
model for the delay functions for scheduled trains. Here, too, the trains are treated in pairs. Sahin [11] analyses dispatchers’ decision processes and emphasises that potential future conflicts have to be taken into consideration when resolving immediate conflicts.

Carey [1], Yuan and Hansen [15] and Wendler [14] present another method, based on pure statistics. Interconnections between trains that use the same track section or platform track are described through convolution. This method is based on assumptions about independence, but still gives a clear picture of the “knock-on” effects in congested situations.

Törnquist and Persson [12] and Törnquist [13] present a heuristic model that can be used to re-schedule traffic according to different objective functions in complicated networks. The interactions between different trains are modelled on a fairly high level and only singular and extensive delays are considered. Mattsson [7] provides a survey of relationships between capacity and delays. He also compares available methods for railway operation analyses.

The work presented in this paper is mostly related to the stochastic analyses performed by Carey [1], Yuan and Hansen [15] and Wendler [14] based on the same assumptions about independent train runs. A major difference compared to most other research, however, is that both infrastructure and timetable appear as variables. It thus becomes possible to show how the infrastructure design influences the operation and feasible timetable strategies.

3 The model

The main purpose of SAMFOST is to examine the traffic properties of single-track railway lines. Some powerful assumptions make the model analytical and thereby distinguish it from different kinds of simulation models. These fundamental assumptions open up for a stepwise systematic analysis that makes the results general and understandable. SAMFOST is primarily constructed to handle analyses of fictive, non-existing railway lines where variables of interest can be examined in a simple manner.

3.1 Fundamental assumptions

In a crossing situation the crossing time is determined by a great many factors. The most important of these are infrastructure design, vehicle parameters, timetable crossing point, driver behaviour, adhesion, arrival delays, and congestion (influence of other trains). The first three can be regarded as deterministic and are easy to model. In many simulation models these parameters are more or less static input data. The last four factors are much more stochastic.

An analytical model can easily handle one stochastic variable. If more variables are to be stochastic the model will lose transparency and one comes closer to a simulation model where it is much more difficult to recognise the effect of a single variable. For this reason only the arrival delays are modelled as stochastic variables in SAMFOST. Driver behaviour and adhesion are assumed to be deterministic.
Another important delimitation is that no other trains than the two crossing ones are modelled. All kinds of congestion situations, where more than two trains interact, are thus left out of the analysis. This assumption is very powerful since it makes analytical calculations possible, where arrival delays are the only stochastic variables. The crossing time is then unequivocally determined by the first three factors together with the delay difference, which is given by the arrival delays for the crossing trains. In a real case, with several interacting trains, a given value of the delay difference would give different crossing times due to the positions and delays of surrounding trains.

The assumption of independence of other trains is only valid when congestion effects in real traffic are small. For a given line the congestion-free capacity is strongly dependent on punctuality. SAMFOS’s precision is therefore dependent on punctuality as well as the frequency of the train services.

A second, less restricting, assumption is that crossing trains are independent before a crossing course starts. It has been shown empirically that this assumption holds. Following this assumption, the two stochastic variables up-going and down-going arrival delay can easily be combined into the very important stochastic variable arrival delay difference.

### 3.2 Structure

The two fundamental assumptions about crossing trains’ independence of each other and other traffic makes it possible to obtain a plain, simple model structure, see figure 1. The length of the modelled line is determined by arrival punctuality. In the first step, dimensioning delay distributions are chosen, preferably distributions with a high variance. The resulting distribution for arrival delay difference is calculated through (numerical) convolution. This distribution shows how the theoretical crossing point is distributed.
To determine the minimum length of the infrastructure model needed, an upper and a lower limit are taken from the delay difference distribution in such a way that a large proportion (99.5%) of the crossings fall within the limits. The chosen length is then filled with infrastructure that has the properties to be examined. Several variables are defined: speed limits on line and station tracks, station parameters such as position, track configuration, track length, points length, speed restrictions at points, interlocking time etc. Double-tracks are modelled as long stations.

Vehicles are defined by parameters that are needed to model running times and signalling: acceleration curves (simplified), braking rate, maximum speed and vehicle length.

Given infrastructure and vehicle data, the model calculates a crossing time function between the time limits for each traffic direction. The crossing time function shows the time needed to perform a crossing as a function of the arrival delay difference. The function, that is independent of timetable and punctuality, gives a character for the studied line. The crossing time function is presented in an aggregated way showing the sum of the crossing times of the two crossing trains. Two examples are shown in chapter 5 (figure 7).

In the final step the crossing time function is combined with the distribution for the arrival delay difference, thereby giving the distribution for the crossing time. One crossing time function may be combined with several different distributions for arrival delay difference. This is possible as long as at least the prescribed part of the crossing situations falls between the limits derived from the dimensioning distributions.
4 Validation

SAMFOST has high validity when the frequency of train services is moderate and/or punctuality is high. An important part of the validation is to estimate how validity changes with frequency and/or punctuality. Before such a validation is performed it is necessary to validate the model under the assumptions described above.

4.1 Given fundamental assumptions

Under the fundamental assumptions of independence one crossing situation at a time may be studied, since trains following in one direction are so far from each other that no interference occurs. Building on this, SAMFOST contains a number of functions:

- **Basic functions**: run time calculations, i.e. passing, deceleration and acceleration courses and arrival delays.

- **Dispatching functions**: station disposition and track disposition.

- **Signalling functions**: modelling crossing courses.

These functions were validated against the simulation tool RailSys [10]. A symmetrical reference line was used having a uniform speed restriction of 200 km/h. So the up-trains and down-trains could be timetabled on different tracks at each station and the need to choose track was eliminated. This was appropriate, since RailSys does not model track disposition in the required way.

4.1.1 Basic functions

Three types of running time calculations were validated: passing time at constant speed, deceleration and acceleration. The validation of passing time was limited to one random sample that showed that the models calculate this type of running times in a similar way. Three series of random samples were used to validate deceleration and acceleration courses:

- **Deceleration**: braking from 200 km/h down to a given target speed.

- **Acceleration**: acceleration from a given initial speed to 200 km/h.

- **Passing of a crossing station with speed reduction**: deceleration from 200 km/h to a given speed restriction, constant speed through a station, followed by acceleration back to 200 km/h.

Each series contained ten randomly sampled values of target speed, initial speed and speed restriction respectively taken on the interval (0, 200) km/h. For each sample the crossing time was calculated using SAMFOST and RailSys respectively. Since SAMFOST and RailSys model the acceleration differently, a small, systematic difference occurred. In order to minimise this difference, a correction factor was introduced in SAMFOST. Figure 2 shows the results after this correction.
The comparison is made more difficult by the fact that the output from RailSys is rounded to complete seconds, while SAMFOST keeps the decimals. For this reason, only points outside the interval [-0.5, 0.5] indicate a systematic difference between the models. All points in the “speed restriction” series lie within the interval. It is noticeable that the series “Deceleration” has several points outside the interval. Both models use constant deceleration to calculate braking courses.

The acceleration series has four points that stand out. One explanation for the first three points could be that their initial speeds are low and acceleration therefore takes place over many speed intervals. At each speed interval, SAMFOST overrates the acceleration and so the crossing time becomes shorter than in RailSys.

![Graph showing the difference in crossing time between SAMFOST and RailSys for three series of random samples.](image)

**Figure 2:** Difference in crossing time between SAMFOST and RailSys for three series of random samples.

Another basic function is the generation of arrival delays. In order to show that RailSys generates the same arrival delays, delay values for 4,000 crossings were generated in RailSys. The resulting distribution followed the expected negative exponential function very well.

<table>
<thead>
<tr>
<th>Target, initial speed and speed restriction [km/h]</th>
<th>Difference between SAMFOST and RailSys [s]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deceleration</td>
<td>![Deceleration symbol]</td>
</tr>
<tr>
<td>Acceleration</td>
<td>![Acceleration symbol]</td>
</tr>
<tr>
<td>Speed restriction</td>
<td>![Speed restriction symbol]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Target speed, initial speed and speed restriction level [km/h]</th>
<th>Difference between SAMFOST and RailSys [s]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>![0 symbol]</td>
</tr>
<tr>
<td>50</td>
<td>![50 symbol]</td>
</tr>
<tr>
<td>100</td>
<td>![100 symbol]</td>
</tr>
<tr>
<td>150</td>
<td>![150 symbol]</td>
</tr>
<tr>
<td>200</td>
<td>![200 symbol]</td>
</tr>
</tbody>
</table>

Table 1: Arrival delays generated by RailSys

<table>
<thead>
<tr>
<th></th>
<th>SAMFOST</th>
<th>RailSys (sample)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean [s]</td>
<td>150</td>
<td>152</td>
</tr>
<tr>
<td>Standard deviation [s]</td>
<td>150</td>
<td>154</td>
</tr>
</tbody>
</table>
Since the arrival delays show good conformity, the subsequent validation could be performed with optional arrival delay distributions. This is an advantage when it comes to validating dispatching and signalling functions since a higher variance increases the possibilities to test the model. Hence, uniform distributions were used in the following validations.

4.1.2 Dispatching and signalling functions

The dispatching and signalling functions were also validated against RailSys. Two important dispatching situations occur on single-track lines: choice of crossing station and track disposition at the chosen station. When a crossing station has been chosen and track disposition decided, signalling functions are crucial for the trains’ interaction during the crossing course. The time for unlocking the entrance train path and locking the exit train path are of special interest.

In order to validate dispatching and signalling functions, some adjustments needed to be made in SAMFOST due to some fundamental differences in model construction, as shown in table 2.

Table 2: Model comparison

<table>
<thead>
<tr>
<th>Model step</th>
<th>SAMFOST</th>
<th>RailSys</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Define infrastructure</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Choose timetable crossing point (where delay difference is defined zero)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Choose values for arrival delay differences, for instance every 5 seconds in [-2100, 2100]</td>
<td>Random sample of arrival delays following predefined distributions.</td>
</tr>
<tr>
<td>4</td>
<td>Calculate crossing time for chosen values of arrival delay</td>
<td>Simulation</td>
</tr>
<tr>
<td>5</td>
<td>Calculate distribution for the crossing time</td>
<td>Compile distribution for crossing time</td>
</tr>
</tbody>
</table>

A fundamental difference between the models is that SAMFOST systematically calculates the crossing time for a number of traffic cases and then, in the final step, takes the distribution into consideration. RailSys, on the other hand, uses the distribution directly to sample traffic cases.

