Performance analysis of initial stabling plans of railway yards subjected to demand variations

Master's thesis

Regino Blankenzee

July 2023

Performance analysis of initial stabling plans of railway yards subjected to demand variations

By

Regino Blankenzee

in partial fulfilment of the requirements for the degree of

Master of Science

in Civil Engineering

at the Delft University of Technology,

to be defended publicly at 11:00 on July 6th, 2023.

Supervisors:	Prof. dr. R.M.P. Goverde,	TU Delft, CT Transport & Planning (chair)
	Dr. Ir. N. van Oort,	TU Delft, CT Transport & Planning
	Ir. M.B. Duinkerken,	TU Delft, 3mE Marine & Transport Technology
	Dr. J. Boeijink,	NS, R&D Hub Logistics

An electronic version of this thesis is available at http://repository.tudelft.nl/.

Cover image by: *Meneer en Mevrouw Trein*

"Plans are worthless, planning is everything." – Dwight D. Eisenhower

Preface

Dear reader,

This master's thesis that lies before you, is the final hurdle in my academic education, and is a testament to the six years I have spent learning more about myself and the things that interest me the most. This thesis is the conclusion of my Master of Science study in Civil Engineering – Transport & Planning at the TU Delft. Being a fan of railways since I was a little boy, it seemed only fitting that I would conduct my research at the Dutch Railways, NS. Between November 2022 and July 2023, I have delved deep into stabling yard operations as an intern at the department of R&D Hub Logistics to investigate how the robustness of a stabling plan could be quantified and measured. Performing my thesis at NS has been a joy. I have learned a lot of valuable things and have grown as a person in the process.

I would first of all like to thank my supervisors from the TU Delft: Rob Goverde, Niels van Oort, and Mark Duinkerken. Your feedback has been extremely helpful in getting the quality of this thesis to a higher standard. From NS, I would like to thank everyone at the department of R&D Hub Logistics and everyone else I have had the pleasure to speak with for making me feel welcome and provide me with useful knowledge. More specifically, I would like to thank Jord Boeijink, for the supervision of my research and the fruitful discussions we have had about stabling yard robustness, and Bob Huisman, for getting me an internship position at NS and for filling in for Jord during his paternity leave, our discussions were of great importance for the direction of this thesis and the final product.

Furthermore, I would like to show my gratitude for my study mates, and above all, friends Youri, Elise, and Peter. The start of our bachelor study Civil Engineering felt like ages ago, but that makes me appreciate the memories and friendship we have made even more. Here's to many more memories.

To my parents, Silvester and Mariska, thank you for your unconditional love and encouragement, and also their partners Lucenda and Edwin, as well as the rest of my family, thank you. Your support and love have not gone unnoticed.

Last but not least, I would like to thank my fiancée Rebecca for putting up with me during these busy times. I am sorry that I could not invest as much time into planning our wedding as we maybe both wanted, but my thesis is now finished. Let's get married!

Regino Blankenzee
De Zilk, July 2023

Abstract

The planning of railway operations is a very complex process, in which timetables and other logistical plans need to be both reliable and robust to effectively prevent and cope with disturbances. However, research in the robustness of initial stabling plans, designed in an earlier planning phase, during the following planning phases, has been lacking, with the models created in research in the Train Unit Shunting Problem (TUSP) being generally deterministic in nature, even though stabling plans could quickly become infeasible when the stabling demand in the form of arriving and departing trains changes.

This thesis therefore proposes a definition of the robustness of an initial stabling plan to changes in the stabling demand, such as changes in train lengths, as well as provide an assessment method of said robustness. The created Robustness Assessment Model (RAM) first generates an initial stabling demand and stabling plan, then performs a Monte Carlo simulation to generate a set of stabling plans created for variations of the initial stabling demand, based on changes in stabling demand such as train length. Finally, the RAM estimates the robustness of the initial stabling plan by analysing the differences between the initial and these generated stabling plan variations and how efficiently the initial plan is able to change to these plans to optimally facilitate the variations of the initial stabling demand.

Running this model across three locations, each with three capacity utilisation scenarios, has shown that the model is able to estimate the robustness of an initial stabling with acceptable confidence, and furthermore is able to give insight into how capable the stabling plan creation method is in generating a robust solution. Furthermore, the RAM can also be used to investigate patterns in stabling plans which could predict the robustness of said stabling plans. Current shortcomings to the RAM are its relatively long runtime, the simple stabling plan creation process used in the RAM, and that only one-sided stabling yards without servicing scheduling have been incorporated. Further research is therefore recommended for extending the RAM with other stabling yard layouts and the scheduling of services, as well as improve the stabling plan generation process to take more constraints into account.

Executive Summary

The Dutch railway network is very extensive and complex, and just like any other public transport network, its schedules are subject to a lot of stochasticity. A lot of hard work is put in to ensure that these schedules are both reliable and robust to disturbances such as delays. To achieve this, a lot of logistical problems need to be solved. One of such tasks is the creation of a stabling plan for each stabling yard, where trains are stabled overnight when train services are stopped. Trains arriving to the stabling yard in the evening will need to be matched to the departing services the following morning, and might need to be split or combined with other train units, cleaned, or inspected. The problem of finding a feasible stabling- and shunt plan is referred to as the Train Unit Shunting Problem (TUSP).
While this problem has seen big improvements both in extensiveness and solution performance, this problem is still subject to a deterministic mindset, meaning that the input to the solution models is fixed. In reality, however, the arriving and departing train services composing the stabling demand can still vary significantly after the initial stabling plan has been created due to stochastic events, such as a change in composition length. As a result, this initial stabling plan can very quickly become infeasible, requiring adaptations to the stabling plan or even a completely new stabling plan needs to be designed, which will take a significant amount of extra labour or computational power, making the stabling plan design process less efficient. The planning process of the Dutch railway operator, NS, contains multiple planning phases, in which plans are continuously updated or re-created. As these stochastic events occur during each planning phase, it is therefore crucial that the initial stabling plan is not only feasible in the theoretical, deterministic environment of a single planning phase, but is also able to effectively cope with the stochasticity of information in the planning phase until the operational phase, reducing the required workload to adjust the initial stabling plan in these later phases.

This thesis provides both a definition for the robustness of a stabling plan during the early planning phase, as well as a proposal for a model which assesses this robustness for a stabling plan. Therefore, the main research question of this thesis reads:

How can the robustness of a stabling plan to stochastic events in the planning phase be defined and assessed?

The literature review found that although there has been a lot of research performed regarding the robustness of railway schedules, the robustness of stabling plans has been out of scope. As a result, no researchers have gone the lengths to give a detailed definition of the robustness of a stabling plan and how this could be assessed. Furthermore, although the TUSP has been heavily researched, only few researchers have started to incorporate some forms of stochasticity into their models, however these are still lacking in the stochasticity of the rolling stock composing the stabling demand.

Besides the parking of trains, stabling yards can offer services such as cleaning and routine technical inspections. The scope of this thesis has been narrowed down to stabling yards with tracks only accessible from one side, and servicing taking place at the stabling tracks itself, rather than special servicing tracks. The scheduling of servicing has also been out of scope for this thesis.

The NS has split up their planning phase into four separate phases:

- Basic Hour Pattern (BHP)
- Basic Day (BD)
- BD update (BDu)
- Specific Day (SD)

The majority of the planning uncertainties in the stabling demand occur in the transition from the BDu phase to the SD phase. Therefore, the model which assesses the robustness of an initial stabling plan is focussed on these two phases. The included planning uncertainties which were found to be most influential to the risk of having to adjust the stabling plan at a later phase are:

- Presence of out of service trains at the stabling yard.
- Change in scheduled arrival or departure time.
- Change in train unit variant.
- Change in train unit type.
- Addition or removal of a train unit from a composition.

Using the insights gained from the planning process and these planning uncertainties, the robustness of an initial stabling plan has been defined as follows:

"The robustness of an initial stabling plan is defined as the effectiveness at which the stabling plan is able to cope with the stochasticity of events during the later planning phase with the least number of changes compared to other stabling plans."

Besides a definition for initial stabling plan robustness, this thesis also proposes the Robustness Assessment Model (RAM), which uses a Monte Carlo simulation to provide a method to assess the robustness of initially created BDu stabling plan when subjected to stochastic changes in the stabling demand. Figure 1 below shows the scheme of the RAM.

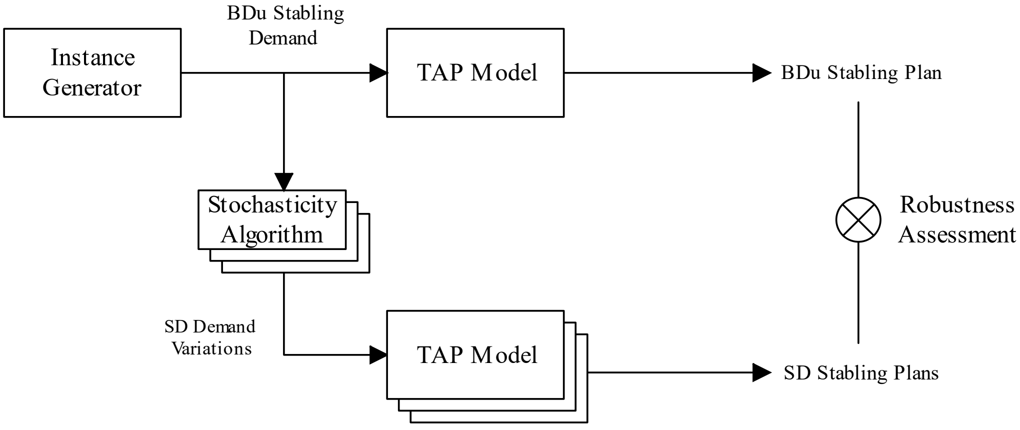


Figure 1: Robustness Assessment Model scheme.

