Improving a railway timetable by implementing technologies for increased automation

Identifying corridors in the Dutch railway network for implementations of technologies

Robbert Abbink

University of Twente

April 2024

Industrial Engineering and Management

Faculty of Behavioural and Management Sciences

This report is intended for Nederlandse Spoorwegen (NS) as an advice on developing implementation strategies for new technologies to improve the national timetable.

University of Twente, Hallenweg 17, 7522 NH Enschede

N.S. Groep N.V., Laan van Puntenburg 100, 3511 ER Utrecht

Paper Title: Improving a railway timetable by implementing technologies for increased automation
Subtitle: Identifying corridors in the Dutch railway network for implementations of technologies.

Author: Robbert Abbink

Supervisors University of Twente: Dr. A. Trivella, Dr. S.M. Meisel

Supervisor Nederlandse Spoorwegen: P. Looij

Date of publishing: 19 May 2024

Number of pages: 74

This report was written as part of the master assignment for the study Industrial Engineering and Management at the University of Twente.

Preface

Dear reader,

Before you lies my master thesis 'Improving a railway timetable by implementing technologies for increased automation'. The goal of this research was to find a way to identify the corridors where technologies can be implemented to decrease planning norms that benefit the timetable of the whole network the most. To this extend, I have designed a new model where new technologies can be tested for their influence on the network timetable. To conduct this research, I worked at NS from March 2023 until April 2024.

I would like to thank all people who assisted me during my research the past year. Firstly, I would like to thank my supervisor from NS, Patrick Looij, for all his time and interest to supervise this research. Patrick has guided me in my research, by explaining the benefits and potential of my research and has introduced me to the complex world of train timetable designs. Our regular meetings resulted in a clear approach for the finalization of my thesis and provided me with valuable feedback. Secondly, I want to thank Alessio Trivella, my first supervisor from the UT. Our discussions pushed me to think more critically about my research and provided me with different angles from an academic point of view. Furthermore, I would like to thank the different employees at NS that helped me during my master thesis, especially Gabor Maroti, who took extra time to provide different perspectives on my decision-making.

In particular I would like to thank everyone involved in this research, my friends and my family, for supporting me during the past year. As a result of this support, I can now be proud to present this thesis. I hope that you may enjoy reading this thesis and that it provides an interesting insight in how new technologies in the railway sector can be tested for their effects on timetable quality.

Kind regards,

Robbert Abbink

Enschede, April 2024

Executive summary

NS (Nederlandse Spoorwegen) is the biggest railway operator in the Netherlands and provides public transport services across the country. Part of the vision and strategy of NS is to utilize the existing infrastructure more efficiently, which is one of the motives for conducting this research. We can utilize the existing infrastructure more efficiently by reducing the planning norms (e.g. running times and headways) as was already studied internally at NS. Passengers will face shorter waiting and travel times and the timetable will also improve in terms of transfers and line frequencies. This results in an overall improvement of the passengers' convenience and leads to an increase in passenger demand.

Planning norms can be reduced, without loss of punctuality, by implementing new technologies on either the rolling stock or the infrastructure. For each of these technologies (we consider ERTMS Level 2 and Hybrid Level 3, ATO, TMS and DCO), investments are necessary, which would be too expensive to implement directly on a national level. Secondly, current training capacity is low, while all planning norm decreasing technologies require new trainings for the train drivers. Furthermore, it is desired to first investigate the effects of the technologies on a small number of corridors before possibly implementing them nationally. To provide NS with an implementation strategy for the technologies, this thesis aims to answer the main research question:

How can NS identify the corridors where technologies can be implemented to decrease planning norms that benefit the timetable of the whole network the most?

The research starts with a context analysis that describes the different planning norms and the way they are setup and investigates the current processes in the timetable design process at NS as well as the tools that the different departments use to analyse the timetables. This is followed by a literature study to research the different technologies and their effects on the timetable designs. The second part of the literature study focuses on the existing models that optimize and analyse train timetables. The last part discusses methods to identify bottlenecks in a train timetable. The findings of the literature study can also be divided into these three topics. First, we found that literature on the technical aspects of the technologies is available but there is still much uncertainty about the effects of these technologies on planning norms and train timetable design. No simulations or real-life experiments have been conducted, showing the need to research the assumed time savings for these technologies, as they currently rely on expert judgment.

In the second part of the literature study, several models were discussed to design and optimize train timetables. We found that there does not exist an optimization model that considers the implementation of new technologies and the location in which they should be implemented jointly. So far, researchers have manually changed the parameters of the model but this task would be too time-consuming to answer the research question. Therefore, we find the need for a tool that can optimize a train timetable by implementing new technologies that decrease planning norms, which was designed as part of this research. In the last section of the literature study, bottleneck identification methods were discussed, as it seems obvious that train timetables can be improved by solving its bottlenecks. The methods that were discussed still required expert knowledge on bottleneck identification or improving timetables, which underlines the desire for a model that does this automatically and jointly.

To solve the research question and fill the literature gap, a new model was designed. This extended Periodic Event Scheduling Problem (PESP) model has the ability to implement new technologies in an existing timetable. An investment budget allows to research different scenarios and the model was first solved for a single-line network and later for a medium-sized network. The timetables were optimized for the in-vehicle time of all passengers in the network, using the Gurobi MIP solver.

Aside from the implementation of new technologies, this model also introduces parallel tracks and platforms jointly, whereas current PESP models focus on either parallel tracks or parallel platforms. Moreover, the connections between tracks and platforms are introduced in the model as constraints, which was not considered in existing models either. As train timetabling problems are known to be NP-hard, computation time for the proposed model increases exponentially with model size and therefore larger network sizes are expected to have high computation times. Furthermore, these instances cannot be solved by a standard MIP solver.

For both a single-line and a medium-sized network, a budget-improvement curve could be created that shows that the optimal investment budget for which the timetable can already be improved significantly. The percental increase of the investment budget after this optimum is higher than the percental improvement of the timetable, resulting in diminishing returns. However, the timetable can still be improved after this point. For both model sizes we found a total potential improvement of around 5.2%, by implementing all technologies on all corridors and trains. The optimal points for both network sizes lie around 20% of the maximum investment budget, which results in a percental improvement in the timetable of around 60% of the total potential improvement. By showing the timetable of the solution to the models for both one corridor and one train series, more insight was gained into the improvements of the timetable in our model.

The initial experiments showed that the model can account for network effects and shows an implementation strategy for NS. In both cases, the implementation strategy would be to prioritize the combination of ERTMS Level 2 and ATO, followed by TMS and when there is budget left, DCO should be implemented as well. Additional experiments were conducted to analyse the effects of the individual technologies and combinations with ERTMS Level 2. Here we found that TMS shows the biggest improvement in the timetable when compared to the investment costs. However, the combination of ERTMS Level 2 and ATO shows an improvement that is closest to the optimal curve, which confirms the implementation strategy that was found for the single-line and medium-sized network.

For the national network, the computation time increases significantly and consequently, the model could not be solved. Conservative estimates predict a computation time of more than 124 days. This could be decreased by using a more powerful computer or by introducing heuristics that improve the root node and/or the branch-and-bound algorithm, but this was not researched as part of this thesis. It is also noted that the decision of implementing new technologies takes place in the long-term project and thus computation times of at most a few weeks can be considered acceptable.

The main finding of this study is that the new model can correctly identify the corridors where technologies can be implemented to decrease planning norms that benefit the timetable of the whole network the most, which answers the research question. Given the assumptions in the model and its limitations, we find that the model has a great potential for NS to develop and analyse implementation strategies.

Table of Contents

Reader's guide

Chapter 1: Introduction

In this chapter, the company and its operations are introduced. The relevance of the problem is explained as well as the problem-solving approach. The research question and the sub-research questions are mentioned and the deliverables are discussed. Overall, this chapter explains why and how the research is conducted.

Chapter 2: Context Analysis

This chapter provides an elaborate analysis of the planning norms and their effects on the train timetable. Furthermore, the design process at NS is discussed and a description is given of the models and programs that are used in this process.

Chapter 3: Literature study

The literature study is separated into three topics: the technologies, the optimization models and the bottleneck identification methods. We find the literature gap that results in the need for a new model that is designed in this research.

Chapter 4: Model Approach

Chapter 4 explains the mathematical formulation of the new model. First the sets, parameters and decision variables are introduced after which the objective function and restrictions of the model are given.

Chapter 5: Numerical study

In this chapter, several experiments are conducted to assess the applicability of the new model for the Dutch railway network. First, the network specific input parameters are provided. Then, the model is solved for a single-line network and a medium-sized network. Later, additional experiments are conducted to solve different scenarios for individual technologies as well as a combination of technologies.

Chapter 6: Conclusions and recommendations

The conclusion mentions what the result from this research is and why this result answers the main research question. In the last section, recommendations are made to further improve the applicability of the proposed model and suggestions for future research are provided.

Terminology

Corridor (p. 13)

Railway section of one or multiple parallel tracks between two stations.

Block (p.17)

A corridor is divided in multiple blocks, indicated by signals on both sides of the block). Blocks can vary in length.

Train conductor (p. 13)

Onboard staff who check the validity of the train tickets and assist the train driver with the closing of the doors (among other tasks). After the implementation of DCO the train conductors will not do the closing of the doors anymore.

Planning norm (p. 13)

Fixed values for different decisions in timetable design. When referring to planning norms, we include minimum running time, running time supplement, headway, dwelling time and buffer time.

Minimum running time (p. 16)

The minimum amount of time it takes to run a specific type of train from one point in the network to another, typically determined per block or corridor.

Running time supplement (p. 16)

A percentage of the minimum running time that is added to the total running time to account for differences due to human interaction and other delays in the different processes with the aim of designing a punctual and robust timetable.

Headway (p. 17)

The minimum amount of time that needs to be planned between to succeeding trains, determined by the amount of time that the signalling system needs to acknowledge that a specific length of track is free for the succeeding train.

Dwelling time (p. 18)

Time that needs to be planned to stop at a station. Dwelling times depend on the number of passengers that is boarding or alighting and how fast they are able to do so. Included in the dwelling time is also the process of opening and closing the doors as well as the departure process itself.

Buffer time (p. 18)

Planning norm to minimize the delay of the succeeding train as a result of the delay in the processes of the preceding train.

ERTMS (European Rail Traffic Management System) (p. 14)

A European train protection system that will allow for shorter headways and faster running times and a more uniform railway network. Two types of ERTMS are ERTMS Level 2 and ERTMS Hybrid Level 3.

TMS (Traffic Management System (p.14)

Technology that enables communication between trains in such a way that trains can adapt their driving according to the behaviour of other trains.

ATO (Automatic Train Operation) (p.14)

Technology that makes processes of a train automatic. Which processes are automatic, depends on the Grade of Automation (GOA).

GOA (Grade of Automation) (p.14)

Used to indicate to what extend ATO is implemented to automate processes. Higher GoA levels result in a lower factor of human error and a more consistent driving behaviour of trains.

DCO (Driver Controlled Operations) (p.13)

Technology that allows the train drivers to open and close all doors themselves. Without DCO the train conductors do this.

PESP (Periodic Event Scheduling Problem) (p.35)
Mathematical model that aims to solve a train timetabling problem that follows a periodic pattern. PESP solves one period in this pattern, which can then be duplicated for a longer time horizon.

In-vehicle time (p. 46)
The time a passenger needs to travel in a train to get from one station to another.

MIQCP (Mixed Integer Quadratically Constraint Program) (p.43)
Mathematical program where some or all decision variables are required to be integer and which includes quadratic constraints.

1. Introduction

This research has been conducted as a master thesis for the study of Industrial Engineering and Management in the faculty of Behavioural, Management and Social Sciences at the University of Twente. The master assignment has been formulated in collaboration with a supervisor within the department of Netwerkontwikkeling en Ontwerp at Nederlandse Spoorwegen. In the first chapter, the company, department and problem context are described in order to get a better picture of the operations of this specific department at Nederlandse Spoorwegen and the cause that let to this particular assignment.

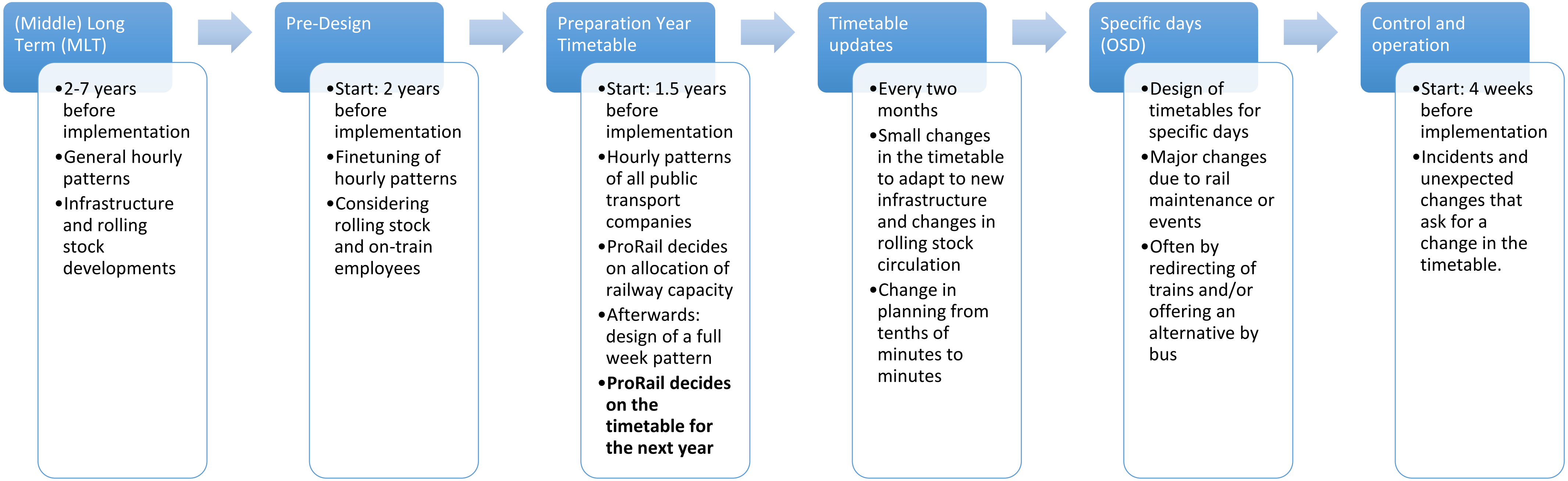
1.1. Nederlandse Spoorwegen

De Nederlandse Spoorwegen (English: Dutch Railways, from now on referred to as “NS”) is the biggest railway operator in the Netherlands and was founded through a merger in 1938. It provides public transport for almost one million passengers per day, by operating around 4800 train runs over a rail network of around 2100 kilometres (Nederlandse Spoorwegen, 2023). The rail network in the Netherlands is managed by ProRail, which itself also has its origins as a department within NS. ProRail is responsible for the maintenance and development of the railway infrastructure but also decides on the allocation of the available railway capacity. To this extent, the network is divided into smaller concession zones and one main rail network (Hoofdrailnet) since 1995. The company that wins the concession has the exclusive right to operate in that area and is responsible for the transport of passengers under the restrictions that are determined within the concession, such as frequencies (number of trains per time period), capacities (total number of passengers that can be transported using public transport) and facilities (seating, toilets, Wi-Fi, etc.) for a predetermined time period. The public transport company that operates in a certain concession zone is obligated to achieve and maintain a certain level of quality (e.g., timetable accuracy, capacity, facilities) which can be fined heavily when the company is unable to do so.

The main rail network is determined by the Dutch government as the part of the national rail network that should be regulated by the national government at all times. NS is concessionaire of the main rail network since 1995 and until at least 2033. To operate the main rail network, NS has a fleet of 761 trains and more than 19,040 employees in the Netherlands (Nederlandse Spoorwegen, 2022). Part of the vision and strategy of NS is to utilize the existing infrastructure more efficiently, which is one of the causes leading to the topic of this research. This relation is further explained in the remainder of this chapter.

1.2. Network Development and Design

As ProRail is responsible for the allocation of the available railway capacity, NS and the other railway companies need to communicate their plans with them. Depending on this request and the capacity of the infrastructure, ProRail decides on the final timetable. To prepare for these decisions, NS has several departments that contribute to the timetables that NS wants to operate. Timetables are in development years before they take into effect to be able to design the best timetable possible that will be approved by ProRail. The department within NS that is responsible for the development of new timetables is “Netwerkontwikkeling en Ontwerp” (from now on referred to in English as “Network Development and Design”). This department itself is split into multiple smaller departments with their own expertise and planning horizon (see also Figure 1 on the next page for a simplified process diagram).



(Middle) Long Term (MLT)

- 2-7 years before implementation
- General hourly patterns
- Infrastructure and rolling stock developments

Pre-Design

- Start: 2 years before implementation
- Finetuning of hourly patterns
- Considering rolling stock and on-train employees

Preparation Year Timetable

- Start: 1.5 years before implementation
- Hourly patterns of all public transport companies
- ProRail decides on allocation of railway capacity
- Afterwards: design of a full week pattern
- **ProRail decides on the timetable for the next year**

Timetable updates

- Every two months
- Small changes in the timetable to adapt to new infrastructure and changes in rolling stock circulation
- Change in planning from tenths of minutes to minutes

Specific days (OSD)

- Design of timetables for specific days
- Major changes due to rail maintenance or events
- Often by redirecting of trains and/or offering an alternative by bus

Control and operation

- Start: 4 weeks before implementation
- Incidents and unexpected changes that ask for a change in the timetable.

Figure 1: Simplified process of timetable development at the department of Network Development and Design

Timetables for a full year (the first three steps in Figure 1) are designed by the department “Integraal Product Ontwerp” (IPO), while timetable changes for specific days, for instance because of rail maintenance or concerts and events, are designed by the department “Ontwerp Specifieke Dagen” (OSD). Where OSD is restricted in its options due to capacity of the current infrastructure and the availability of trains and personnel, long term designs of IPO (MLT) have a lot more freedom. IPO is able to experiment with changes in the infrastructure and can assume new developments in the capacity and availability. Typically, IPO develops timetables for a long period leading to the year of implementation and OSD changes this timetable design for the specific days that the timetable might no longer be feasible to achieve certain goals and keep KPIs at their desired levels.

1.3. Problem introduction

Following the general vision of NS to better utilize the current infrastructure, several studies have been conducted by the different departments. One of these studies concerns planning norms. Planning norms are fixed values for different decisions in timetable design. For example, NS uses a fixed minimum running time from one station to another when using a specific type of rolling stock. Another example is the headway between trains. Fixed planning norms decrease the number of decision variables in timetable design, but this is not the main reason that NS has implemented them. The reason planning norms are integrated in the timetable design is twofold. First, NS wants to ensure a robust timetable that is able to absorb small delays and guarantees a stable and safe operation. In this case, the timetable design minimizes the delay of other trains as a result of the delay of one train. Second, planning norms are necessary to function as a buffer to be able to provide a punctual timetable for the passengers. Punctuality means that the delays in one process of a train do not lead to delays in the timetable of that train.

An in-depth explanation of planning norms, their dynamics and their implications are given in Section 2.1 but they consist of several norms that contribute to different kinds of headway or buffers. Ideally, the planning norms would be as small as possible to be able to operate as many trains as possible on a certain corridor (railway section of one or multiple tracks between two stations), while still providing a robust and predictable operation. An internal study of NS concerning these planning norms concluded that decreasing the planning norms in the timetable nationally, would have a significant financial benefit due to an increase in ticket sales and a more efficient utilization of the resources, while not having to invest in building expensive new infrastructure.

One of the ways to decrease planning norms (while maintaining high punctuality) is by implementing new technologies. Examples of these planning norm decreasing technologies are “Driver Controlled Operation” (DCO), “European Rail Traffic Management System” (ERTMS), “Automatic Train Operation” (ATO) and “Traffic Management System” (TMS). These technologies are briefly introduced here, a more elaborate description and concrete effects on the planning norms can be found in Section 3.1, where we also discuss briefly several other technologies that were left out of the scope of this research.

- DCO refers to the opening and closing of the doors of the trains during dwelling. The train driver currently does this after the train conductor has already closed every door except its own and after the train conductor has checked that all doors have been closed. From now on we refer to train conductors as the onboard staff who check the validity of the train tickets and assist the train driver with the closing of the doors (among other tasks) (NL: conducteur). When implementing DCO, the train drivers will be able to open and close all doors themselves, eliminating the “door check”, resulting in a lower dwelling time. Moreover, by eliminating the interaction between conductor and driver, the dwelling process will become more predictable, which makes it possible to reduce the buffer time that accounts for short delays in the process.

- ERTMS is a security system that will allow for shorter headways and faster running times. There are several levels of ERTMS, that refer to a higher incorporation of several modules that relate to ERTMS. The higher the incorporation of the model, the higher the benefit in terms of planning norms. The new security system makes it possible to make headways dependent on the position and speed of a train instead of the infrastructure (such as signs). This way shorter headways can be realized. Moreover, ERTMS provides the train driver with information on maximum speeds and ideal breaking curves. When these speeds and curves are followed by the train drivers, the running time of the trains can also be decreased.
- ATO is also divided into different levels, indicated by the "Grade of Automation" (GoA). GoA0, for example, is used to indicate that all processes (operating, opening/closing of doors, acceleration/deceleration) are controlled by the train driver or conductor. The highest level of ATO, GoA 4, refers to a system where all processes are controlled automatically and autonomously by a computer, without supervision of a train driver or conductor. Higher GoA levels result in a lower factor of human error and a more consistent driving behaviour of trains. Since a higher consistency leads to higher predictability, this will enable planners to use lower planning norms in the timetable design.
- The implementation of TMS will enable communication between trains in such a way that trains can adapt their driving according to the behaviour of other trains. For instance, when one train is delayed this might normally affect the following train on that corridor. When using TMS, the following train will know the exact location and speed of the first train (a long time before the encounter) and may decide to decelerate or stop the train. This can be done with high accuracy, limiting the impact on all following trains as much as possible. This makes it possible to assume lower planning norms in the timetable design.

For each of these technologies, hardware and software investments are necessary, which would be too expensive to implement directly on a national level. Secondly, current training capacity is low, while all planning norm decreasing technologies require new trainings for the train drivers. Furthermore, it is desired to first investigate the effects of the technologies on a small number of corridors before possibly implementing them nationally. Since these technologies are relatively new and impacts (of implementing one technology on one corridor) on the network as a whole are still uncertain, there is the need for a model or tool that can evaluate these effects on the different parts of the whole network timetable and identify corridors that would be most promising to be outfitted with (one of) these new technologies. The design of this model is the main purpose of this research.

1.4. Research questions

To structure this research, the above-mentioned problem is formulated in one central research questions and multiple sub research questions. The central research question is formulated as follows:

How can NS identify the corridors where technologies can be implemented to decrease planning norms that benefit the timetable of the whole network the most?

To solve the central research question the following sub research questions are formulated:

1. How do planning norms influence the timetable design at NS?
2. What models and methods are currently used by NS to assess the effects in the timetable of decreasing the planning norms?
3. What planning norm decreasing technologies exist that can be implemented on corridors in the network of NS?
4. What techniques are currently used to identify corridors where lower planning norms are desired?

5. What models and programs exist that can experiment with the characteristics of corridors and trains?
6. What models and methods exist that can optimize timetables?
7. What is the best solution out of the chosen models and methods?
8. What objective function leads to the identification of the most promising corridors for the implementation of planning norm decreasing technologies?
9. Which technologies need to be implemented on which corridors to benefit the timetable of the whole network the most?
10. How can the solution be implemented in the timetable design process at NS to decide where which technologies should be implemented?

1.5. Reader's guide

In Table 1, a reader's guide can be found, to see which chapters answer which of the aforementioned sub research questions. In the last column, a brief description of the content of that chapter is given.

Table 1: Reader's guide

Chapter title	Related sub research question(s)	Content
2. Context Analysis	1,2	Elaborate analysis of planning norms and their effects on the timetable. Layout of the current timetable design process and description of the models and programs that are currently used in this process.
3. Literature study	3,4,5,6	Explanation of the different planning norm decreasing technologies and their effects on timetable design. Research on different timetabling methods as well as models that can identify corridors by their characteristics.
4. Model Approach	7,8	Design of the mathematical model, its restrictions, assumptions and limitations, and the formulation of the objective function
5. Numerical study	8,9	Implications and results of implementing the proposed approach
6. Conclusion and recommendations	10	Description of how the proposed approach can be implemented at NS and how future research can improve the applicability of the model.

2. Context Analysis

In this chapter, the aforementioned problem is further identified. First, a description of the different types of planning norms is given as well as a brief explanation on how they are determined and to what extent they can be decreased. Then, the outline of the current timetable design process is introduced to determine in which part of the process the research problem exists and what time horizon should be considered when developing a solution. This will be followed by a brief summary of the main programs, models and tools that are currently used by the planning department for this time horizon. Finally, the usefulness for these department resources and the consequences for setup of this research are discussed.

2.1. Planning norms

In timetable design, planning norms together contribute to the arrival and departure times at the corresponding points in the network. In this section, the different types of planning norms are introduced and their use in timetabling is explained. It is defined how each of these norms are determined and in what ways and to what extent they can be decreased to optimize the train s.

2.1.1. Minimum running time

The minimum running time is the most straightforward planning norm. It is determined as the minimum amount of time it takes to run a train from one point in the network to another. Factors that influence the minimum running time are (among others) the maximum speed, acceleration and deceleration of the rolling stock, the maximum speed on that corridor and the distance between the two points. Different types of trains may have different properties and some corridors allow higher speeds than others, so the minimum running time is determined for every train on every corridor. As the minimum running time is restricted by at least one of the factors mentioned above, it can be decreased by changing the restricting factor. Infrastructure development may change the properties of the corridor to allow for higher speeds and newer trains may be able to accelerate or decelerate faster or reach a higher maximum speed.