In order to validate dispatching and signalling functions it is preferable to use the same values of the arrival delay difference in both models. By doing so, a direct comparison becomes possible. The distribution for the arrival delay difference does not have to be “the real one”. Instead, it is preferable to use a distribution with a high variance so that the sample has many unique values of delay difference.

4,000 crossings were therefore sampled in RailSys, using uniform distribution for both directions on [0, 2100] s. The resulting distribution for arrival delay difference was then triangular on [-2100, 2100] s. About 2,350 of the sampled crossings proved
to have a unique delay difference, which gives a good coverage. Every unique crossing case was simulated in RailSys and then run in SAMFOST.

Figure 3 shows that the differences in crossing times between SAMFOST and RailSys lie in the interval (-1.43, 0.47) s. The differences only vary in those cases when both trains arrive at a station at the same time. These cases correspond to the minimum areas of the crossing time function. Between the areas with variations the difference is constant. This is because the delay difference is always an integer in RailSys. In intervals with constant difference between the models, one of the trains has stopped to wait to cross. This means that the decimal part of the crossing time always becomes the same as soon as a crossing only gives rise to crossing time for one of the trains.

The fluctuations are due to differences in run time calculations, mostly accelerations, and rounding differences. The mean difference is below zero, indicating that SAMFOST has higher acceleration ratios. The diagram shows the difference in total crossing time, and so the rounding difference gives an interval (-1, 1) s within which no conclusions can be drawn. The dispatching and signalling functions are very similar and so it is shown that SAMFOST models the crossing situations very similarly to RailSys, as long as the fundamental assumptions about independence hold.

Introducing passenger stops makes it more difficult to dispatch the crossing situations. A special validation was therefore performed just for situations with passenger stops. The results showed almost the same pattern as presented above, except at two points, giving a difference in crossing time of 7.5 s. These singularities were due to the models choosing different crossing stations and do not influence the overall results. SAMFOST thus also handles passenger stops properly.

![Figure 3: Difference in crossing time between SAMFOST and RailSys.](image)

4.2 Congestion: high frequency of train services and low punctuality

The results from SAMFOST are only valid as long as only the two crossing trains influence the crossing. Other trains must be far enough from the crossing area so as not to disturb the choice of crossing station or the track disposition. If other trains in fact influence a crossing there is a congested situation. At least four factors are important for the occurrence of congestion:

- Infrastructure design, including passenger stop locations etc.
- Vehicle parameters
- Frequency of train services
- Arrival delays (punctuality)

Infrastructure design and vehicle parameters are crucial for the theoretical capacity, i.e. capacity during total punctuality. This capacity is easily calculated from the crossing time function. With total punctuality two trains can follow at a distance of two inter-station distances. The run time for these distances must be supplemented with the crossing time for one crossing. This is shown in figure 4.

For asymmetrical railway lines, with passenger stops and/or uneven inter-station distances etc, the inter-station distance cycle that is longest in time is dimensioning for the theoretical capacity. These values can also be found in the crossing time function.

Train services must never be so frequent that trains run closer than the theoretical capacity permits. In real railway operation, trains may come closer through the influence of arrival delays, even though timetable frequency meets capacity restrictions. Congestion is therefore unavoidable in real operation. A lower frequency of train services decreases the number of congestion situations.

The actual time distance between two following trains depends on the distribution of their delay difference. The timetabled time distance is the mean of this distribution and its shape is calculated through convolution (provided that independence can be assumed).
Figure 4: Minimum time distance for following trains.

Figure 5: Distributions for arrival delay difference at high and low punctuality.

Figure 5 shows two examples. The steeper curve represents a situation with high punctuality (mean arrival delay is 50 s), which gives a low variance in delay difference, while the flatter curve results from lower punctuality (mean arrival delay is 350 s). Congestion occurs for small time distances and so the right tail in each distribution could generate problems.

Increasing the timetabled time distance between two trains (decreased frequency) reduces the influence of arrival delays. The level of this statistical buffer time can be
calculated after determining a tolerance level of congestion (for instance 0.5%). In figure 5 a buffer time of 5 minutes would be sufficient for the high punctuality case whereas the low punctuality case would require 20 minutes' buffer time. These values are then added to the minimum technical time separation for the line in question to obtain the minimum time separation that avoids congestion.

This simplified way to determine the non-congested capacity implies that the trains are handled pair-wise in adjacent order. If trains numbers 3, 5 and 7 follow train number 1, the only interference risk taken into account is between 1 and 3. The interference risk between the pair 1 and 5 and the pair 1 and 7 is ignored. This is accepted since the probability is very small.

If congestion is accepted in 0.5% of the crossings, the congestion-free capacity in trains/h/direction is as shown in figure 6. Three different traffic situations, e.g. combinations of infrastructure design, vehicle properties and passenger stop are shown. The upper curve (dotted) represents an infrastructure with high crossing station density (7.5 km between stations) and high acceleration vehicles without any stop for passenger exchange. The middle curve represents a line with moderate crossing station distances (15 km between stations) and high acceleration vehicles without any stop for passenger exchange. The lower curve (solid) represents a line with moderate crossing station density (15 km between stations) and low acceleration vehicles with a passenger stop at the timetabled crossing station.

Figure 6: Congestion-free capacity for three different combinations of infrastructure, vehicle and passenger stop.

Figure 6 shows the number of trains per hour and direction that lead to congestion in 0.5% of the crossing situations. The congestion-free capacity decreases as punctuality deteriorates and more statistical buffer time is needed. The worst situation arises when passenger stop is combined with a slow acceleration vehicle.
since the inter-station distances adjacent to the passenger stop station get a much longer run time. A shorter inter-station distance increases the congestion-free capacity considerably, especially when punctuality is high.

Figure 6 also shows that SAMFOST's validity depends on infrastructure design (the most important factor being the inter-station distances), occurrence of passenger stops and vehicle type. In general, the congestion effects are small below each curve and SAMFOST thus has a high validity at those points. In the area above the curves, congestion increases and SAMFOST therefore underestimates the crossing times if it is applied in this area.

5 Applications

An important advantage of analytical models such as SAMFOST is that selected variables can be examined systematically and “automatically”. The variables can be tested for a great many values. It is also possible to find interactions between variables. Examples of variables that can be examined in this way are inter-station distances, length of partial double-tracks, stop patterns, vehicle parameters, the timetable and arrival punctuality. Together, different values of these variables create fictive combinations of infrastructure designs and traffic patterns.

5.1 Combined crossing and passenger stop

In [6] SAMFOST is used to analyse partial double-tracks and inter-station distances respectively. In none of the analysed cases is the crossing combined with a passenger stop on the modelled line. The crossing time thus consists of time for deceleration, waiting and acceleration. If a crossing is planned at a station where the trains have a regular passenger stop, the time for deceleration and acceleration, as well as most of the waiting time, is useful time that does not become crossing time.

Figure 7 shows two examples of crossing time functions. The crossing time function indicates how long the crossing time will be as a function of arrival delay difference. The examples shown represent a line without any passenger stop (dotted) and a line where the trains have a passenger stop in the middle station (solid). The inter-station distance is 15 km, the track length is 750 m at all stations and the speed restriction at points is 100 km/h. The stop position is at one end of the station (train length is 160 m) and so the speed restriction at the farther point generates a low crossing time.
Figure 7: Crossing time function for a line with 9 crossing stations (15 km inter-station distances). Dotted line is without passenger stop, solid line with passenger stop.

Figure 8: Crossing time function (solid) and probability density function for arrival delay difference (dashed).

The combination with passenger stop gives the following effects:

- The shortest crossing time falls from 140 s to almost 0 s. A time interval occurs within which the shortest crossing time is possible. The length of this interval depends on the timetabled dwell time.
• The time distance between the two adjacent maxima that surround the passenger stop increases and so the time interval within which the crossing station is used increases. This follows from the fact that the run time increases when a stop is introduced.

• The maxima that surround the stop position become higher. This is due to the fact that the run time on the single-track sections surrounding the station where the trains stop increases.

The crossing time function shows what the crossing time will be for a given arrival delay difference between the crossing trains; the crossing time during complete punctuality (no delays) can thus be read directly in the crossing time function. Since the actual arrival delay difference follows a distribution, the actual crossing time varies from one crossing to the other. An example is shown in figure 8. Combining the probability density function and the crossing time function results in a distribution for the crossing time. This distribution is most conveniently represented by the cumulative probability function.

Figure 9 shows some examples. Here, the inter-station distances vary from 15 km down to 3 km (2 km step length) giving rise to 7 different crossing time functions. When these functions are combined with two different examples of distributions for the arrival delay difference, two groups of cumulative probability functions are created. These are the curves shown in figure 9. Solid curves represent a higher punctuality, Exp(150)-distributed arrival delays, and the dashed curves represent a lower punctuality, Exp(350)-distributed arrival delays. The lower curve in each group has an inter-station distance of 15 km, the upper curve 3 km.

![Cumulative distributions for 7 different inter-station distances. Distributions are shown for two levels of arrival punctuality: high (solid) and low (dashed).](image-url)
In each group, the curves coincide up to a crossing time of 100 s and so up to this level the inter-station distance is of no interest. This shows (implicitly) that only a combined crossing and passenger stop results in a crossing time lower than 100 s. With high punctuality, as many as 70% of the crossings take place with a shorter crossing time than 100 s.

Crossing stations without a passenger stop normally give a (shortest) crossing time of about 140 s. However, curves for the shortest inter-station distances “leave” the other curves below this level. This indicates that stations, in this case, are located within the acceleration and deceleration distance of the passenger stop location. This is one example of interaction between inter-station distance and occurrence of passenger stop.

Figure 10 shows how the mean crossing time and its standard deviation vary with inter-station distance when a passenger stop is located at the timetabled crossing station. The five solid curves represent mean crossing time for different levels of punctuality and the five dashed lines are standard deviation. Punctuality varies from Exp(350)-distributed arrival delays (upper curves) to Exp(150)-distributed arrival delays (lower curves) and step length is 50 s.