The Instance Generator, the only module of the RAM provided by NS and not expanded in this thesis, generates realistic stabling demand, which then, together with the lengths of the stabling tracks, is used to create an initial stabling plan from the BDu phase in the created Track Assignment Problem (TAP) model. The same BDu stabling demand is then fed through a Monte Carlo simulation. For each iteration in the simulation, a SD stabling demand is created by running the BDu stabling demand through the Stochasticity Algorithm (SA), which alters the stabling demand based on the chosen planning uncertainties and their probabilities of occurring. This SD stabling demand is then fed through the same TAP model as used for the creation of the BDu stabling plan, to create a SD stabling plan. This simulation results in a set of SD stabling plans created for SD stabling demand based on the BDu stabling demand. Then, the robustness of the BDu stabling plan is estimated by comparing it to the portion of viable SD plans, where a SD plan is viable if all trains could be stabled. How different two stabling plans are from each other is determined using a comparison algorithm which returns a difference score. With this algorithm, the robustness of a BDu stabling plan is estimated using two metrics. The first metric is the offset, which is the average difference score between the BDu plan and the viable SD plans. The second metric is the spread distribution, which composes of the average difference score between each viable SD plan and every other SD plan. This distribution gives insight in how efficiently a stabling plan could be able to cope with changes in stabling demand. The robustness of the initial stabling plan is then estimated as the percentage of SD plans having a higher estimated average difference score compared to the offset of the initial stabling plan.

The RAM has been run on three case study locations: Carthesiusweg, Nijmegen, and Zwolle. Each of these three locations have three stabling demand scenarios: 40%, 60%, and 80% capacity utilisation. The estimated offset of the BDu plan generally sits at the lower end of the spread distribution, indicating that the TAP Model in the RAM is able to generate fairly robust initial stabling plans, with an estimated robustness generally sitting between 80% and 90%. The best and worst performing stabling plans of three scenarios have been analysed on patterns which could predict their robustness. It was found that an even spread of compositions across the stabling yard, as well as minimising the consequences of the most bottlenecked stabling tracks could improve stabling plan robustness. Furthermore, a sensitivity analysis of the RAM has been performed by increasing and decreasing the probabilities of the stochastic events in the model. It was found that changes in the probabilities only had marginal effects at most, indicating that the model is not sensitive to small changes in probabilities. Of the planning uncertainties, changes in arrival or departure time, and addition or removal of a train unit from a composition were the most influential for the robustness estimation.

However, the TAP model is rather simple, resulting in multiple solutions having the same optimal objective function value, which means that rerunning the RAM, especially with lower capacity utilisation, can lead to significantly different results than before, as a different initial stabling plan will result in a different offset estimation. The runtime of the RAM is however its biggest drawback, especially when desiring to run the RAM often in short succession.

The definition of the robustness of an initial stabling plan and the RAM has opened the door to optimising initial stabling plan creation for robustness, aiming at reducing the average number of changes required in later planning phases and subsequently reducing workload for human planners or computer-run solution models.

By extending the RAM to the carousel layout and servicing method, as well as improving the TAP model, the RAM could be used to either assess a single or set of candidate initial stabling plans on robustness or be used to assess the performance of the underlying TAP model to generate robust initial stabling plans. Furthermore, the RAM could be used to find patterns in the generated stabling plans which could predict the robustness of a stabling plan.

Abbreviations

ATB	*Automatische Treinbeïnvloeding* (Automatic Train Protection)
BD	*Basis Dag* (Standard Day)
Ctw	Utrecht Carthesiusweg stabling yard
DD-AR	*Dubbeldeksagglomeratiematerieel* (Double-Decker Agglomeration Rolling Stock)
(m)DDM	*motorwagen Dubbeldeksmaterieel* (motorised Double-Decker Rolling Stock)
DDZ	*Dubbeldekker Zonering* (Double-Decker Zoning)
EMU	Electric Multiple Unit
FIFO	First In First Out
FLIRT	Fast Light Intercity- and Regional-Train
IC	InterCity
ICMm	*InterCityMaterieel modern* (Modernised InterCity Rolling Stock)
ICNG	*InterCity Nieuwe Generatie* (InterCity New Generation)
ICR	*InterCity Rijtuig* (InterCity Carriage)
IM	Infrastructure Manager
LIFO	Last In First Out
(M)IP	(Mixed) Integer Programming
Nm	Nijmegen stabling yard
NS	*Nederlandse Spoorwegen* (Dutch Railways)
RAM	Robustness Assessment Model
RU	Railway Undertaking
SA	Stochasticity Algorithm
SD	*Specifieke Dag* (Specific Day)
SLT	Sprinter LightTrain
SNG	Sprinter Nieuwe Generatie (Sprinter New Generation)
SPR	Sprinter
TAP	Track Assignment Problem
TMP	Train Matching Problem
TUSP	Train Unit Shunting Problem
VIRM	*Verlengd InterRegio Materieel* (Extended Inter-Regional Rolling Stock)
Zl	Zwolle stabling yard

List of Figures

List of Tables

Table of Contents

1 Introduction

1.1 Background

The transportation sector is a core pillar of everyday life, and the performance of the transportation network is heavily connected to the socio-economic strength of a country or region. A country with a low-quality transport network will not thrive in (international) trade and general attractiveness, resulting in low economic growth and as a further consequence reduce attractiveness to settle or do business in this country even more. The circle of Wegener, shown in Figure 1.1, captures the circular effect of the performance of a transport network very well. A strongly performing transport network increases the attractiveness of the area, which results in more businesses and families moving to the area, growing the (local) economy. As a result, trip generation in the area increases, requiring the transport network to expand further and as a result increase attractiveness even more. The transportation system is similar to the network of blood vessels of a human body. It allows the inhabitants of the area to move themselves to work, shopping, family or friends, and the most important destination: home. Furthermore, cargo can also be transported via these networks from their origin to their desired destination. Moreso, the country or local area strongly depends on transportation network performing the way it is intended to. When it is not performing well, economic growth and social welfare is threatened when businesses and residents move to other areas where accessibility is higher.

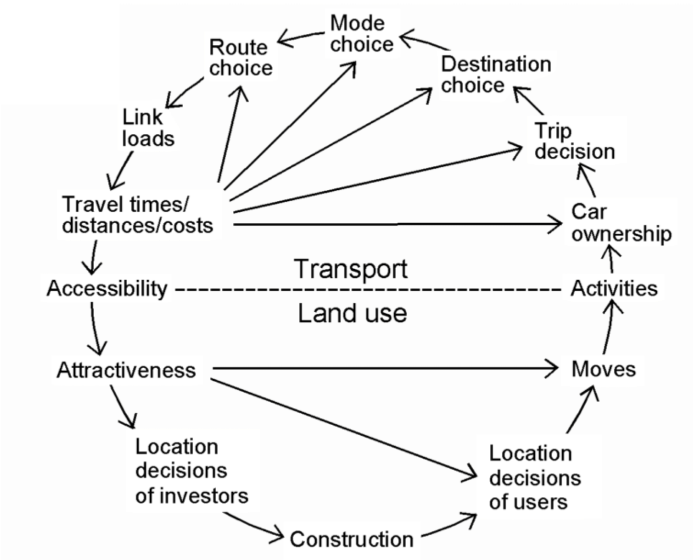


Figure 1.1: Wegener's circle of land use and transport (Wegener, 2004).

An important part of the modern transportation network is the public transport network. Ever since the first train line was founded in the start of the 19th century after the invention of the steam locomotive, public transportation has grown and evolved into larger, denser, versatile, and more efficient networks, together with the private modes of transport. An increase in mobility has led to economic growth and then a further increase in transport demand. The main benefit of (conventional) public transport as opposed to private means of transport is its ability to cluster trips along predefined 'lines' at which travellers can board and alight at predefined 'stops' against a certain price. It is able to transport more people over a link in fewer vehicles, reducing the load on said transport link and therefore increasing the efficiency of the transportation network. As a further benefit, it allows residents who are unable to use or afford a private transport vehicle, such as a car, to have a similar level of mobility compared to residents who have access to a car. Furthermore, public transport is better for the environment compared to cars, especially since the electrification of public transport modes other than already electrified rail vehicles, as shown in Figure 1.2.