2.1.2. Running time supplement

To account for delays due to a slower acceleration or deceleration, or a speed that is slightly lower than the maximum speed, a running time supplement (NL: rijtijdtoeslag) is added. Typically, this planning norm is determined as a percentage of the minimum running time on that corridor. The general value that is used by NS is 7%. In some specific cases, the percentage may be lower due to practical evidence that this is possible. In rare cases, a higher percentage is used for the running time supplements, however, this is not a result from stricter planning norms, but rather from implications elsewhere in the timetable that requires longer running times to realize a feasible schedule.

For a big part, running time supplements account for differences in running times due to human interaction and other delays in the different processes. One train driver accelerates faster than another, for example, or dwelling may take longer unexpectedly because of higher passengers' demand. Running time supplements can therefore easily be decreased by eliminating the (effect of) human interaction in the process. When the process is more automated or standardized, this will result in processes that behave more similarly and running times that come closer to the determined minimum running time. Consequently, it is possible to decrease planning norms for running time supplements. In timetable design, the planning norm of the running time is predetermined as the minimum running time plus the running time supplement, so the two components are not used independently anymore. However, it remains interesting to investigate the possibility to decrease the two planning norms separately, as all time savings contribute to shorter running times.

2.1.3. Headway

To prevent two succeeding trains (two trains that use the same infrastructure successively) from colliding, a headway is used. In the Netherlands, railway security dictates that only one train is allowed in one “block” (Weeda, 2005). A block is a part of the corridor, typically between two signals and can vary in length. Before a block, the signal shows a red (train stop), yellow (put on the brakes) or green signal (drive) to indicate that a train is present in the following block, a train is present in the block after the following, or no train is present in the following two blocks, respectively. To make sure that a train can maintain its top speed, it should only encounter green signals, which is only possible if a block is free for a significant amount of time before the train reaches that block. Then, the train needs to drive the total distance of the block plus its own length to clear the block, before the signal turns yellow and finally green again. To clarify this process, suppose the situation as shown in Figure 2 from Hansen and Pachl (2014). A block needs to be declared ‘free’ and a train driver needs to be able to see and react to the colour of the signal. Since the colour orange would be the first signal a train driver would see before approaching an occupied block, this needs to be shown a block in advance, which is indicated by the approach time. Then, the train has to travel the block, indicated by the running time, and clear the block for the next train, which is shown in the figure as the clearing time. Finally, traffic control needs to receive the information that the block is cleared, which is indicated by the release time. Thus, the time (note that the headway is always calculated using time and not distance) that is needed to reserve a block (blocking time) is determined by adding up all these time segments, which is much longer than the actual occupation time of a block. Since block distances differ and this affects blocking times, headways may vary within one corridor. The headway between two succeeding trains is then determined by the longest blocking time of all the blocks.

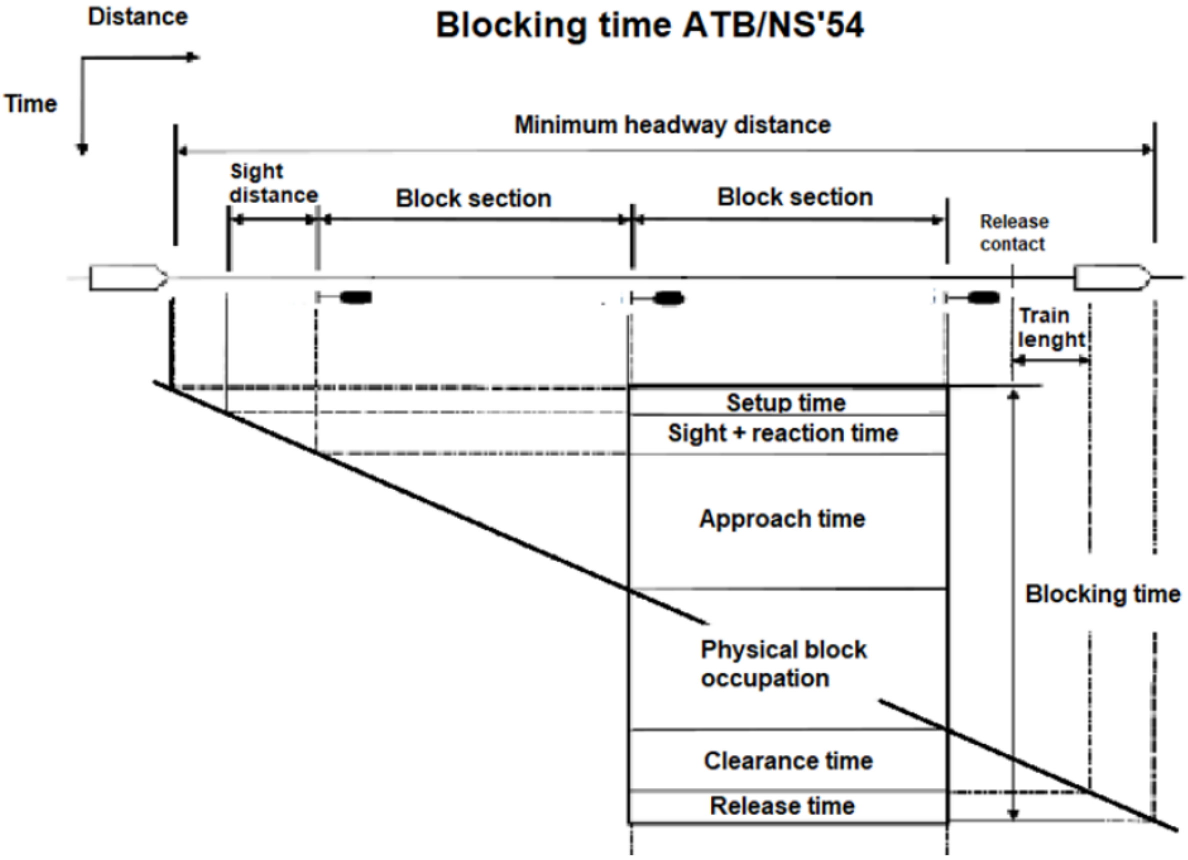


Figure 2: Blocking time for a running train (Hansen & Pachl, 2014)

Headways can be decreased when a new security system is introduced that does not depend on the colour of the signs, but on the actual distance between trains (for instance, through the implementation of ERTMS). When trains can communicate their positions and speeds, you can compute the minimum distance that needs to be maintained and act accordingly without having a signalling system in place. This way, shorter distances and thus shorter headways can be realized. Another possibility is decreasing the block length in such a way that trains can drive with shorter headways (TwynstraGudde, 2022). In the Netherlands, block lengths are already decreased in the past for many corridors, but due to the expected implementation of ERTMS (see chapter 3), this can also be realized at the other corridors that still have longer block lengths.

2.1.4. Buffer time

As processes are subject to several factors (both internal and external), delays in the processes may occur. To prevent that a delay in the processes of one train leads to a delay of the succeeding train, a buffer time is used in the timetable design. The duration of a process follows a certain probability distribution; some values occur more than others. The value of the buffer time is determined in such a way that, in a certain percentage of all cases, the process is finished before the next starts. In other words, in a certain percentage of all cases, a train will only encounter green signs and can run without any delays as a result of the behaviour of a previous train.

Buffer times can be decreased when processes can be planned and operated more accurately and delays in the process are less common. Again, this can be realized by eliminating the human factor in train operation. This way, the probability distribution will be more centred around the average process duration and thus a shorter buffer time is needed to obtain the same percentage of processes that are not affected by a delay in the previous process.

2.1.5. Dwelling time

So far, we considered process times for traveling between two stations or between departure and arrival. However, a train stops at the station for some time as well, which is indicated by the dwelling time. Dwelling times depend on the number of passengers that is boarding or alighting and how fast they are able to do so. Included in the dwelling time is also the process of opening and closing the doors. Sometimes dwelling time supplements are included as well to account for delays in the dwelling process, but they can be used interchangeably with running time supplements (Weeda, 2005).

Dwelling times can be reduced in two ways: either boarding and alighting can be done faster or the process of opening and closing doors can be done more efficiently. Passengers that want to board a train can be instructed not to stand in front of the doors, to make way for the alighting passengers. Wider doors can contribute to a faster process as well, but at the loss of seating capacity. The dwelling time can also be significantly reduced by automating the opening and closing of the doors or operating them from the train cabin.

2.1.6. Combination of planning norms in the timetable design

The aforementioned planning norms are all used in the timetable design. The running time determines the duration between departure and arrival, the headway is used as a restriction for the departure and arrival times of succeeding trains and the dwelling time shows the minimum duration between arrival and departure. As mentioned, the planning norms vary depending on which corridor, block and train it applies to, but its values may also depend on the type of processes. For example, the time difference between two departing trains is allowed to be shorter than the headway between two arriving trains or an arriving and a departing train. In this research we exclude this dynamic, as is further explained when discussing the input parameters of the model in Chapter 5.

While all planning norms can be decreased, not every time saving will lead to shorter travel times. For instance, if we are able to significantly reduce the dwelling time of a certain train, its departure time may still be restricted by the departure time of the preceding train, because of the required headway. The opposite is also true. By saving just a few seconds on each of the individual planning norms, they can together contribute to a significantly lower travel time, possibly even for another train. This research focusses on the effective implementation of new technologies to reduce planning norms and consequently, improve the timetable of the whole network.

2.2. The design process

To understand to which part of the design process this research contributes, it is necessary to gain more insight into the design process as a whole, which is described in this section. The first stage of timetable design starts around seven years before actual implementation of the timetable or even earlier. Since the process from decision making to actual implementation of technologies in practice will take several years, our model can be used in this phase of the timetable design process. In the processes that follow, the timetables are continuously changed and altered until the final implementation on the day itself. The design process is also described in Planting (2016a), but since then, the structure is slightly changed. This section is structured chronologically, starting with the long-term design (at least seven years in the future) and ending with the operational planning (last-minute changes while operating the timetable).

2.2.1. (Middle) Long Term

Before the design of specific timetables starts, NS wants to conduct several studies to evaluate effects in infrastructure (construction or new technologies) and rolling stock (circulation, new train types or new technologies). While in this thesis the name (Middle) Long Term or (M)LT is used, projects can be divided in two separate groups: Long Term and Middle Long Term.

Long Term projects typically start many years before the year of implementation and end around seven years before that. At seven years, the project can be taken over as a Middle Long-Term project, but it can also be finished, depending on the topic of the project. The research questions in these projects often relate to the frequency of lines and the availability of public transport to satisfy (predicted) demand and concern substantial changes in the infrastructure, such as building new railway sections, improving corridor characteristics or introducing faster trains. As ProRail is responsible for the infrastructure, NS needs to be in close contact with them. ProRail will communicate foreseeable infrastructure improvements with NS, but NS could also advise or ask for certain changes. As these are decisions for the Long Term, multiple timetables are developed to show the possible effects of the corresponding implementation.

Middle Long-Term projects include smaller projects with a time horizon of two to seven years before implementation. Multiple timetables are designed to show the difference between the variants and to show how this impacts the goals and KPIs of NS. While the time horizon may be much shorter than that of the Long-Term projects, there is still room to research impactful changes to the infrastructure and rolling stock.

Since our research concerns possible changes in the infrastructure and rolling stock for years in the future, it is part of the studies that are conducted in the (M)LT phase. However, to understand the importance of this phase and how it leads to a final operatable timetable, the rest of the design phase is discussed in the remainder of this section.

2.2.2. Preliminary Design (VO-phase)

In the next phase, the first preliminary designs (NL: voorontwerp or VO) are developed. They form the foundation of the timetable designs of the following phases and are based on several studies on possibilities and limitations on capacity and availability. Preliminary designs are made for a specific year and this part of the process starts one and a half years before the implementation of the timetable, which is in December before that year (i.e., the timetable for 2024 is implemented in December 2023 and the VO-phase starts in June 2021). When no major changes in the infrastructure are scheduled for that year (leading from decisions that are made in the previous stage), the Preliminary Design team uses the timetable of the previous year. Otherwise, a timetable is delivered by the (M)LT. The goal of this stage is to test timetable changes (e.g., total number of operating hours, changes in rolling stock circulation, freight trains and other major development plans) from the MLT in more detail. Using the software of DONNA (a microscopic timetable design tool, see section 2.3.3), the Preliminary Design team is able to evaluate these changes internally and make decisions on changes that are necessary to include in the following phases (which include external communication), while modelling timetables in the format that will be used in the succeeding stages.

2.2.3. Preparation Yearly Timetable (VJD-phase)

In the Preparation Yearly Timetable (NL: voorbereiding jaardienst or VJD) phase, a Basic Hourly Pattern is developed. A Basic Hourly Pattern is a timetable that shows the hourly pattern of all trains, whether they run the entire day or not. In the Netherlands, train schedules are designed with a cyclic hourly pattern for passengers' convenience but it is also necessary to develop such timetables because ProRail demands that all self-planning public transport companies deliver an hourly pattern to realize a combined hourly pattern. Self-planning companies include NS Reizigers, but also local public transport companies and freight companies. There are also companies that do not plan their timetables themselves. For these companies (often freight but also public transport) and for themselves (for maintenance and development), ProRail provides the hourly pattern.

An hourly pattern for one train does not mean that this train operates every hour. The hourly pattern is determined on the behaviour of the train when it operates. As a result, Basic Hourly Patterns are not perfect, but that is not essential. Two trains might show a conflict in the hourly pattern but when one of them only operates in peak hours on Mondays and the other only operates on the weekends, the conflict does not occur in practice.

As the second part of this phase, ProRail leads the process to obtain a final timetable. After combining all hourly patterns of the different public transport companies, this will show their collective request for railway capacity. The aim of this phase is to, collectively, obtain an hourly pattern that honours the wishes of the different railway companies and ProRail, preferably without conflicts. When no such timetable can be found and conflicts remain, a so-called 'agree-to-disagree' occurs. This disagreement will be solved in the next phase, which regards the definitive capacity allocation.

2.2.4. Basic Daily Pattern (BD-phase)

The BD-phase encompasses both the capacity request (NL: capaciteitsaanvraag or CA) and the capacity allocation (capaciteitsverdeling or CV). Based on the hourly pattern in the previous phase, companies can start to develop a timetable for whole days and for the whole week (7x24h). This includes differences between days (Monday, Wednesday or weekends) and hours (peak or off-peak hours). Moreover, passenger organisations are consulted to provide their input for the timetables. Every operator models this Basic Daily Pattern individually (only self-planning companies as ProRail does this for the other companies) and sends a new request to ProRail in April before implementation. This part of the BD-phase concludes the capacity request.

ProRail will now decide on the capacity allocation based on predetermined decision rules. The capacity allocation is finalized in August before implementation. Recall that new yearly timetables are implemented the first of December ahead of that year, so the final capacity allocation is known half a year before the implementation. After this decision is made, ProRail will not accept changes in the timetable that cause conflicts. However, other changes in the timetable are still possible in the next stages. These changes are approved on a 'first come, first served' basis. As a result, changes that would be possible based on the timetable determined after capacity allocation, may not be possible anymore if another company requested a change earlier.

2.2.5. Basic Daily Pattern Update (BDu-phase)

The Basic Daily Pattern Update (BDu) concerns structural changes to the Basic Daily Pattern. There are several BDu's per year, but the one that affects the timetable the first time is the BDu-December. In the BDu-December phase, the Basic Daily Pattern, as decided by ProRail, is examined for finetuning. Here, the timetables are tested for feasibility in terms of employee availability and the implications of the available rolling stock. To this extent, a complete employee and rolling stock plan are designed to accommodate for the final timetable. Furthermore, the Basic Daily Pattern determines arrivals and departures specific to six seconds (tenth of a minute), while passengers want to the specific minute that their train arrives or departs. These minor changes can be implemented in the final timetable if it does not result in conflicts.

Throughout the year, these types of changes can be implemented approximately every two months, starting at the day of implementation in December (February, April, June, September and October). Changes that are incorporated in these BDu's are often improvements that are permanently included in the timetable starting the day of implementation of that BDu or alterations in timetable or rolling stock availabilities that occur during large parts of a BDu.

As mentioned, BDu's consider a time period of approximately two months, but there are exceptions such as the BDu-September, which is only several weeks long. Furthermore, TBDu's can be designed if necessary. These Temporary Basic Daily Pattern Updates are "extra" BDu's that include major infrastructure projects, for example. These TBDu's are also only several weeks long.

2.2.6. Specific Days (SD-phase)

Resulting from the previous stages in the design process, we have a timetable for approximately two months (depending on the BDu) that specifies every movement of every train on any particular point in time, with a weekly pattern (e.g., every Monday is the same). The next step is to determine necessary changes in this timetable to account for differences in travel behaviour on specific days. These changes apply to (sports) events, concerts but also infrastructure maintenance. In these cases, NS may need to run a higher frequency on a certain line or run longer trains (particularly in the case of events), reroute trains or provide bus transport between two stations (in the case of railway maintenance).

The SD-phase is different from the BDu-phase as it includes temporary changes in the timetable for a specific day, whereas the BDu-phase includes structural changes in the timetable that affect every Monday for example. These alterations are communicated with ProRail and the other public transport companies and can be modelled in the shared microscopic timetable in DONNA (see Section 2.3.3) until 56 hours before operation. If companies have requests after this deadline, it must be sent to Traffic Control. From this point onwards, timetable changes are not requested by subdepartments of the Network Development and Design department, which is because they only include incidental changes that cannot be foreseen by the timetable designers.

2.2.7. Preparing Operation (BO)

These incidental changes are managed by the team Preparing Operation (NL: Besturing Operatie or BO). Examples for these changes are defective overhead wires resulting in rerouting or cancelation of trains, or changes in crew availability resulting in other schedules to account for a change in train driver take-overs. Preparing Operation applies these changes until one day before the operation of the timetable and sends it to Traffic Control.

2.2.8. Traffic Control

The timetable that is sent to Traffic Control shows a theoretically feasible and manageable schedule. While systems are in place to prevent train delays as much as possible (such as the planning norms), there will always occur unexpected situations that result in necessary changes to the timetable. Examples include defective trains, last minute insertion of freight trains, excessive delays or accidents. Traffic Control makes sure that these changes will impact the rest of the timetable as least as possible, while also maintaining railway security.

2.3. Resources and models

All the departments in the previous section use various types of resources to design the (intermediate) timetables. From the previous section, we can derive that our research concerns a problem for the (Middle) Long Term. As a result, the resources that are used in this phase are possibly the most useful for this research. For this reason, the models, programs and software that are used by the other departments are not mentioned in this section. There is also a department (called "Kenniscentrum") that is not a part of Network Development and Design but works closely together to evaluate the timetables for several key performance indicators (KPIs). This evaluation process and the tool that is used in this process are also described at the end of this section. We start this section with a distinction between macroscopic and microscopic modelling to better understand the capabilities of the various models.

2.3.1. Macroscopic versus Microscopic

NS uses two different types of modelling when designing their timetables: macroscopic and microscopic. The difference between them can best be explained by the use of Figure 3 on the next page. Microscopic models split the network into blocks with homogeneous behaviour (speed, acceleration, etc.) and nodes between them where this behaviour changes (Planting, 2016a). Moreover, microscopic models include the possibilities and limitations related to track switches and multiple platforms at a station. Macroscopic models have a much lower level of detail. Multiple platforms are reduced to one node and parallel tracks are combined into one arc between these stations. Furthermore, macroscopic models do not include the behaviour in a specific block, but rather of a series of blocks between critical timetable points such as stations and junctions.

The difference between these two approaches lies in the feasibility of the timetable as well. Since macroscopic models have a lower level of detail, timetables are feasible when no conflicts occur on the important nodes. The same timetable might not be feasible for a microscopic model due to characteristics of the underlying blocks and different behaviour of the trains between them. However, microscopic models are often used to test whether a macroscopic model can be changed in such a way that is also feasible for every block. For long term issues, it is not necessary to describe the behaviour of an individual train and an individual block, because we want to know the general implications of a certain change in infrastructure or rolling stock. The closer we approach the implementation of a timetable, the more important it will be that conflicts on individual blocks are solved (Planting, 2016b).

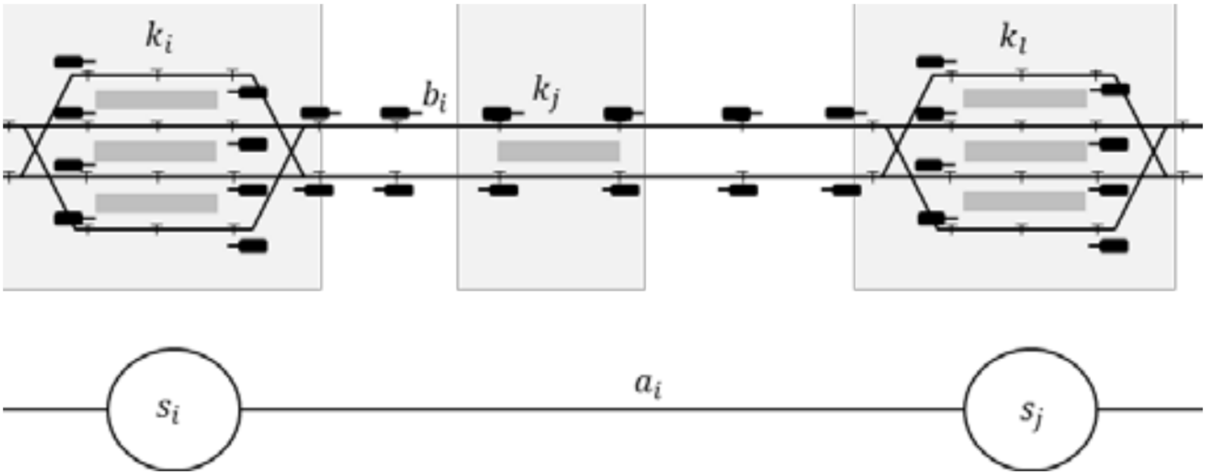


Figure 3: Microscopic (top) and Macroscopic (bottom) design of infrastructure (Planting, 2016a)

In the (M)LT-phase, macroscopic models are used for the design of timetables for the long term and microscopic models are used for Middle Long-Term projects. In the phases from Preliminary Design onwards, microscopic models are used as well, some designed to model a small aspect of a timetable. For instance, NS uses the programs TAM and CREWS to assess the feasibility of a timetable in terms of rolling stock and staff availability, respectively. As our solution does not restrict itself with these constraints, they are left out of consideration and not mentioned here.

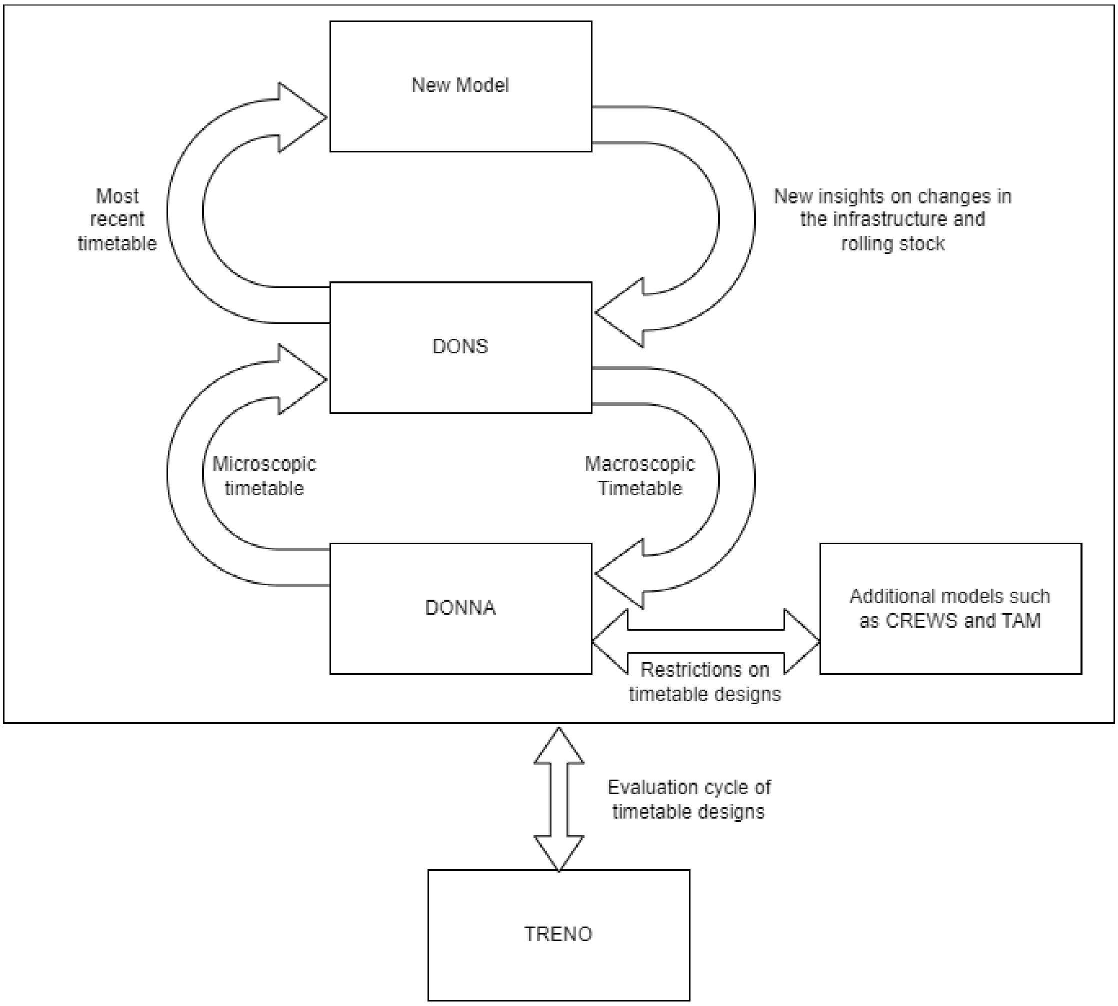


Figure 4: Relationship diagram for timetable design tools

2.3.2. Designer of Network Schedules (DONS)

A macroscopic modelling tool that is used by the department of Network Development and Design, is called DONS (Designer of Network Schedules) (Weeda, 2005). Important timetable points (stations and some junctions) are connected by arcs that represent the corridors (combination of blocks and parallel tracks) between these points. The characteristics of these arcs are given by general data for the infrastructure (e.g., distances and maximum speeds) and general train characteristics are also included (e.g., mass, acceleration, maximum speeds). The user of DONS can also change these input parameters manually per train or corridor and analyse the corresponding effects.