Arrival punctuality is of great importance. This is due to the fact that the stations are not identical, as was the case in [6], hence the timetabled crossing station (with passenger stop) cannot be replaced by any other station without greatly increasing the crossing time. High punctuality gives a higher utilisation of the timetabled station and so a lower mean. Decreasing the inter-station distances is thus a more effective measure when punctuality is low.

---

Figure 10: Mean crossing time (solid) and standard deviation (dashed) for crossing time as a function of inter-station distance. Five levels of arrival punctuality.
So far the timetabled crossing point has been a constant, located at a crossing station. This location is natural, since it gives a minimum crossing time. In practice it is not always possible, or desirable, to choose the point that results in the lowest crossing time. Different kinds of timetable dependencies often make it impossible to obtain the optimal timetable crossing point for every crossing. It is therefore of great interest to examine how the crossing time is influenced by the location of the timetable crossing point. The *timetable crossing point* is here defined as the (theoretical) point where the crossing would take place at arrival difference equal to zero if the line was double-track.

In SAMFOST, timetable crossing point can be varied. To do this the location of the distribution of the arrival delay difference is moved stepwise along the crossing time function. Each step corresponds to a timetable crossing point located where the arrival delay difference is zero. Figure 11 shows an example. In each step, the product of the probability density function (dashed and dotted) and the crossing time function (solid) is integrated giving the mean of crossing time. The standard deviation is also calculated in each step.

Figure 11: Different timetable crossing points can be analysed by moving the probability density function.
Figure 12: Mean (solid lines) and standard deviation (dashed) for crossing time as a function of location for timetable crossing point.

Figure 12 shows how mean crossing time and standard deviation depend on the location of timetable crossing point for 7 different inter-station distances. Solid lines are mean values and dashed lines are standard deviations. Inter-station distances vary from 15 km (upper curves) to 3 km (lower curves) with a step length of 2 km. Underlying punctuality corresponds to Exp(150)-distributed arrival delays.

The curves are cyclic and the period depends on the inter-station distance. Each valley on the mean curves corresponds to a crossing station, giving a (local) minimal crossing time. A passenger stop is located at the middle crossing station, giving the absolute minimum mean crossing time. Far out from the middle, the influence of the passenger stop is small but increases the closer to the middle the timetabled crossing point is located. The mean crossing time always depends on the inter-station distance, but this dependency is weaker when crossing is combined with a passenger stop.

A passenger stop influences the mean crossing time along a longer section than the extension of the actual crossing station. For a rather wide time (distance) interval, the mean crossing time is lower than for adjacent crossing stations. Asymmetrical curves show that the passenger stop is located at one station end, giving a different influence on different sides. The impact is highest on the side that is located opposite to the stopping location, since the station (750 m) acts as a short double-track.

Standard deviation curves have the same order as the mean curves. A long inter-station distance gives a higher standard deviation. The influence of the passenger stop is very clear. Since one local section has much shorter crossing times the standard deviation increases strongly when the difference between low and high crossing times increases. The high standard deviation results in higher disturbance.
sensitivity. Higher time supplements are therefore needed in the timetable when a
crossing is combined with a passenger stop and so part of the time gain is lost.

Infrastructure and traffic in the examples shown above contained two important
symmetries that do not exist in real cases: even inter-station distances and same
arrival delay distribution for both traffic directions. The next example shows analyses
of a asymmetric infrastructure design with asymmetric traffic conditions.

5.2 An existing line

As an example of an existing line the Svealand line between Södertälje and
Eskilstuna, south-east of Stockholm in Sweden was chosen. This line section
contains a partial double-track as well as ordinary crossing stations with and without
passenger stop. This section of line thus shows several properties that are of interest
for single-track lines.

![Figure 13: Infrastructure design of the Svealand line.](image)

Figure 14 shows the crossing time function. Eskilstuna is located to the left and
Södertälje to the right. From the left three stations without passenger stop can be
recognised, the partial double-track (Lg – Ryb) and finally a station with passenger
stop. Between Härad and Malmby the services have a passenger stop without any
possibility for simultaneous crossing, giving these adjacent crossing stations special
properties compared to Kjula, since they are located within acceleration/deceleration
distance.

Due to the passenger stop between Härad and Malmby the run time between these
stations is increased (maximum crossing time of 480 s). Using Härad or Malmby as
timetable crossing point therefore gives a high mean crossing time since a great part
of the probability density function covers the area where the crossing time is high.
Figure 14: Crossing time function and empirical probability density function for arrival delay difference on Svealand line.

Figure 15: Crossing time function (solid thin). Mean (solid bold) and standard deviation (dashed) for crossing time. Two different levels of arrival punctuality are shown.

Figure 14 also shows the probability density function derived from empirical delay data (dashed curve). The curve is located at the timetable crossing point used in the current timetable (Jan 2007). It can be seen that the margin against secondary delays is 6 minutes in one direction and 3 minutes in the other. The density curve shows a high degree of asymmetry since punctuality from Södertälje is far lower than from Eskilstuna.
If the timetabled crossing point is varied along the studied line section, the mean and the standard deviation of the crossing times are as shown in figure 15. In this case, two different levels of punctuality were used. Solid curves are means and dashed curves standard deviations. The mean curve showing the highest amplitude corresponds to the highest punctuality and the curve with the lowest amplitude has the lowest punctuality (representing actual punctuality). Higher punctuality generally gives a lower standard deviation.

The figure clearly shows that higher punctuality would decrease the crossing times (both mean and standard deviation). But higher punctuality also means a greater difference between different timetable crossing points! When punctuality increases the mean time function converges towards the crossing time function and with absolute punctuality the functions coincide.

Of special interest is the fact that the highest punctuality would give timetable crossing points on the partial double-track with really short crossing times (mean and standard deviation). The figure also shows the connection between punctuality and the need for time supplements in the timetable. Not even the highest punctuality level would be enough to avoid a time supplement when a crossing is located at a crossing station where the trains have a passenger stop (Nykvärs).

Another result of the figure is that special caution is needed when a crossing is planned at a border station between a double-track and a single-track line section. This is the case in Södertälje, where the double-track makes the crossing time function follow the zero line. Due to the adjacent single-track section both the mean curve and the standard deviation curve show much higher time values. The timetabled crossing point should therefore be located on the double-track section a time distance from the border point in order to obtain time efficient crossings.

6 Conclusions

The occurrence of crossing time is probably the most important difference between single-track and double-track railway lines and it has a great influence on factors such as capacity, run times and overall punctuality. To handle these factors an understanding of crossing times is necessary. The crossing time is strongly dependent on punctuality (arrival delays). Since delays behave stochastically the crossing time will also behave stochastically. This fact increases the peculiarity of single-tracks even more.

Single-track lines have properties that, in certain circumstances, allow analytical modelling. For instance, the train services are generally separated in time and the number of interacting trains is often limited. Using analytical models several variables can be handled at the same time and results become transparent and understandable. Analytical modelling, however, involves simplifications of which one has to be aware.

In this work only two stochastic variables have been taken into account: arrival delays for up-going and down-going services. This is an important delimitation and simplification. Defining arrival delay as the time deviation from timetable at the moment when a crossing course starts implies that most factors that influence
punctuality are covered. Given this definition, only stochastic factors that act during the actual crossing course are excluded. Similar assumptions are made in many other traffic models.

However, the definition above makes it difficult to compare the distributions in the model with empirical delay distributions that result from measurements in predefined physical points. Under the assumption that no other delay sources exist on the line section where most (99.5%) of the crossings take place, the two distribution types are identical.

Passenger stop is one important example of a non-negligible source of delays. Passenger stops often induce changes of delay distributions. If a crossing is combined with, or located near to, a passenger stop, the validity of SAMFOST is probably lower, since the model does not handle this kind of superposed variances.

The assumption that the arrival delays of two crossing trains are independent makes it easy to replace the two stochastic variables with one: arrival delay difference. This new variable plays a very central role.

An even more important assumption is that of independence of other trains, which implies that one value of the arrival delay difference always gives the same crossing time. This assumption implies that the model cannot handle congested situations. However, it can be argued that in cases where congestion has a major impact on the overall punctuality the actual line is overloaded and measures need to be taken to change the timetable and/or infrastructure design. The crossing time that is added due to congestion is in a way a second order crossing time.

Due to its simplicity SAMFOST gives a clear view of traffic properties. A limitation is that the model, in its present form, does not handle series of crossings. In a crossing the delay distributions change. Empirical data shows that it is difficult to recover the distributions using static time supplements in the timetable. An assumption that the distributions can be restored to their original shape is therefore an approximation.

Iterative calculations could solve this problem with a more complicated, and more expensive, model. Moreover, the important factor “time supplement” must then be introduced. For this type of more sophisticated analyses simulation should be considered. SAMFOST could then be used during the preparations for the simulation work.

One important result of this work is the “crossing time function”. Given infrastructure design and vehicle parameters this function shows traffic properties of a line and so a first analysis can be performed without any knowledge of timetable and punctuality. The crossing time function, as it is shown in this paper, displays the total crossing time for both trains. This gives a clear system perspective that can be somewhat misleading when the traffic directions are to be analysed separately.

When the arrival punctuality is known, the mean crossing time function and the associated standard deviation can be calculated. The analysis is still independent of
timetable since the calculated functions show expected outcome of all possible
timetable crossing points (given that the condition of non-congestion holds).

6.1 Validation
In this project a special kind of validation has been performed. Instead of validating
against real traffic situations, SAMFOST has been validated against an accepted
simulation model. A high degree of agreement shows that SAMFOST has the same
validity as RailSys, given the fundamental assumptions of independence.

The validation was performed with respect to the crossing time function. In
SAMFOST the calculations of the crossing time function are the most complex part
where interacting train runs and signalling system are modelled. An alternative would
have been to perform the validation after the statistical preparation, i.e. after the last
model step. A more complete validation would have included both variants.

The validation of SAMFOST during varying frequency of services and arrival
punctuality was performed as an analytical calculation of congestion limits. These
were calculated for different infrastructure designs and different punctuality levels. A
severe disadvantage of this method is the risk of underestimating the area of use. The
method does not say anything about how SAMFOST gradually underestimates
crossing times when congestion increases. An alternative way of examining this
would have been to use a simulation tool to simulate successively increasing
congestion. Such a validation would also have given more knowledge about how
capacity depends on punctuality etc.