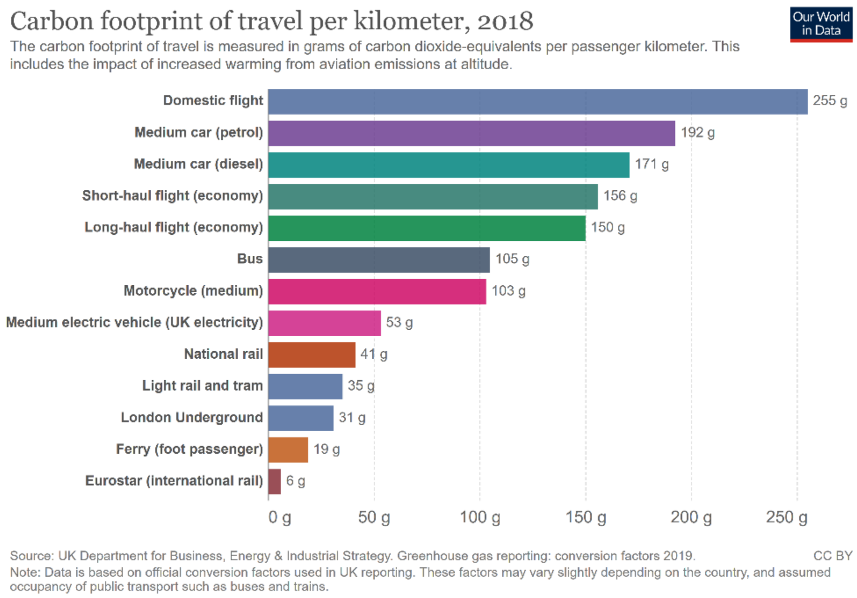


Figure 1.2: Carbon footprint of several transportation modes, visualisation by OurWorldinData.org (UK Department for Business, Energy & Industrial Strategy, 2019).

The Netherlands has a very advanced public transport network compared to other countries. In fact, the average number of daily travellers by train alone was around 1.3 million in 2019, before the start of the COVID pandemic, which equates to ca. 475 million passengers per year (NS, 2019) and a modal share of around 11% for all trips in the Netherlands (Centraal Bureau voor de Statistiek, 2022). Compared to other EU members regarding the modal share of train travel, the Netherlands only has a lower modal share than Sweden, Austria, and Switzerland (Eurostat, 2022).

The main railroad operator in the Netherlands is the *Nederlandse Spoorwegen* (NS, Dutch Railways), and is fully owned by the Dutch government. The NS operates 417 Intercity train units and 307 Sprinter (regional) train units across a rail network of around over 2100

kilometres of track (NS, n.d.), which is called the *Hoofdrailnet* (Main Rail Network).
To ensure that the train services are able to operate according to schedule, a lot of logistical problems need to be solved. A crucial logistical task is the creation of a stabling plan for the stabling yards. In the late evening, when services begin to end, numerous trains arrive at their designated stabling yard, where multiple tasks need to be executed. Arriving trains might need to be serviced (e.g., cleaning, maintenance checks), coupled and/or decoupled, parked overnight, and matched to departing services the following morning. The problem of finding a feasible stabling plan for these tasks is referred to as the Train Unit Shunting Problem (TUSP). This problem has been widely researched by NS themselves (van den Broek, 2016) (Beerthuizen, 2018) (van Leeuwen, 2021) as well as worldwide (Rusca, et al., 2016) (Wang, Zhao, Gronalt, Lin, & Wu, 2022), and has seen big improvements in solvability and performance compared to ealier solution methods over the past years, ever since its first mention by Freling et al. (2005).
Carriages, train units, and compositions are terms recurring throughout this thesis. Figure 1.3 illustrates what these terms are with relation to each other. A train unit consists of multiple carriages. The majority of the trains on the Dutch railway network are Electric Multiple Units, which consist of a predefined number of carriages with on-board electric motors, therefore not requiring separate locomotives at either end. These train units cannot be shortened or lengthened by desire; however, they can be coupled to other train units. A train composition is the full train, which consists of one or more train units.

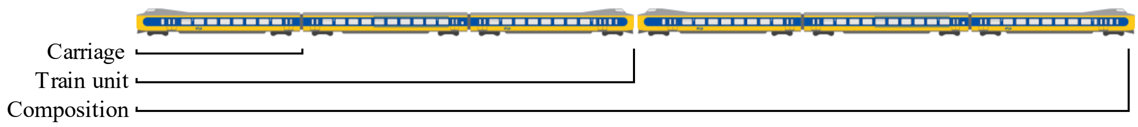


Figure 1.3: Components of a train composition.

1.2 Problem Statement

The planning and timetabling of public transport operations is still subject to a deterministic mindset and starts long before its intended start of execution and is very much so an iterative process. Figure 1.4 shows a general overview of these planning phases up until the moment of operation. Starting with a Basic Hour Pattern and the timetable for a standard day around one year beforehand, the timetable planning is continuously updated to adjust the schedule when additional timetabling constraints become known. Such constraints could for instance be infrastructural maintenance or regional or national events. As a result, the basic day planning acts as a blueprint for the creation of the planning on specific days of operation, which have all been tweaked to remain viable during the infrastructural and operational constraints for that day. These types of variations occur during every planning stage, up until and during the operational day itself, at which rolling stock or crew availability could change last minute, or the occurrence of disturbances requiring further timetabling adjustments. As these uncertainties occur during every planning phase, it is therefore important that an initial plan or timetable is not only feasible in the theoretical, deterministic environment at that planning stage, but is also resistant to the occurrences of stochastic events during later phases. For the performance of a public transport network in practice, an important performance indicator is *robustness*. Robustness is generally defined as the ability of a system to maintain an acceptable level of performance when subjected to disturbances (Norrbin, 2016). Although this general definition is broadly agreed upon, the exact definition varies greatly between different fields, and even between individual researchers within the same field (Lusby, Larsen, & Bull, 2018).

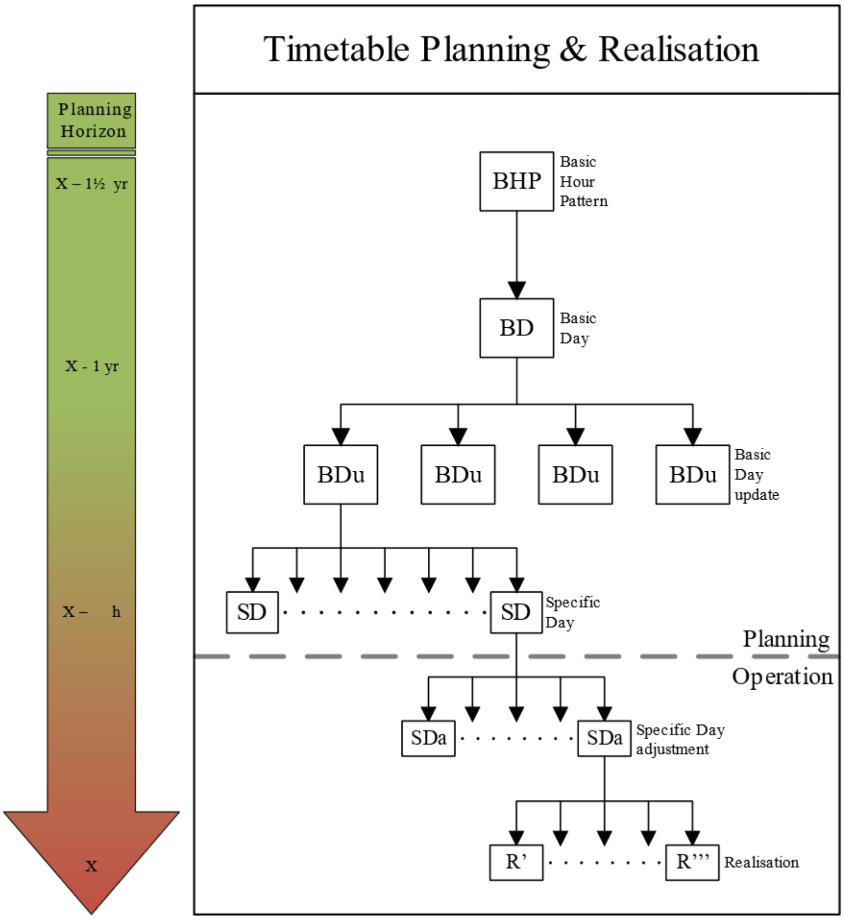


Figure 1.4: Timetable planning and realisation scheme of NS.

Although there has been a lot of research performed on robustness of public transport operations (Goverde, 2007), robustness regarding the stochasticity of the operations at stabling yards and their stabling plans have enjoyed less attention, even though these operations are crucial to ensure services are able to start their timetables for the following day accordingly. Furthermore, it is also unhelpful that until now there is no clear definition in what makes stabling plans robust, as proper definitions of these indicators are necessary in order to provide good assessment methods.

Currently, NS is developing the *Hybride Integrale Planmethode* (Hybrid Integral Planning method, HIP) to aid planning staff in scheduling train movements around a node, where multiple train lines come together at a large station with generally one or more stabling yards in the vicinity. Besides planning the train movements around the station itself, staff also create a stabling plan and schedules the required shunting movements around the stabling yards for trains to undergo servicing and to be parked overnight. This is a very complex puzzle to solve, and HIP aims to alleviate some of the heavy-duty thinking of the human planners by performing a Local Search in combination with a Mixed Integer Programming model to find feasible shunting and stabling plans. However, HIP is still deterministic when it comes to input data to the model, such as the stabling demand in the form of arriving and departing trains, and therefore does not consider the robustness of the plans it returns.