Using a separate module of DONS, a DONNA timetable (see next section) can be imported. This relationship is also shown in Figure 4. Since DONNA returns a microscopic timetable, DONS can now provide a list of conflicts between trains. These conflicts can also be visualized. Figure 5, on the next page, shows such a visualization. On the x-axis, we can find the stations in the area that is visualized (in this case the corridors between Zwolle and Groningen) and on the y-axis the time from 8.00 until 9.00. In the DONS diagram, we can see the train paths between the stations in this period, indicated by the individual lines. Around stations (such as Mp: Meppel), we can see that the lines curve to a vertical line and afterwards, more horizontal again. This shows the deceleration, dwelling and acceleration pattern and already tells us that this is a microscopic visualization.

In the figure, blocking times are added for all trains in the direction of Groningen and these are used to detect conflicts in the timetable. When the blocking times overlap (shown in red) a conflict occurs. Examples of conflicts can be found at station Meppel (Mp) and on the right side at station Groningen (Gn). Using the conflict lists and output graphs given by DONS, the user can solve these conflicts by manually changing the characteristics of one or multiple train series. As the Dutch rail network is considered to be very busy and complicated, this is a process that typically requires several iterations before a feasible timetable is found. The final timetable is specific for every tenth of a minute (six seconds). DONS is a tool that is developed by NS and ProRail, where both parties can work on timetables separately while input data for infrastructure and rolling stock is generalised across the two companies (Planting, 2016b). This way, timetables that are designed by NS are easily communicated with ProRail.

While DONS provides the possibility to design multiple alternative timetables, these can only be developed through iteration of conflict detection and manual conflict solving. In other words, the solution is restricted by the characteristics of the infrastructure, while for the research question, it is desired to determine promising implementations of technology in the infrastructure and/or rolling stock to realise a more efficient timetable. As these changes can be made on every location in the network, this would mean that the user has to design a lot of timetables per manual alteration of the infrastructure and/or rolling stock characteristics as well as a combination of these changes. This would be very time consuming and the computations could be done faster by an automated program. On the other hand, in- and output of DONS can be used to setup a new tool. The input (infrastructure and rolling stock characteristics) can be used as parameters to setup the restrictions of our model and the output (train series with hourly patterns) can be used to determine the required number of trains that need to be scheduled. Furthermore, the output of the new model can also be used as input for DONS to test the feasibility of the timetables and infrastructure alterations.

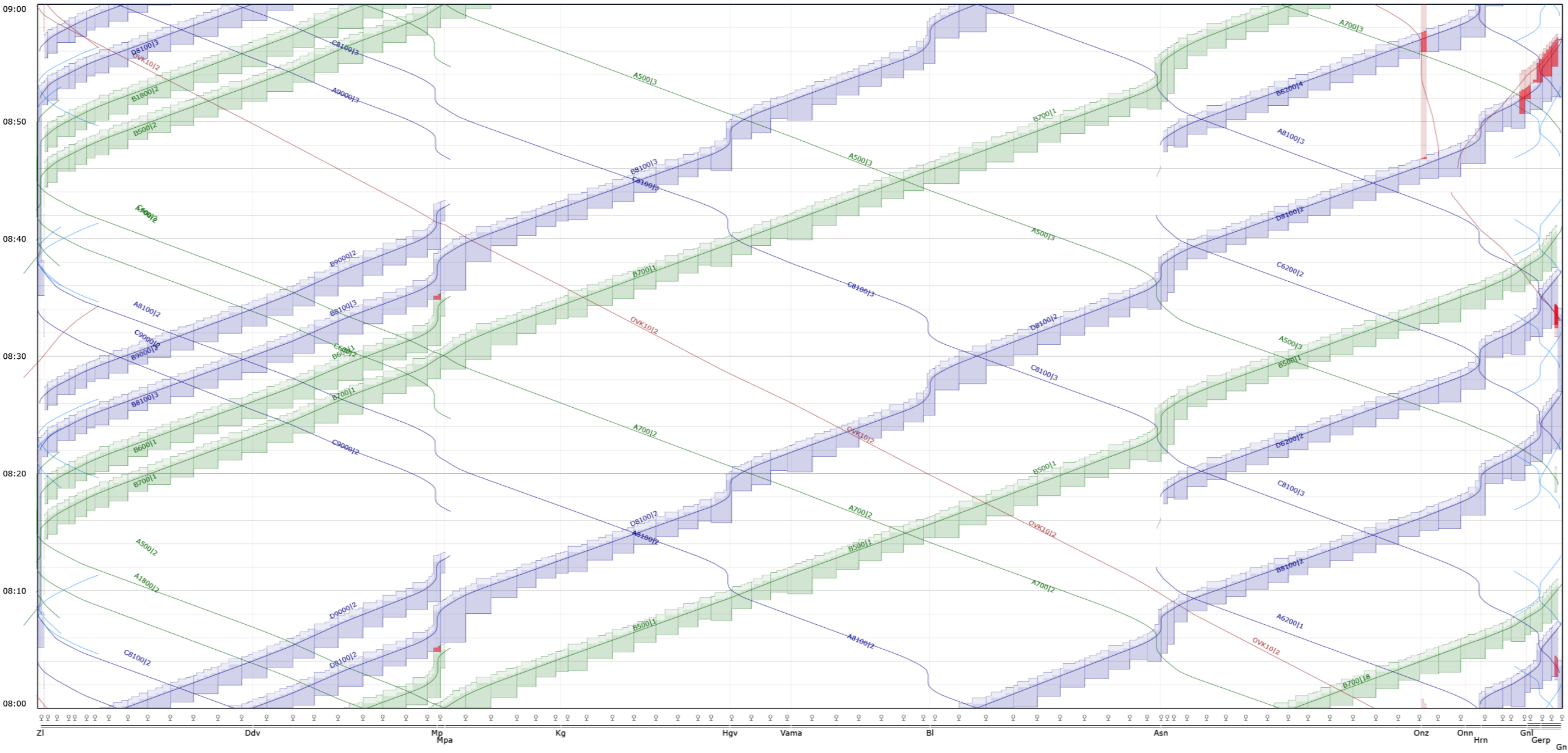


Figure 5: Space-time output diagram of DONS (with conflicts) for the network between Zwolle (Zl) and Groningen (Gn)

2.3.3. Designer of Network Schedules for National Use (DONNA)

Where DONS is only used by NS and ProRail, DONNA is used by every company that is involved in timetable design in the Netherlands. DONNA is a microscopic model where separate files can be created for different periods. The files can be adjusted by multiple users at the same time (Planting, 2016b). Since we are not interested in the characteristics of individual blocks, a microscopic model would provide too much detail for our research. Furthermore, because the proposed solution would be a model that can experiment with changes in the infrastructure and rolling stock, it is not desired that these experiments can be seen by other parties, or even be influenced by them. Similar to DONS, the timetable designed in DONNA is restricted by the infrastructure, while the research aims for a model that can investigate a certain degree of freedom to these restrictions. The microscopic nature of DONNA, results in high computation times when testing different combinations of restrictions for the infrastructure.

2.3.4. Treinen, Reizigers en Euro's voor Netwerkontwikkeling en Ontwerp (TRENO)

TRENO is another program that is used only by NS. This model is not used as a design tool but as an evaluation tool. Timetables are tested for several indicators that assess the value of a certain timetable. Indicators include costs, customer satisfaction (enough seating), robustness and timeliness. TRENO gives a value to these indicators for every timetable that they assess and compare this to the goals that are set by NS or by the government in the concession agreements.

The goal of this research is "to identify promising corridors where technologies can be implemented to decrease planning norms that benefit the timetable of the whole network the best." To this extent, we must also determine which timetable is the better out of the alternatives. An essential step in this process is determining an objective function that maximizes or minimizes a certain value. For instance, an objective function can be aimed at minimizing costs or maximizing customer value. The calculations used to determine the values of the indicators in TRENO can provide important insights to develop the (multi-) objective function for our model.

2.4. Conclusion of the context analysis

In this chapter, we have provided a description of the current timetable design process as well as the planning norms that are used in this process. We have found that the current models that are used to manually design timetables can be divided into macro- and microscopic tools. In Figure 4, the interactions between the different models can be found, which shows that the timetable design process is an iterative process. We have also included the position of our new model. While the existing tools are able to assess the timetable feasibility and identify conflicts, they are less useful to experiment with loosening restrictions in the infrastructure or rolling stock, since every alteration and assessment is done manually. This requires manual changes in the input data for every location and train, resulting in high computation times. Furthermore, the models indicate whether a timetable is feasible but do not provide the most optimal timetable based on an objective. These restrictions in the existing models result in the desire for a new model that is able to investigate a large number of changes in infrastructure and rolling stock and provide the most optimal timetable based on a given objective function.

3. Literature review

There are several technological innovations that can be implemented on the infrastructure in the Netherlands or the rolling stock of NS and that are able to decrease the planning norms in timetable design. In this chapter, the most relevant options are discussed using existing literature. After this, a few other options are mentioned, that are outside the scope of the solution for this research. Note that we are interested in the effects on the planning norms or the timetable and less in the technological features such as software and hardware that are involved in implementing the technologies.

To determine the effects of the planning norm decreasing technologies on train timetable design, we must also understand how train timetables are designed and optimized. This literature review introduces four categories of techniques and methods that concern train timetable design and optimization: Deterministic models, Optimization techniques, Stochastic models and Real-Time Rescheduling models.

Furthermore, we discuss models that are able to identify bottlenecks in train timetables and railway networks, which can be useful for developing a model that can do this for the technologies that we want to analyse. We address the literature gap that is found in this literature review. At the end of this chapter, a conclusion of the literature review is given to address the literature gap and to summarize the most interesting findings that may be useful when developing our model.

3.1. Planning norm decreasing technologies

The first part of this literature review concerns technologies and innovations that are able to decrease one or more planning norms as discussed in the previous chapter: minimum running time, running time supplement, headway, buffer time and dwelling time. The design of these technologies is discussed as well as their effect on the planning norm(s). Finally, we explain why several other technologies are left out of the scope of this paper.

3.1.1. Driver Controlled Operation (DCO)

The first technology that we discuss is Driver Controlled Operation (DCO) (TwynstraGudde, 2022). DCO refers to the automatic closing of the train doors. To explain the effect of DCO on the dwelling time, let us first consider the current subprocesses of dwelling. At the moment that a train arrives at a platform, passengers are able to open the doors using buttons on or next to the doors. The doors remain open for passengers to board or alight the train. Train conductors close all the doors (except for the door that they use to board the train) at the time that is scheduled for this subprocess, based on the departure time of the train. The train conductor does a final check to see if all the doors are closed and communicates this with the train driver. Then, the train driver closes the last door and blocks all the doors. Once this last subprocess is finished, the train is able to depart (Buchmueller, Weidmann, & Nash, 2008).

DCO eliminates the final check that is done by the train conductor. This way, the train driver can close all doors simultaneously and depart once all doors are blocked. The time savings are two-fold: on the one hand, the check is eliminated, on the other hand, the communication between train conductor and driver is eliminated. As a result, a time saving of twelve seconds per stop can be realized (Sigger, 2023). Furthermore, because of the elimination of the interaction between train conductor and driver, the duration of the dwelling process will become more predictable, enabling NS to schedule shorter running time supplements, as these account for delays in the processes.

The investments that are necessary to be able to implement DCO in the daily operations of NS, depend on the type of train. Some Sprinters are already prepared for an upgrade to DCO or will be prepared for an upgrade to DCO after an upgrade to ERTMS (see next section), for others it is more complex to do so. Moreover, technology design also plays a large role in the complexity of the implementation. For one corridor, NS has decided that the trains will be equipped with cameras on the outside. While these cameras are not necessary for a successful implementation, it may be desirable, but this will also increase the investment costs. Public transport companies in other countries have already chosen the option of DCO without cameras, which also works and leads to lower investment costs.

Finally, there is also a restriction on the number of training hours/days that is available. Train drivers need to follow a training for their new role. During this training, they cannot be scheduled to drive trains. As NS already struggles with a shortage in train drivers, there will not be a lot of availability for the number of training moments, resulting in a long process of educating train drivers and thus a long time before final implementation of DCO (Sigger, 2023). For DCO, the training period is only half a day, but in combination with longer training periods for ERTMS for instance, this may still block fast implementation of DCO.

3.1.2. European Rail Traffic Management System (ERTMS)

The second potential technology that can be implemented to decrease planning norms is ERTMS. This technology is a combination of ETCS (European Train Control System) and GSM-R (Global System for Mobile Communications – Railways) or FRMCS (see section 3.1.5) and is created by the European Union with the goal to ensure a higher degree of interoperability of train security systems between all European countries (Li, 2012). There are three different levels of ERTMS:

Level 1

ERTMS Level 1 can be implemented without changing the signalling system that is in place (Li, 2012). In the Netherlands, ERTMS Level 1 works as follows. A Lineside Electronics Unit (LEU) sends information about movement authority (obtained from the interlocking system, called IXL) to so-called Eurobalises. These Eurobalises are situated in the track and the information is collected by the train when it drives over the Eurobalise. An on-board computer equipped with ETCS uses this information to determine the maximum speed, maximum distance and the braking curve and can act when these values are exceeded.

As a train has to be positioned directly above the Eurobalise to obtain the information, it may occur that a train stops on the track when there is no Eurobalise present. To prevent this from happening, trains stop just before a Eurobalise where a so-called Euroloop is installed. This Euroloop sends a continuous signal to the train to inform when the next Eurobalise will provide the train with new movement authority information. These Euroloops are typically tens of meters long to ensure data transfer with a stopped train. Alternatively, GSM-R or signals alongside the track can be used instead of Euroloops. When the existing signalling system remains in place a ‘dual signalling’ system can be realized, which enables the mixed use of trains with and without ERTMS equipment on the same corridors.

Level 2.
With the extension to ERTMS Level 2, information about movement authority is not send from the IXL to a LEU but to the RBC (Radio Block Centre) which continuously monitors the positions of trains (Li, 2012). This way the RBC can send the correct data to the corresponding train, which is done via GSM-R. Again, the train is equipped with ETCS to maintain the correct maximum speed and braking curvature. The train sends the location of the front to the RBC via GSM-R, but this data does not provide information on train integrity (in case of multiple carriages) and thus the IXL does not know if a block is free for the succeeding train. To obtain this piece of information, trackside equipment is still necessary in ERTMS Level 2. The benefit in comparison with Level 1 is faster communication and less trackside equipment (train integrity detectors are needed instead of Euroloops). Eurobalises are still used in Level 2 as it functions as a system to check the digital data and as a division of the tracks in blocks (Bersani, Qiu, Sacile, Sallak, & Schön, 2015).

Hybrid Level 3.
With ERTMS Hybrid Level 3, the fixed block security system is eliminated and replaced with so-called ‘moving blocks’ (Li, 2012). In this level, trains also have on-board equipment that determines train integrity, so the trackside detection equipment is not necessary anymore. For most passenger trains of NS this is already realized (since they cannot be split) and the rest of the fleet is easily adapted as well. This may be more difficult for freight trains and other trains with separate wagons. By sending information about location, speed, acceleration and train integrity, the RBC knows which section of the track is free at all times (Bersani, Qiu, Sacile, Sallak, & Schön, 2015). As a result, succeeding trains do not have to wait before a whole block is free but rather a section of the track. Consequently, we obtain moving blocks as well as shorter block distances, which has a positive effect on the headway planning norm. The reason that this level is often referred to as ERTMS Hybrid Level 3, is that the infrastructure that operates with Hybrid Level 3 should also be usable by trains that are not equipped with Hybrid Level 3 technology as well. This will absolutely be necessary during a transition period, before all trains are upgraded, but may also be obligatory when other train operators have not upgraded their trains yet to ensure interoperability of the railway network.

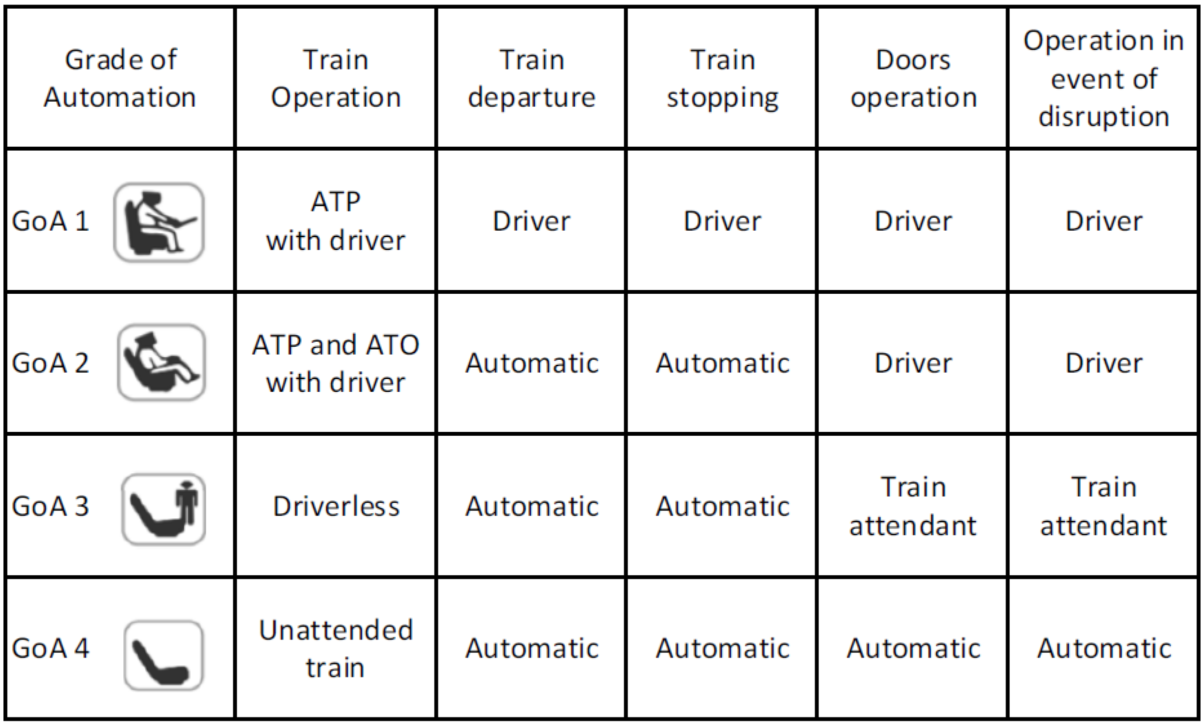
In the Netherlands, ERTMS Level 2 is currently being installed at corridors where this is deemed beneficial and in the coming years more corridors will follow. As briefly mentioned above, with ERTMS Level 2, trains will follow a more automated process to maintain a maximum speed and follow a calculated braking curve (Li, 2012). The first aspect, maintaining maximum speeds, makes running times more predictable, decreasing the running time supplement as a result. Following a calculated braking curve also makes the running time more predictable but a more significant effect is the possibility to brake later and faster. This way trains can maintain their maximum speeds for a longer period of time and over a longer distance, clearing the blocks faster and thus making more room for other trains. A third aspects relates to the faster communication with the IXL. Since information about freeing a block is provided to the IXL faster than before, succeeding trains will also be provided with a new movement authority faster. This allows trains to maintain a shorter headway. The corresponding change in the planning norms by implementing ERTMS Level 2 is a shorter running time, a shorter running time supplement and a short headway.

ERTMS Hybrid Level 3 offers the added benefit of moving blocks (under Level 3) (Bersani, Qiu, Sacile, Sallak, & Schön, 2015) as well as shorter block distances (under Hybrid Level 3) (Li, 2012). This can also be realized under Level 2, but this will require a lot more investments as more equipment is necessary to check if a certain block is cleared (because more blocks are created). Moving blocks are virtual blocks that correspond to the position of the train as well as the absolute braking distance (Li, 2012). They are designed in such a way that the distances between trains are large enough to ensure a safe braking distance. Since the position and the speed of the train change, the block in front of the train to realize this absolute braking distance also changes continuously. Furthermore, shorter block distances can be used without an increase in investment costs since this can be done virtually. Both aspects of ERTMS Hybrid Level 3 make it possible to decrease the headway.

For both ERTMS levels, implementation is bounded by the number of available training days for train drivers, just as is the case with DCO. Other investments include the Eurobalises although these investments can be avoided by combining ERTMS Level 2 with ERTMS Hybrid Level 3 as the latter does not need the Eurobalises to function (TwynstraGudde, 2022). Furthermore, new equipment for ETCS is necessary on the trains as well as the design of the new security system on every block.

3.1.3. Automatic Train Operation (ATO)

Automatic Train Operation or ATO makes all processes of a train automatic (driving and dwelling). As is the case with ERTMS, there are several levels of ATO ranging from GoA1 to GoA4 (GoA = Grade of Automation), as shown in Figure 6 (Dimitrova & Tomov, 2021). GoA1 is the current level of ATO in the Netherlands, where all operations are carried out by a train driver. ATP in Figure 6 refers to Automatic Train Protection which is a system that automatically activates an emergency brake if the train driver disobeys a red or orange signal or when the trains speed increases above the maximum allowable speed. The current security system in the Netherlands is a form of ATP, as is ERTMS (in any level).



Grade of Automation	Train Operation	Train departure	Train stopping	Doors operation	Operation in event of disruption
GoA 1	ATP with driver	Driver	Driver	Driver	Driver
GoA 2	ATP and ATO with driver	Automatic	Automatic	Driver	Driver
GoA 3	Driverless	Automatic	Automatic	Train attendant	Train attendant
GoA 4	Unattended train	Automatic	Automatic	Automatic	Automatic

Figure 6: Grade of Automation (GoA) for Automatic Train Operation (ATO)

GoA2 elaborates on this system by automating the driving processes. The onboard computer receives information from trackside equipment (such as movement authorities and maximum speeds) to depart on time, drive to the next station and stop there again (Lieskovský, Myslivec, & Žemlička, 2020). A train driver is still present on the train to guard the computer and operate the train in case of an error or malfunction. The train driver carries the final responsibility for the driving behaviour of the train.

GoA2 can be expanded to GoA3 (no driver but still train personnel on board for door closing/opening and in case of malfunctions) and GoA4 (no personnel on board), but since implementation of these technologies nationwide is not expected within the next decade(s), we limit this research to GoA2. The major benefits in terms of planning norms are already realized at GoA2: optimal speeds and braking curves lead to smaller minimum running times and shorter running time supplements and real-time information about train positions and movement authorities reduce the need for longer headway norms. For this reason, NS also focuses on this level (TwynstraGudde, 2022).

ATO in railway operations are not new: in many cities, closed railway systems like subways and metros are already equipped with ATO (Lieskovský, Myslivec, & Žemlička, 2020). In the Czech Republic, ATO is also implemented on shared tracks with non-ATO trains since 1991 and more countries are starting to implement ATO on a larger scale, though still primarily in closed circuits (Sigger, 2023). To realize ATO on corridors in the Netherlands, the ATP needs to be upgraded with an ATO module. ATO can be implemented on smaller networks or even single corridors as well as nationwide, but the latter would require a more automated process at the traffic management department (TwynstraGudde, 2022).

With the implementation of ATO on a corridor, running times become more predictable and also shorter (due to faster acceleration and deceleration). A more predictable process will allow for shorter buffers in the timetable design such as running time supplements and buffer times (TwynstraGudde, 2022). ATO also shortens braking distances, which results in shorter headways (Dimitrova & Tomov, 2021).

3.1.4. Traffic Management System (TMS)

Rail Traffic Management Systems (TMS) are designed to be able to manage all trains on the whole rail network by evaluating the real-time situation on the network (Davey, 2012). Currently, this is a manual task, which can only be carried out for a handful of trains on a lower level of detail. With TMS, it is possible to make real-time decisions that improve the timetable performance, based on the output of an algorithm. This algorithm is provided the actual information of the network, such as the position of the trains and their speed, the available infrastructure and the desired timetable. Based on this information, the algorithm can detect delays and conflicts and make decisions for the whole network (i.e., rerouting or delaying succeeding trains). Because all information is available, the logistic planning and alternative plans can be calculated automatically and smarter, making it possible to react in real-time and prevent further delays (TwynstraGudde, 2022). The biggest improvement that is possible with the implementation of TMS in the Netherlands is that it will be possible to plan in seconds whereas DONNA (see section 2.3.3) currently plans per six seconds (1/10th of a minute). Planning in seconds decreases the buffer time and running time supplements, providing extra time on a corridor for more trains (Davey, 2012). While this may provide major benefits for the timetable of the whole network, it is questionable whether the implementation of TMS will actually lead to planning in seconds. It would mean a change in or of timetable design software but more importantly, signals are changed every minute and not operated per second and passenger information is also provided in minutes. The actual improvement in the timetable design may therefore not have any influence on the capacity of the railway network. This effect is already visible in the current timetable design per six seconds, which is still dependent on the timetable in minutes for ProRail and passenger information.

Buffer times can also be decreased as the TMS is better at determining the course of action after a malfunction or delay. Fewer trains are affected with fewer delays as a result (TwynstraGudde, 2022). Since this can all be done in real-time, passenger information is also improved (Davey, 2012). The continuous analysis of the current state of the network and communication with the trains also allows for shorter headways between trains. To summarize, for the planning norms, the following improvements can be realized: shorter buffer times, running time supplements and headways.

The implementation of TMS does not ask for hardware investments such as trackside or onboard equipment. An algorithm needs to be programmed to safely guide all trains and this algorithm should be provided decision rules and processes in case of incidents and delays. Moreover, a fast communication system should be in place to realize real-time operations management. At the moment, NS uses a form of such a system called TIMTIM, but this system only advises the train driver based on its own position and timetable and disregards the rest of the network.

3.1.5. Other technologies and innovations

There are three other technologies that may be promising with respect to decreasing the planning norms at NS but are left out of the scope of this research. In this section, these technologies are briefly described and a reason for excluding them is provided.