6.2 Modelled infrastructure
Different kinds of infrastructure measures that can be taken in order to decrease the
crossing time and/or its variance are of special interest in this project. To evaluate
these measures fictive model lines were designed and examined. All these lines have a
high degree of symmetry in station design, inter-station distances, vehicles, arrival
delays and traffic patterns. The symmetries help to isolate the effects of a few
examined variables.

Inter-station distance is a variable that does not show any symmetries on real railway
lines. Stations are often located where it is easy to construct them, where the services
have a passenger stop etc, resulting in varying inter-station distances. The station
design, i.e. track lengths, locations of signals, and interlocking systems are also
specific to the individual station.

The analyses of fictive model lines were supplemented by the analysis of an existing
line that lacks symmetries. These analyses proved to be very useful since several
asymmetries could be recognised from existing experience from the analyzed line.
6.3 Future work

The work that has been carried out in this project opens up for further work in different directions. Further studies of railways' technical traffic properties, in particular all types of variances and deviations from the well structured timetable, have the potential to extend the understanding of the railway as a technical system.

Even if the crossing time is a central parameter for operation of single-track lines, methods that translate it into other important factors such as capacity and travelling and transport times etc must be developed further. To succeed in such a translation more knowledge of time supplements and how they work is needed. When time supplements are introduced into the analysis it is possible (and desirable) to introduce new stochastic variables. Additional delay at a passenger stop is one important variable. Additional sources of disturbances open up for a more realistic model of time supplements, since their purpose is to reduce the effect of the sum of all disturbances.

One way of introducing new stochastic variables is to perform simulation in an existing simulation tool. The research method then resembles laboratory experiments. It is possible to perform only a few experiments where the variables are set in advance. After the simulation, as general results as possible must be drawn. To succeed in this a developed methodology, similar to statistic experimental design theory has to be applied. An alternative continuation of this work is to develop such methods.

This work can be applied in strategic planning of new (or upgraded) infrastructure as well as timetabling on existing infrastructure. Timetabling is thus a subset of strategic planning since the timetable itself is one of the most important variables when new infrastructure is planned.

7 Acknowledgements

The Swedish National Rail Administration provided financial support for the project and the work was supported by a reference group consisting of representatives of the Royal Institute of Technology in Stockholm and the Swedish National Rail Administration.
References


Paper III

Olov Lindfeldt

“Crossing times on single-track railway lines – dependencies of different infrastructure and traffic factors”

CROSSING TIMES ON SINGLE TRACK RAILWAY LINES – DEPENDENCIES OF DIFFERENT INFRASTRUCTURE AND TRAFFIC FACTORS

O. Lindfeldt MSc
Royal Institute of Technology
Transport and Economics
Teknikringen 72
100 44 Stockholm
Sweden
olovl@infra.kth.se

KEYWORDS: Single-track, crossing, railway operation, delay, timetable flexibility

Abstract
Operation of single-track railway lines is characterised by crossings. Due to disturbances the crossing time is stochastic. Important properties such as capacity, run times and overall punctuality are strongly dependent on the crossing time. In non-congested situations the crossing time can be modelled analytically. SAMFOST (Simplified Analytical Model For Single-tracks) is a model that shows in a transparent way how the crossing time varies along a railway line.

This paper shows how the mean crossing time function can be used to examine timetable flexibility, i.e. possibilities to change the timetable given constraints on the mean crossing time. Three different measures for timetable flexibility are proposed, showing how the available crossing points are spread along the line and the spread of mean crossing time for available crossing points.

Many factors such as infrastructure parameters, vehicle parameters, timetable, delays etc affect the crossing time. In many cases these factors interact in complicated ways. To show this, the results of a 2²-factorial experiment are presented. Partial double-tracks and passenger stops at timetable crossing points are examples of measures that give geographically local effects with strong interactions, whereas inter-station distance, vehicle type and arrival punctuality are factors with a more general impact and weaker interactions.

1 Introduction
The railway is a complex traffic system. Factors such as infrastructure design, timetable and punctuality interact in a complicated way. The ongoing increase in utilisation of the railway system calls for a better understanding of its traffic properties. In order to display and describe some of the fundamental characteristics of single-track railway lines, a Simplified Analytical Model For Single-tracks, SAMFOST, has been developed.
This paper will describe some of the fundamental operation properties of single-track railway lines. A short background will be given, followed by a description of the SAMFOST model. The concept of timetable flexibility will be discussed and results from a factorial experiment will be presented at the end of the paper.

About 80% of Swedish railway lines are single-track. Single-track lines have lower capacity, lower average speeds and higher disturbance sensitivity than double-track lines. All these differences are due to train crossings where the two traffic directions become interdependent. The crossing time is the time that is needed to accomplish the crossing. The crossing time is thus the extra time that has to be added to the run time because of the crossing and it is dependent on such factors as infrastructure design, timetabled crossing point, vehicle parameters, driver behaviour, punctuality etc. The last two are more or less stochastic, which also makes the crossing time stochastic. The crossing time will therefore have a distribution.

This distribution directly influences traffic properties such as capacity, running times, secondary delays, timetable flexibility etc. Since the crossing time is such an important property of single-tracks it is natural to examine it carefully. In Lindfeldt (2006) and Lindfeldt (2007) it was shown how the crossing time is dependent on inter-station distances, length of station track (occurrence of partial double-tracks), primary delays etc.

The occurrence of crossings implies time separation of the services, which in turn opens up for powerful mathematical models to be constructed for single-track lines. SAMFOST is a simple and transparent mathematical model that can be used to increase understanding of single-track railway traffic. It handles both infrastructure parameters and timetable and punctuality. In SAMFOST, the analyses are performed in steps and the first two are independent of the timetable crossing point.

A major advantage of mathematical models is that several variables can be handled systematically at the same time. Interactions between different variables can be examined. A theoretical alternative to using a mathematical model like SAMFOST is to perform a large number of simulations. In practice this is not possible since today’s simulation tools are not suited to automatic variable change and so this method would be very time-consuming. It can therefore be argued that mathematical models can be very useful in many cases.

2 Related research

The research on railway operation, including scheduling, re-scheduling (disturbance management) and infrastructure planning is extensive. Scheduling and re-scheduling seem to be a better-covered area than infrastructure configuration and its influences on traffic operation and operation possibilities. This is in a way natural since the infrastructure is much more static and more difficult to change. Nevertheless, knowledge of infrastructure properties is essential when robust timetables are to be constructed.

Higgins et al (1997) proposed an optimization model to determine the required number and location of crossing stations. The crossing time is decomposed into
“conflict delay” and “risk of delays”, which correspond to the lowest crossing time at a station and the stochastic delay propagation part of the total crossing time respectively. The model has several similarities to SAMFOST but both the train runs and the delays are modelled in less detail.

Petersen (1974) presents an analytical model for mean delays due to crossings and overtakings as a function of traffic density. All departures are random, the model handles the trains pair-wise and the delays due to interferences are calculated roughly. Higgins et al (1996) propose a model to optimise the timetable on a single-track stretch of railway, disregarding initial delays. Chen and Harker (1990) propose a somewhat more developed model for the delay functions for scheduled trains. Here, too, the trains are treated in pairs. Sahin (1999) analyses dispatchers’ decision processes and emphasises that potential future conflicts have to be taken into consideration when resolving immediate conflicts.

Carey (1994), Yuan and Hansen (2007) and Wendler (2007) present another method, based on pure statistics. Interconnections between trains that use the same track section or platform track are described through convolution. This method is based on assumptions about independence, but still gives a clear picture of the “knock-on” effects in congested situations.

Törnquist and Persson (2007) and Törnquist (2007) present a heuristic model that can be used to re-schedule traffic according to different objective functions in complicated networks. The interactions between different trains are modelled on a fairly high level and only singular and extensive delays are considered. Mattsson (2007) provides a survey of relationships between capacity and delays. He also compares available methods for railway operation analyses.

The work presented in this paper is mostly related to the stochastic analyses performed by Carey (1994), Yuan and Hansen (2007) and Wendler (2007) based on the same assumptions about independent train runs. A major difference compared to most other research, however, is that both infrastructure and timetable appear as variables. It thus becomes possible to show how the infrastructure design influences the operation and feasible timetable strategies.

Most literature concerns one or just a few factors. Not even simple methods such as factorial design can be found in literature. This work therefore contributes to a deeper understanding, also as regards different types of interaction effects that are revealed by the factorial experiment.

The concept of timetable flexibility is rarely encountered in the literature. This seems very strange since one of the railways’ most important disadvantages is their lack of flexibility. When investments in new infrastructure are made, timetable flexibility ought to be one of the most interesting factors to take into account. It is well known that the demand for rail services varies and changes much faster than the infrastructure.

3
3 The model

The main purpose of SAMFOST is to examine the traffic properties of single-track railway lines. Some powerful assumptions make the model analytical and thereby distinguish it from different kinds of simulation models. These fundamental assumptions open up for a stepwise systematic analysis that makes the results general and understandable. SAMFOST is primarily constructed to handle analyses of fictive, non-existent railway lines where variables of interest can be examined in a simple manner.

3.1 Fundamental assumptions

In a crossing situation the crossing time is determined by a great many factors. The most important of these are infrastructure design, vehicle parameters, timetable crossing point, driver behaviour, adhesion, arrival delays, and congestion (influence of other trains). The first three can be regarded as deterministic and are easy to model. In many simulation models these parameters are more or less static input data. The last four factors are much more stochastic.

An analytical model can easily handle one stochastic variable. If more variables are to be stochastic the model will lose transparency and one comes closer to a simulation model where it is much more difficult to recognise the effect of a single variable. For this reason, only the arrival delays are modelled as stochastic variables in SAMFOST. Driver behaviour and adhesion are assumed to be deterministic.

Another important delimitation is that no other trains than the two crossing ones are modelled. All kinds of congestion situations, where more than two trains interact, are thus left out of the analysis. This assumption is very powerful since it makes analytical calculations possible, where arrival delays are the only stochastic variables. The crossing time is then unequivocally determined by the first three factors together with the delay difference, which is given by the arrival delays for the crossing trains. In a real case, with several interacting trains, a given value of the delay difference would give different crossing times due to the positions and delays of surrounding trains.

The assumption of independence of other trains is only valid when congestion effects in real traffic are small. For a given line the congestion-free capacity is strongly dependent on punctuality. SAMFOST’s precision is therefore dependent on punctuality as well as the frequency of the train services.