When robustness is not taken into account in the design of a stabling plan in the early design phase, there is a high probability that this initial plan will immediately become infeasible when information regarding the rolling stock changes, such as changes in arrival time or composition. Whenever that is the case, the stabling plan needs to be adjusted or a completely new plan needs to be designed, which is very time consuming, both for human planners as for HIP itself. Therefore, it is crucial that when HIP designs a plan, it should be robust enough to cope with small changes in input, which in this case is the stabling demand in the form of arriving and departing trains. Ideally, a highly robust stabling plan is able to deal with changes in stabling demand with only minimal changes to the original stabling plan and could also be reused as a blueprint in different scenarios. This has the potential to reduce the running time of HIP as it can first search through a database of previous stabling plans with sufficient robustness to start creating a stabling plan from, rather than starting from scratch again.

However, before being able to optimise HIP for robustness of the generated stabling plans, first a definition of this robustness of a stabling plan and an assessment method of said robustness needs to be formed.

1.3 Research Questions

Following the arguments made in the previous section, the central aim for this research is to both define the robustness of an initial stabling plan, as well as give an assessment method to estimate the robustness of such a stabling plan. The main research question of this thesis will therefore be:

How can the robustness of a stabling plan to stochastic events in the planning phase be defined and assessed?

In order to properly be able to answer this main research question, there will be the following accompanying sub questions:

1. *What knowledge is currently available in literature regarding stabling yard operations and, if applicable, how do they take stochasticity of such operations into account?*
2. *Which methods can be used to assess robustness?*
3. *What are the planned operations at NS stabling yards?*
4. *What are the biggest uncertainties in the initial planning stages regarding the feasibility of the stabling plan and how can the robustness of an initial stabling plan be defined?*
5. *How can a stabling plan be assessed on robustness against stochastic events?*
6. *How do the case study stabling plans perform on robustness?*

1.4 Research Scope

The scope of this thesis is restricted to the stabling plan of a stabling yard, thus leaving out planned movements around the yards or stations. The scope is further limited to Last In First Out (LIFO) types of rail vehicle depots, where trains are able to approach the stabling yard from just on side. This limitation has been chosen due to the even higher degree of complexity of free shunt yards, as trains can approach from both sides here, which increases the number of variables and constraints in the model. By restricting this research to the 'simpler' cases, a more solid foundation can be built regarding defining and assessing the robustness of an initial stabling plan. In this thesis, stabling yards for intercity and regional trains of NS are used, but the findings in this research can be applied to stabling yards of other rail vehicles such as metros and trams which are subjected to the same stochasticities of the stabling demand as in this thesis. The data of this research is fully acquired from NS, as this thesis is performed under their supervision. The robustness assessment will be done based on the stabling plan in the initial planning stage, as in this stage there is still a lot of uncertainty regarding the arriving and departing train compositions.

1.5 Thesis Structure

The structure of this report is as follows. Chapter 2 will consist of the literature review regarding the research into robustness in logistical and transport problems, the research into stabling operations, and assessment methods for robustness. Chapter 3 discusses the methodology that will be used in order to answer the remaining research questions of this thesis, as well as elaborate on the data required for the model which will be built. Chapter 4 will focus on the stabling yard operations, starting with their possible layouts, the rolling stock, and the offered services. This is followed by a discussion regarding the planning process for stabling plans and what the influential factors are for the creation of such a stabling plan in different planning phases. Finally, the stabling plan robustness will be defined in this chapter. In Chapter 5 the models used will be discussed. The robustness of different initial stabling plans will be assessed and analysed in Chapter 6. Furthermore, options are explored to improve the robustness of the initial stabling plans. Finally, Chapter 7 will close out this thesis with the conclusions and recommendations. Table 1.1 shows the outline of this thesis, as well as in which chapter each sub question will be answered.

Table 1.1: Thesis outline.

Chapter	Title	Sub question
1	Introduction	-
2	Literature Review	1, 2
3	Methodology	
4	Stabling Yard Operations	3, 4
5	Robustness Assessment Model	5
6	Robustness Assessment	6
7	Conclusions and Recommendations	-

2 Literature Review

In this chapter, a selection of scientific literature useful for this thesis is reviewed. The literature review will provide an answer to the first to research questions of this thesis. The collected literature has been grouped across three themes, which can be found in Sections 2.1, 2.2, and 2.3:

1. Research in logistic/public transport robustness.
2. Research in stabling yard operations.
3. Research in assessment methods for robustness.

The first literature theme will discuss the varying definitions of robustness across researchers and discuss (possible) reasons for these differences. This part of the literature review will help defining the robustness of a stabling plan in Section 4.5.
The second literature theme will analyse current knowledge in stabling yard operations (Train Unit Shunting Problem, TUSP), and whether or not researchers take stochasticity of events into account and, when given, their reasons to incorporate this or not. This section will help understand stabling yard processes, which in turn can aid in finding influential factors to the robustness of a stabling plan, which will be discussed in Chapter 4, and create the assessment model in Chapter 5. This section will play a key role in defining the research gaps.
The third theme will focus on literature regarding the performance assessment method of logistical/transport problems. This will lay a foundation for the argumentation of which method will have to be used to build the robustness assessment model in Chapter 5.
Finally, after these three themes, Section 2.4 will discuss the discovered research gaps from the literature review.

2.1 Robustness in Transport

Robustness is an important indicator for the performance of a logistical network, such as the public transport network, when subjected to disturbances. However, as this indicator is used in numerous fields of expertise, each field has their own variation of the precise definition of robustness, with even variations in definitions between researchers in the same fields. Furthermore, some researchers use 'disturbances' and 'disruptions' interchangeably, even though these do not mean the same. Disturbances in public transport networks are seen as small deviations to the schedule, such as delays, whereas a disruption is seen as an event which has a much larger impact on the network requiring radical changes in the schedule, such as the closure of a railway line between two busy stations.

Generally, robustness can be defined as the ability of a system to maintain a desired level of performance despite disturbances (Norrbin, 2016). Lusby et al. (2018) argue that there is no true definition for robustness, as its definition heavily relies on what the involved parties want to show with the robustness assessment. They further note that the central pillar of a robustness definition is the presence of stochastic events. As robustness is the main theme of this thesis, this section of the literature review focusses mostly on comparing robustness definitions across literature in the logistical and transportation sectors.

Robustness is not only a standalone performance indicator, but it is also very much intertwined with other indicators, such as reliability. Van Oort (2011) mentions that network robustness must be increased in order further improve service reliability, with Tahmasseby et al. (2008) stating the same in their research earlier, also indicating that when reliability and robustness is included in the design phase, the final network design can look very different compared to a network designed without these indicators kept in mind. By performing a case study on the tram network in The Hague, the writers conclude that infrastructural measures such as bypasses can greatly increase network robustness and reliability, underlining the fact that decisions in the strategical and tactical planning stages can greatly affect the operational performance. This is echoed by the findings of Ge et al. (2022), who mention that the robustness of a network is connected to its connectivity. Furthermore, they state that robustness is also strongly linked with resilience, as the resilience indicates how quickly a system is able to recover from a disruption and return to normal operations again. They further added that disturbances can be divided into different types regarding their characteristics. Examples of such characteristics are their probability of occurrence, impact, duration, size, and frequency. Morga (2013) has also stated this in their research, linking a high robustness indicator with low recovery costs when under disrupted situations. They further add that optimisation of scheduling can be categorised in four different methods:

1. Deterministic: seeking an optimal solution when all information is known and there exists no stochasticity.
2. Stochastic: seeking best performing solution on average, given multiple scenarios with stochastic elements.
3. Robust: seeking the best performing solution in the worst-case scenario.
4. Recoverable robust: seeking the solution which is able to recover from disturbances with the least effort as possible, given a number of recovery strategies.

Lusby et al. (2018) have performed a literature survey on robustness in public transport operations and state that robustness in public transport can be assessed across three planning stages: network design and line planning, timetabling, and rolling stock and crew planning. They themselves have investigated different literature to find a definition of robustness in this stage of planning. In this stage robustness generally can be viewed from a passenger or operator perspective. Where for passengers the robustness is related to minimising the additional travel time or missed connection in the case of a disruption, the robustness from an operator's perspective is aimed at minimising additional operation costs as a consequence of the disruption. What is notable is that these two metrics are inversely related; when optimising from a passenger's perspective, disruption costs increase substantially, while when optimising by reducing disruption costs, robustness seems to be best, but it worsens the performance from a passenger's point of view. Although they found some useful literature for this planning stage, they state that there is relatively few research available, which is a possible reason that there is no clear consensus of the robustness definition at this stage as of yet. Furthermore, they state that increasing robustness in one planning stage or planning problem will only result in an increase of robustness of the complete system if all other stages and problems are able to match this increased level of robustness themselves, as these are all interconnected. Otherwise, the increased robustness of this specific stage or problem could never be reached during disruptions or disturbances, as in that case another stage or planning problem would have already failed.