Remote operations

In Remote Operations, we look at a specific implementation of ATO/GoA4. Instead of using GoA4 in every process of the timetable, we only use its functionality when shunting trains. GoA4 enables automatic shunting, which means that shunting trains can be done remotely, also enabling the parking of trains anywhere on the track (TwynstraGudde, 2022). Besides the promising nature of this specific aspect of GoA4, benefits in terms of planning norms are not foreseen, which is why this feature is excluded in this research. Moreover, improved passenger satisfaction is also not expected as it does not improve the timetable for the passengers.

FRMCS

The Future Railway Mobile Communication System (FRMCS) will be the successor of Global System for Mobile Communication – Railways (GSM-R). Where GSM-R uses 2G communication, FRMCS uses 5G which allows for faster communication (Sigger, 2023). The shift to FRMCS is mandatory since GSM-R will be gradually phased out by 2030. FRMCS will also become a mandatory communication system for Europe since this will also improve interoperability on the railway. Benefits of FRMCS are limited to faster communication, but this might be a necessary technology to implement other technologies that require fast communication, such as TMS and ATO. However, since FRMCS does not provide improvement of the planning norms it is left out of the scope of this thesis. Moreover, since rules about the design of FRMCS are unclear at the moment of writing, there are no companies that provide this service yet and thus the final effect of implementing FRMCS is still uncertain (TwynstraGudde, 2022).

3kV

3kV means that instead of using overhead wires that have a voltage of 1.5kV, the overhead wires are upgraded to have a voltage of 3kV. This will enable faster acceleration and higher speeds as well as a reduction in energy loss from the overhead wires (TwynstraGudde, 2022). Besides the necessary investments in the infrastructure (overhead wires and understations), also extra onboard equipment is needed. This equipment is much heavier, which also results in much heavier trains (Sigger, 2023). To accommodate the heavier trains, the infrastructure needs to be reinforced as well, which will take a long time (not expected to be finished before 2040). This makes 3kV an expensive technology to implement. So, while 3kV may be promising in terms of planning norms (shorter running times, headways and buffer times), the costs may not outweigh the benefits (TwynstraGudde, 2022).

3.1.6. Combination of technologies

In Table 2, the aforementioned technologies and their expected effect on the planning norms are summarised. The last three technologies in the table are left out of the scope of this research, which is indicated by the lighter text colour. The values in the table are derived from interviews with employees at NS. From this part of the literature review, it is clear that there is much information on the technical aspects of the technologies and the potential it has to decrease the planning norms. However, for numerical values, we still rely on expert judgement as there are no real-life (or simulation) experiments that back these expectations. For this reason, we assume that the effects can be added cumulatively for all different combinations. We assume values for the separate implementation of the planning norms, which comes with its own limitations on the model and introducing the combination effects of technologies may increase uncertainty in the input and therefore also limits the validity of the output even further.

Table 2: Technologies and their (expected) effect on the planning norms

Technology	Running time	Running time supplement	Headway	Dwelling time	Buffer time
DCO	Not affected	Decreases (1%) because of more reliable processes	Not affected	12 seconds shorter for Sprinters	Not affected
ERTMS Level 2	Decreases slightly because of higher speeds and shorter braking curves	Decreases (1%) because of more reliable processes	Decreases by 30 seconds because of faster movement authorities	Not affected	Not affected
ERTMS Hybrid Level 3	No additional benefits	No additional benefits	Decreases by 15 seconds because of moving blocks and/or shorter block distances	Not affected	Not affected
ATO	Decreases slightly because of faster acceleration and deceleration	Decreases by 2% for Intercitys and 3% for Sprinters because of more reliable processes	Decreases slightly because of continuous communication about positions and movements	Not affected	Decreases slightly because of more reliable processes
TMS	Not affected	Decreases by 1% because of the possibility to plan per second	Not affected	Not affected	Decreases by 15 seconds because of faster communication between trains and traffic management

Remote operations	*Not affected*	*Not affected*	*Not affected*	*Not affected*	*Not affected*
FRMCS	*Not affected*	*Not affected*	*Maybe necessary to obtain faster communication to enable TMS and ATO*	*Not affected*	*Not affected*
3kV	*Decreases because of faster acceleration and higher speeds*	*Not affected*	*Not affected*	*Not affected*	*Not affected*

3.2. Timetable design and optimization models

Train timetable design and train timetable optimization are often two separate processes that together lead to a final timetable that is implemented in practice. At NS, this is no different: timetable design is done with programs such as DONS and DONNA, while timetable optimization follows an iterative approach by using evaluation software like TRENO. Train timetable design is often a process that follows from expert knowledge to adapt new changes by manually updating the existing timetable, while the evaluation tools use input and output to determine the value of that design. In this section of the literature review, we introduce combinatorial models that design train timetables according to one or multiple objective functions to solve for optimality. This part is divided into four subsections, based on the type of modelling that is used: Deterministic models, Optimization techniques, Stochastic models and Real-Time Rescheduling models. These models cannot be used interchangeably but have different functions in the current practice of train timetable design and optimization.

3.2.1. Deterministic models

The first type of train timetable design models is called deterministic. This means that input values, or planning norms, are fixed as a design choice and given a set of constraints, the model will design a train timetable that satisfies them in such a way that we obtain an optimal objective value. Constraints in this case are given by modelling the planning norms and an objective function is created depending on the goal of the programmer. Examples of objective functions include number of trains or passengers per time period, travel times for the passengers, train operating time, cost of the timetable design and robustness of the timetable (Hansen I. A., 2009). Deterministic models are often solved using Mixed Integer Linear Programming (MILP; typically for small instances) or heuristics (when the network grows too big).

A disadvantage of deterministic models is that the input variables (planning norms) may limit a guaranteed feasible timetable and therefore, a thorough analysis is necessary to determine their values. Furthermore, since the Train Timetabling Problem (TTP; the umbrella term for deterministic models in railway timetabling) is NP-hard, computation times increase exponentially for bigger instances and therefore networkwide implementation cannot be solved via an MILP solver and efficient algorithms and heuristics need to be developed (Hansen I. A., 2009). The advantage is that input and output provide fixed values, which makes the model and its outcome easy to interpret. However, this is also a disadvantage as fixed values do not or barely incorporate stochasticity such as delays due to human interaction or malfunctions (Dotoli, et al., 2013).

Deterministic train timetable scheduling models can be divided into six subcategories in terms of periodicity, which can be found in Figure 7. The categories where most literature focuses on are periodic timetables, often referred to as the Periodic Event Scheduling Problem (PESP), and individual trip timetables. Partial periodic (some trains are periodic and other not) and individual trip (every train can be scheduled at any time without periodicity) timetables are outside the scope of this research since timetables are of cyclic design in the Netherlands. In the remainder of this section, four modelling types are explained and their corresponding applicability for NS is discussed.

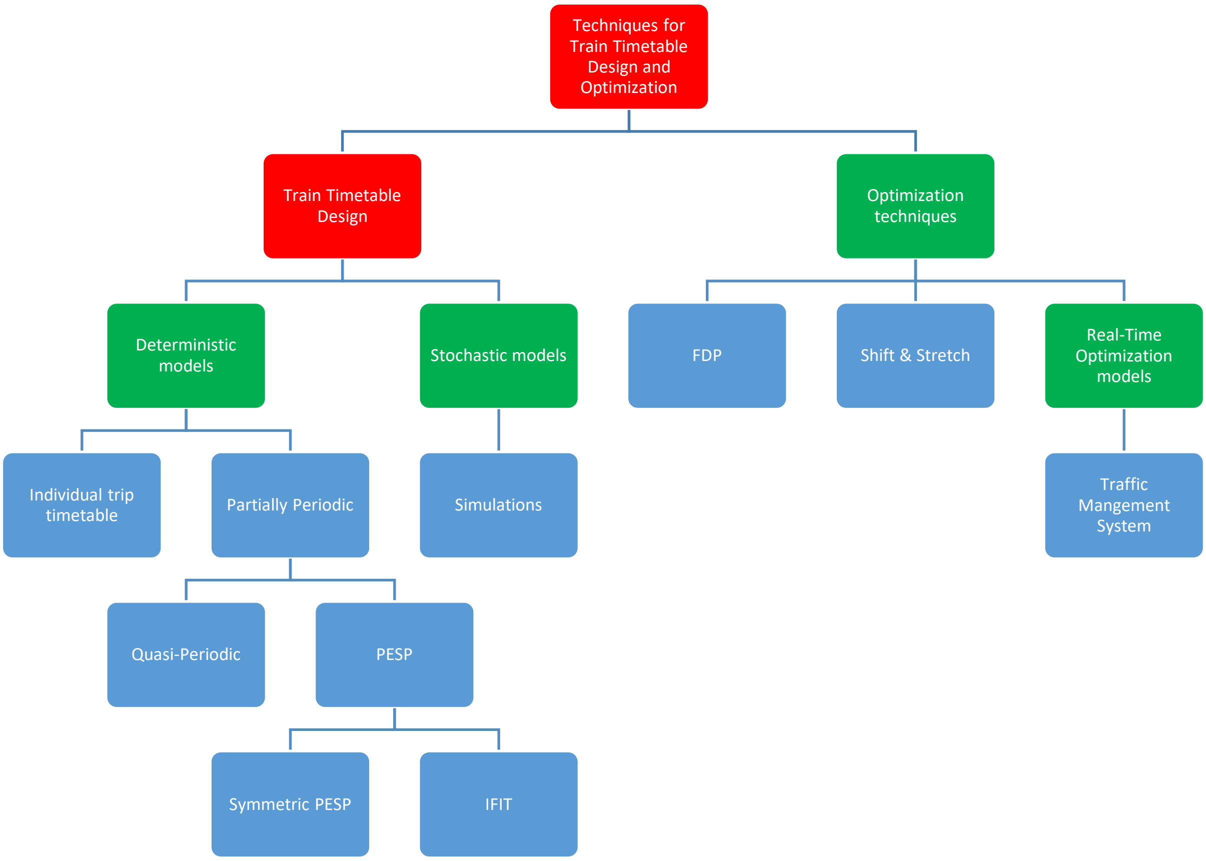


Figure 7: Timetabling techniques

Periodic Event Scheduling Problem (PESP)

The main discipline for train timetable design in the Netherlands focuses around PESP, which is why this modelling type is discussed first. It is widely used since periodicity improves passenger satisfaction due to clear and consistent timetables, e.g., passengers know that a train will depart every 15 minutes (Dotoli, et al., 2013). PESP is also the basis for DONS, the design tool for macroscopic timetables at NS. With DONS, basic hourly patterns are designed, which in turn form the basis for the final timetable. With PESP, only one predetermined pattern can be designed, which is assumed to run indefinitely. This is often assumed to be one hour. However, at night and outside of peak hours, some trains may run on a lower frequency, resulting in a different pattern per hour. Furthermore, we do not want to create timetables where trains run every 7 minutes, for example, as this would not fit within a cycle of one hour. When a daily timetable will then be made, extra trains are necessary to start the new hourly pattern, because the cycle is not finished yet.

While PESP is often modelled to minimize travel times and waiting times or maximize frequencies and passenger kilometres, another approach can be to maximize robustness and stability, which is done in the research by Sparing and Goverde (2017). Another option is to make the objective function dependent on the number of passengers on that specific corridor. This can be done by using the Origin Destination Aware PESP (ODPESP) (Siebert & Goerigk, 2013). This way the travel time or waiting time in the objective function can be weighted over the number of passengers and the optimization can be done, calculating the impact on the largest number of passengers.

Symmetric timetable

To simplify the differences depending on the time of the day, a specific instance of PESP, symmetric PESP, is developed. A symmetric train timetable means that on the same minute of every hour a train or a set of trains that operate on the same corridor will have the exact same position and the opposite direction deviates by exactly the same amount of minutes (Caimi, Kroon, & Liebchen, 2017). For example, if a train departs from a station in one direction at 8.20, the train in the opposite direction will depart at 8.40, both deviating twenty minutes from the whole hour (or ten minutes from the half hour). A direct result from this type of modelling is the fact that all trains will run on an hourly pattern. This is also what happens in DONS, where the basic hourly pattern shows cyclic patterns for all train sets with a least common multiple of 60 minutes (or one hour).

In Figure 8, a symmetric DONS output is shown together with the symmetry line (the dashed line at 8.30). In the figure, you can clearly see that the train paths in one direction (top left to bottom right: Groningen-Zwolle), deviate exactly as much from the symmetry line as the train paths in the other direction (top right to bottom left: Zwolle-Groningen), with some exceptions for freight trains and other irregular paths. For instance, trains A8100/2 and B8100/3 (indicated by the arrows), that follow the same route but in opposite directions, depart from Meppel at 8.22 and 8.38, respectively and thus both deviate exactly 8 minutes from the half hour symmetry line. This behaviour still remains when we look at other stations in the train routes.

Quasi Periodic Event Scheduling Problem

In the Quasi Periodic Event Scheduling Problem (QPESP), the periodicity is not fixed for (a subset of) the trains (Sartor, Mannino, Nygreen, & Bach, 2023). Where in PESP, trains are scheduled exactly 15 minutes apart, in QPESP, the periodicity is maintained (4 trains per hour), but the intervals may differ (e.g., intervals of 14 and 16 minutes). NS also designs the timetable in this way, mostly on the corridors with high frequencies. The necessity for this type of modelling lies in the fact that periodicity comes at the cost of railway capacity, while it improves passenger convenience (Sartor, Mannino, Nygreen, & Bach, 2023). Because passenger convenience is an important factor in timetable design, the objective function of QPESP penalizes the variations in interval times to provide the most benefit for the passengers. Another disadvantage is the longer computation time of QPESP, because a higher degree of freedom is realized by relaxing the periodicity constraints (Sartor, Mannino, Nygreen, & Bach, 2023). While QPESP can be seen as a partially periodic timetable, it has a higher degree of periodicity and approaches PESP as we assume periodicity as much as possible.

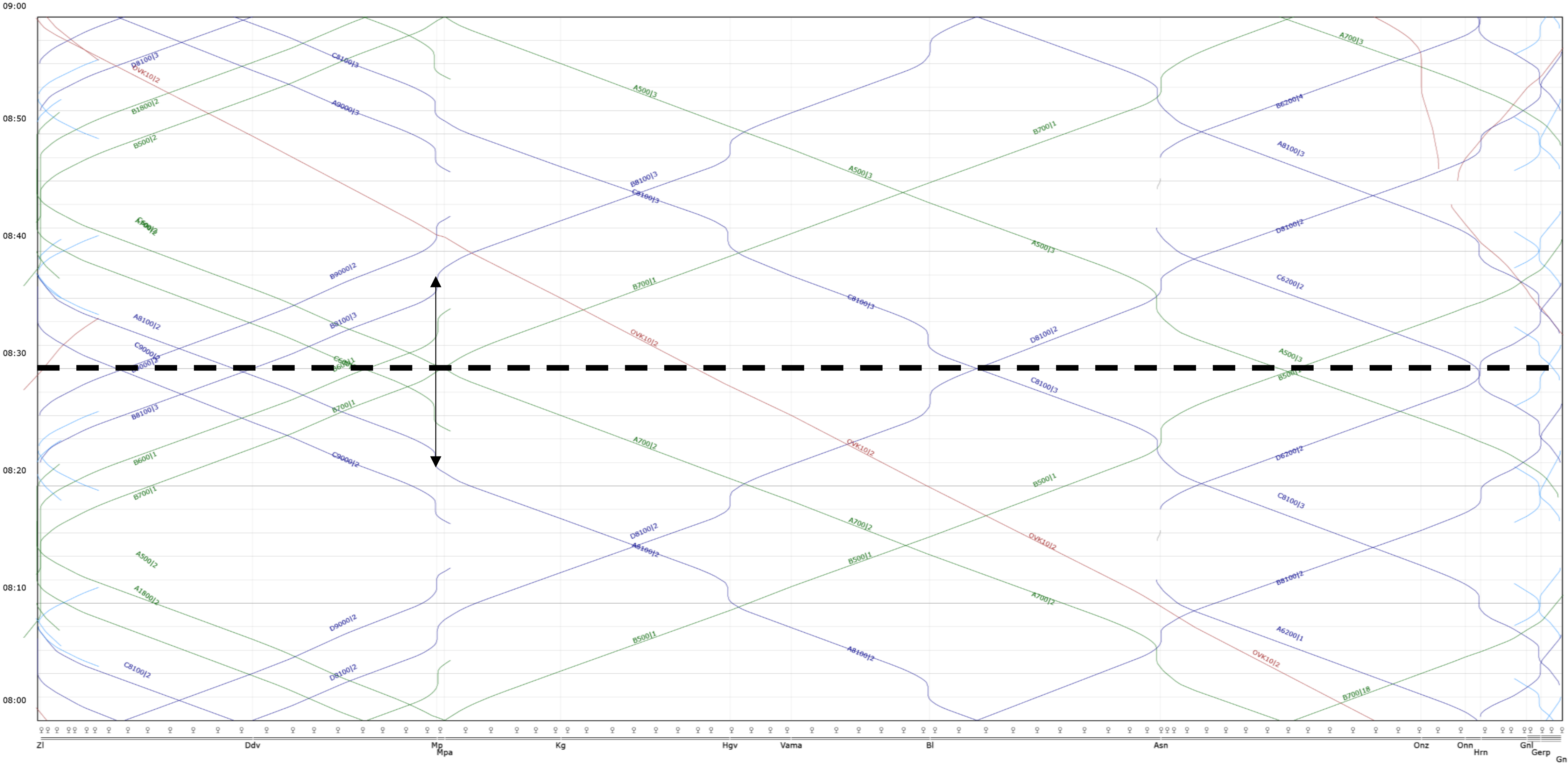


Figure 8: Symmetric output diagram of DONS (symmetry around the dashed line) for the network between Zwolle (Zl) and Groningen (Gn)

Integrated Fixed Interval Timetable

IFIT (Integrated Fixed Interval Timetable) is an extension of the symmetric timetable where trains of different lines meet each other on so-called IFIT-hubs (Caimi, Kroon, & Liebchen, 2017). This way periodicity is ensured on these IFIT-hubs and therefore elsewhere in the network. The difference with symmetric timetabling is that the timetable is not symmetric on time instances but on locations, while this often results in symmetry on time instances as well. The advantage of such a system is that passengers are offered excellent transfer possibilities, but disadvantages include longer travel times and capacity problems on IFIT-hubs that cannot accommodate the number of trains on the different lines (Caimi, Kroon, & Liebchen, 2017). As a result, this is generally not seen as a promising design approach.

3.2.2. Optimization Techniques

Until now, we have discussed various models that design train timetables themselves, but there are also optimization techniques that are used to reduce the computation time or that can change existing train timetables to more optimal ones or infeasible designs to feasible designs. These are discussed in this section.

Feasible Differential Problem

In the Feasible Differential Problem (FDP), some parts of the train timetable are fixed. For instance, we may fix the arrival time of one particular train in the morning. This limits the solution space of the problem and therefore, significantly reduces the computation time. On the other hand, by fixing the arrival time of one particular train on one particular moment of the day, you fix the arrival time of the trains in that train series, due to periodicity (Caimi, Kroon, & Liebchen, 2017). As a result, the final timetable may be less efficient and optimal than a model with a larger degree of freedom.

Shift and Stretch

Shift and Stretch is an optimization technique that can optimize an existing train timetable or make an infeasible train timetable feasible. In this type of modelling, trains may be shifted, where the whole operation of the train is delayed with a certain amount of time, which is often done to create a larger buffer between separate operations (Caimi, Kroon, & Liebchen, 2017). The other option, stretching, adds extra time to part of the operation, for example, the running time between two specific timetable points or stations. This is also done to create a larger buffer, but in this case, between two subprocess of separate operations (for instance, extra time between two arrivals of two different trains on the same station). In some extreme cases, trains can be deleted from the timetable when no feasible timetable can be found (Caimi, Kroon, & Liebchen, 2017).

3.2.3. Stochastic models

A disadvantage of deterministic models is the inability to account for delays and varying processing times. In stochastic models (such as simulations), these variations can be modelled, leading to a better understanding of the robustness and stability of the train timetable. Another function of stochastic models is to estimate waiting times and capacity consumption of corridors and stations (Hansen I. A., 2009). While deterministic models try to exclude the possibility of delays and knock-on delays (delays of succeeding trains) through the use of planning norms such as buffer times and running time supplements, there always remains a part of stochasticity that can only be modelled by using stochastic modelling.

The disadvantage of using a stochastic model is that they require more computation time to detect conflicts on a microscopic level and that it requires manual interaction with the model to solve these conflicts based on blocking time graphs (Hansen I. A., 2009). Furthermore, speed variations and behaviour of train drivers are difficult to model and are therefore not widely incorporated in most stochastic models yet. The microscopic nature, high computation times, high complexity and necessity for the effects of manual interaction is the reason that stochastic models are not considered feasible options to solve the research question of this thesis.

3.2.4. Real-Time Rescheduling

Another type of optimization models is called real-time rescheduling. In this case, a train timetable design has been made and is currently operated on the network, but new information, such as incidents or delays, make us decide to make real-time alterations to minimize the knock-on delays and maintain punctuality as much as possible. Traffic Management Systems are a form of real-time rescheduling that can be implemented in practice. Since this is not something you can plan on the long-term and does not involve the design, but only the changing of a timetable, these models will not provide a suitable solution for this research. However, it may be possible that heuristics from these optimization models, can also be used in deterministic modelling. For example, Tabu-search algorithms can be used to escape local optima when finding the optimal timetable (Hansen I. A., 2009).

3.3. Bottleneck identification methods

It seems straightforward to solve the bottlenecks to improve and optimize a train timetable, because they are often a restrictive factor on the train timetable design. However, as we find later in this research, it may be possible to solve bottlenecks by implementing technologies on another corridor elsewhere in the network. Still, for this research, it may be beneficial to understand how we can identify bottlenecks in a train timetable. While there is no consensus about the definition of a bottleneck in train timetabling, there are two main approaches to identify a bottleneck. One focuses on the characteristics of the corridors and block sections and the effect on the whole network, while the other evaluates the delays of trains on the block sections. In this section, both strategies are explained.

3.3.1. Operational risk index

The operational risk index is “the expected value of the negative impacts caused by the occurrence of disturbances on the specific block section” (Zhao, Martin, Cui, & Liang, 2017). In other words, the operational risk index (RI) is calculated by looking at all the negative impacts that are caused by a delay on that block section. To determine the negative impacts, the total weighted waiting time as a result of the disturbance is calculated. Because disturbances cannot be predicted or planned, this is done by simulating different types of disturbances ranging in magnitude and inserting them in an existing train timetable. The average of all these scenarios is then normalized to obtain the final RI for every block section. This results in a final list that sorts all block sections from highest to lowest RI that easily shows which block sections are the biggest bottlenecks for the whole system (Zhao, Martin, Cui, & Liang, 2017). While this approach may help to identify the bottlenecks in the network when a delay occurs, it does not say anything about the restrictions the block sections lay on the network timetable design. Besides, we do not know which actions should be taken to solve the bottleneck and if this would result in an improvement of the timetable.

3.3.2. TNV-Conflict and TNV-Statistics

Where the operational risk index focuses on off-line analysis and focuses on block section characteristics, the combination of TNV-Conflict and TNV-Statistics allows for analysis of realized train timetables and their actual arrival and departure times as well as blocking times. Here, both software systems are described, but it should be noted that this is only in place in the Netherlands and is therefore not applicable to other countries as well. For more information on the implementation of the systems, we refer to the existing literature.

First, we start with the tool TNV-Conflict. The name TNV comes from the TNV-Systems, which contains TNV-logfiles. TNV-logfiles log all infrastructure messages and generated train number messages chronologically (Goverde & Meng, 2011), where TNV stands for TreinNummer-Volgsysteem (Train Number Following System). TNV-Conflict is a data mining tool which uses the TNV-logfiles to analyse information such as arrival, departure and blocking times. Recall that the blocking time is the time that is reserved for one train on a single block.

Based in this information, TNV-Conflict provides route conflicts, conflicting trains, knock-on delays and blocking time diagrams (Goverde, Daamen, & Hansen, 2008). Most importantly, TNV-Conflict can be used to identify structural conflicts that are the result of a fault in the timetable design, which can either be because of infeasible headways, too small buffers or even excessive running time supplements (Goverde, Daamen, & Hansen, 2008). This is particularly interesting since a big part of the problem in conflict detection lies in the data gathering. Measurements are often only done at station level, which makes it hard, if not impossible, to analyse the location and cause of the delay (Goverde & Meng, 2011). Consequently, train timetable designers and analysts are unable to produce solutions to prevent the delays.

A limitation to TNV-Conflict is that a train timetable designer still has to interpret the output of TNV-Conflict and apply their own expert knowledge to identify the most significant bottleneck(s). To solve this, TNV-Conflict received an add-on called TNV-Statistics. TNV-Statistics can combine the output of TNV-Conflict into multiple lists, defining “the top signals with most conflicts, the top delayed trains, and the top delayed train lines” (Goverde & Meng, 2011). As a result, the analyst can easily see the most important and significant bottlenecks and start to solve these (by manually trying different solutions in the train timetable design software).

3.4. Literature gap

Since this literature review is designed around three separate topics, the literature gap can also be structured around this approach. Firstly, literature on planning norm decreasing technologies focus only on the assumed potential of these technologies. As most technologies are not implemented on large networks or researched in simulations, it is difficult to express the benefits in terms of numerical examples. While it would be beneficial to research the actual potential and benefits, this will take years, as this asks for implementation of the technology on one corridor, then multiple corridors and then the whole network. Therefore, we do not aim to give a numerical value to the technologies but rather design a model that can be provided the expected improvement and search for the most promising places to implement it. To investigate the chance of different values for the assumed benefits, a sensitivity analysis can be conducted using the model, containing multiple experiments with various input parameters for the expected time savings. As a result, we can show the benefits for the whole train timetable, based on the different combinations of expected time savings per technology and clearly see the range of the effects that implementing a technology may have. From this, it is already possible to quantify the timetable improvements by implementing new technologies based on the assumed benefits.