A second, less restricting, assumption is that crossing trains are independent before a crossing course starts. It has been shown empirically that this assumption holds. Following this assumption, the two stochastic variables up-going and down-going arrival delay can easily be combined into the very important stochastic variable arrival delay difference.
3.2 Structure

The two fundamental assumptions about crossing trains’ independence of each other and other traffic makes it possible to obtain a plain, simple model structure, see figure 1. The length of the modelled line is determined by arrival punctuality. In the first step, dimensioning delay distributions are chosen, preferably distributions with a high variance. The resulting distribution for arrival delay difference is calculated through (numerical) convolution. This distribution shows how the theoretical crossing point is distributed.

To determine the minimum length of the infrastructure model needed, an upper and a lower limit are taken from the delay difference distribution in such a way that a large proportion (99.5%) of the crossings fall within the limits. The chosen length is then filled with infrastructure that has the properties to be examined. Several variables are defined: speed limits on the line and station tracks, station parameters such as position, track configuration, track length, points length, speed restrictions at points, interlocking time etc. Double-tracks are modelled as long stations.

Vehicles are defined by parameters that are needed to model running times and signalling: acceleration curves (simplified), braking rate, maximum speed and vehicle length. Given infrastructure and vehicle data, the model calculates a crossing time function between the time limits for each traffic direction. The crossing time function shows the time needed to perform a crossing as a function of the arrival delay difference. The function, that is independent of timetable and punctuality, gives a character for the studied line. The crossing time function is presented in an aggregated way showing the sum of the crossing times of the two crossing trains. An example is shown in figure 2.

Figure 1: SAMFOST’s structure.
Figure 2: Crossing time function for a line with 9 crossing stations (15 km inter-station distances). The dotted curve is without passenger stop, the solid curve with passenger stop.

The figure shows two examples of crossing time functions. In both cases the inter-station distance is a constant 15 km, which induces the cyclic shape of the functions and determines the period of time between two adjacent maxima. Each low-level area represents a crossing station. When a passenger stop is introduced at the middle station, it becomes possible to accomplish a crossing with a very short crossing time.

In the final step the crossing time function is combined with the distribution for the arrival delay difference, thereby giving the distribution for the crossing time. One crossing time function may be combined with several different distributions for arrival delay difference. This is possible as long as at least the prescribed part of the crossing situations falls between the limits derived from the dimensioning distributions.

SAMFOST has been successfully validated against the accepted simulation tool RailSys. The validation shows that the model has an accepted accuracy as long as the crossing time depends only on the two crossing trains (non-congested situations); see Lindfeldt (2007) for details.

4 Timetable variants and timetable flexibility

In figure 3 the timetabled crossing point is located at the middle station, corresponding to a delay difference equal to zero. As already mentioned, the delay difference is stochastic, depending on the arrival delays of up-going and down-going trains respectively. The actual outcome of a timetable will therefore depend both on the crossing time function and on the location of the timetabled crossing point and the distribution of the arrival delay difference.
In figure 3 the timetabled crossing point is still chosen to be the middle crossing station, indicated by the maximum density of arrival delay difference for the crossing trains. The arrival delay for each traffic direction here is \( \text{Exp}(150) \)-distributed, which on Swedish railways corresponds to a line with relatively high punctuality (mean delay 150 seconds).

### 4.1 Timetables

It is natural to locate the timetable crossing point at a station, such as figure 3 shows, since it gives a minimum crossing time. In practice it is not always possible, or desirable, to choose the point that results in the lowest crossing time. Different kinds of timetable dependencies often make it impossible to obtain the optimal timetable crossing point for every crossing. It is therefore of great interest to examine how the crossing time is influenced by the location of the timetable crossing point. The *timetable crossing point* is here defined as the (theoretical) point where the crossing would take place at an arrival difference equal to zero if the line was double-track.

In SAMFOST, the timetable crossing point can be varied. To do this the location of the distribution of the arrival delay difference is moved stepwise along the crossing time function. Each step corresponds to a timetable crossing point located where the arrival delay difference is zero. Figure 4 shows an example.
Figure 4: An evaluation of different timetable crossing points means that the probability function is moved stepwise from one end of the line to the other.

Figure 5: Mean crossing time function (solid) and standard deviation (dashed) curve for different timetable crossing points. Arrival punctuality corresponds to $\text{Exp}(150)$-distributions.

In each step, the product of the probability density function (dashed and dotted) and the crossing time function (solid) is integrated to give the mean crossing time. The standard deviation is also calculated in each step. The resulting mean and standard deviation functions are shown in figure 5. The upper curve represents the mean crossing time for different timetable crossing points. Far out from the middle the mean curve is completely cyclic since the impact of the differing station in the middle
is very small. The frequency depends on the inter-station distance while the amplitude corresponds to the punctuality as will be shown later. Note that the standard deviation curve (dotted line) is also cyclic with min-values for the same timetable variants as the mean-curve. This is a valuable fact since it makes it easy to choose timetable point in regions with cyclic mean and standard deviation functions.

Closer to the middle of figure 5 the mean crossing time falls to an absolute minimum. The shortest mean crossing time is thus achieved if the timetable crossing point is chosen to be the station where the trains stop for passenger exchange. This is natural and only reflects the shape of the crossing time function. However, the situation is actually much more complicated since the standard deviation increases dramatically when the timetable crossing point is located at the station where the trains have an ordinary stop. The combination of timetable crossing point and passenger stop thus results in a much more disturbance-sensitive operation and calls for more timetable supplements in order to keep the overall punctuality at an acceptable level.

The observant reader will also note that the curves are left-right-asymmetrical. This is due to the fact that the stop position for the passenger stop is located at one end (left) of the middle station. Since each station has a track length of 750 m the stop position influences the crossing time as the tolerance is different in different directions.

In the example shown in figure 5 the underlying punctuality corresponds to \( \text{Exp}(150) \)-distributions for both traffic directions. The punctuality affects the amplitude of the mean-function; the higher the punctuality, the higher the amplitude. When punctuality is complete and all trains run on time the mean-function coincides with the crossing time function!

![Figure 6: Mean crossing time functions (solid) and standard deviations (dashed and dotted) for two different levels of arrival punctuality. Greater punctuality results in higher amplitude both in mean and standard deviation.](image-url)
Figure 6 shows mean crossing time functions (solid) and corresponding standard deviation functions (dashed and dotted) for two different levels of punctuality. This time the analysed line consists only of crossing stations without a passenger stop, which makes all curves completely cyclic. The higher punctuality, Exp(150)-distributed arrival delays, gives a much higher amplitude of the mean-function while a lower punctuality, Exp(350)-distributed arrival delays, decreases the amplitude and thereby the difference between different timetable crossing points. The standard deviation is affected less than the mean. A cyclic function (dotted) with distinct maxima and minima becomes a function that does not vary very much when punctuality falls (dashed).

An interesting result of this is that the choice of timetable crossing point is much more important when punctuality is high. In those cases the gain from a correctly located crossing point is higher and the tolerances lower. In figure 6 the length of the analysed lines differs. This is because the line length is adjusted according to the spread in arrival delay difference. Note that the two curves have the same frequency, which indicates that the inter-station distance is the same and constant (giving the cyclic form).

In the presented examples the delay situation is symmetric; both traffic directions follow the same distribution. The method becomes even more useful in asymmetric cases where the traffic directions follow different distributions. An example is given in Lindfeldt (2007).

Figure 7 shows what happens when the inter-station distance is decreased.

Figure 7: Different inter-station distances result in different frequencies, amplitudes and overall levels of the mean time function (solid) as well as of the standard deviation curve (dashed). Here the inter-station distance varies from 15 km (upper curves) to 3 km (lower curves) with a step length of 2 km.
As before the upper group of curves (solid) represent the mean crossing time while the lower group show standard deviation of crossing time. The curves come from 7 different infrastructure designs, i.e. 7 different inter-station distances from 15 km (upper curve) to 3 km (lowest curve). In all cases the punctuality corresponds to an Exp(150)-distribution in both traffic directions. Each infrastructure variant has its own crossing time function, which is reflected in the mean functions shown in the figure.

A decrease in inter-station distance affects the mean crossing time function in three different ways:

- **Higher frequency.** Shorter distance between adjacent minima.
- **Lower amplitude.** Although the minimum crossing time is equal for all stations (they are identical), the maximum crossing time, i.e. worst case when one train has to wait a maximum time, decreases when inter-station distance is decreased.
- **Lower mean-level.** Decreasing the inter-station distance implies that the probability of the minimum crossing time increases. The mean-level of the mean-functions therefore also decreases.

All three changes can be seen when comparing the mean crossing time functions shown in figure 7. For all seven variants the stations have the same minimum level characteristics in the crossing time function as the stations without passenger stop in figure 2. Since every station has the same characteristic independent of the inter-station distance the mean crossing time functions converge to this minimum level of about 140 s. This level is determined by the infrastructure design of the stations, interlocking system and vehicle performance.

The mean crossing time function and/or standard deviation function can be used to compare different alternatives when choosing timetable crossing point. As will be shown in the next section such choices are fundamental when trying to describe and estimate timetable flexibility.

### 4.2 Choice criteria for acceptable timetables

The previous section showed that different points along a line work differently when used as timetable crossing point. This is a fundamental property of single-track railway lines. In order to evaluate timetable flexibility the available points have to be compared to some sort of choice criteria.

Two requirements for a timetable crossing point are quite natural:

- As low a mean crossing time as possible
- As low a crossing time variance as possible

Following these criteria the crossings become as time efficient as possible and the disturbance sensitivity is kept low. For lines that have equal inter-station distances and crossing stations with the same properties (regarding passenger stop, track length, infrastructure design etc.), both the mean crossing time function and the
standard deviation function are entirely cyclic, varying at the same phase (see figure 8). In these cases it is fairly easy to choose the accepted timetable crossing points.

Figure 8: Mean crossing time function (bold solid) and corresponding standard deviation curve (dotted). A tolerance line (thin solid) marks the accepted timetable crossing points located below it.

Since minimum mean and minimum standard deviation occur for the same locations of timetable crossing point a basic rule of choice is easily applied. By drawing a horizontal line at a tolerance distance above the minima of the mean function the accepted timetable crossing points are chosen. All timetable crossing points that have a crossing time below the line are accepted, regarded as useful and thereby contributing to the timetable flexibility. These points may be the subject of different kinds of flexibility evaluation, as proposed in the next section.