Most of the research regarding robustness has been performed in the field of timetabling. The general consensus in this field is that robustness is strongly related to delay propagation, where buffer- or slack times are seen as a method to reduce delay propagation and increase timetable robustness. Vromans et al. (2006) state that a timetable can only be robust to small disturbances, as it would be impossible for a timetable to be able to withstand large disruptions with just small timetable modifications. Goverde (2007) has addressed the issue of timetable stability using max-plus algebra to identify the so-called critical path, being a series of dependent events with the least amount of slack time in one time cycle (e.g., one hour). Goverde further proposes a model to determine delay propagation and relates this to the robustness of the timetable.
Finally, Lusby et al. (2018) state that for rolling stock planning the most prevalent factor of robustness is minimising or restricting composition changes. Abbink et al. (2004) have further stated that robustness is influenced by the homogeneity of rolling stock types, with a low homogeneity resulting in a lower robustness.

As mentioned earlier in this section, the presence of stochastic events is the central pillar of robustness. Haahr et al. (2015) and Norrbin (2016) both agree on this notion, stating that in order to increase robustness, the aim should be to reduce the stochasticity of events. Norrbin further adds that robustness is a part of the resilience of a system, with rapidity – how quickly a system can recover back to normal performance after a disruption – being the other factor. In general, however, Norrbin states that robustness and resilience are similar terms, with the latter being popular in ecological research. Haahr et al. further distinguish two types of robustness: proactive and reactive robustness. Examples of proactive robustness are buffer times or standby crew, whereas reactive robustness focusses on the recoverability of a system following a disruption by altering the original plan to return to normal operations as quickly as possible, all the while following the original plan as closely as possible. Haahr et al. further repeat the statements made by Abbink et al. (2004) regarding the effect of homogeneity of the rolling stock on the robustness and flexibility of the scheduling of rolling stock.

Other researchers have defined robustness more in an analytical way rather than descriptive. Landex and Jensen (2013) have analysed station capacity and operational robustness with analytical models based on the track- and timetable complexity at station areas, which is inversely related to robustness. For timetable complexity, they propose two methods: one using buffer times between conflicting routes, and one using probability distributions of potential delays. For the buffer time method, they determine all conflicting routes and their buffer times. The timetable complexity is then equal to the division of the number of conflicting routes with a buffer time lower than a predefined threshold by the total number of conflicting routes. For the method with delay probability distributions, the timetable complexity is equal to the total probability of a delayed train resulting in knock-on delays for following trains on a conflicting route divided by the total probability of train delays.
Cicerone et al. (2007) and Liebchen et al. (2009) have analytically defined the recoverable robustness category by Morga (2013), stating that, given an original plan O, the possible disruption scenarios S, and the recovery algorithms R, the plan is recovery robust if and only if there is a feasible solution x for all possible disruption scenarios S when using recovery algorithms R. This means that the writers envision recoverable robustness as a plan

characteristic, rather than a performance indicator like other researchers, since a plan is either recoverable robust or it is not.

In short, robustness is a heavily researched field with regard to the stochasticity of operation, however, this has been primarily focussed on timetable robustness. The most effective method to increase the robustness of a system is to reduce the occurrence and severity of the stochasticity of events in the system. These improvements can be made in two ways: proactive and reactive, with proactive robustness focussing on pre-emptively increasing reserve space or time in the system to contain the disturbance, whereas reactive robustness focusses on increasing the recoverability of the system after the disturbance has occurred.
Even though all researchers from this section have their own exact definitions of robustness and the methods to assess and improve it, to the best of our knowledge, stabling yards have been out of scope regarding defining schedule robustness. Although most of these definitions could in theory be used to assess the robustness of a stabling plan, these definitions are not finetuned to the operations at a stabling yard and its influential factors, leaving out key characteristics of the performance indicator for these specific types of operations.

2.2 Shunting and Stochasticity

The shunting operations in stabling yards pose a big challenge in finding a solution for the matching of arriving and departing train services, servicing of trains, allocating parking locations, and the movements of trains between their scheduled events. Both within and outside of NS, multiple researchers and graduates have been performing research in attempting to improve and extend the solution methods of the TUSP.

The first researchers to mention the TUSP in its form were Freling et al. (2005), where they have used a combination of integer programming and column generation to solve the TUSP. As they have been the first researchers in the TUSP, their model is not as extensive as other, later created models. Their model consists of matching arriving and departing train compositions – the Train Matching Problem (TMP) – also taking into account the ability to couple and decouple train units and assigning these units to stabling tracks – the Track assignment Problem (TAP). The model is not only able to solve this for LIFO stabling yards, but free stabling yards as well. However, they do not incorporate servicing needs and the routing of train units within the stabling yard. Furthermore, their model is also fully deterministic, with no mention of the stochasticity of events in practice.
Kroon et al. (2008) have continued on researching the TUSP and the model of Freling et al. by first of all improving the efficiency of the model by removing a large number of superfluous constraints. Compared to the work of Freling et al. the proposed model now is able to solve the matching and parking problems simultaneously instead of sequentially. The writers do mention that a stabling plan needs to be robust, as in practice disturbances will occur, requiring adjustments of the original plan, however they do not elaborate on how this robustness can be defined. The objective function of the model is aimed minimising splitting of compositions and having different train types on the same track, which is claimed by the writers that this will increase robustness of the stabling plan, as units of the same type and subtype can be used interchangeably.

Haahr et al. (2015) have analysed and benchmarked multiple solution methods for the TUSP across different scenarios. The tested methods consisted of serval types of Mixed Integer Programming (MIP), a randomised heuristic method, and the column generation method as used by Freling et al (2005). However, the latter method was dropped from the analysis very quickly, as it showed far inferior performance compared to the other methods. Benchmark results showed that the Randomised Greedy Construction Heuristic method and the Two-Stage MIP Method had the best performance overall, although each of the tested methods have their own strengths and weaknesses. Haahr et al. further propose a method to solve the matching and parking of train units simultaneously using MIP with a Branch-and-Cut framework. Testing showed that integrating the TMP and TAP into one model is only justifiable when a (random) parking order has already been given. In the models, no servicing and stochasticity in the input data has been implemented.

Van den Broek (2016) has continued the work of extending the TUSP and optimising its performance. They state that mathematical models have become too complex to solve the TUSP within a reasonable timeframe, and that dividing the problem into sub-tasks will not be effective, as there are strong dependencies and interactions between elements of the various sub-tasks. Van den Broek therefore has developed a Local Search model for finding the optimal stabling plan, given the layout of the stabling yard, the arrival and departure times of train units, and a list of service tasks. By using Simulated Annealing, the 'neighbouring' stabling plan similar to the original is selected and their objective function values compared. If the candidate solution is an improvement to the original plan, the candidate is immediately accepted. If the candidate solution does not perform better, the candidate still has a probability on being selected, with the probability decreasing with the difference between the objective function value of the candidate and that of the currently accepted solution. This method has showed to be an improvement over the planning tools used at the time for determining the service capacity of the service location. Van den Broek mentions in his conclusion that a stabling plan is only usable in the real world if the plan is also robust to disturbances. He states that for further research it would be fruitful to define stabling plan robustness and create an assessment method by running stochastic simulations with variable arrival time and task durations. He further adds that the proposed local search method could also be of use in the case of robustness assessment, as it might be able to find alternative stabling plans in case of disturbances to make the original plan feasible again.

Further research has been performed regarding the measurement and optimisation of the capacity of a stabling yard. Van Marsbergen (2018) has created an assessment model of the capacity of an SL when utilising the 100% servicing concept, where trains undergo service checks more frequently. Although the model is fully deterministic, variability of event times and occurrences is mentioned, stating that these variabilities could result in significant different between estimated and actual service capacity.

Gilg et al. (2018) have proposed two IP models for solving the TAP without the use of shunting for LIFO, First In First Out (FIFO), and free types of stabling tracks, given that the TMP has been solved beforehand. Although the model further does not incorporate (de)coupling and servicing, the writers have made an addition to the model which takes into account the robustness of the plan. They state that increasing the robustness of a stabling plan has mainly been restricted to aiming at minimising the number of shunting movements required. However, for their model they incorporate stochasticity in the form of delays in the arrival time of trains.

The model applies a penalty for a pair of trains based on the probability that the following train arrives earlier than the leading train, assuming an exponential distribution of the train delays. The model then tries to maximise the number of assigned trains and minimise the conflicts caused by the delays by computing the Pareto-frontier using the weighted sum method. By incorporating stochasticity in the solution model, the writers state that the stability of the TAP solution is increased.

Van Hövell (2019) has proposed an analytical model to increase the usage of SLs by implementing daytime servicing of train units between the morning and evening peaks. The model is able to match arriving and departing units and is able to schedule the servicing tasks by allowing trains to wait between arriving at the SL and undergoing servicing in an addition to the initial model. However, the model does not take into account routing between tracks and parking locations, as only the vehicles are modelled, not the infrastructural characteristics of the SL. Furthermore, the model assumes no coupling or decoupling is taking place. Van Hövell's model is fully deterministic and, besides a quick mention in the description of the planning process, there is no mention of stochasticity in the model input.