Second, we find that the literature on different train timetable design models is extensive. However, none of the models introduce the possibility to test the implementation of new technologies or changes in corridor characteristics. The models that were found in the literature require manual input for these changes in infrastructure and rolling stock, but this can be done automatically by changes in the model design. Therefore, this thesis may add to the existing literature by designing a model that combines existing optimization models and new variables, parameters and constraints to account for the implementation of new technologies. This leads to a model that can be used by all railway companies to optimize their timetables by implementing new technologies that decrease planning norms.

Finally, bottleneck identification is already done in two different ways: by focusing on the characteristics of the corridor and/or blocks or by focusing on the delays of the trains on corridors and/or blocks. The literature gap that was found in this section is that the current bottleneck identification methods still require expert knowledge to correctly identify and solve the bottlenecks. Identifying and solving bottlenecks should be done simultaneously without the requirement of expert knowledge to optimize the train timetables. Only this way will the design department know which bottlenecks exist, how these can be solved and what the actual improvement of the timetable is. The model from this thesis integrates the bottleneck identification into the train timetable design process to simultaneously improve capacity utilization. This fills the gap in the literature and makes way for a train timetable design process that focuses on the potential of the infrastructure and rolling stock instead of the capacity limitations.

3.5. Conclusion of the literature review

This literature review was divided into three topics: planning norm decreasing technologies, train timetable design and optimization models, and bottleneck identification methods. In the first section, we have introduced four technologies that can be implemented in either the rolling stock of NS or the infrastructure to be able to decrease the planning norms: DCO, ERTMS, ATO and TMS. We have also identified three other technologies that were left out of the scope of this research paper. The literature that is available for these technologies currently only covers the technical aspects that are needed to implement them and do not concern simulations or real-life experiments to determine the effects on timetable designs. This stresses the importance of good assumptions for the input for our model. Moreover, the model from this research can be used as a tool to determine the expected benefits for the timetable corresponding to different input values for the technologies. This way, we will know what the benefits for the timetable are if the planning norms cannot be decreased as much as we assumed initially.

Second, we reviewed the available literature on train timetable design and optimization models. Deterministic models, such as PESP or symmetric PESP show similarities with the current timetable design process at NS, but a useful objective function needs to be developed to fulfil its full potential. Furthermore, existing models can only solve smaller instances for subway lines or small networks (cities or regions) with fewer trains operating on them than on the Dutch national network. Our model is a combination of the existing techniques to develop train timetables and an extension to solve larger train timetable problems. We have also looked at stochastic models and their usefulness for this research. Stochastic models include more variation in process times and intervals. However, they often come with high computation times and high complexity, which is why they are more useful for the evaluation of train timetables. The third type of model was identified as Real-Time Rescheduling, which is also left out of the scope of this paper, since this focuses on operational planning instead of strategic or tactical planning.

Finally, we described two methods to identify bottlenecks in either the train timetable design, through the use of operational risk analysis, or bottlenecks in the realized train timetable, by using TNV-Conflict and its add-on TNV Statistics. Both approaches result in a list of bottlenecks, but it is still up to the designer to assess the bottlenecks and to provide solutions by manually trying them out in the design software.

In the remainder of this paper, a deterministic (MIQCP) model is introduced which can identify and solve bottlenecks that affect the network timetable, while aiming to maximise the benefits for the timetable of the whole network in terms of passenger travel times.

4. Formal description of the model

In the previous chapter, we found that literature does not provide us with a model that has the potential of identifying corridors where planning norm decreasing technologies can be implemented to improve the train timetable of a bigger networks. Here, we propose such a model, by extending an existing PESP model (Dotoli, et al., 2013) for the implementation of new technologies. The model we propose is a Mixed Integer Quadratically Constrained Program (MIQCP), which means that some or all variables must be integer and that there are quadratic constraints present in the model. As mentioned, the PESP model is a Train Timetabling Problem, which is considered NP-hard, so computation times increase exponentially for bigger instances. Since this model includes quadratic constraints the computation time also increases compared to Mixed Integer Linear Programs (MILP). As a result, networkwide implementation cannot be solved via a MILP solver and efficient algorithms and heuristics need to be developed (Hansen I. A., 2009). For smaller instances, the proposed model can be solved in polynomial time. First, the sets, indices, parameters and decision variables are introduced, after which the model itself is discussed. In this chapter, we provide a mathematical formulation of the model, while Chapter 5 focuses on the specific case of solving this model for the Dutch railway network.

4.1. Sets, Indices, Parameters and Decision Variables

To understand the formulation of the model and its restrictions, first, the notation of the model must be defined. While there are many similarities with PESP and other Train Timetabling Problems, we provide the full notation here so that there is no room for misinterpretations in the formulation of the model itself. The sets, indices and parameters are given here but the corresponding data sets and numerical values that are used in the experiments are discussed in Chapter 5.1.

Sets	
$\boldsymbol{T}$	Set of all train series present in the chosen time period and region
$\boldsymbol{T_{INT}} \subseteq \boldsymbol{T}$	Set of Intercity trains
$\boldsymbol{T_{SPR}} \subseteq \boldsymbol{T}$	Set of Sprinter trains
$\boldsymbol{S}$	Set of all station codes that lie in the chosen region
$\boldsymbol{R_t} \subseteq \boldsymbol{S}$	Set of station codes that lie on the route of train $t \in T$
$\boldsymbol{Stop_t} \subseteq \boldsymbol{R_t}$	Set of station codes in the stopping pattern of train $t \in T$
$\boldsymbol{N_s} \subseteq \boldsymbol{S}$	Set of direct neighbour stations of station $s \in S$
$\boldsymbol{P_{s,n}}$	Set of tracks from station $s \in S$ in the direction of $n \in N_s$
$\boldsymbol{P_{s,0}}$	Set of tracks on station $s \in S$
$\boldsymbol{U_s}$	Set of platforms on station $s \in S$
$\boldsymbol{Q_{s,n,p}} \subseteq \boldsymbol{P_{n,0}}$	Set of tracks on station $n \in N_s$ that can be reached from track $p \in P_{s,0}$
$\boldsymbol{C} = \{\boldsymbol{DCO},$ $\boldsymbol{ERTMSL2}$ $\boldsymbol{ERTMSHL3},$ $\boldsymbol{ATO}, \boldsymbol{TMS}\}$	Set of technologies that can be implemented

In the first set, the term train series is introduced. Train series refers to a single line that runs periodically. At NS train series are indicated by a letter (A, B, C, etc.) depending on the direction and frequency. For instance, a line that runs four times in one hour will have four train series with letters A, B, C and D. This letter is then followed by a series specific code to indicate the line. Similarly, stations have their own name, abbreviation and code and because of programming convenience, this model uses station codes instead of names. For every train series, there is a route and a stopping pattern. The latter only includes the stations where the train series stops, while the route also includes stations that are passed. Tracks and platforms have their own codes to distinguish them as well.

Indices	
$\boldsymbol{t \in T}$	Train series index
$\boldsymbol{b \in T}$	Preceding or succeeding train series index
$\boldsymbol{s \in S}$	Station index (indicated with the station code)
$\boldsymbol{n \in S}$	Direct neighbour station index (indicated with the station code)
$\boldsymbol{c \in C}$	Technology index

The indices for preceding or succeeding train series are necessary for the constraints where the headways between two trains are determined and where the order of trains on tracks and platforms is set.

Parameters	
$\boldsymbol{dMin_{t,s}}$	Minimum dwelling time of train $t \in T$ at station $s \in Stop_t$ in seconds
$\boldsymbol{dMax}$	Maximum dwelling time in seconds (true for all trains and stations)
$\boldsymbol{sd_{DCO}}$	Expected time savings in seconds for the dwelling time when implementing DCO
$\boldsymbol{r_{t,s}}$	Minimum running time in seconds for train $t \in T$ to station $s \in R_t$ from the previous station in R_t
$\boldsymbol{rMax}$	Maximum running time factor (true for all trains and corridors) (percentage)
$\boldsymbol{rSup}$	General running time supplement (true for all trains and corridors) (percentage)
$\boldsymbol{sSupINT_c}$	Expected percental time savings for the running time supplement when implementing technology $c \in C$ in an Intercity train
$\boldsymbol{sSupSPR_c}$	Expected percental time savings for the running time supplement when implementing technology $c \in C$ in a Sprinter train
$\boldsymbol{hTrack}$	Headway in seconds between two trains on the tracks
$\boldsymbol{hPlat}$	Headway in seconds between two trains on the platforms
$\boldsymbol{shTrack_c}$	Expected time savings in seconds for the headway between two trains on the tracks when implementing technology $c \in C$ on that track
$\boldsymbol{shPlat_c}$	Expected time savings in seconds for the headway between two trains on the platforms when implementing technology $c \in C$ on that station
$\boldsymbol{Period}$	Duration of the time horizon for periodicity of the model in seconds
$\boldsymbol{Demand_{s,n}}$	Average number of passengers travelling from station $s \in S$ to station $n \in N_s$, per train
$\boldsymbol{Dwelling_s}$	Average number of passengers per train that do not alight at station $s \in S$ because they are traveling to another station
$\boldsymbol{MaxInv}$	Maximum budget that can be used to implement technologies on a train series or on a corridor
$\boldsymbol{K_c}$	Cost factor for the implementation of technology c on a corridor (for $c \in C - \{DCO\}$) or on a train series for $c = DCO$.
$\boldsymbol{B_t = \lvert R_t \rvert - 1}$	Number of corridors that train $t \in T$ travels in its route
$\boldsymbol{M}$	Used to operate conditional constraints.

The model includes two parameters for the maximum running and dwelling times. This is necessary because of periodicity. To exemplify this, suppose that we have a period of one hour and the departure of a specific train from the first station is scheduled at ten minutes in this period. If there would not be a maximum running time, the model could set the arrival time of the same train for the second station to nine minutes in the period, showing a running time that is lower than the minimum running time, namely minus one minute. This is not possible in reality of course, which is why the maximum running and dwelling times are necessary in this model.

Decision variables	
$\boldsymbol{a_{t,s}} \in \mathbb{Z}$	Arrival time of train $t \in T$ at station $s \in S$
$\boldsymbol{d_{t,s}} \in \mathbb{Z}$	Departure time of train $t \in T$ from station $s \in S$
$\boldsymbol{ya_{t,s}} \in \mathbb{Z}$	Cycle restricted arrival time of train $t \in T$ at station $s \in S$
$\boldsymbol{yd_{t,s}} \in \mathbb{Z}$	Cycle restricted departure time of train $t \in T$ from station $s \in S$
$\boldsymbol{ka_{t,s}} \in \mathbb{Z}$	Auxiliary variable for the creation of the cycle restricted arrival time of train $t \in T$ at station $s \in S$
$\boldsymbol{kd_{t,s}} \in \mathbb{Z}$	Auxiliary variable for the creation of the cycle restricted departure time of train $t \in T$ from station $s \in S$
$\boldsymbol{YA_{t,b,s}} = \begin{cases} \mathbf{1} \\ \mathbf{0} \end{cases}$	If train $t \in T$ arrives later than train $b \in T - \{t\}$ on station $s \in R_t \cap R_b$ Otherwise
$\boldsymbol{YD_{t,b,s}} = \begin{cases} \mathbf{1} \\ \mathbf{0} \end{cases}$	If train $t \in T$ departs later than train $b \in T - \{t\}$ from station $s \in R_t \cap R_b$ Otherwise
$\boldsymbol{YAV_{t,s}} = \begin{cases} \mathbf{1} \\ \mathbf{0} \end{cases}$	If $ya_{t,s} \geq yd_{t,s}$ because of periodicity Otherwise
$\boldsymbol{X_{t,s,n,p}} = \begin{cases} \mathbf{1} \\ \mathbf{0} \end{cases}$	If train $t \in T$ uses track $p \in P_{s,n}$ Otherwise
$\boldsymbol{X_{t,s,0,p}} = \begin{cases} \mathbf{1} \\ \mathbf{0} \end{cases}$	If train $t \in T$ uses platform $p \in P_{s,0}$ Otherwise
$\boldsymbol{Dir_{s,n,p}} = \begin{cases} \mathbf{1} \\ \mathbf{0} \end{cases}$	If track $p \in P_{s,n}$ is used for departures in the direction of $n \in N_S$ Otherwise (used for arrivals)
$\boldsymbol{Z_{c,s,n}} = \begin{cases} \mathbf{1} \\ \mathbf{0} \end{cases}$	If technology $c \in C - \{DCO\}$ is implemented on the corridor between $s \in S$ and $n \in N_s$ Otherwise
$\boldsymbol{Z_{c,s}} = \begin{cases} \mathbf{1} \\ \mathbf{0} \end{cases}$	If technology $c \in C - \{DCO\}$ is implemented on station $s \in S$ Otherwise
$\boldsymbol{DCO_t} = \begin{cases} \mathbf{1} \\ \mathbf{0} \end{cases}$	If train $t \in T$ is outfitted with DCO Otherwise

Here, we make the distinction between technologies that are implemented on the corridors (ERTMS Level 2 and 3, ATO and TMS) and the technology that is implemented on a train (DCO). For this reason, the corridor specific technologies are indicated with station codes and DCO is indicated per train. For experiments where more technologies are researched, another possibility to make this distinction is to make two subsets of technologies, one for the corridors and one for the train, but it was chosen not to do that in this case, since there is only one technology that is implemented on a train. To determine the investment costs per corridor, we determine the number of corridors that a train travels on its route (see also the parameters of the model).

4.2. Objective Function

The objective function of the proposed model consists of two parts: (1b) the travel time on a certain corridor multiplied by the average number of passengers per train on that corridor, and (1a) the dwelling time on a station multiplied by the average number of passengers per train that do not alight the train on that station (because they are traveling to another station further along the route). The objective function represents the travel time of all passengers in the network, where travel time consists of the travel time between stations as well as the dwelling time on the stations itself. From the remainder of this thesis, we refer to the objective function as the in-vehicle time, which combines the travel time and the dwelling time for passengers.

Objective Function

$$\boldsymbol{Minimize} \qquad \sum_{t \in T} \sum_{s \in Stops_t} \boldsymbol{Dwelling_s * (d_{t,s} - a_{t,s})} \qquad (1a)$$

$$+ \sum_{t \in T} \sum_{n \in Stops_t, if\ s\ is\ predecessor\ of\ n} \frac{\boldsymbol{Demand_{s,n}}}{\boldsymbol{NrTrains_{s,n}}} * \boldsymbol{(a_{t,n} - d_{t,s})} \qquad (1b)$$

4.3. Model Constraints

The constraints are divided into groups with the same aim or topic within the model. We start with the initialization of the model to implement the basic rules for train timetable designs (dwelling time and running time). Then, we introduce the restrictions that make sure that all trains are assigned to a platform or track. These are followed by constraints concerning the headway between the trains, ending with the restriction for implementing the technologies (such as the investment budget).

Initialization Constraints

$$\boldsymbol{yd_{t,s} \geq 0} \qquad \forall t \in T, \forall s \in R_t \qquad (2)$$

$$\boldsymbol{yd_{t,s} \leq Period - 1} \qquad \forall t \in T, \forall s \in R_t \qquad (3)$$

$$\boldsymbol{ya_{t,s} \geq 0} \qquad \forall t \in T, \forall s \in R_t \qquad (4)$$

$$\boldsymbol{ya_{t,s} \leq Period - 1} \qquad \forall t \in T, \forall s \in R_t \qquad (5)$$

$$\boldsymbol{d_{t,s} = Period * kd_{t,s} + yd_{t,s}} \qquad \forall t \in T, \forall s \in R_t \qquad (6)$$

$$\boldsymbol{a_{t,s} = Period * ka_{t,s} + ya_{t,s}} \qquad \forall t \in T, \forall s \in R_t \qquad (7)$$

The cycle restricted departure and arrival times need to be scheduled within the time period (2-5). To develop departure and arrival times that can be used in the objective function (1), restrictions (6) and (7) are needed. The auxiliary variable restricts the two types of departure or arrival times to be exactly a multiple of the period duration apart from each other.

Dwelling time Constraints

$$\boldsymbol{d_{t,s} \geq a_{t,s} + dMin_{t,s} - sd_{DCO} * DCO_t} \qquad \forall t \in T, \forall s \in Stop_t \qquad (8)$$

$$\boldsymbol{d_{t,s} \leq a_{t,s} + dMax} \qquad \forall t \in T, \forall s \in Stop_t \qquad (9)$$

$$\boldsymbol{d_{t,s} = a_{t,s}} \qquad \forall t \in T, \forall s \in R_t - Stop_t \qquad (10)$$

The dwelling time at the stations must lie between the minimum dwelling time at the station, minus the time savings of DCO if it is implemented on that train (8), and the maximum dwelling time (9). When a station is not included in the stopping pattern of the train, the departure time equals the arrival time on that station (10).

Running time Constraints		
$a_{t,s} \geq d_{t,n} + \left(rSup - \sum_{c\in C-\{DCO\}}\left(sSupINT_c * Z_{c,n,s}\right) - DCO_t * sSupINT_{DCO}\right) * r_{t,s}$	$\forall t \in T_{INT}, \forall s \in R_t, \forall n \in N_s$, such that n is the direct predecessor of s in R_t	(11)
$a_{t,s} \geq d_{t,n} + \left(rSup - \sum_{c\in C-\{DCO\}}\left(sSupSPR_c * Z_{c,n,s}\right) - DCO_t * sSupSPR_{DCO}\right) * r_{t,s}$	$\forall t \in T_{SPR}, \forall s \in R_t, \forall n \in N_s$, such that n is the direct predecessor of s in R_t	(12)
$a_{t,s} \leq d_{t,n} + (1 + rSup) * rMax * r_{t,s}$	$\forall t \in T, \forall s \in R_t, \forall n \in N_s$, such that n is the direct predecessor of s in R_t	(13)

The running time between two stations must lie between the minimum and maximum travel time for that corridor. The minimum travel time is determined by the minimum running time plus a percentage of the minimum running time for the running time supplement and minus the savings in running time supplement through the implementation of new technologies, its value depending on whether the train is an Intercity or Sprinter (11-12). The maximum travel time is determined by the minimum running time plus the running time supplement as well as an extra maximum running time factor (13).

Track and Platform assignment Constraints		
$\sum_{p\in P_{s,n}} X_{t,s,n,p} = 1$	$\forall t \in T, \forall s \in R_t, \forall n \in R_t \cap N_s$	(14)
$\sum_{p\in P_{s,0}} X_{t,s,0,p} = 1$	$\forall t \in T, \forall s \in R_t$	(15)
$\sum_{u\in U_s} X_{t,s,0,u} = 1$	$\forall t \in T, \forall s \in Stop_t$	(16)
$X_{t,s,n,p} = X_{t,n,s,p}$	$\forall t \in T, \forall s \in R_t, \forall n \in R_t \cap N_s, \forall p \in P_{s,n}$	(17)

Every train must be assigned a track on the corridors and stations in its route (14-15). When a train stops at a station, a platform must be assigned to that train (16). Once a track on a corridor is chosen for a certain train, the same track must be chosen on the other side of that corridor (17). The reason for this last constraint is that corridors lead from one station to another and while in practice trains may be allowed to switch tracks between those stations, this model concerns a macroscopic timetable where such dynamics are not considered. Without this constraint, trains might switch tracks to avoid headway norms on either side of the corridor, but we cannot determine whether the headway norms were complied when switching from tracks. Therefore, we assume in this model that trains stay on the same track for the whole length of the corridor.

Order of Arrival and Departure Constraints		
$yd_{t,s} - yd_{b,s} \leq M * YD_{t,b,s}$	$\forall t \in T, \forall b \in T - \{t\}, \forall s \in R_t \cap R_b$	(17)
$ya_{t,s} - ya_{b,s} \leq M * YA_{t,b,s}$	$\forall t \in T, \forall b \in T - \{t\}, \forall s \in R_t \cap R_b$	(18)
$YD_{t,b,s} + YD_{b,t,s} = 1$	$\forall t \in T, \forall b \in T - \{t\}, \forall s \in R_t \cap R_b$	(19)
$YA_{t,b,s} + YA_{b,t,s} = 1$	$\forall t \in T, \forall b \in T - \{t\}, \forall s \in R_t \cap R_b$	(20)

This set of constraints is necessary to determine and restrict the order of arrivals and departures of two trains on the same station (17-20). This order is not predetermined but the model needs to know the order to ensure the headway norms between trains.

Headway Constraints for Tracks and Platforms

$$\boldsymbol{yd_{t,s} - yd_{b,s} - M * YD_{t,b,s} \leq Period - (Period + hTrack - \textstyle\sum_{c \in C-\{DCO\}} shTrack_c * Z_{c,n,s}) * (X_{t,s,n,p} + X_{b,s,n,p} - 1)} \quad \forall t \in T, \forall b \in T - \{t\}, \forall s \in R_t \cap R_b, \forall n \in N_s \cap R_t \cap R_b, \forall p \in P_{s,n} \tag{21}$$

$$\boldsymbol{ya_{t,s} - ya_{b,s} - M * YA_{t,b,s} \leq Period - (Period + hTrack - \textstyle\sum_{c \in C-\{DCO\}} shTrack_c * Z_{c,n,s}) * (X_{t,s,n,p} + X_{b,s,n,p} - 1)} \quad \forall t \in T, \forall b \in T - \{t\}, \forall s \in R_t \cap R_b, \forall n \in N_s \cap R_t \cap R_b, \forall p \in P_{s,n} \tag{22}$$

$$\boldsymbol{yd_{t,s} - yd_{b,s} - M * YD_{t,b,s} \leq Period - (Period + hTrack - \textstyle\sum_{c \in C-\{DCO\}} shTrack_c * Z_{c,n,s}) * (X_{t,s,0,p} + X_{b,s,0,p} - 1)} \quad \forall t \in T, \forall b \in T - \{t\}, \forall s \in R_t \cap R_b, \forall n \in N_s \cap R_t \cap R_b, \forall p \in P_{s,0} \tag{23}$$

$$\boldsymbol{ya_{t,s} - ya_{b,s} - M * YA_{t,b,s} \leq Period - (Period + hTrack - \textstyle\sum_{c \in C-\{DCO\}} shTrack_c * Z_{c,n,s}) * (X_{t,s,n,p} + X_{b,s,n,p} - 1)} \quad \forall t \in T, \forall b \in T - \{t\}, \forall s \in R_t \cap R_b, \forall n \in N_s \cap R_t \cap R_b, \forall p \in P_{s,0} \tag{24}$$

$$\boldsymbol{ya_{t,s} - yd_{b,s} - M * YD_{t,b,s} \geq -Period + (Period + hPlat - \textstyle\sum_{c \in C-\{DCO\}} shPlat_c * Z_{c,s}) * (X_{t,s,0,u} + X_{b,s,0,u} - 1)} \quad \forall t \in T, \forall b \in T - \{t\}, \forall s \in R_t \cap R_b, \forall n \in N_s \cap R_t \cap R_b, \forall u \in U_s \tag{25}$$

$$\boldsymbol{ya_{t,s} - yd_{b,s} - M * YA_{t,b,s} \geq -Period + (Period + hPlat - \textstyle\sum_{c \in C} shPlat_{c-\{DCO\}} * Z_{c,s}) * (X_{t,s,0,u} + X_{b,s,0,u} - 1)} \quad \forall t \in T, \forall b \in T - \{t\}, \forall s \in R_t \cap R_b, \forall n \in N_s \cap R_t \cap R_b, \forall u \in U_s \tag{26}$$

Headway norms apply to all trains on every track on every corridor (21-22) and every station (23-24) and all trains on every platform on every station (25-26). The headways between trains are determined based on the order of the trains that use the same part of the infrastructure (track and/or platform), the initial headway norms and the time savings when new technologies are implemented on a corridor or station.

Periodicity Constraints for Headways

$$\boldsymbol{ya_{t,s} - yd_{t,s} \leq M * YAV_{t,s}} \quad \forall t \in T, \forall s \in R_t \tag{27}$$

$$\boldsymbol{\textstyle\sum_{t \in T, if\ s \in R_t} YAV_{t,s} * X_{t,s,0,p} \leq 1} \quad \forall s \in S, \forall p \in P_{s,0} \tag{28}$$

$$\boldsymbol{ya_{t,s} - yd_{b,s} + M * \left(1 - YAV_{t,s}\right) \geq -Period + Period * (X_{t,s,0,p} + X_{b,s,0,p} - 1)} \quad \forall t \in T, \forall b \in T - \{t\}, \forall s \in S, \forall p \in P_{s,0} \tag{29}$$

$$\boldsymbol{ya_{b,s} - yd_{t,s} + M * \left(1 - YAV_{t,s}\right) \geq -Period + Period * (X_{t,s,0,p} + X_{b,s,0,p} - 1)} \quad \forall t \in T, \forall b \in T - \{t\}, \forall s \in S, \forall p \in P_{s,0} \tag{30}$$

Due to periodicity, it is possible that the cycle restricted arrival time is scheduled later than the cycle restricted departure time on the same station. To account for this effect, an auxiliary variable is introduced that indicates whether or not the departure time lies before the arrival time (27). This can only be the case for at most one train for every track on every station (28). Constraints (29-30) are additional constraints for train orders, comparable to constraints (23-24) but accounting for this periodicity effect.

Direction and Routing Constraints

$$\boldsymbol{X_{t,s,0,p} \leq \sum_{q \in Q_{s,n,p}} X_{t,n,0,q}} \qquad \forall t \in T, \forall s \in R_t, \forall n \in N_s \cap R_t, \text{ such that } s \text{ is direct predecessor of } n \text{ in } R_t \tag{31}$$

$$\boldsymbol{\sum_{t \in T, if\ s \in R_t\ and\ n\ is\ direct\ successor} \left((d_{t,s} + 1) * X_{t,s,n,p} \right) \leq M * Dir_{s,n,p}} \qquad \forall s \in S, \forall n \in N_s, \forall p \in P_{s,n} \tag{32}$$

$$\boldsymbol{\sum_{t \in T, if\ s \in R_t\ and\ n\ is\ direct\ successor} \left((a_{t,n} + 1) * X_{t,s,n,p} \right) \leq M * Dir_{s,n,p}} \qquad \forall s \in S, \forall n \in N_s, \forall p \in P_{s,n} \tag{33}$$

$$\boldsymbol{Dir_{s,n,p} + Dir_{n,s,p} = 1} \qquad \forall s \in S, \forall n \in N_s, \forall p \in P_{s,n} \tag{34}$$

Some tracks on stations are only accessible from specific tracks on neighbouring stations and constraint (31) introduces these connections. To avoid that one track can be used in both directions, constraints (32-34) are necessary. In available literature on PESP, there are models that consider parallel tracks with train orders and there are other models that consider parallel platforms, but this model combines these two to model networks with both parallel tracks and platforms, with restrictions in connectivity between them.