However, real railway lines seldom have all the symmetries that give synchronised mean and standard deviation functions. Station properties differ. For instance stations with and without passenger stop are often mixed on the same line and infrastructure parameters such as length of station-tracks vary. Figure 9 shows a simple example.

In this case a line with 11 stations is evaluated. At the middle station the trains stop for passenger exchange, which implies a dramatic drop in mean crossing time (solid curve) for timetables using that station as a timetable crossing point. As the figure shows, the passenger stop influences the mean crossing time and its standard deviation over a long time interval. Following this influence the line can be divided into five segments:
Figure 9: A passenger stop results in a strong, local decrease in mean crossing time (solid). At the same time the standard deviation (dashed) increases strongly. In order to determine accepted timetable crossing points the line may be divided into five sections.

A One section on each side that is located at such a distance from the passenger stop that the influence is negligible.

B One section on each side of the passenger stop within which both the mean crossing time and the standard deviation change rapidly with the distance to the passenger stop.

C The crossing station where the trains stop for passenger exchange.

In order to identify the accepted timetable crossing points, the first question is whether to accept a mean crossing time that is higher than the absolute minimum for the line as a whole. If the answer is no, only timetable crossing points in section C can be accepted and the basic rule of choice described above is still applicable (using a tolerance level of about 90 seconds).

In most practical cases one has to accept timetable crossing points other than those giving the lowest mean crossing times. The method of choice is then not that simple.

For sections of type A the basic rule may be applied, see the tolerance line in figure 9. Finding a simple rule of choice for points within sections of type B is not that easy. Even though the mean function lies under the tolerance line for A sections, the standard deviation is substantially higher for B sections. One way would be to take the standard deviation into account when choosing the accepted timetable crossing points. Doing so might disqualify points corresponding to the slopes adjacent to section C. Such a method is nonetheless probably a good way to translate the influence of a passenger stop into a change in timetable flexibility.
For section C it is possible to apply the basic rule of choice, although one has to keep in mind that the high standard deviation level will require a greater time supplement in the timetable in order to maintain overall punctuality after the crossing.

In this first study of timetable flexibility the choice rules have been kept very simple. The basic rule has been applied over the whole line. This means that the crossing station having the highest mean crossing time is used as the norm station for the tolerance line. Thus, all points along the line that have a lower mean crossing time become accepted timetable crossing points. This will probably overestimate the effect of a passenger stop or other properties that decrease the crossing time locally. This simple rule nonetheless works as a first attempt to describe timetable flexibility. An adjustment of the rule in order to take the variance into account is possible in future work.

Another problem that is also left to later research is the influences of different arrival punctuality. As shown in figure 6 arrival punctuality also influences the mean and standard deviation functions. Two principle different strategies are possible:

- **The tolerance line is absolute**, independent of the punctuality. This method clearly visualises that higher punctuality is advantageous as it gives more accepted timetable crossing points. For really low punctuality no timetable crossing points would be accepted. This would imply that punctuality is too low for any crossing to be allowed according to the set tolerance level!

- **The tolerance line is allowed to change with punctuality**. In this case the tolerance line is allowed to “float” at a specified distance above a minimum of mean crossing time that results from a given punctuality.

The two methods work differently. The first method is a much more offensive approach clearly showing the importance of high punctuality, while the second method will actually give more accepted timetable crossing points at low punctuality, at a higher tolerance level, i.e. give more time-consuming crossings.

### 4.3 Measures for timetable flexibility

After choosing accepted timetable crossing points according to some distinct rule the properties of the chosen group may be examined. Studying figure 9, for instance, at least three different measures concerning the group of accepted timetable crossing points can be seen:

- Share of accepted points
- Spread in position (horizontal axis) – Position variance for accepted timetable crossing points
- Spread in crossing time (vertical axis) – Crossing time variance for accepted timetable crossing points

Assuming that all accepted timetable crossing points have the same utility the simple timetable flexibility can be calculated as the share of accepted timetable crossing points:
Simple flexibility tells us the share of the line that is accepted for timetable crossing. If all points are accepted the simple flexibility is total and equal to 1. A double-track line is a good example of a system that has total simple flexibility. Depending on tolerance level and punctuality, single-track lines have a simple flexibility between 0 and 1.

In order to calculate simple flexibility the total number of points must be defined. It is appropriate to do this by first assigning a frequency of train services and then examining all alternative crossing points that are possible for that frequency. The time interval for crossings is half the period time for the services, so for one-train-an-hour traffic a 30 minute time interval has to be examined.

The simple flexibility tells us how great a share of a line can be accepted (and used) as a timetable crossing point. It does not, however, say anything about how the accepted points are distributed geographically over the line. Depending on the design of adjacent infrastructure and the traffic situation on adjacent parts of the railway system, it is sometimes advantageous either to have the accepted timetable crossing points concentrated near each other or more spread over the line. It is therefore natural to introduce a second flexibility measure:

\[ \text{Position spread of accepted timetable crossing points} = \sqrt{\frac{\sum (t_{i_p} - \mu_p)^2}{n-1}} = \sqrt{\frac{\sum t_{i_p}^2}{n-1}} \]

\( t_{i_p} \): time location for accepted timetable crossing point \( i \)  
\( \mu_p \): mean of time location  
\( n \): number of accepted timetable crossing points

Since all the infrastructure designs that are examined in this work are symmetric, or almost symmetric, \( \mu_p \) is 0 and the equation is simplified to the right hand expression. A high value of this flexibility measure implies that the accepted timetable crossing points are spread whereas a low value indicates a higher concentration of accepted points.

The accepted points may imply different mean crossing time, i.e. they may be located at different vertical distances to the tolerance line. This calls for a third measure of flexibility that covers the spread of crossing time:
Together these flexibility measures show the degrees of freedom that are available in the timetable construction.

5 Factorial experiment

In the previous sections the crossing time was proposed as one of the most important operation properties of single-track lines. Both infrastructure design such as inter-station distances and properties of crossing stations and operative factors like vehicle parameters, occurrence of passenger stops and arrival punctuality strongly influence the crossing time.

A common way to analyse railway operation properties is to examine only one factor at a time, keeping all other factors constant. Using this method it is impossible to find important interaction effects between interesting factors. When two factors interact it is impossible to know anything about the effect of one factor without simultaneous knowledge of the interacting factor’s state.

Experience from railway operation tells us that interactions are common. To show this, six different factors were examined in a $2^6$ factorial experiment. In such an experiment two values (levels) for each factor are chosen. The experiment is then run for all combinations of factor levels, in this case 64 combinations. A major advantage of factorial experiments, besides the interaction effects, is that the analysis becomes more general. General results are valuable since the operation condition changes over the lifetime of infrastructure investments.

In the proposed kind of experiment linearity is assumed. This is of course a simplification, but within a limited area (between the levels) this assumption is most often accepted.

5.1 Factors, levels and response variables

In SAMFOST variables are assigned before each run. Every variable may thus take different values in different runs. A complete $2^n$-experiment is therefore easy to perform. From earlier experience six factors were found to be of special interest:

Crossing time spread of accepted timetable crossing points $= \sqrt{\frac{\sum(t_{ni} - \mu_n)^2}{n-1}}$

$ t_{ni} :$ crossing time (mean) for accepted timetable point $i$

$ \mu_n :$ mean of crossing time (mean) in accepted timetable crossing points

$n :$ number of accepted timetable crossing points

\[(3)\]
Table 1: Factors and level-values for the 2⁵-experiment.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Description</th>
<th>Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Track length at station / on double-track where timetable crossing point is located</td>
<td>0.75 km, 10 km</td>
</tr>
<tr>
<td>B</td>
<td>Inter-station distance</td>
<td>15 km, 7.5 km</td>
</tr>
<tr>
<td>C</td>
<td>Passenger stop</td>
<td>No stop, 60 s</td>
</tr>
<tr>
<td>D</td>
<td>Speed restriction at points at station where timetable crossing point is located</td>
<td>100 km/h, 160 km/h</td>
</tr>
<tr>
<td>E</td>
<td>Vehicle type</td>
<td>X2, X50</td>
</tr>
<tr>
<td>F</td>
<td>Arrival punctuality</td>
<td>200 s, 100 s</td>
</tr>
</tbody>
</table>

For all factors the minus-level was chosen to give a longer crossing time than the plus-level.

**Factor A** is the track length at the station or on the partial double-track where the timetable crossing point is located. The minus-level, 0.75 km, corresponds to an ordinary crossing station whereas the plus-level, 10 km, corresponds to a partial double-track of intermediate length.

**Factor B** is the inter-station distance. At the minus-level it is set to 15 km (a long inter-station distance) and at the plus-level 7.5 km (a short distance compared to practice). These levels were chosen to obtain the same total line length and the same positions for the farthest out located stations, and to fit the plus-level of factor A. Figure 10 shows the four combinations of factors A and B. Factor B controls the distance between stations surrounding the middle station/partial double-track, whereas factors A and B together control the single-tracked length on each side of the middle station/partial double-track.

![Figure 10: The four combinations of factor A (double-track length at middle station) and factor B (inter-station distance).](image-url)
**Factor C** is a stop for passenger exchange. Minus-level implies that the trains do not stop for passenger exchange at any station on the modelled line. Plus-level implies that the trains stop at the mid-station at 60 s for passenger exchange.

**Factor D** is the speed restriction at the points at the middle station. The minus-level is 100 km/h and the plus-level is 160 km/h. Note that the speed restriction at all other stations is 100 km/h independent of the level of factor D.

**Factor E** is the vehicle type (same in both directions). At the minus-level the Swedish long distance train X2 is used and at plus-level the Swedish regional train X50 with faster acceleration values is used. Both vehicles have the same top speed of 200 km/h.

**Factor F** is arrival punctuality. At minus-level arrival punctuality (in both directions) corresponds to Exp(200)-distributions, i.e. mean arrival delay is 200 seconds. At plus-level arrival punctuality corresponds to Exp(100)-distributions.