Beerthuizen (2018) has also looked in optimising the utilisation rate by using In Residence Time Strategy to find feasible solutions to the TUSP and compare its solution performance to the Local Search method from Van den Broek (2016) and the OPG tool of NS, which determined parking locations and routing. Although the method of Beerthuizen does not perform as well as both the Local Search and the OPG tool, it is much simpler and therefore has a significantly lower solution time compared to the other methods. They further note that their method is able to generate a robust solution, but it lacks the flexibility to further optimise the solution. They conclude that the proposed method is performing very well compared to the Local Search and OPG, and yields satisfactory results given the lower solution time. Beerthuizen has also tried to draw parallels between the TUSP and the container stacking problem, however fails to find many significant similarities other than the stochasticity in arrival times, the presence of a pickup schedule for containers similar to a shunt plan, and containers blocking the exit of their stack for containers below them, similar to a train at the end of a storage track blocking the exit of the yard for the trains behind it. Main differences have been found to be the variation in train unit types and dimensions compared to the mostly fixed dimensions of containers, the self-propelling ability and disability of Electric Multiple Units (EMUs) and containers respectively, and the ability of having a storage track be accessible from two sides as opposed to a container stack, which is only accessible from the top.

Bao (2018) has attempted to solve the matching, parking, and routing problems of the TUSP by using a sequential algorithm based on Decision Trees and measuring its robustness against delays of train arrivals. The performance of the model is assessed by the number of solution instances found, as well as their robustness. The algorithms for the Decision Trees are being trained by data generated by an instance generator and the Local Search algorithm from Van den Broek and NS, after which their performance is tested against a heuristic solution method and the Local Search method. Bao defines the robustness of the algorithm as the ratio between the number of unique solutions it has to create and the number of problem instances, or 'scenarios', the lower the better. In this research 200 problem instances were used of ten train units, of which the Decision Tree algorithms were able to solve on average 113 instances with three of four unique solutions, resulting in a ratio of around three percent. Although these results show promise, it also indicates that the algorithms were only able to solve for little over half of the problem instances. The heuristic and Local Search methods, on the other hand, are able to solve 187 and all 200 problem instances with a unique solution ratio of 84% and 100% respectively, indicating that these solutions would be less robust to changing conditions. Regarding the low solvability of the Decision Trees algorithms, Bao states possible reasons for this are randomness in training data, error accumulation during scheduling, and the lack of detail of the algorithms. The solvability further reduces significantly when using problem instances of 12 train units. However, a combination of Local Search and Decision Trees might offer a satisfactory middle ground between solvability and robustness with a 100% solvability rate with the unique solution ratio down to 45.5% and 76.5% for 10 – and 12 train unit problem instances. Bao finally mentions that the Local Search method might be able to produce robust solutions itself when its starting point is a robust solution itself.

An important part of the TUSP is the scheduling and performing of servicing or maintenance checks. Over the last number of years, more research has been performed regarding implementing this in solutions methods for the TUSP, as well as aiming to optimise the efficiency of SLs regarding the execution of said checks.
Huizingh (2018) has created an IP model to find an optimal service planning at NS service locations by considering the planning problem as an adaptation of the flow shop problem, in which a series of operations have to be performed in a determined sequence, as opposed to the regular job shop problem, in which the order of operations can differ between jobs. By applying the flow shop problem instead of the regular job shop problem, the number of different possible plans that are considered is reduced. Huizingh's research is limited to the scheduling of servicing, and does not take into account the matching, parking, routing, (de)coupling and staff planning of the TUSP. The created model aims to minimise the tardiness, or lateness, of the trains in the schedule, but is fully deterministic in arrival time, servicing needs, and other input data, which in reality could be stochastic of nature. The model is able to find optimal servicing plans for up to 16 train units and for the majority of the problem instances without any tardiness, with the writer claiming any tardiness in the results in due to tight margins in the time planning and the additional waiting time for trains due to the fixed order of servicing operations.
Mira et al. (2020) have also proposed an IP model that determines an optimal maintenance schedule for trains on a single train line by minimising dead-head and shunting costs, taking into account the number of trains required to be in operation to satisfy travel demand. Furthermore, contrary to the work of Huizingh, the researchers have researched and incorporated the stochasticity of the operations at a stabling yard by performing a Monte Carlo Simulation to assess the performance of the maintenance schedule when subjected to stochasticity in maintenance task durations.
Van Leeuwen (2021) further worked on optimising the scheduling of maintenance tasks in an SL by incorporating resource requirements for maintenance in the form of mechanics with different qualifications. By using a combination of a Genetic Algorithm and Simulated Annealing, the proposed model aims to maximise the number of maintenance tasks that can be performed within a specific time horizon. Although the model is fully deterministic, van Leeuwen does mention that there are forms of stochasticity involved in the scheduling of maintenance tasks, such as the arrival time of trains at the SL, the workload of maintenance per train unit, and availability of mechanics. It is further mentioned that additional research in the effects of stochasticity on the schedule would be able to improve schedule optimisation.
Wang et al. (2022) have also looked into optimising the utilisation of maintenance tracks, in this case at a shunt yard for high-speed rail vehicles. Their Integer Programming model is able

to solve the parking, servicing, and routing tasks of the TUSP by minimising the shunt costs. At the shunt yard, two different types of EMUs are parked, which have to undergo maintenance and washing before they can head out again the next morning. The major constraints to the problem are said to be the capacity of the tracks, the conflicting routes, and the run time of a shunting route. The arrival and service times are fully deterministic, and the assumptions are made that only one train is allowed movement at the shunt yard at a time and that the depot is empty at the start and the end of the planning horizon.

Besides the matching, servicing, and parking, the routing of trains between tracks is also a significant part of the TUSP. Where a majority of the researchers have not incorporated this in their model or made large assumptions to simplify this portion of the TUSP, a few researchers have gone the lengths to model and optimise the routing of train units inside a stabling yard. Van Cuilenborg (2020) states that the local search method of NS and Van den Broek (2016) has trouble with finding a feasible alternative path for a train when its original path is being obstructed by another train. Van Cuilenborg has proposed a Multi-Agent Pathfinding model to solve the TUSP by using a Conflict-Bases Search. This model is able to solve the matching, parking, and routing of the TUSP and its performance is compared to the HIP algorithm used by NS. However, the model does not take servicing and (de)coupling into account and does not mention any forms of stochasticity. Comparing the performance of the MAPF to HIP, the MAPF lags behind in runtime and solution performance.
Borecka (2021) has approached improving the pathfinding methods of the TUSP from a different angle, namely by using Deep Reinforcement Learning. They state that the local search method by Van den Broek (2016) and NS cannot account for the stochasticity in the arrival and departure time of trains, as well as the variability in service task duration. The DRL model has shown to increase solvability and reduce unnecessary movements but does this far slower compared to HIP due to the long training time required for the model to recognise and apply patterns.

Finally, van den Broek has further continued working on the TUSP himself as well. In his PhD thesis (2022) an extension to the local search model is proposed that includes staff (drivers, mechanics, cleaners) assignment to service and movement tasks, also taking into account walking time between subsequent tasks for crew. The two alternatives for the staff scheduling are a List Scheduling Policy, which assigns staff to the activity list, and a Decomposition Heuristic, which splits the problem up into the part of assigning staff to activities and the part of determining the start time of activities and solves them sequentially. Comparing these two methods, experiments showed that the first method performs better compared to the second in the likes of solvability as well as computational efficiency.
Of the staff assignment problem, van den Broek states that the assignment of drivers is the main bottleneck, as train movements take place in a variety of locations, whereas cleaning and maintenance happens on a relatively small set of predetermined locations. They propose two MIP models, one which allows drivers to ride along with another driver to travel to the start of their task to save on walking time, and one who does not allow this. The models aim at minimising the required number of drivers for the set of tasks, while robustness of the schedule can be increased by maximising the slack time between activities. Compared to the previous model for staff assignment, the two proposed driver assignment MIP models showed that less drivers are necessary to ensure execution of the scheduled activities. Furthermore, allowing multiple drivers on the same train does reduce the walking time for drivers between tasks but

the gains are not significant. On the other hand, the additional computational strain is said to be small, so this addition to the MIP model could be used for sites which are very spread out. Van den Broek has also researched the effectiveness of a number of robustness measures to estimate the robustness of a shunting plan (2022) (2018). He states that in the early planning stages, the creation of a robust shunting plan is a challenge, as there is no distribution of the uncertainty. Van den Broek uses an MC simulation to estimate the robustness of a shunting plan based on the fraction of trains departing later than scheduled. The results of the MC simulation trials are then compared to the calculated values of certain robustness measures to investigate correlation and therefore the robustness estimation performance of said robustness measures. The simulation model takes stochasticity of events into account in the likes of arrival times, which is based on a uniform distribution, and the service and movement times, based on a lognormal distribution with the nominal activity times as the mean and the standard deviation equal to ten percent of the mean. The assessed robustness measures are:

1) The sum of total slack times.
2) The sum of free slacks.
3) The minimum total slack.
4) Sum of the fractions between free slacks and respective activity times.
5) Maximise assignment intervals.
6) Maximise the minimum assignment interval.
7) Minimum path slack times divided by the number of activities.
8) Minimum probability that path of activities is completed within the deadline.
9) Probability that final activity falls outside of the deadline.

The MC simulation followed by statistical analysis showed that the best correlation robustness measures are measures eight and nine. These two measures greatly benefit from the known variance in uncertainty. When this variance in uncertainty is not known, only nominal times can be used. In that case, measures three, six, and seven show the best correlation. Measures two, four, and five showed no correlation.