Technology and Budget Constraints

$$\boldsymbol{Z_{c,s,n} = Z_{c,n,s}} \qquad \forall c \in C - \{DCO\}, \forall s \in S, \forall n \in N_s \tag{35}$$

$$\boldsymbol{Z_{c,s} \leq Z_{c,s,n}} \qquad \forall c \in C - \{DCO\}, \forall s \in S, \forall n \in N_s \tag{36}$$

$$\boldsymbol{Z_{c,s,n} \leq Z_{ERTMSL2,s,n}} \qquad \forall c \in \{ERTMSHL3, ATO\}, \forall s \in S, \forall n \in N_s$$

$$\boldsymbol{\sum_{c \in C - \{DCO\}} \sum_{s \in S} \sum_{n \in N_s} \left(\frac{Z_{c,s,n}}{2} * K_c \right) + \sum_t \left(DCO_t * B_t * K_{DCO} \right) \leq MaxInv} \tag{37}$$

If a technology is implemented on a corridor, it should be done in both directions (35) and the stations can only benefit from the implementation of a technology when it concerns adjacent corridors (36). The number of investments is restricted to a budget (37).

Technological Constraints

$$\boldsymbol{Z_{ATO,s,n} \leq Z_{ERTMSL2,s,n}} \qquad \forall s \in S, \forall n \in N_s \tag{38}$$

$$\boldsymbol{Z_{ERTMSHL3,s,n} \leq Z_{ERTMSL2,s,n}} \qquad \forall s \in S, \forall n \in N_s \tag{39}$$

$$\boldsymbol{Z_{ATO,s} \leq Z_{ERTMSL2,s}} \qquad \forall s \in S \tag{40}$$

$$\boldsymbol{Z_{ERTMSHL3,s} \leq Z_{ERTMSL2,s}} \qquad \forall s \in S \tag{41}$$

In addition to the investment costs, there are also two technological restrictions. ERTMS Level 2 needs to be implemented before ATO or ERTMS Hybrid Level 3 can be implemented. For this reason, we add the following constraints to the model, where (38-39) refer to the corridors and (40-41) refer to the stations.

For this model, we exclude the possibility to change line routings and line frequencies. While they are a significant factor in the quality of a train timetable and affect passenger convenience, the aim is to improve the current timetable, which ultimately could also lead to the possibility of changing line routing and line frequencies. For the same reason we do not include train activities such as reversing, (un-)coupling and shunting, as well as transfer times for the passengers. For examples on these types of models, we refer to existing literature, such as Fuchs and Corman (2019) or Fuchs, Trivella and Corman (2022).

As mentioned, larger network sizes cannot be solved in polynomial time by using a standard MIP solver. For smaller network sizes we can find feasible and optimal solutions using a MIP solver with reasonable computation times, as is discussed in the next chapter.

5. Numerical study

In this chapter, the model from Chapter 4 is solved for the Dutch railway network. First, the network specific parameters are discussed as well as the assumed benefits for implementing specific technologies. Then, the model is solved for a single line in the national network, a medium-sized subnetwork and finally the national network for different sizes of the investment budget. To this extend, the model was coded in Python and solved by using the MIP solver Gurobi (version 9.5.0), which is also able to solve MIQCP models. The solutions to these experiments and other insights are discussed, after which extra experiments are conducted for the medium-sized subnetwork to draw more conclusions on isolating or combining technologies in different scenarios. The aim of these experiments is to see whether the model can be solved for different network sizes and scenarios, as well as to identify an investment strategy for implementing new technologies on corridors and trains in these different settings. Additional findings such as the percental improvement of the timetable are discussed too. In the final section of this chapter, general conclusions on the model and its applications are provided.

5.1. Input parameters

Before we can solve the new model for the Dutch railway network, we need to provide the input values for the parameters, insert the correct train series and model the infrastructure. Furthermore, we need values for the time savings per planning norm for each of the technologies that we wish to include in our model.

5.1.1. General parameters

We start this section by providing the values for the initialization parameters. These are general parameters that are not dependent on the timetable or the implementation of technologies. Their values can be found in Table 3. For the maximum dwelling time, we have taken a number of seconds that is sufficiently large such that the current timetable can be run in the model (i.e., there does not exist a train series that has a dwelling time greater than or close to the maximum dwelling time) , while being sufficiently small to account for excessive dwelling times that would never be used in practice. The maximum running time is configurated with the same idea to avoid trains with very low speeds on corridors. Both parameters need to be significantly smaller than the period to make sure that train series cannot have a negative dwelling or running time. The percentage for the running time supplement is the same as the one that is used in the timetable design process at NS. In reality, the timetable designers sometimes change this value in some exceptions, but we do not consider those cases in our model. The values for the headway norms are the general headway norms used by the timetable designers at NS. In practice, these norms vary depending on the type of activity that the two subsequent trains carry out (arriving, departing, short stops, passing, etc.).

The choice for the period that is used in the model is determined by the timetable that is used to setup the model (see section 5.1.2). Most models are based on a period of one hour, but in this case, we can decrease the number of variables by using a period of half an hour without excluding train series. The reason for this is that every train series has a frequency of two per hour (for a frequency of four per hour, the series is split up into two series with a frequency of two per hour). Lastly, we give a value to the parameter M that is sufficiently large to find feasible timetables, but small enough to decrease the feasible region and reduce the computation time.

Table 3: Input parameters

Parameters	
$dMax = 600$	Assumption
$rMax = 1.1$	Assumption
$rSup = 1.07$	Used in timetabling by NS (7%)
$hTrack = 180$	General headway norm
$hPlat = 240$	General headway norm on platforms
$Period = 1800$	Model specific parameter
$M = 10^5$	Necessary parameter for conditional constraints

5.1.2. Infrastructure and timetable specific parameters

For the initialization of the model, a current timetable was used as input. This timetable (of December 2023) is known to be a feasible timetable for a representative hourly pattern in the peak hours (not altered for specific events or situations). From this timetable, the individual train series could be identified as well as the routes, stopping patterns, minimum running times and the number of corridors that are included in the route. Note that these times are dependent on the type and length of the rolling stock that is fixed for this timetable. For this model, we excluded international and freight trains due to their irregular pattern and/or low frequencies. Using the available infrastructure data set, the station codes could be translated to the station names so that the input of and solution to the model would be better to interpret.

The minimum dwelling time is determined for every train type (Intercity or Sprinter) and every station. The general norm for the minimum dwelling time is 54 seconds for Intercitys and 42 seconds for Sprinters. Some stations are bigger than others and thus specific minimum dwelling times are necessary to ensure connections between trains. Moreover, these stations typically have high passenger demand and thus longer dwelling times are necessary for alighting and boarding the trains. These values are all extracted from the actual norms used in the timetable design process at NS. Using the realized passenger data, an average demand could be calculated for all corridors. Furthermore, the average number of passengers in a train when dwelling could be determined per station and per train. It should be noted that these values are extracted from realized data for one specific train timetable and that the timetable design influences the passenger demand.

Table 4: Characteristics of the model for different network sizes

	Single-line network	Medium-sized network	National network
Number of Trains	14	24	119
Number of Stations	9	37	263
Number of Corridors	8	38	287
Number of Binary Variables	3238	8482	77301
Number of Integer Variables	168	492	4016
Number of Linear Constraints	15211	43572	335846
Number of Quadratic Constraints	28875	78911	602757

With the input of the infrastructure, we can determine the size of the model for the different network sizes in this research, as shown in Table 4. As network size increase, so does the model size and as a result, computation times increase as well. Because of this effect, the single-line network was chosen as a starting point for the model that is a good representative for the rest of the network. This single-line network was then extended to a medium-sized network to research the solvability of the model for larger networks and gain more insights into network effects in the model. The national network shows all trains, stations, corridors and the corresponding number of variables and constraints in the model. All three network sizes are discussed in the upcoming sections.

5.1.3. Technology specific parameters

As mentioned in section 3.1.6, the time savings per planning norm and per technology are based on expert judgment as there are no real-life experiments (or simulations) that back these expectations. In Table 5, we recall the time savings as provided by interviewing these experts. Note that for TMS and ATO the headway reduction of 15 seconds follows from a 15 second reduction of the buffer time, but to include this effect in the model, we have assigned it to the headway norm, to better compare it with the time savings of ERTMS Level 2 and Hybrid Level 3. Furthermore, the time savings for the running time supplements follows partially from the decrease in the minimum running time, but for the same reason of comparing it to other technologies, it was chosen to model as time savings in the running time supplement.

Table 5: Expected time savings per technology

	Time savings per planning norm			
	Running time supplement (Intercity)	Running time Supplement (Sprinter)	Dwelling time (seconds)	Headway (seconds)
DCO	1%		12	0
ERTMS Level 2	1%		0	30
ERTMS Hybrid Level 3	0		0	15
ATO	2%	3%	0	15
TMS	1%		0	15

There are several ways to determine the investment costs for the technologies. The first way is to express a financial cost per technology. However, since these technologies are novel, it is uncertain what the financial costs will be for implementing them. Furthermore, it is unclear whether the financial costs are to be paid by NS or ProRail (or both). Therefore, we look at the investment costs in terms of the capacity to change. For example, NS has a limited capacity to train personnel or upgrade rolling stock and ProRail has a limited capacity to upgrade the infrastructure. Since DCO needs no capacity to upgrade the infrastructure and personnel does not need to do extra trainings for ERTMS Hybrid Level 3, we choose to express the investment costs in terms of change in the rolling stock.

All five technologies are assigned a weight based on the relative costs of the available capacity and these weights can be found in Table 6. All weights are calculated per corridor. As mentioned, for DCO (which is implemented on the train), we determine the number of corridors in the train route and multiply the weight by that number. All technologies contribute to an improvement in the timetable, but by assigning weights, we can compare the investment costs and the improvements between the different technologies. This finalizes the setup of the parameters for the model to be solved for the Dutch railway network.

Table 6: Relative investment weight per corridor per technology

Technology	Relative Investment Weight per Corridor
DCO	1
ERTMS Level 2	3
ERTMS Hybrid Level 3	1
ATO	2
TMS	2

5.2. Single-line network

By finalizing the setup of the new model, we solve it for a single line in the Dutch railway network first. This first initial single-line network should be a representative corridor for the rest of the network and serve as a good starting point for increasing the network size. For this reason, the line Amsterdam-Castricum was chosen (see Figure 9 for a graphical representation). This line accommodates both Intercitys and Sprinters and includes corridors with high passengers' demand. Furthermore, the network can be extended to a subnetwork for the province of Noord-Holland and from there to the whole national network.

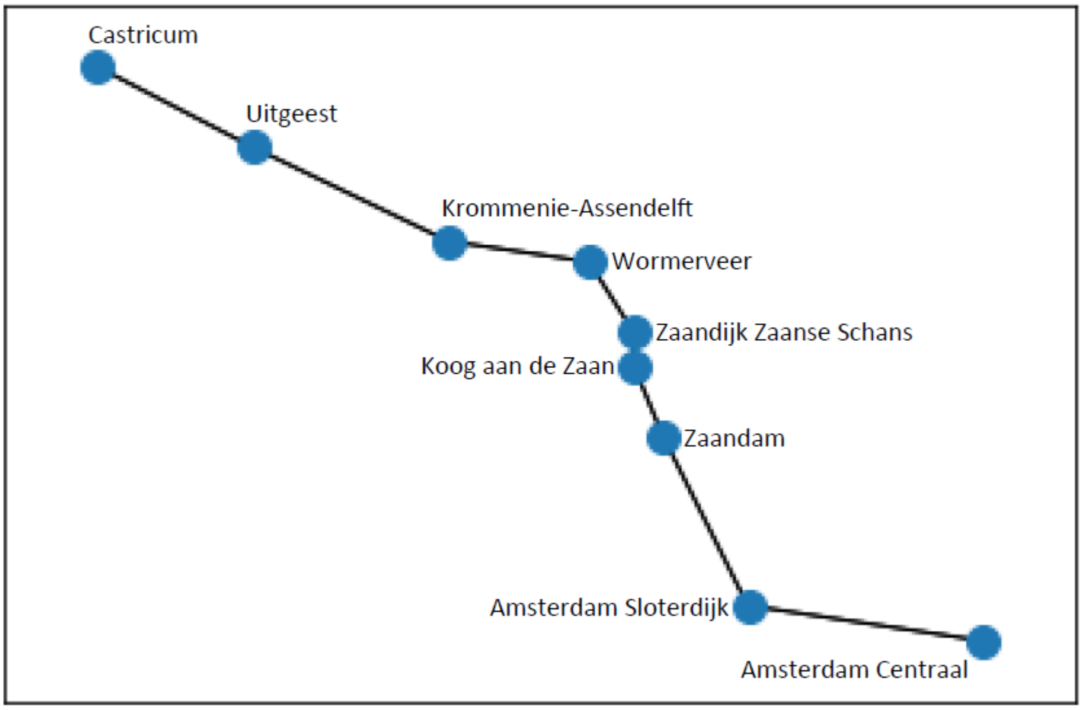


Figure 9: Graphical representation of the single-line network

For this network, we take the subset of the trains in the hourly pattern that run on at least one of the corridors within the line network. We start with two experiments to set the baseline for the model: one experiment with no budget to implement new technologies and one with unlimited budget. The model is given the possibility not to implement all technologies on all corridors and trains, to see whether it the optimal solution would return this strategy or if all benefits would be received at a lower degree of implementation. From these two experiments we find that the maximum improvement for the line network is approximately 5.1% in terms of in-vehicle time, when implementing all technologies on all corridors and trains. Due to the small number of trains in this network size, improvements are not very clear in visualizations of the timetable. Therefore, we will discuss visualizations of the timetable for the medium-sized network in the next section.

Improvement curve

With the baseline, we are now able to calculate the percentage of improvement for the experiments that have a budget in between the two extreme points. Figure 10 shows the relation between the percentage of budget that is invested in the line network and the percentage of the total improvement potential of implementing all technologies. To exemplify this, suppose we decide to invest only 20% of the maximum budget (the green point in the graph). From the figure, we can see that this investment would lead to an improvement of more than 60% of the maximum potential for this line network. Since we know the objective value of the baseline, we know that this translates to an improvement of 3.2% in terms of in-vehicle time.

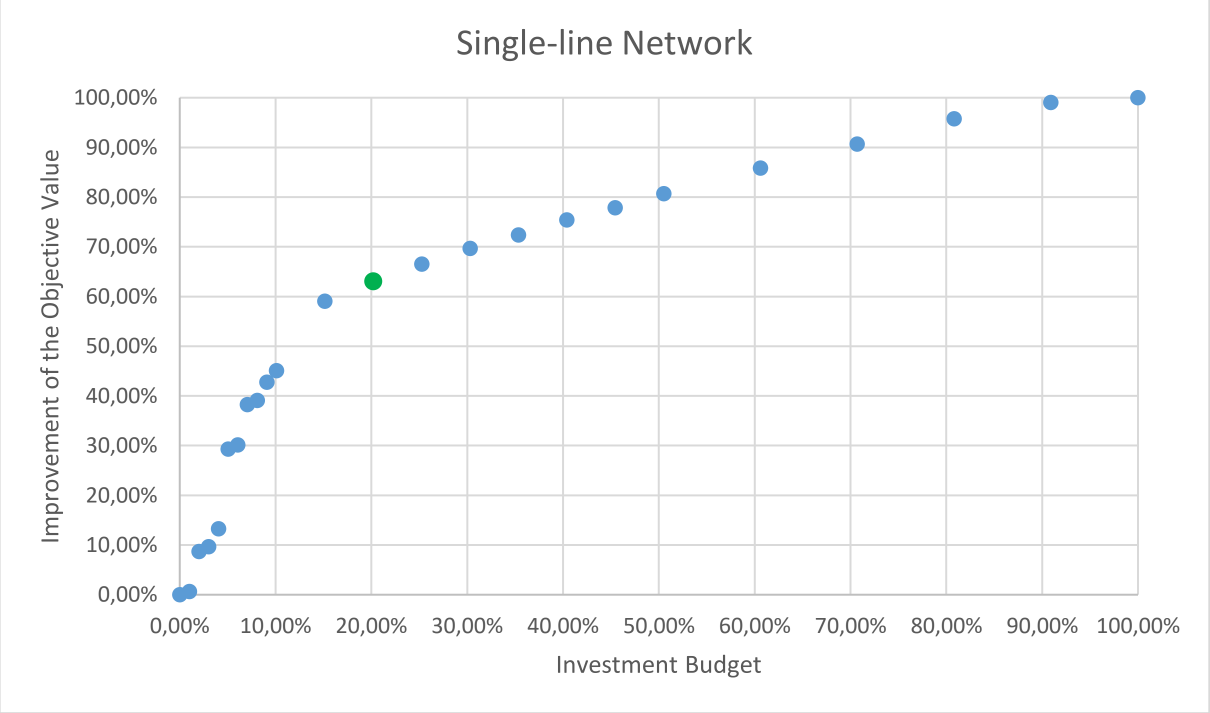


Figure 10: Budget-Improvement curve for the single-line network.

Now that we know the improvement pattern of implementing the technologies on the line network, we might want to know how much budget we should assign to the implementation of new technologies. To answer this question, we can look at the ratio between the difference in budget and the difference in the objective value, when adding more budget. Table 7 shows the values for the experiments and from the last column we find that the budget of 15.15% is the last experiment with a positive ratio (≥1, highlighted in green), meaning that the percental increase in budget is lower than the percental increase of the objective value. Therefore, increasing the budget even more, will result in a lower percentage of improvement of the timetable relative to the budget. It is important to note here that the ratio remains positive, meaning that the timetable will always improve when adding more budget, but the percental change is lower with respect to the budget (also referred to as diminishing returns).

Table 7: Ratio between the percentage of change in the investment budget versus the improvement in the objective value

Budget	Change in the Investment Budget since the previous experiment	Change in the timetable improvement since the previous experiment	Ratio
1	1,01%	0,68%	0,673654
2	1,01%	8,05%	7,973635
3	1,01%	0,97%	0,96081
4	1,01%	3,61%	3,574371
5	1,01%	16,02%	15,86177
6	1,01%	0,85%	0,842686
7	1,01%	8,08%	7,995382
8	1,01%	0,85%	0,842686
9	1,01%	3,67%	3,633186
10	1,01%	2,31%	2,284395
15	5,05%	14,00%	2,772708
20	5,05%	3,96%	0,784958
25	5,05%	3,45%	0,683638
30	5,05%	3,16%	0,62512
35	5,05%	2,72%	0,538034
40	5,05%	3,05%	0,604856
45	5,05%	2,40%	0,476056
50	5,05%	2,88%	0,56927
60	10,10%	5,11%	0,505908
70	10,10%	4,84%	0,479367
80	10,10%	5,09%	0,503486
90	10,10%	3,27%	0,323433
99	9,09%	0,97%	0,106921

Implementation strategy

The solutions to the model, specifically the optimal values for the decision variables, also show us in which locations we should implement which technologies to come to the improved timetable. Figure 11, shows the implementation of the technologies for the experiment with a budget of 20% of the maximum budget (the green point in Figure 10). The numbers next to the corridor indicate the number of trains that pass on that corridor in a half hour timeframe (in both directions), to show how busy the corridors are in comparison to each other. From the figure, we can see that ERTMS Level 2 and ATO are implemented on the corridors between Amsterdam and Koog aan de Zaan, TMS is implemented on the corridors between Amsterdam and Zaandam and DCO is implemented only on train series 3300 (Amsterdam Sloterdijk-Zaandam). While the information is very concrete, it is not necessarily the case that by increasing the budget, these corridors and this train series will still be outfitted with new technologies. Train series 3300, for example, only runs on one corridor (Amsterdam Sloterdijk-Zaandam) and therefore the investment costs for this train are low (weight=1). For this reason, its implementation is easier to fit in the budget than that of a more expensive technology (such as ERTMS Level 2), while the benefits of that technology may be bigger.

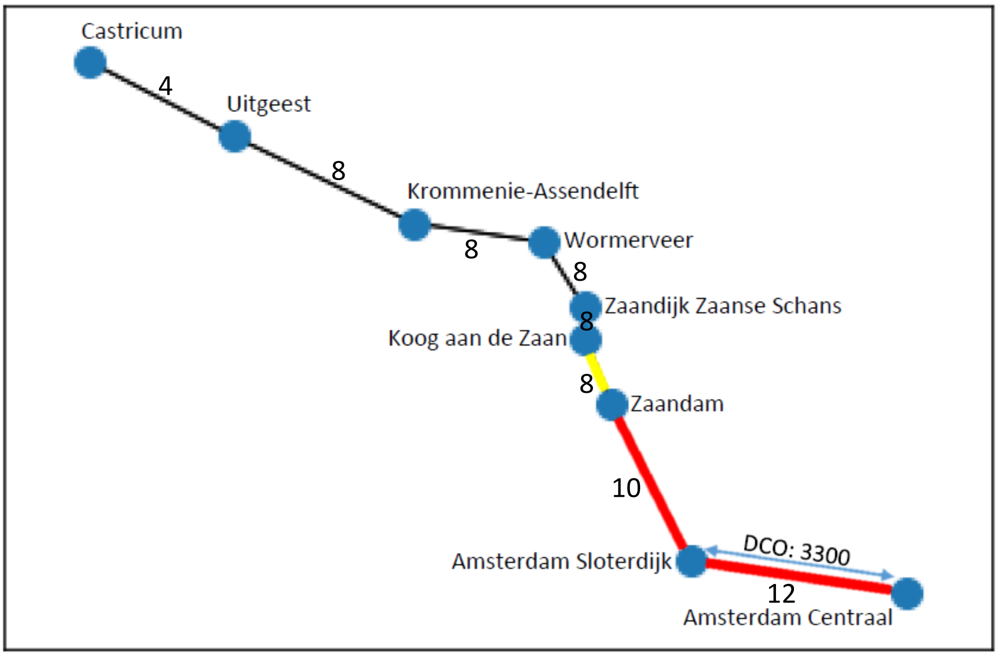


Figure 11: Implementation of technologies for the single-line network with a 20% budget (Red = ERTMS Level 2, ATO and TMS, Yellow = ERTMS Level 2 and ATO, numbers indicate the number of trains on that run on that corridor within a half hour timeframe)

For an implementation strategy of the new technologies, more insight can be gained by looking at the decisions that the model makes at different budgets. Table 8 shows the decisions for the first ten experiments and from this overview we can conclude that the model wants to implement ERTMS Level 2, ATO and later TMS jointly as much as possible, starting at the busiest corridors. When there is not enough budget (left), the model decides to implement TMS on an extra corridor and otherwise DCO for a train that runs on only one corridor. The decision to implement DCO on a single train that runs only on one corridor makes sense as a solution to the model, though, it would not be a sensible strategy in practice. Higher budgets lead to reversing implementations of these technologies. This does not only lead to higher costs (lost investment costs and reversing costs), but it also leads to passengers' inconvenience when the benefits for the passengers on that corridor (e.g., shorter in-vehicle times) are also reversed. Imagine passengers that hear that their daily train trip will take shorter because of new technologies, but one year later they hear that the time savings are reversed and the old trip times are reinstated. To avoid this, an implementation strategy can only be formed by conducting multiple experiments with different budgets to find consistent decisions across the scenarios.

Computation time

The computation time for the single-line network is several seconds or minutes depending on the budget scenario. The experiments with 100% of the investment budget are relatively fast to solve because the model is allowed to implement every technology and this contributes to a model that is less restricted by planning norms. As a result, both a feasible and an optimal timetable are easier to find. The other extreme, where we assign 0% of the investment budget is also fast. While the feasible region is smaller because of the restrictive planning norms, the model does not have to check which technologies are implemented (because none are allowed) and is only dedicated to finding a feasible and optimal solution. For the intermediate points in the budget-improvement curves, the computation rapidly increases. The average computation time of the different scenarios for the single-line network is around twenty minutes, while the maximum computation time is just over two hours for the scenario of 45% of the maximum budget. This shows us that for larger networks, solving the model will take multiple hours to solve.

Table 8: Technology implementations per investment budget scenario

Budget	Investments
1	DCO 3300
2	TMS Amsterdam Centraal – Amsterdam Sloterdijk
3	TMS Amsterdam Centraal – Amsterdam Sloterdijk and DCO 3300
4	TMS Amsterdam Centraal - Zaandam
5	ERTMS Level 2 and ATO Amsterdam Centraal – Amsterdam Sloterdijk
6	ERTMS Level 2 and ATO Amsterdam Centraal – Amsterdam Sloterdijk and DCO 3300
7	ERTMS Level 2, ATO and TMS Amsterdam Centraal – Amsterdam Sloterdijk
8	ERTMS Level 2, ATO and TMS Amsterdam Centraal – Amsterdam Sloterdijk and DCO 3300
9	ERTMS Level 2, ATO and TMS Amsterdam Centraal – Amsterdam Sloterdijk and TMS Amsterdam Centraal – Zaandam
10	ERTMS Level 2 and ATO Amsterdam Centraal – Zaandam

5.3. Medium-sized network

Now that we know that the proposed model works for a single-line network and gives us intelligible solutions, we expand the infrastructure and the train set to a medium-sized network for the province of Noord-Holland. A graphical representation of this network can be found in Figure 12. Again, we start with two experiments to establish the baseline. We find that for the medium-sized network an improvement of the objective value of approximately 5.2% can be achieved (when no budget limit is enforced), which is similar to what we found for the single-line network. Since we found that the relation between budget and objective value follows a certain curve for the single-line network, we conduct several experiments to see if the same is true for the medium-sized network. In Figure 13, the curve for the medium-sized network is shown. As mentioned, the computation time increases for larger networks, so bigger intervals were chosen to make this graph.