The effect of these six factors was examined with regard to the optimal timetable crossing point as well as the possibility to construct other timetables. Altogether five response variables were used:

**Centred timetable**, i.e. the crossing centred at the mid-station:

- Mean crossing time
- Standard deviation for the crossing time

**Possibility to change the timetable:**

- Simple flexibility
- Position spread of accepted timetable crossing points
- Crossing time spread of accepted timetable crossing points

The last three measures are all calculated according to the definitions in the previous section.

The complete design matrix is shown in appendix 1. Each row corresponds to one experimental condition, i.e. a run with a certain combination of levels of the factors. SAMFOST has calculated the values of the five response variables for each row. Between the runs the settings were updated according to the table of signs. Finally, after all SAMFOST calculations, the design matrix was used to calculate the six main effects and all 57 interactions.

Together, the first five factors form 32 different combinations of infrastructure designs and vehicle properties generate the same number of crossing time functions. These crossing time functions are then combined with two different distributions for arrival delay difference, corresponding to the levels of factor F, which results in 64 different values of the five response variables. For detailed information about these calculations the reader is referred to literature on factorial design.
5.2 Results – centred timetable

Two response variables reflect the properties when the timetable crossing point is located at the crossing station or partial double-track in the middle of the modelled line: mean crossing time and standard deviation for the crossing time. In the following sections the main effects and the most important two-way interactions are shown.

Each main effect is defined as half the mean change, relative to the total mean for all 64 experiments, in the response variable when the factor goes from the lower level to the higher level.

Each two-way interaction effect is calculated as a difference between the effect of changing the first factor from low to high level when the second factor is low and the effect of changing the first factor from low to high level when the second factor is high. This figure is finally standardised by division.

A simple example, consisting of the two factors T and K results in the following model.

\[ \theta_{ij} = \mu(\pm) \bar{T}(\pm) \bar{K}(\pm)(\pm) TK \] (4)

In this example \( \theta \) denotes the response variable, \( i \) and \( j \) are indices representing the levels (– or +) for the factor T and K respectively and \( \mu \) is the total mean value.

Depending on the levels the main effects \( \bar{T} \) and \( \bar{K} \) are added or subtracted from the total mean value. The effect of the two way interaction \( \bar{TK} \) depends on both levels, i.e. same level results in addition and different levels results in subtraction.

The corresponding equation for the six-factor case includes a \( \mu \) (total mean), six main effects (A-F) and 57 interaction effects. Altogether the expression consists of 64 different terms. In the following the estimated values of these 64 parameters are omitted. Instead, only the main effects are shown together with the most important interaction effects.

5.2.1 Mean crossing time

In figure 11 the main effects, i.e. those caused by a change from low to high level in each factor, are shown by dark bars. Light bars are interaction effects. The effects are sorted according to size and shown absolute value.

The most important factor is the passenger stop (C). If a timetabled crossing is combined with a passenger stop, the crossing time falls by approximately 90 seconds. This is a mean value and the actual value is dependent on interactions with other factors. If no interactions exist, the change is independent of the other factors. In
such cases the shown change in crossing time will be achieved as soon as a passenger stop is introduced. If, on the other hand, interactions are present, the actual effect of a passenger stop will depend on the interaction! Figure 11 clearly shows that the passenger stop factor (C) interacts with both the length factor (AC interaction), vehicle type (CE interaction) and arrival punctuality (CF interaction).

Using only figure 11 it is therefore difficult to draw any general conclusions about the effect of a passenger stop. One important result, however, is in fact that a passenger stop has a major impact on the crossing time, that this effect is dependent on three other factors and the corresponding two way interactions should be examined in more detail.

The double-track length (A) also has a major impact on the crossing time. This effect, however, is strongly dependent on the occurrence of passenger stop (AC), the speed restriction at the points (AD) and the inter-station distance (AB).

Hereafter follow arrival punctuality (F), vehicle type (E), speed restrictions at points (D) and the least important factor, inter-station distance (B). All these factors show interactions with other factors. Inter-station distance (B) and arrival punctuality (F) interact least.

The speed restriction at points (D) affects the crossing time surprisingly much. The main effect (D), however, is only slightly higher than the interaction effect with double-track length (AD). This indicates that a higher speed at the station points is only of interest when combined with a partial double-track (A at a high level), which will be shown below.
Most interaction effects (44 in total) are zero or almost zero and have been omitted in figure 11.

Figure 12: Two-way interaction effects on mean crossing time, AC-interaction (left) and AD-interaction (right).

Figure 12 shows the AC- and AD-interaction respectively. The first factor is displayed on the horizontal axis whereas the second factor is represented by two lines: low level (dotted) and high level (solid). The interaction is recognised directly by the fact that the two lines are not parallel. The higher the difference in inclination, the stronger the interaction.

The interaction between double-track length (A) and passenger stop (C) is strong. Figure 12 clearly shows that the effect of a 10 km long partial double-track is highest when it is located at a place where the trains do not already stop for passenger exchange. It is also obvious that passenger stop as a single measure gives a lower crossing time than a partial double-track within the examined interval of punctuality (100 – 200 seconds mean arrival delay).

Figure 12 also shows the already mentioned interaction between double-track length (A) and speed restriction at points (D). Note that when the double-track is only an ordinary crossing station (0.75 km) there is almost no difference between different point speeds. Note also that the lines start at approximately 135 seconds, which is a mean of the starting values in the AC-interaction figure. When calculating the AD-interaction, the passenger stop factor (C) was therefore at an intermediate level.

A large difference in inclination implies that a high speed is essential at entrance and exit points on partial double-tracks. Much of the desirable effect of partial double-tracks is thus dependent on a high speed at these points.
5.2.2 Standard deviation for crossing time

The six factors also affect the crossing time variance in different ways. Figure 13 shows this, dark bars corresponding to main effects and light bars to interaction effects. As before, one has to be aware of the interaction effects, which make it difficult to say anything about the main effects separately.

![Figure 13: Main (dark bars) and interaction (light bars) effects on standard deviation of crossing time, sorted by absolute value (some effects are negative).](image)

The inter-station distance (B) is the most important single factor. A shorter inter-station distance gives a lower crossing time variance. This result is quite natural since the inter-station distance is a factor that affects the crossing time function generally and not only locally decreases the crossing time for situations with a low arrival delay difference. Note that the effect is dependent on the double-track length (AB interaction) and the arrival punctuality (BF interaction) and so these factors must be taken into consideration when evaluating the inter-station distance factor.

Arrival punctuality (F) is the second most important main effect. Greater punctuality gives a lower crossing time variance. This factor also works generally over the whole line. A passenger stop (C) also affects the crossing time variance. When a passenger stop is introduced, the variance increases dramatically. This is due to an increase in difference between the highest and lowest possible crossing time that occurs when very time efficient crossings are made possible for small arrival delay differences. This increase may be eliminated if a passenger stop (C) is combined with a partial double-track (A), due to a strong interaction (AC).

The speed restriction at the points (D) also affects the crossing time variance. A higher speed implies a higher variance since more time efficient crossings are made possible, but only for some values on the arrival delay difference. The interaction
between speed restriction (D) and double-track length (A) is strong. Without a double-track the point speed (on these levels) is not very interesting. The vehicle type (E) is the single factor that affects the crossing time variance least.

Figure 14 shows the four most important two way interactions:

- Double-track length – passenger stop (AC)
- Double-track length – speed restriction at points (AD)
- Double-track length – inter-station distance (AB)
- Passenger stop – arrival punctuality (CF)

The double-track length (A) is the most important interaction factor, which clearly shows that partial double-tracks have to be constructed with care. Moreover, the three interactions work in different ways, making it even more difficult to draw general conclusions about partial double-tracks.

The interaction between double-track length (A) and passenger stop (C) is strong. A partial double-track implies a large decrease in variance when it is combined with a passenger stop. Without a passenger stop a partial double-track implies an increase in the crossing time variance. This fact is not intuitive, but follows from the fact that both partial double-tracks and passenger stops open for very time efficient crossings.
in situations with a small arrival delay difference. The difference in crossing time between different arrival delay situations thereby increases, which in turn implies a higher variance. The figure shows how a partial double-track reduces most of the variance that is induced by a passenger stop (solid line).

The double-track length (A) also interacts with the speed restriction in the points (D). The figure shows that the speed restriction does not affect the crossing time variance at ordinary crossing stations very much (lines coincide when A is on its low level). When a partial double-track is introduced a low permitted speed at the points implies a higher decrease in crossing time variance than a high permitted speed. A low permitted speed at the points tends to reduce disturbance sensitivity since each train has a longer timetabled time for passing the partial double-track.

The double-track length (A) also shows an important interaction with the inter-station distance (B). The effect of a partial double-track is greater if the inter-station distance is long (on lower level).

A passenger stop (in the way it is treated in this study) implies a local change of the crossing time function making it possible to achieve very time efficient crossings when the arrival delay difference is low (the local effect). The introduction of a passenger stop therefore increases the crossing time variance. As shown in figure 14, the increase depends on arrival punctuality (F). High punctuality implies a higher utilization of the crossing station where the crossing time is low, while a lower punctuality spreads the crossings over the adjacent parts of the crossing time function where crossing times are higher.

5.3 Results – timetable flexibility

In the previous part of the factorial experiment the timetable crossing point has been fixed at a crossing station or a partial double-track. This location implies that only the timetable crossing point that gives the absolute minimum mean crossing time has been analysed. The six studied factors also affect the possibility to use other timetable crossing points. The factorial experiment was therefore also carried out using the three measures of timetable flexibility as response variables.

5.3.1 Simple flexibility

Simple flexibility is a general measure of timetable flexibility. The measure is based on the mean crossing time function and all points that give a lower mean crossing time than a tolerance level are accepted as timetable crossing points. In this example the tolerance level is defined as five seconds over the local minimum that has the maximum mean crossing time (see figure 9 for an example). All points that lie below the tolerance level are accepted, regardless of station design, passenger stop and partial double-tracks, which often implies much lower mean crossing times for some timetable crossing points.

Figure 15 shows how simple timetable flexibility is affected by the factors and their interactions. The three most important factors, inter-station distance (B), double-
track length (A) and occurrence of passenger stop (C), all show important interactions, whereas arrival punctuality (F) and vehicle type (E) lack interactions.

The inter-station distance (B) is the predominant factor. A halving of the inter-station distance from 15 km to 7.5 km means a strong increase in simple flexibility. A partial double-track (A) also increases flexibility and so does a passenger stop (C). In both cases, flexibility consists of a few points with really low mean crossing times and adjacent points with a higher mean crossing time (and variance), similar to figure 9.