To summarise, the TUSP has been heavily researched the past two decades, with the primary focus being on improving the performance of the solution methods and extending the TUSP with more sub-problems, such as crew allocation. However, slowly researchers have been adding forms of stochasticity into their models, such as changes in maintenance durations and arrival or departure times, but this stochasticity has rarely been the central pillar in the respective research.

2.3 Robustness Assessment

Performance assessment methods can be categorised into two groups: analytical, and simulation. The analytical method offers a way to quickly find an exact solution in a generic predefined environment, whereas simulation aims to estimate the performance by mimicking a system which is subject to time. For robustness assessment, researchers have both attempted to analyse this with analytical and simulation methods.

Bao (2018), whose research has been mentioned in this literature review already, has assessed the robustness of the shunting process in an analytical manner by using decision trees

constructed with the help of Machine Learning. By generating a number of problem instances with variable arrival times, the number of required unique solutions to solve all these problem instances is calculated.
Goverde (2007) has linked robustness to the total amount of slack time for a connected series of activities within a specified periodic cycle time, with the critical path indicating the cycle with the lowest amount of slack time. These calculations are done analytically and deterministically using max-plus algebra. However, one could argue that this slack time on its own is not a good indicator for robustness, as it omits the stochasticity of railway event. For instance, a path with a very low slack time may have much better reliability with less severe delays, resulting in less secondary delays, while a path with a very high slack time but low reliability and on average very high delays, which would result in very frequent occurrence of secondary delays. The latter might have a higher stability compared to the former, but its robustness can be seen as very low compared to the path with less slack time, as it requires timetabling adjustments much more often.
Goverde et al. (2009) have expanded this deterministic analysis model by implementing stochasticity in the activity times in the determination of delay propagation. The writers state that this dynamic behaviour could be assessed with the means of simulation, performing this network-wide would result in extreme computational workload, whereas the max-plus algebra model allows this analysis to be done analytically using a number of algorithms, while being much quicker in the process. This method could then be used to assess critical path cycle times with varying values for mean and standard deviation values for the delays.

However, when assessing robustness measurements in complex logistical network problems which are influenced by the presence of stochastic events, exact or analytical methods begin to fall short in a majority of the cases due to the increasingly high complexity of these network problems, leading to very large computational workload. In these situations, simulation methods are generally chosen as assessment method, as simulation is able to assess these metrics quicker and more efficiently compared to the analytical methods. Furthermore, it was found that the performance of a network cannot be determined by deterministic parameters alone, and that therefore probabilistic parameters should also be taken into account (Ozkan & Kilic, 2019).

Both Van den Broek (2022) and Mira et al. (2020) have both used Monte Carlo (MC) simulation, in which stochastic events can be accounted for by running a large number of trials to capture all reasonably possible scenarios. This has enabled the writers to assess the reliability of their schedules based on the probability that the schedule is feasible under the effects of stochasticity in the scheduled events.
Ozkan and Kilic (2019) have also used a MC simulation in their study, in this case to estimate the reliability of a supply chain network. Performance against stochastic events can be calculated in three ways: exact, best-/worst-case scenario calculation, and estimation methods. With MC simulation being an estimation method, it has already been stated as to why this method performs better than exact methods. Furthermore, estimation methods yield more insightful results for the performance assessment, as the best-/worst-case scenario calculation method only returns the upper- and lower bounds of the performance. For complex networks, the probability that such a scenario will occur approach zero, meaning that these results do not give any insight in how the network is performing in general. An estimation method is able to give more useful insight in the performance, as it captures the performance for all reasonable

possible scenarios, making it possible to not only retrieve a single performance value, but the entire performance distribution, of which it is then possible to determine a confidence interval (Guimarães, Leal Junior, & da Silva, 2018).
Simulation is therefore one of the go-to methods of estimation, either for measuring the performance of a network under the influence of stochasticity or for comparing the performance of different scenarios. Chang and Wang (2021) have created a simulation model in order to determine the exit capacity of an urban rail transit depot. This simulation model allowed the researchers to investigate the performance of different operation modes. Rusca et al. (2016) have done something similar as well, in this case for the transit capacity of a port-side shunt yard. They firstly used a deterministic model to determine the theoretical capacity, then a simulation model is created to evaluate the performance of different shunting processes under the influence of stochasticity and interaction between events into account. Zieger et al. (2018) have also used MC simulation to assess capacity. Using the simulation, the researchers were able to assess the effect of changes in buffer time distributions to the knock-on delays, which are related to the capacity of the rail line. They found that the severity of knock-on delays can significantly differ if buffer time distributions are minimally altered.
Zinser et al. (2019) have shown in their research that an MC simulation does not always have to be extremely time consuming compared to analytical approaches. The writers have been able to create a model that simulates a day of the German rail network in less than a minute. This has been achieved by reducing the resolution of the simulation from a microscopic level to a macroscopic level of detail simulating delays by the means of a sampling train delays from a histogram of historical train delay data, instead of simulating the delays themselves accurately. Although this reduction in resolution will result in a lower similarity to reality, these effects can be acceptable when such high precision is not required for the specific set of answers the model is trying to answer.

Concluding the literature review of this theme, analytical methods begin to fall short for robustness assessment when the system becomes more complex and more stochastic events are added, resulting in much higher computational workload. MC simulation is able to tackle these issues and capture the stochasticity of events more efficiently by generating a large sample of scenarios which can then be evaluated.

2.4 Research Gaps

In Section 1.3, the main research question and the sub questions have been discussed. The first sub question has been defined as:

What knowledge is currently available in literature regarding stabling yard operations and, if applicable, how do they take stochasticity of such operations into account?

This section will summarise the findings of this literature review regarding the three determined literature themes in order to provide an answer to this sub question. Furthermore, the research gaps found in this literature review will be discussed as well, which will verify the scientific contributions of this thesis.

After reviewing the collected literature, it becomes clear that there is still a significant gap of knowledge in robustness of stabling yard plans, both in definition and assessment.
For the robustness definition, it has become clear that, to the best of our knowledge, there is no true definition for robustness, let alone a definition for robustness in initial stabling yard plans. In fact, robustness of a stabling plan has only rarely been mentioned in literature, and the writers who did, have not gone the lengths to actually define this robustness and what this definition entails, nor provide an assessment or measurement method for this robustness. Before any assessment model of stabling plan robustness can be created, it is therefore of critical importance that first of all the robustness is precisely defined.

Furthermore, the research focus of robustness in public transport is generally aimed at 'operational' robustness, that is, robustness of the schedule against the stochasticity of events at the moment of operation itself, such as delays. However, research has been lacking in the 'planning' robustness, which is focussed on the stochasticity of events between the planning phases of the schedule creation, for example, a change in train composition. Figure 2.1 visualises the general timetable planning phases at NS and also shows the difference in focus of these two types of robustness. What each planning process entails is discussed in more detail in Section 4.3, however for now it is only important to note how these planning steps are linked together up until and during the realisation of the respective timetable itself in the operational phase.

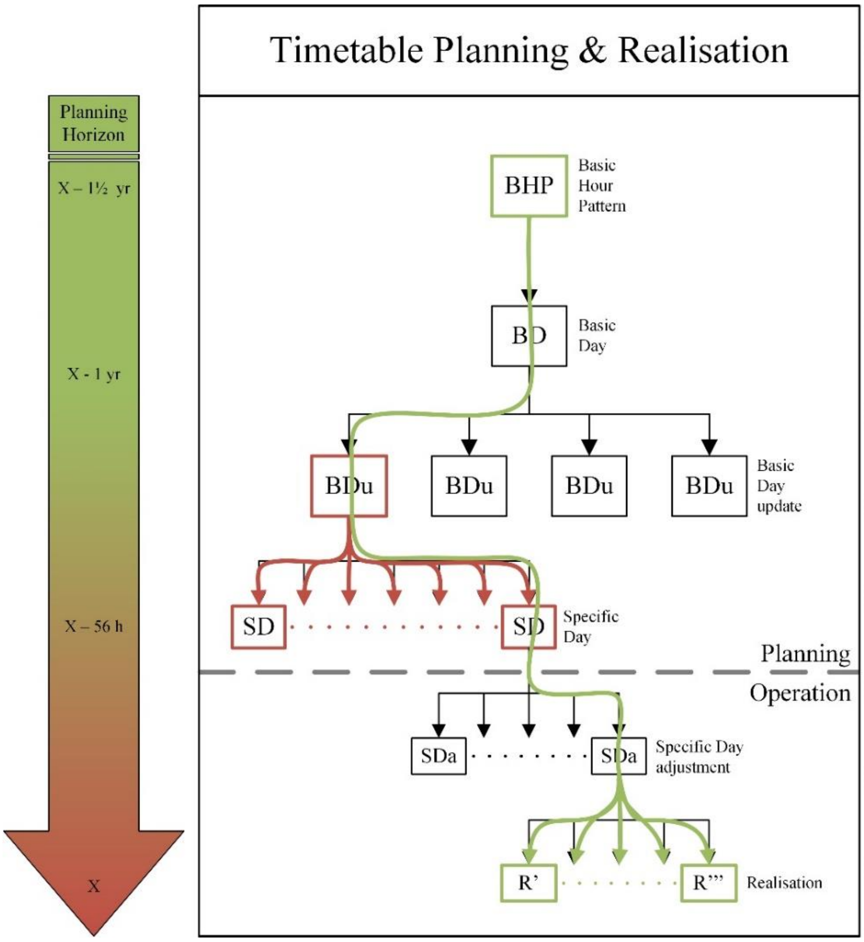


Figure 2.1: Current robustness research planning phases (green) and the robustness research area in this thesis (red).