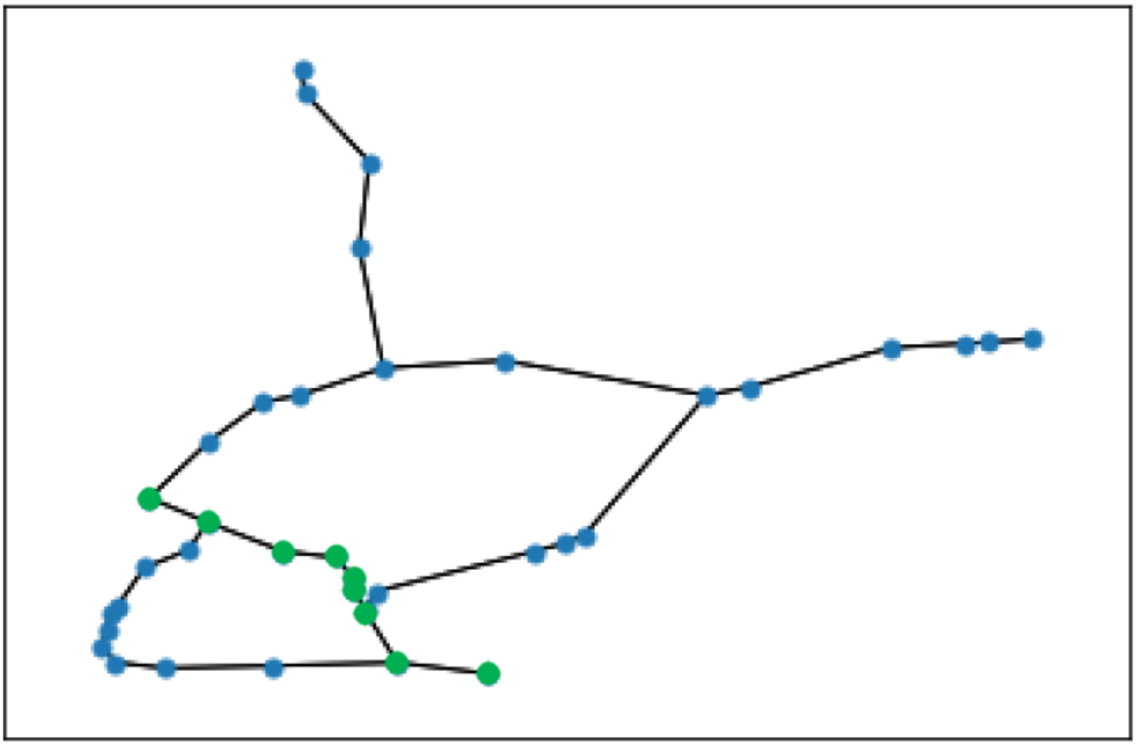


Figure 12: Graphical representation of the medium-sized network with the single-line network indicated in red

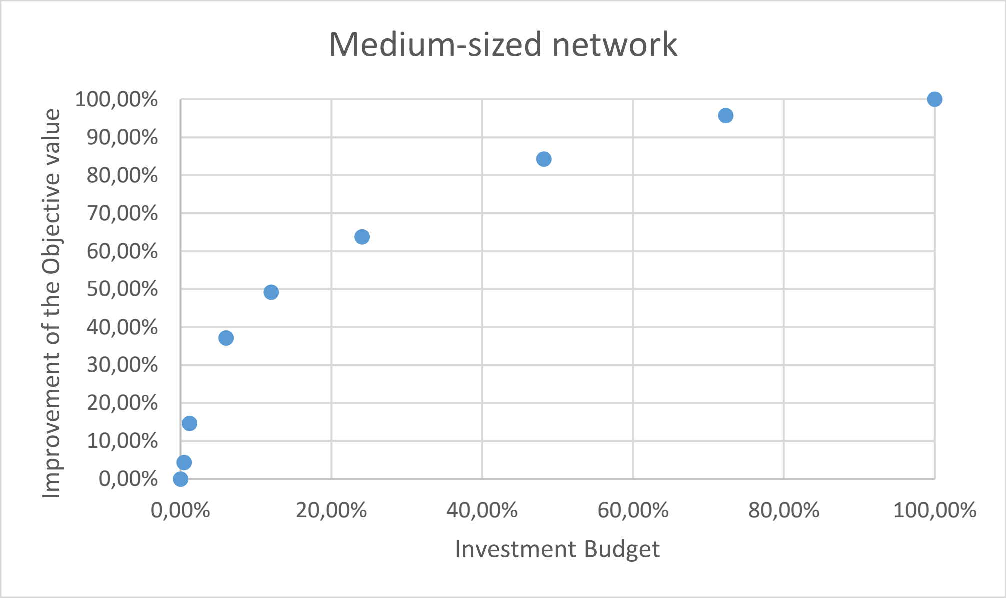


Figure 13: Budget-Improvement curve for the medium-sized network

Improvement curve

To compare the curve to that of the single-line network, both graphs have been included in Figure 14. We find that the graphs of both network sizes follow a similar curve. However, the curve of the medium-sized network starts slightly steeper and a higher ratio is maintained for larger budgets. It is possible that this is due to the network size. Time savings on one corridor affect more train series than in the smaller network and as a result, timetabling problems elsewhere in the network can be solved as well. This way, a smaller percentage of technology implementations can already lead to a higher percentage of improvements. Then, when a majority of the bottlenecks is solved, the timetable can only be improved slightly by implementing new technologies, but this still affects a higher number of trains. The right tail of the graphs is very similar, showing that there is an optimum at an investment of around 20% of the maximum budget. The point where the ratio becomes negative (≤1) is also situated around the same point as in the single-line network.

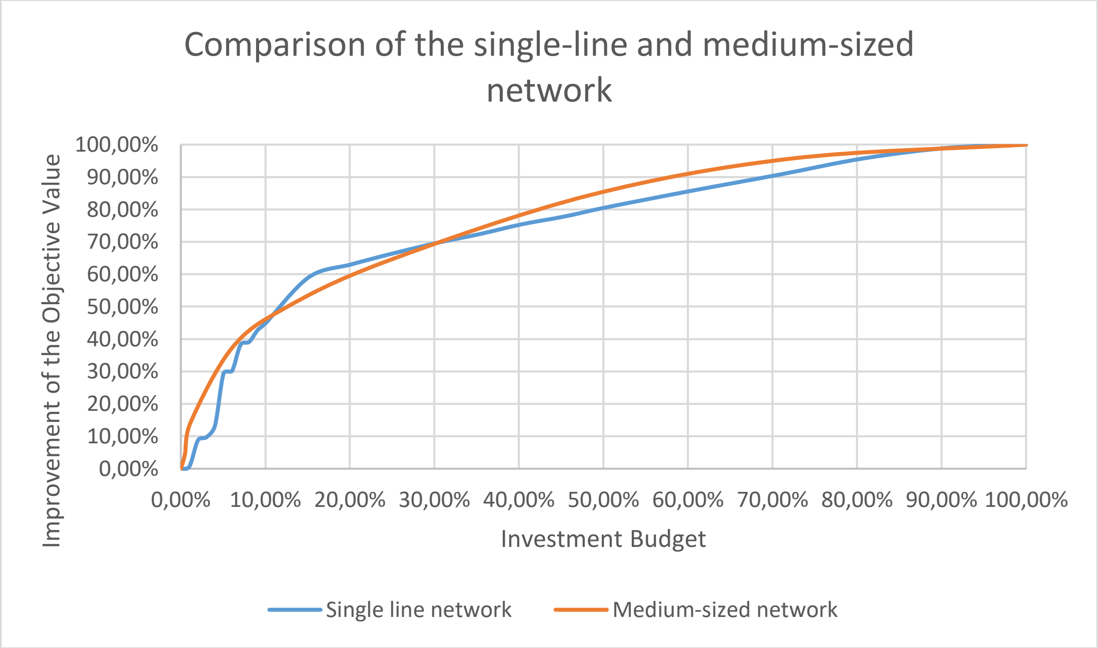


Figure 14: Budget-Improvement curves for the single-line and medium-sized network

Improvements in the timetable

To exemplify the improvements in the timetable, let us look at the differences between the two graphs in Figure 15 and Figure 16. Blocks in the graphs show the time difference between the departure time at station Amsterdam Centraal and the arrival time at station Amsterdam Sloterdijk. These blocks then indicate the travel times between the two stations for different trains. The x-axis shows the time period of the model (half an hour equals 1800 seconds) and the y-axis shows the different tracks on the corridor between the two stations and the trains that run on that corridor. For example, in the first figure, "0, A3000" refers to the train series A3000 that runs on track number 0 (since lists in Python start at 0) and the same colour is used for the same train in the second figure. The horizontal black lines indicate which trains run on the same tracks. While it may seem that overlapping blocks for the same track is impossible, this is not true in this case. The block indicates that the train is running on the track for a period of time but it does not say anything about the location of the train. The train with the earliest departure date has already travelled some distance in the direction of Amsterdam Sloterdijk and thus the first part of the track is free for a new train to depart. The difference between the departure and arrival times will always be bigger than the headway norms.

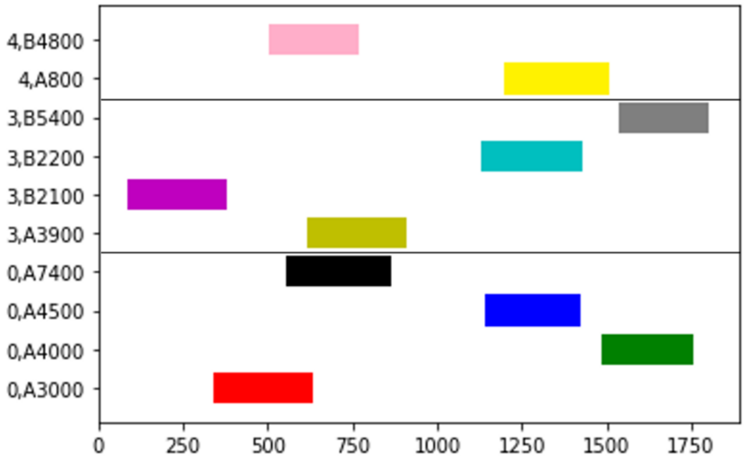


Figure 15: Timetable of the solution to the model for 0% of the budget

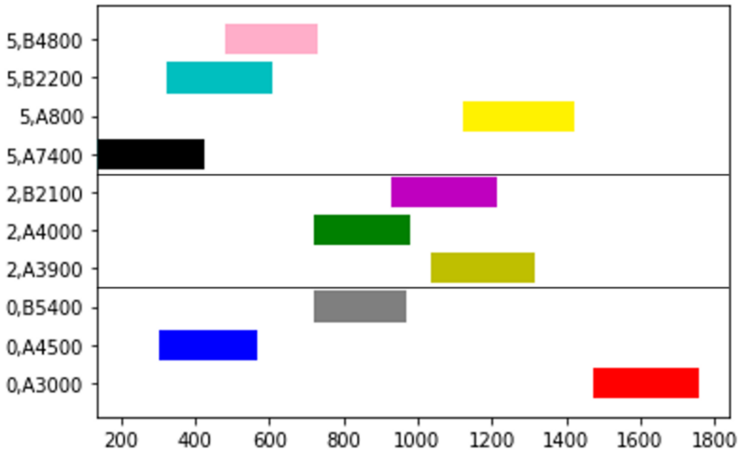


Figure 16: Timetable of the solution to the model for the maximum budget

The first thing that can be noticed from comparing these figures is that the same train may run on a different track (e.g., train series A7400 runs on track 0 in the first graph but on track 5 in the second) and that other tracks are used (0,3 and 4 in the first scenario, 0,2 and 5 in the second). This is because the model designs a new timetable every time a new scenario is given as input and thus other tracks may be used.

Now that we understand what the graphs represent, we can look at the differences in the timetables of the two scenarios. For this, it is important to look at comparable tracks with the same number of trains. For instance, let us look at track 3 in Figure 15 and track 5 in Figure 16. On both tracks four trains are scheduled over a half hour period. In Figure 15, the trains are evenly spread over this time period without any overlaps, while in Figure 16, some of the trains (A7400, B4800 and B2200) are clustered around the same time and overlap. This is now possible because headway norms are decreased at different places and thus trains can run within shorter distances of each other. This effect is what we expected to see in these graphs and also shows the clearest benefit of improving the train timetable of NS. The "empty spaces" between the block on the same track show the possibility to run extra trains in that time period. The closer trains are scheduled to each other, the more room there is to run extra trains and move more passengers. This extra room can also be used to realize certain transfer possibilities.

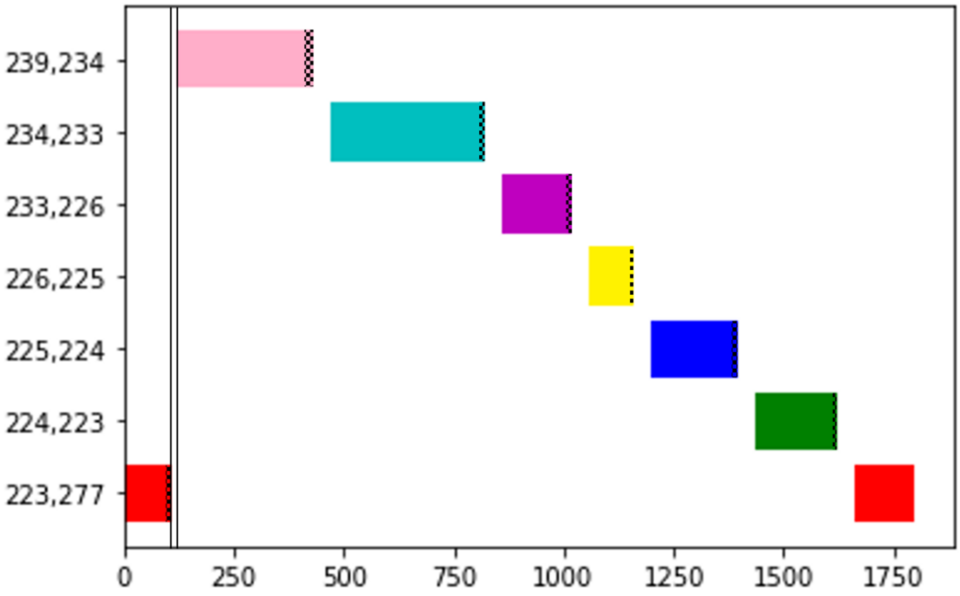


Figure 17: Timetable of the A7400 train series from Amsterdam Centraal (239) to Uitgeest (277) from the solution to the model for 0% of the budget

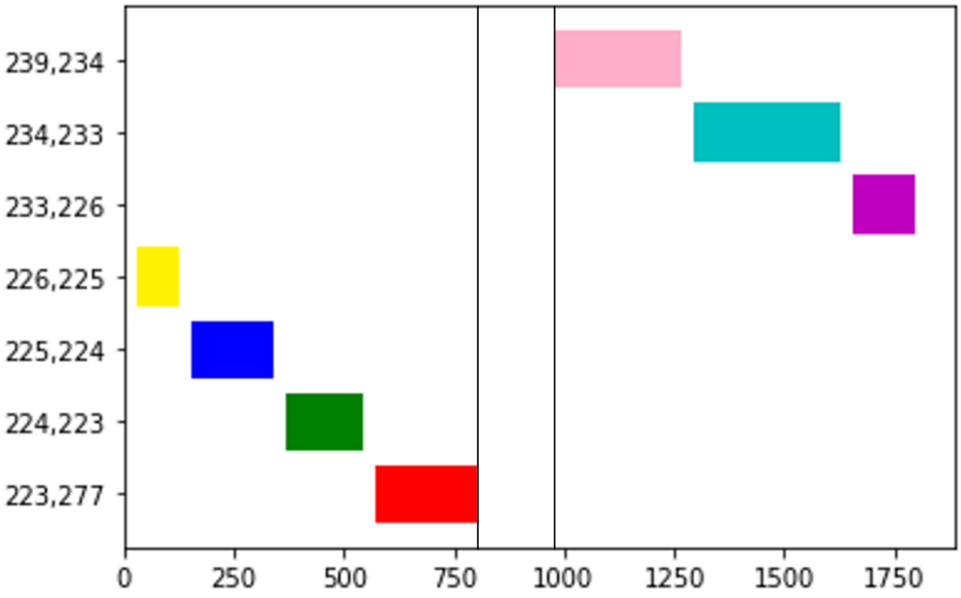


Figure 18: Timetable of the A7400 train series from Amsterdam Centraal (239) to Uitgeest (277) from the solution to the model for the maximum budget

A second option to explain the improvements in the timetable is by looking at the timetable for a single train in the network. In Figure 17 and Figure 18, the two timetables for the train series A7400 are shown. This train runs the same route as the single-line network with the exception of Castricum and this is indicated by the station codes on the y-axis. Amsterdam Centraal (239) is shown at the top of the graph and Uitgeest (277) at the bottom. Again, the blocks show the travel times between one station and the next in the route and we see that a new timetable was designed, because none of the blocks are in the same time period in both graphs. In Figure 17, we can also clearly see the periodicity in the model since the travel time from station Krommenie-Assendelft (223) to Uitgeest (277) is separated in two parts, where the first (starting from t=0) continues from the second (ending at t=1800).

In Figure 17, the black grids show the time savings in the travel times between the stations that were realized in Figure 18. The dwelling times (the empty spaces between two subsequent blocks) are also smaller in Figure 18, but since time savings of implementing DCO is only 12 seconds, this cannot be seen very clearly in the graphs. While the time savings are seemingly small, together they also contribute to a faster overall travel time for the A7400 train series (shown by the vertical black lines). In the first figure, the total travel time equals 1787 seconds and in the second figure, 1628 seconds. The total time savings in the travel time from Amsterdam Centraal to Uitgeest equals 159 seconds, around two and a half minutes. Considering that every train could achieve this time savings and this train also runs on corridors outside the medium-sized network, these minutes time savings per train can add up and make enough room for higher frequencies and extra trains, attracting even more passengers.

Implementation strategy for the medium-sized network

The solution to the model for 24% of the maximum potential budget, as shown in Figure 19, tells us more about the network effects in the model. We find that ERTMS Level 2 and ATO are implemented jointly, as was the case in the single-line network. However, the technologies are implemented in different locations instead of one single line and not necessarily on the busiest corridors. It shows us that implementing a technology on one corridor can lead to significant improvements for the timetable of the whole network. This marks a stark contrast with current implementation strategies, where a snowball effect is preferred: start at one end of the network and gradually spread out from there. Furthermore, we see that implementing DCO on certain trains receives a higher priority than the implementation of the combination of ERTMS Level 2 and ATO on an additional corridor, even though budget is still available for the latter. This shows us that smaller time savings on parts of the network that are already outfitted with other technologies are preferred over larger time savings on other parts of the network, in order to provide benefits for the timetable of the whole network.

Computation time

By comparing the experiments of the single-line and the medium-sized network, we found that the computation time increased significantly. Where the single-line network model can be run within minutes or even seconds, the medium-sized network model needs a few hours in most cases. For the intermediate points in the budget-improvement curves, the computation rapidly increases, reaching multiple hours or sometimes even days. The average computation time for the medium-sized network is around 19 hours and the maximum (for the chosen intervals) is around three days.

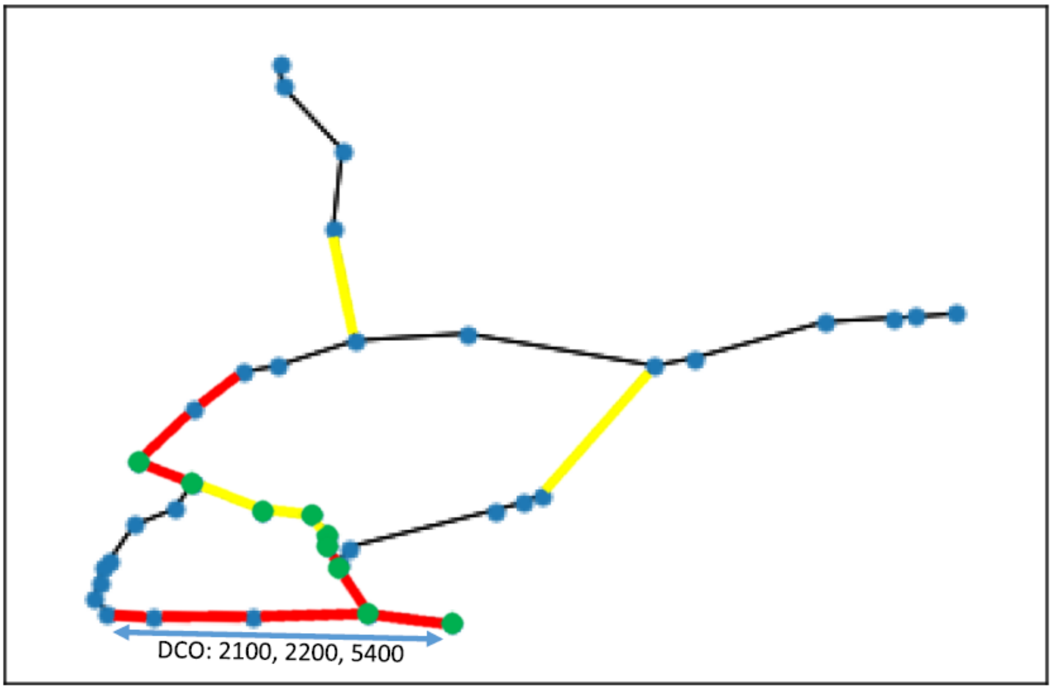


Figure 19: Implementation of technologies for the medium-sized network with a 24% budget (Red = ERTMS Level 2, ATO and TMS, Yellow = ERTMS Level 2 and ATO)

5.4. National network

The next step for the model would be to expand to a national network. The computation time would certainly increase if we solved the model for a national network, but there is another problem when trying to find a feasible solution for the national network model. The national network shows a significant increase in the number of trains, stations and corridors (see Figure 20 for a graphical representation of the national network), which is further translated into many more variables and (quadratic) constraints (see also Table 9).

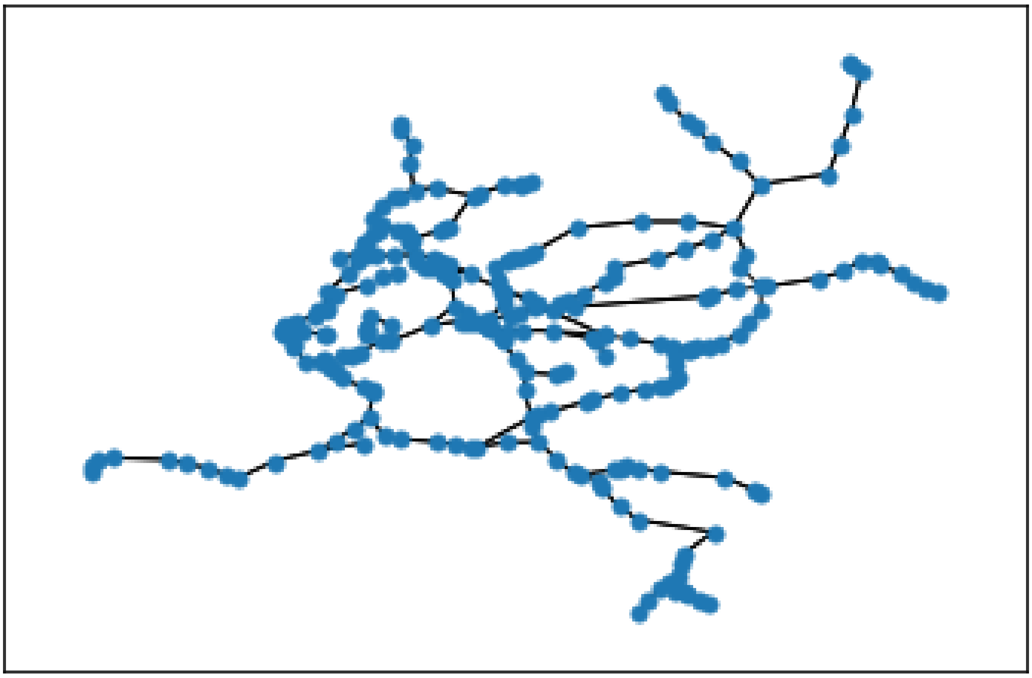


Figure 20: Graphical representation of the national network

Table 9: Model size for the proposed and the simplified model

	Proposed model	Simplified model
Number of Trains	119	119
Number of Stations	263	263
Number of Corridors	287	287
Number of Binary Variables	77301	73834
Number of Integer Variables	4016	4016
Number of Linear Constraints	335846	930172
Number of Quadratic Constraints	602757	3289
Total number of Constraints	938603	933461

This increase has led to a poorer pre-solving phase of the Gurobi model solver. In the single-line and medium-sized network model, the pre-solver was able to find an initial best bound that was very close to the final optimal bound. In the national network model, however, the best bound after the pre-solving phase returns a negative number. Since the objective function calculates the in-vehicle times, the final optimal bound will have to be positive. Furthermore, the objective value should be higher than that of the single-line and medium-sized network because we add extra trains, stations and corridors and we take the sum of all in-vehicle times (instead of the average).

The reason for this phenomenon is that the solver uses an integrality tolerance, which allows the integral variables not to be integral to some extend. For example, with an integrality tolerance of 10^{-5} (the default for the solver), binary variables may take the value of 0.00000999 and this may lead to other variables taking values that should not be possible under the given constraints of the model, for instance, a departure time on a station that lies before the arrival time on that same station, resulting in a negative objective value. Solving the model as a standard train timetabling program with the corresponding time savings would eliminate the variables and constraints related to implementing technologies and could make the model size smaller (see Table 9 for the specific details). However, this approach is not sufficient enough to improve the root node and reduce the computation time as the same problem occurs.