Figure 15: Main (dark bars) and interaction (light bars) effects on simple flexibility, sorted by absolute value (some effects are negative).

Increased punctuality (F) has the opposite effect, since the crossing time is more sensitive to the location of the timetable crossing point when operation precision is higher. For the same reason, a higher vehicle performance (E) results in a decrease in simple flexibility.

Figure 16 shows the two most important interactions where the inter-station distance (B) interacts with the double-track length (A) and passenger stop (C) respectively. Note that a short inter-station distance (B at a high level) always results in high flexibility (95 %), regardless of the other factors. This can also be seen in figure 7, where the third mean-curve (solid) from the bottom has low amplitude.

The left part of figure 16 shows that a partial double-track induces many more new accepted timetable crossing points when inter-station distance is long (low level). This is natural since a short inter-station distance does not leave very many unaccepted timetable crossing points. The right part of figure 16 shows the same picture. A passenger stop induces more accepted timetable crossing points when inter-station distance is long (low level).
5.3.2 Position spread of accepted timetable crossing points

The accepted timetable crossing points have different positions along the studied line. In the scheduling work it is in many cases a great advantage to have a high spread of useful timetable crossing points. The factorial experiment was therefore also performed with position spread of accepted timetable crossing points as response variable.

![Graph showing position spread of accepted timetable crossing points](image)

**Figure 16:** Two-way interaction effects on simple flexibility, AB-interaction (left) and BC-interaction (right).

![Graph showing position spread of accepted timetable crossing points](image)

**Figure 17:** Main (dark bars) and interaction (light bars) effects on position spread of accepted timetable crossing points, sorted by absolute value (some effects are negative).
Figure 17 shows how the factors and their interactions affect the position standard deviation of accepted timetable crossing points. Inter-station distance (B) is the most important factor. Shortening the inter-station distance is the most efficient way to obtain a higher position variance. This is natural since the inter-station distance factor works over the whole line, whereas the other factors (except for the vehicle type) only work on a limited section in the middle of the line.

The great impact of arrival punctuality (F) is not intuitive. Greater punctuality implies a lower position variance. On and around crossing stations and partial double-tracks the accepted crossing points are located in groups (intervals). When punctuality improves, the size of these groups (the length of the intervals) decreases, thereby also lowering the position variance. The same phenomenon implies that a better vehicle type (E) also gives a lower position variance.

The effect of a passenger stop (C) is difficult to generalise since two interactions (BC and AC) are stronger than the main effect. The same holds for a partial double-track (A). The speed restriction at the points (D) influences the position variance very little.

Figure 18 shows that a shorter inter-station distance (B) gives a higher position variance and that the increase is larger if the trains have no passenger stop. This is due to the fact that the new accepted timetable crossing points that are generated by a passenger stop (as modelled in this example) are located close to the middle of the line. This results in a low contribution to the position variance.

Similarly, a partial double-track (A) creates a concentration of accepted timetable crossing points, which implies a decrease in position variance if the trains do not stop for passenger exchange (C). If a passenger stop is introduced, the position variance increases since the most distant accepted timetable crossing points contribute to an increase in position variance.

Figure 18: Two-way interaction effects position on spread of accepted timetable crossing points, BC-interaction (left) and AC-interaction (right).
5.3.3 Crossing time spread of accepted timetable crossing points

Different accepted timetable crossing points may have different mean crossing times. For instance, the mean crossing time at a station where the trains have a passenger stop or on a partial double-track is lower than at an ordinary station without a passenger stop. It is therefore interesting to determine how the factors influence the mean crossing time for the accepted timetable crossing points. This is measured by the standard deviation of mean crossing time for accepted timetable crossing points.

Figure 19: Main (dark bars) and interaction (light bars) effects on crossing time spread of accepted timetable crossing points, sorted by absolute value (some effects are negative).

A low deviation means that the accepted timetable crossing points give approximately the same mean crossing time, whereas a high value indicates that the accepted timetable crossing points differ in mean crossing time.

Figure 19 shows that a passenger stop (C) and a partial double-track (A) induce differences between accepted timetable crossing points. This reflects the fact that these factors imply the creation of a few points with a much lower mean crossing time than can be achieved with ordinary crossing stations without a passenger stop.

If punctuality (F) increases the differences in mean crossing time also increase. This follows from figure 6 where higher punctuality is shown to give sharper minima (where the accepted timetable crossing points are located). A higher speed at the points (D) also results in greater differences in mean crossing time, since some of the accepted timetable crossing points then have a lower mean crossing time.

The inter-station distance (B) and the vehicle type (E) influence the crossing time variance in the opposite way. A shorter inter-station distance reduces the differences
in mean crossing time. This follows from the fact that a shorter inter-station distance implies that the accepted timetable crossing point having the highest mean crossing time falls to a lower level. This is clearly seen in figure 7, where a shorter inter-station distance gives a curve at a lower level. This results in a lower spread, especially when the trains have a passenger stop. A relatively strong BC-interaction supports this.

The double-track length (A) and the passenger stop (C) interact in the same way as for the position variance. If the trains do not stop for passenger exchange a partial double-track means a substantial increase in crossing time variance since a small number of points with a really low mean crossing time are created. Similarly, a passenger stop alone implies a pronounced increase in crossing time variance. When the two measures are combined the interval with a low crossing time is extended further. This extension does not, however, contribute to the same extent as one measure alone.

Note that the crossing time variance is very low in cases where the line has neither a passenger stop nor a partial double-track. In such cases, all accepted timetable crossing points have roughly the same properties.

The effect of a partial double-track (A) is also dependent on the speed restriction at the points (D). A high speed in the points permits timetable crossing points with a lower crossing time, which in turn implies a higher crossing time variance. Note that the point speed, as before, is not important when the points are located at an ordinary crossing station with 750 m track length.
6 Conclusions

Operation of single-track railways is severely constrained by crossings that can only take place at special locations (crossing stations) along the line. Crossings usually imply that at least one of the two trains loses time compared to a crossing-free situation. To achieve a minimum crossing time the operation must have very high precision (punctuality). As soon as one of two crossing trains is more delayed than the other, the resulting crossing time increases through delay propagation (could be handled by time supplements which increase the scheduled crossing time). The stochastic factor delay difference is therefore very important.

The location of the timetable crossing point influences the crossing time (distribution, mean and variance) since different locations mean that the distribution of the delay difference is positioned differently relative to the crossing time function (shown in figure 4). The resulting mean crossing time function, that shows the mean crossing time for different locations of the timetable crossing point, is of great importance where timetable flexibility is concerned.

The idea of timetable flexibility is to describe the possibilities for future timetable changes. Such changes could be induced by demand changes on the studied part of the rail network and/or adjacent lines as well as major changes in adjacent infrastructure that call for timetable adjustments. The concept of timetable flexibility may also be extended to describe capacity and the dependencies between capacity and punctuality. This extension, however, is not treated in this paper.

Using the mean crossing time function timetable flexibility may be defined in a very straightforward way. This is advantageous since both the properties of the infrastructure and vehicles and punctuality affect the mean crossing time function. Thus, the local influences of passenger stops and partial double-tracks are modelled in a reasonable way. However, the simplest form of timetable flexibility calculations, using a constant tolerance level, as presented in this paper, is probably too simple to be more than a first step in the development of general measures. A more complicated model that also takes into account the crossing time variance could be an alternative.

Timetable flexibility, as treated in this paper, requires analysis of a group of points (locations) along a line that are chosen to be possible timetable crossing points. This group may be characterised by different measures. The basic tolerance method was found to give three different measures that complement each other. Since both geographical location and crossing time differ between available timetable crossing points, it is necessary to measure spread in location as well as spread in crossing time for the group of accepted timetable crossing points.

The factorial experiment gives a great many interesting results. It becomes obvious that the factors cannot be analysed separately. As in previous analyses punctuality is explicitly shown to be of great importance on single-track lines. One conclusion is that punctuality ought to be taken into account during scheduling when time supplements are to be assigned.
The six factors have different properties and may be divided into two groups according to how they affect the mean crossing time function:

- **Factors with a local effect:** Occurrence of partial double-track and passenger stop.
- **Factors with a more general effect:** Inter-station distance, vehicle type and punctuality.

Partial double-tracks and passenger stops are factors that give a very low crossing time for a limited interval of delay difference. This property results in higher variances in crossing time and higher sensitivity to punctuality. The local effect of each factor is extensive but when a partial double-track is combined with a passenger stop the additional gain is limited which shows an important interaction between these factors.

Partial double-tracks and passenger stops also affect timetable flexibility in a special way. Due to their great (local) impact on the mean time crossing function, they generate new accepted timetable crossing points that are located in a continuous interval. Simple flexibility, position spread, and crossing time spread all hereby increase when a partial double-track or a passenger stop is introduced.

Inter-station distance, vehicle type and punctuality work on the whole line length. In particular, the inter-station distance is a factor of great importance that strongly affects the crossing time variance, while the mean crossing time is not affected so much. The inter-station distance also affects the simple flexibility and the location spread of accepted timetable crossing points (i.e. capacity).

The vehicle type has a surprisingly weak influence on the evaluated parameters. Especially when the timetable crossing point is combined with a passenger stop, the vehicle becomes unimportant since the vehicle characteristics are taken into account in the timetable.

The factorial experiment also shows the importance of punctuality. It becomes obvious that high punctuality leads to more time efficient crossings. The crossing time variance is the response variable that is most sensitive to the punctuality level.

### 7 Acknowledgements

Financial support from Banverket (Swedish National Rail Administration) is gratefully acknowledged. The author would like to thank Bo-Lennart Nelldal (supervisor in the PhD project, KTH) and Lars-Göran Mattsson (assistant supervisor in the PhD project, KTH) for valuable discussions and comments on the manuscript. The author also wishes to thank the other members of the reference group: Magnus Wahlborg (Banverket), Magdalena Grimm (Banverket), Hans Dahlberg (Banverket) and Håkan Berell (Banverket) who are actively supporting the project.
References


## 8 Appendix

### Design matrix

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Main effects</th>
<th>Two way interactions</th>
<th>Three way interactions</th>
<th>Four way interactions</th>
<th>Five way interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

(Continues with 64 experiments)