The solver follows a branch-and-bound algorithm starting at the root node that is found in the pre-solving phase. In the branch-and-bound process, branches are added where these non-integrality issues are solved one-by-one. In theory, we could run the model until we found the optimal timetable for the national network but this would prove to be too time-consuming for this master thesis. If this were a linear model and the train timetabling problem would not be NP-hard (which both are not true), a starting point for estimating the computation time could be to calculate the percental increase in the model size and applying the same factor to the computation time of the smaller model. In that case, we assume that computation time is linearly related to the model size. When applying this rhetoric to our (quadratically constrained and NP-Hard) model, we would find an average computation time of approximately 124 days for the national model (see also Table 10), when we assume the model size to be a multiplication of the number of trains, stations and corridors. Of course, the actual computation time cannot be calculated this way, but it gives an insight into the problem of solving the national model within a reasonable timeframe.

Table 10: Estimated computation times for the different network sizes

Network Name	Model Size (assumed values)	Realized average computation time (days)	Estimated computation time (days)
Single-line	2016	0.014	-
Medium-sized	67488	0.791	0.465
National	17964478	-	123.76

On the other hand, the model is able to use the national railway infrastructure and the current NS timetable to create variables and constraints and a feasible timetable will eventually be found, which shows us that the model can be solved for a national network, if there would be a way to decrease the computation time. In section 6.2, we give recommendation on how to reduce the computation time of the national network model.

5.5. Implementation of individual technologies

In the previous sections, we have proven that the model is able to correctly identify corridors where new technologies can be implemented to reduce planning norms that benefit the timetable of the whole network. This is done for the combination of all technologies, but this is not necessarily the most desired scenario for a railway company. For example, NS or any other railway company may want to focus on one or two technologies to streamline the implementation and assure uniform corridors across the network. Furthermore, current EU policy is to make ERTMS Level 2 mandatory on every corridor. To this end, these two sections (5.5 and 5.6) cover the experiments for, respectively, the implementation of individual technologies (along with a sensitivity analysis) and the combination of ERTMS Level 2 with the other technologies individually. Because of the computation time of the national network, these experiments were conducted for the medium-sized network.

Since ERTMS Hybrid Level 3 and ATO require ERTMS Level 2 to be implemented first, these two technologies and their effects on the timetable cannot be tested individually. ERTMS Level 2, TMS and DCO were isolated by setting the potential benefits of all other technologies to zero. What remains is a model that is allowed to implement the other technologies, but this will not improve the timetable so it will never decide to spend budget on that implementation. The restriction of the investment budget is not included in these experiments to find the full potential of the technologies, thus the model is reduced to a standard timetabling problem but with looser planning norms.

In chapter 3.1.6, we mentioned that the potential benefits are based on expert knowledge and simulations or real-life experiments to prove these values are not conducted yet. Therefore, we do not know for certain that the assumed benefits can be realized and if the corresponding timetable is feasible in reality. To account for these assumptions and research their effects on the potential improvement of the timetable, we provide a sensitivity analysis in these experiments. We conduct one experiment with the assumed values for the benefits, one with slightly more benefits (overestimated scenario) and one with slightly fewer benefits (underestimated scenario). The setup of these experiments can be found in Table 11 on the next page.

Table 11: Assumed potential benefits per technology for different scenarios

		Time savings per planning norm and scenario		
	Scenario	Running Time Supplement	Headway (seconds)	Dwelling Time (Seconds)
DCO	**Underestimated**	0.5%	0	10
	Assumed	1%	0	12
	Overestimated	1.5%	0	14
TMS	**Underestimated**	0.5%	10	0
	Assumed	1%	15	0
	Overestimated	1.5%	20	0
ERTMS Level 2	**Underestimated**	0.5%	25	0
	Assumed	1%	30	0
	Overestimated	1.5%	35	0

Since the time savings for the running time supplement for all three technologies is assumed to be 1%, we set the underestimated and overestimated values to 0.5% and 1.5%, respectively. The planners at NS do not use decimal percentages for the running time supplements but since the experiments for 0% time savings would not be interesting to research, we choose to consider decimals in our scenarios. For the headway time savings, we choose a deviation of five seconds in both directions and for the time savings in the dwelling time, a deviation of two seconds was chosen. The assumed time savings in the dwelling time are very accurate since a process of exactly 12 seconds is eliminated. However, for the sake of the sensitivity analysis, an under- and overestimated value was chosen for possibly unknown side effects of the change in the departure procedure.

This setup results in the graph that we find in Figure 21. The orange line indicates the improvement curve for the medium-sized network as found in section 5.3. The points on this optimal curve are timetables where multiple different technologies are implemented and thus it is not expected that this percental improvement can be achieved by implementing only one technology, even if it is implemented on the whole network. The data points for DCO, TMS and ERTMS Level 2 show approximately the same improvement of the timetable and indeed lie below the optimal curve. However, the investment costs for TMS are significantly lower than that for the other two technologies. So, when we are allowed to implement only one technology, based on the graph, we can conclude that TMS would be most desirable since improvements of the timetable are similar to that of the other technologies but investment costs are much lower. While we might expect that TMS would also get a higher priority from the model for all technologies than others, we do not see this strategy for neither the single-line nor the medium-sized network. The reason for this is explained in the next section, where we discuss the combination with ERTMS Level 2.

Additionally, the findings in Figure 21 show us that the improvements of the running time supplements may be more important than decreasing the headway norms. This insight follows from the difference in benefits between TMS and ERTMS Level 2. While both technologies have the same percentage of time savings for the running time supplements, TMS shows a decrease in the headway norm of 15 seconds, whereas ERTMS Level 2 leads to a headway norm reduction of 30 seconds. The results are approximately the same and thus the extra 15 seconds of headway norm reduction do not contribute to a better timetable. Combining this insight with the implementation strategy of the medium-sized network in section 5.3 (first ERTMS Level 2, ATO and TMS), shows us that the running time supplements are more restrictive to the timetable than the headway norms.

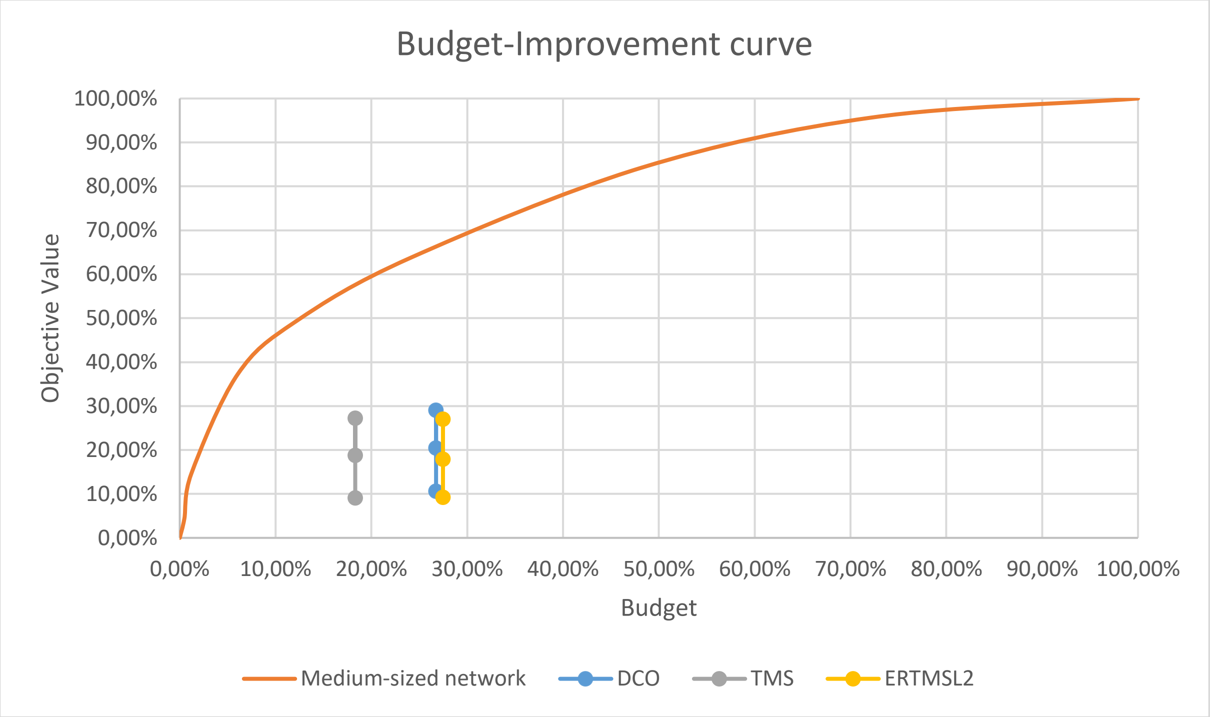


Figure 21: Individual technologies compared to the Budget-Improvement curve for the medium-sized network

Similarly, we find that decreasing the dwelling times may also be more beneficial for the timetable than decreasing the headway norms. Implementing DCO leads to a slightly better timetable than TMS or ERTMS Level 2. The running time supplement reduction is the same but instead of decreasing the headway norms (with 15 or 30 seconds) the dwelling times are decreased (by only 12 seconds). As a result, we find that the running time supplement is the most restrictive planning norm, followed by the dwelling time and lastly, the headway norms.

On the other hand, our model does not consider the changing of line routes or frequencies when we improve the timetable. In some cases, decreasing dwelling times or headway norms may have a bigger influence on those factors, which in their turn can lead to a better timetable for passengers as well. Furthermore, both norms also contribute to transfer times, which are also not incorporated in the model. Therefore, we can only draw the above-mentioned conclusions with regard to these limitations.

Lastly, the margin (indicated by the line between the three points) in the graph in Figure 21 shows us that the assumptions for the benefits affect the potential improvements in the timetable to some degree. The under- and overestimated scenarios show a difference of almost 20 percent in terms of potential improvements for all three technologies. This insight shows us the importance of a better understanding of the technologies and a more accurate estimate of the potential benefits. It will be even more important when a railway company would want to research the implementation of multiple technologies, since this error margin will increase with the number of technologies under consideration.

5.6. Combining ERTMS Level 2

As mentioned in the previous section, uniformity of the technologies across the corridors may be desirable for railway companies as well as (local) governments. To this extend and to improve competition within the EU, the European Commission has the aim to outfit the European railway network with ERTMS Level 2. With this knowledge, we can perform additional experiments that research the second technology that performs best when ERTMS Level 2 is already implemented. The experiments are of similar nature to those of the individual technologies in part since we set all benefits in the model to zero except those of ERTMS Level 2 and the technology that we want to test the combination of. Furthermore, we lose the investment constraint to see the full potential of the combination. The outcomes of these experiments can be found in Figure 22.

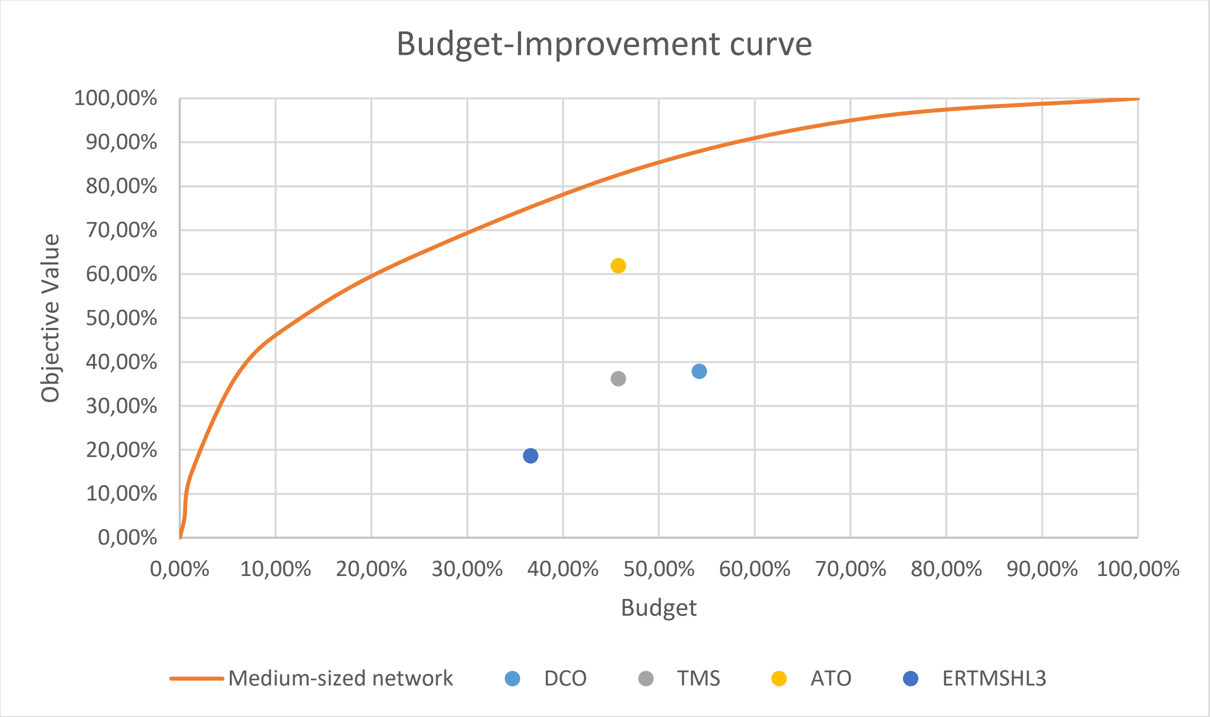


Figure 22: Combinations with ERTMS Level 2 compared to the Budget-Improvement curve for the medium-sized network

The first observation from this graph is that all combined experiments have a higher percentage of improvement than the implementation of ERTMS Level 2 alone. Still, we see that the combination of ERTMS Level 2 and Hybrid Level 3 does not lead to a significantly better timetable. This can be explained due to the low assumed potential benefits of ERTMS Hybrid Level 3, which is only a decrease of 15 seconds in the headway. Of course, this is also reflected in the low investment costs for ERTMS Hybrid Level 3, but in this model, we only focus on the capacity to change. The combination of ERTMS Level 2 and ATO shows the most promising improvement in the timetable. From the single-line and medium-network experiments we find the same dynamic where the implementation of ERTMS Level 2 and ATO are prioritized. For lower budgets, we even find instances where only this combination is implemented and thus the budget-improvement curve for this combination would share a big part of the curve for the model of all technologies combined.

The improvement in the timetable under the combination with TMS and DCO is slightly lower than with ATO. From the previous section, we found that TMS would be the technology with the highest benefits in relation to the investments. However, the above insight shows why TMS is not prioritized in the models where all technologies can be implemented together. The combination of ERTMS Level 2 and ATO shows an improvement in the timetable that is almost 25% higher than the combination with TMS and thus the model will always choose to implement this combination first when the budget allows it. TMS is implemented as a ‘third choice’ when the budget does not allow for an additional corridor to be outfitted with the combination of ERTMS Level 2 and ATO, showing that the individual improvements of TMS are still more significant than that of other individual technologies or combinations. Furthermore, it should be noted that since ATO cannot be implemented without ERTMS Level 2 present, we could not assess the initial improvement of solely implementing ATO as is also the case for ERTMS Hybrid Level 3.

Lastly, from Figure 22, we find that only the combination of ERTMS Level 2 and ATO has a positive ratio (≥1), meaning that the percental improvement of the timetable is higher than the percental increase of the budget. On the other hand, from the single-line and medium-sized network experiments we find that for some budgets TMS is also implemented with a ratio greater than one. The insight that we gain from this is that that for every technology and every combination of technologies there will exist an optimal point where most of the benefits are already achieved and further investments will not lead to significantly better timetables.

5.7. Model conclusions

In this chapter, we have conducted several experiments for varying network sizes. From these experiments we can also draw conclusions for the model itself. First, we find that the model is able to correctly identify corridors where technologies can be implemented to decrease planning norms that benefit the timetable of the whole network the most. From the medium-sized network model specifically, we can conclude that the model is able to account for network effects when improving the timetable. The model can decide to implement technologies on seemingly calm corridors to solve timetabling problems on other corridors or stations.

One drawback of the model is that these network effects are not provided as direct output of the model. Since the model solution returns one single (improved) timetable, we do not know what specific problems were solved by implementing a technology on that particular corridor or train series. For example, by implementing TMS on one corridor, the headway and running time supplements may be decreased for that corridor, but this could lead to even more significant improvements on other parts of the network of which we do not know. For DCO, it is particularly hard to find the specific place on which station the dwelling time was a restrictive factor for an improved timetable.

The experiments in this chapter have also shown us the added potential of the model for testing individual technologies or different combinations. The ability to test individual technologies also shows the potential that the model can have when innovations lead to new technologies, of which the benefits for the timetable should also be tested. Especially for the combination of ERTMS Level 2 and other technologies, the model can provide valuable insights on what technology should be implemented, since ERTMS Level 2 is expected to be implemented (inter-)nationally.

The computation time is the most restrictive factor for the solvability of the model. For smaller network sizes and specific investment budgets (e.g. 0% and 100%), the model is able to find feasible and optimal timetables relatively fast. For other budgets this computation time increases and for the national network, the computation time is too long to make the model a useful tool for the timetable design process at NS. The computation time could be decreased by using a more powerful computer, upgrading to a newer solver version or by implementing heuristics to decrease the improve the root node or decrease the size of the branch-and-bound tree. This would reduce the computation time but it is not expected to be as fast as the single-line or medium-sized network. However, since the implementation of technologies is a long-term decision, high computation times (of at most a few weeks) are acceptable and thus the model is still a viable option to obtain an implementation strategy for NS. In the case that longer computation times remain, the solutions to the model for smaller network sizes may provide enough insight to develop a national implementation strategy.

6. Conclusion and recommendations

This research aims to find a way for NS to implement new technologies in such a way that it benefits the timetable of the whole network the most. In this chapter, the main conclusions from this research are discussed. Finally, we give recommendations for the implementation of the model that was developed as part of this study and provide aspects of the model that might require future research.

6.1. Conclusions

By analysing the current timetable design process and conducting a literature review, we found the necessary requirements for a tool that was able to identify the corridors where the technologies should be implemented. The existing literature includes models with both advantages and disadvantages, so a new model was developed to fit the goal of this research. This extended Periodic Event Scheduling Problem (PESP) model has the added ability to implement new technologies in the existing timetable. An investment budget allows to research different scenarios and this was first solved for a single-line network and later for a medium-sized network. The timetables were optimized for the in-vehicle time of all passengers in the network. For both the single-line and the medium-sized network, a budget-improvement curve could be created that shows the timetable can already be improved significantly, when 20% of the budget is used to implement new technologies. The percental increase of the investment budget after this optimum is higher than the percental improvement of the timetable.

One of the insights from the literature review was that the potential benefits of the technologies have not been tested by simulations or proven by real-life experiments. The added advantage of the new model is that the parameters of the technologies can be changed to conduct a sensitivity analysis. Furthermore, the model can assess the best implementation for individual technologies as well as a combination of different technologies. NS or other railway companies can use the model to assess the effect on the timetable for different assumptions of the potential benefits and different values for the investment budget. This would be particularly interesting when developing a cost-benefit analysis. This same model characteristic is also advantageous when NS wants to research new innovations and technologies that were not studied in this report. Additionally, it could be used for other scenarios, such as temporary speed restrictions (Dutch: Tijdelijke Snelheidsbeperkingen or TSB), for which a strategy can be developed, which restrictions should be solved first or where we would absolutely not want to run trains on lower speeds.

There are two drawbacks to the model. First, the model does not provide us with the insight how the implementation of a technology leads to a better timetable. In other words, we do not know which bottlenecks, train orders or other problems could be fixed due to the implementation of technologies. Second, the computation time of the model increases rapidly with network size. This leads to problems when running experiments for the national network but can be solved by using a more powerful computer or by implementing a new heuristic to improve the root node and the branch-and-bound algorithm of the solver.

The main finding of this study is that the new model can correctly identify the corridors where technologies can be implemented to decrease planning norms that benefit the timetable of the whole network the most, which answers the research question. Given the assumptions in the model and its limitations, we find that the model has a great potential for NS to develop and analyse implementation strategies.

6.2. Recommendations

While the potential of the model is great, there are also limitations to the model that can be investigated further. One of the limitations is that transfers, line routing, and line frequency changes are not incorporated in the model. The model shows significant improvements in the timetable but passenger convenience is also expressed in the transfer possibilities and transfer times, which contribute to the travel time of the passenger. Furthermore, a higher line frequency results in lower waiting times for the passengers. A substantial part of the quality of the timetable lies in these aspects and the model could be altered in such a way that it can account for them.

Another limitation of the model is that international and freight trains are not included in the current timetable. These train series have highly irregular time paths or run with long intervals. Due to this nature, these paths were not included in the timetable, but their effect on the timetable could be researched as well as the effect of the technologies on these train types. Next to this, while the infrastructure of the Dutch railway network is almost directly used as input for the model, the model avoids single-track corridors by considering them as two parallel tracks. Further research could be done to change the model so that it is able to include single-track corridors as well.

It is necessary to improve and prove the assumptions for the potential benefits of the new technologies in order to use the solution to the model to develop an implementation strategy. Without certainty for these potential benefits, the improvement of the timetable is also an assumption. Alternatively, a more extensive sensitivity analysis could be conducted to find the error margin for the improvement of the timetable when implementing multiple technologies at the same time. Similarly, the weights for the investment budget are also assumed values, for which an additional sensitivity analysis could be conducted.

Lastly, the main limitation of the model is the computation time. In Section 5.4, we found that the model has great difficulty to run an experiment for the national network. Therefore, the most important recommendation would be to find a way to significantly reduce the computation time, either by using more powerful computers or by applying a heuristic to improve the root node and branch-and-bound algorithm of the solver. After this, the national network model can be run for different experiments and an implementation strategy can be developed for NS.

References

Bersani, C., Qiu, S., Sacile, R., Sallak, M., & Schön, W. (2015). Rapid, robust, distributed evaluation and control of train scheduling on a single line track. *Control Engineering Practice, 35*, 12-21.

Buchmueller, S., Weidmann, U., & Nash, A. (2008). Development of a dwell time calculation model for timetable planning. *WIT Transactions on The Built Environment*, 525-534.

Caimi, G., Kroon, L., & Liebchen, C. (2017). Models for railway timetable optimization: Applicability and applications in practice. *Journal of Rail Transport Planning & Management, 6*, 285-312.

Davey, E. (2012). *Rail Traffic Management Systems.* London: Thales Transportation Systems.

Dimitrova, E., & Tomov, S. (2021). Automatic Train Operation for Mainline. *13th Electrical Engineering Faculty Conference.* Department of Telecommunications and Signalling; Todor Kableshkov University of Transport.

Dotoli, M., Sciancalepore, F., Epicoco, N., Falagario, M., Turchiano, B., & Constantino, N. (2013). *A Periodic Event Scheduling Approach for Offline Timetable Optimization of Regional Railways.* Bari: Politecnico di Bari.

Fuchs, F., & Corman, F. (2019). An Open Toolbox for Integrated Optimization of Public Transport. *2019 6th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, (pp. pp. 1-7). Cracow, Poland.

Fuchs, F., Trivella, A., & Corman, F. (2022). Enhancing the interaction of railway timetabling and line planning with infrastructure awareness. *Transportation Research Part C: Emerging Technologies, 142*.

Goverde, R. M., & Meng, L. (2011). Advanced monitoring and management information of railway operations. *Journal of Rail Transport Planning & Management, 1*, 69-79.

Goverde, R., Corman, F., & D'Ariano, A. (2013). *Railway line capacity consumption of different railway signalling systems under scheduled and disturbed conditions.* Delft University of Technology, Department of Transport and Planning.

Goverde, R., Daamen, W., & Hansen, I. (2008). Automatic identification of route conflict occurences and their consequences. *Computers in Railways, 103*, 473-482.

Hansen, I. A. (2009). Railway Network Timetabling and Dynamic Traffic Management. *2nd International Conference on Recent Advances in Railway Engineering (ICRARE-2009)* (pp. 135-145). Tehran: Iran University of Science and Technology.

Hansen, I., & Pachl, J. (2014). *Railway Timetabling & Operations.* Hamburg: Eurailpress.

Li, L. (2012). *An Introduction to European Rail Traffic Management System.* School of Transportation and Logistics, Southwest Jiaotong University.

Lieskovský, A., Myslivec, I., & Žemlička, M. (2020). Automatic Train Operation: History and Open Questions. *6th International Conference on Vehicle Technology and Intelligen Transport Systems* (pp. 260-267). Science and Technology Publications, Lda.

Nederlandse Spoorwegen. (2022). *NS Jaarverslag 2022.* Nederlandse Spoorwegen. Retrieved from https://www.nsjaarverslag.nl/FbContent.ashx/pub_1004/downloads/v230223085309/NS-Jaarverslag-2022.pdf

Nederlandse Spoorwegen. (2023). *Over NS: Verantwoordelijkheden*. Retrieved April 20, 2023, from https://www.ns.nl/over-ns/de-spoorsector/verantwoordelijkheden.html

Planting, T. (2016a). *Potential of Microscopic Timetable Design in the Netherlands.* Delft: Delft University of Technology.

Planting, T. (2016b). *Ontwerpmethoden van Dienstregelingen; Verkennend onderzoek naar de ontwerpmethodiek per fase in het planningsproces bij NS&ProRail.* Delft: Technische Universiteit Delft.

Sartor, G., Mannino, C., Nygreen, T., & Bach, L. (2023). A MILP model for quasi periodic strategic train timetabling. *Omega, 116*.

Siebert, M., & Goerigk, M. (2013). An experimental comparison of periodic timetabling models. *Computers & Operations Research, 40*, 2251-2259.

Sigger, B. (2023). *Spoorsysteem van de toekomst - migratieplanning.* Nederlandse Spoorwegen.

Sparing, D., & Goverde, R. M. (2017). A cycle time optimization model for generating stable periodic railway timetables. *Transportation Research Part B, 98*, 198-223.

TwynstraGudde. (2022). *Onderzoek systeemkeuzes spoor.* TwynstraGudde.

Vergroesen, R. (2020). *ERTMS/ETCS Hybrid Level 3 and ATO: A simulation based capacity study for the Dutch railway network.*

Weeda, A. (2005). *Plannen met speling, spelen met planning.* Utrecht: Technische Universiteit Delft.

Zhao, W., Martin, U., Cui, Y., & Liang, J. (2017). Operation risk analysis of block sections in the railway network. *Journal of Rail Transport Planning & Management, 7*, 245-262.