Although this planning robustness at first does not seem to be of any significant value to the operation, researching this robustness could have benefits to the timetable planning process as a whole. When a created timetable or plan in an earlier planning phase does not have the sufficient robustness to cope with the stochasticity of the demand and constraints of the planning problem, there is a high probability that the plan needs to undergo numerous adjustments to become feasible again, or even be totally remade, which is very inefficient.

Regarding the TUSP, it is clear that there is already a lot of knowledge regarding the solution methods of this problem, and this knowledge is continuing to grow, adding additional activities and complexities to the solution models. The vast majority of these models are fully deterministic. However, some aspects of stochasticity are making their way into these models, such as variable arrival times and maintenance durations. Elements of stochasticity that are still missing in current research though are the stochasticity in servicing needs and especially the stochasticity in train composition. When designing the base stabling plan in the early planning phase, the exact trains are still unknown, only the unit type and composition is generally known, although this may still vary. With this stochasticity, it is likely that adjustments are required to the stabling plan when moving to the planning phase focussed on a specific day of operation when the stochasticity of the required information reduces. Ideally, the number of changes required to make the plan feasible again should be minimised, keeping the similarity between the initial plan and adjusted plan as high as possibles. Incorporating these stochasticities in the assessment model of the stabling plan robustness in the initial planning phase could offer opportunities to improve the updating of the initial stabling plan such that the number of adjustments required for specific days of operation are minimised. Consequently, the aim of this increase in similarity between generated plans originating from the initial stabling plan is to increase the human level of acceptance of the generated specific plans.

Table 2.1 shows a comparison of the collected literature regarding the research in the TUSP and this thesis. The literature is compared based on which TUSP operations are incorporated into the model, which stochasticity in the data are taken into account, and the solution method.

Table 2.1: Overview of collected literature on TUSP optimisation.

Author(s)	Operations in model						Stochasticity in data			Solution method
	Matching	Servicing	Parking	Routing	(De) coupling	Staff	Arr./Dep. time	Servicing durations	Composition changes	
Freling et al. (2005)	✓		✓		✓					IP + Col. generation
Kroon et al. (2008)	✓		✓		✓					MIP
Haahr et al. (2015)	✓		✓		✓					MIP
Van den Broek (2016)	✓	✓	✓	✓	✓					Local Search
Van Hövell (2019)	✓	✓								MIP
Huizingh (2018)		✓								IP
Mira et al. (2020)		✓						✓		IP + MC Simulation
Van Leeuwen (2021)		✓				✓				GA + SA
Wang et al. (2022)		✓	✓	✓						IP
Gilg et al. (2018)			✓				✓			IP
Van Cuilenborg (2020)	✓		✓	✓						MAPF
Borecka (2021)	✓	✓	✓	✓	✓		✓	✓		DRL
Van den Broek (2022)	✓	✓	✓	✓	✓	✓	✓	✓		LS + MIP + MC Sim.
Beerthuizen (2018)	✓	✓	✓		✓					IRTS
Bao (2018)	✓		✓	✓			✓			Decision Trees
This thesis	✓		✓		✓		✓		✓	**MIP + MC Sim.**

This thesis aims to contribute to the research regarding the stochasticity of the input data of the TUSP, as the data and models for this problem currently are predominantly deterministic. Over the last few years, some data stochasticity has seeped into models, mostly of variation in arrival time, and in some cases variation in servicing durations as well. However, variations in arriving and departing train compositions have, to the best of our knowledge, not been researched in this form as of yet.

The second research question of this research was:

"Which methods can be used to assess robustness?"

From Section 2.3 it can be deduced that robustness can be assessed in an analytical manner or by using simulation. However, simulation is found to be the best method to assess the robustness of a network, more specifically the use of Monte Carlo Simulation is recommended, which is reaffirmed by the researchers who have incorporated stochasticity in their models, as they predominantly use MC simulation to efficiently generate a large number of scenarios to assess the performance in its totality. A MIP model will be used to solve the TAP section of the TUSP in this research, as the majority of previous researchers have argued that this is the best method to solve the parts of the TUSP.

As this thesis will aim to define and assess the robustness of a stabling plan in the initial planning phase and incorporate the uncertainties present between the initial and the following planning phase, the literature review has shown that this thesis will have a contribution to the knowledge in robustness of stabling plans, as well as the use of simulation to assess this robustness.

3 Methodology

In this chapter, an overview of the expected methods and tools in order to be able to answer the posed research questions is given for each research sub-question. The main research question of this thesis reads:

"How can the robustness of a stabling plan to stochastic events be defined and assessed?"

This main research question can be answered by joining the answers to the sub questions together. Based on the results of the literature review in Chapter 2, the methodology which will be used to answer each of the sub questions is determined in this section. The answers to the first two sub questions have already been investigated and discussed in the literature review. Therefore, this section will focus on the methodology for the remaining sub questions.

What are the planned operations at NS stabling yards?

This sub question will be approached by using literature from theses at NS regarding stabling yard operations in the past years, as well as discussing and 'riding along' with colleagues who work at the department of planning or at the stabling yards themselves. The goal of this sub question is to gain knowledge in the several types of stabling yards, the rolling stock, and the operations at the yards. This sub question will also aid in giving a proper definition to the robustness of a stabling plan.

What are the biggest uncertainties in the initial planning stages regarding the feasibility of the stabling plan and how can the robustness of an initial stabling plan be defined?

This is a key sub question to this research, as the definition of the robustness of a stabling plan directly influences how the assessment of this robustness is modelled. First off, the uncertainties regarding the available information required in the different planning stages for the creation of a stabling plan, need to be determined. These planning uncertainties are identified based on discussions and experiences with NS colleagues working in the involved departments. These uncertainties are important to take into account when defining the robustness of a stabling plan, as there exists a duality in what the robustness reflects and what it takes into account. Therefore, the definition of the robustness will be made by combining the findings from the collected literature of the first theme, which has focussed on the definition of robustness across different logistical and transportation fields, and the identified planning uncertainties.

How can a stabling plan be assessed on robustness against stochastic events?

Before creating the model which assesses an initial stabling plan on robustness against stochastic events, first needs to be decided which stochastic events are incorporated. From the planning uncertainties found in the previous research question, a selection of stochastic events will be made based on the planning period these uncertainties take place, and whether these events occur regularly. In collected research, such as from Borecka (2021), Mira et al. (2020), and Van den Broek (2022), some stochastic events such as changes in arrival or departure times, or servicing durations, have been incorporated. However, changes to the rolling stock itself have not. The planning stage which suffers the most from these types of uncertainties, is the BDu phase when transitioning to the following SD phase. Therefore, the focus in determining which planning uncertainties will be incorporated into the model as stochastic events, is on uncertainties in the rolling stock of the stabling demand.
With the stochastic events decided upon, a model needs to be created that is able to assess the robustness of an initial stabling plan against these events. This thesis proposed the Robustness Assessment Model to estimate the robustness of a BDu stabling plan, which will be created in the Python programming language. The purpose of this model is to investigate how effectively an initially designed BDu stabling plan is able to deal with changes in stabling demand by altering the original plan to fit the new stabling demand in the best way. Morga (2013) has stated that a high robustness of a plan is linked to low recovery costs in a disruptive state. In this situation, the recovery costs of a plan would be how many changes a BDu stabling plan would need to undergo to fit the new, disruptive state, namely the changed stabling demand in the SD phase. Therefore, a highly robust BDu stabling plan would be able to turn into the optimal SD stabling plan for each of the generated SD stabling demands with the least number of changes.
To estimate the robustness of a BDu stabling plan by determining how many changes it needs to undergo to match the ideal SD stabling plan for the SD stabling demand, the model takes three main steps:

1) Create a BDu stabling demand and stabling plan.
2) Create a set of SD stabling demands and stabling plans.
3) Compare the differences between the BDu stabling plan and the SD stabling plans.

<u>BDu stabling demand and stabling plan creation:</u>

First off, the model needs to obtain a BDu stabling demand and accompanying stabling plan. This can either be real, historic stabling demands and plans, or can be generated from scratch. For this research, the BDu demand and plan is generated, as this gives more freedom in creating the desired scenarios. For the creation of the stabling demand, NS have already created a model which does this realistically, namely the Instance Generator. With some extensions to this generator in the form of additional train unit types, the BDu stabling demand will be generated by this model. Given the BDu stabling demand and a set of stabling tracks, a stabling plan needs to be created. For this, a TAP model will be created which is inspired by the model used by Freling et al. (2005). As the scope of this research is limited to stabling yards which follow the LIFO method of stabling, the TAP model should ensure that the stabling demand is stabled accordingly to this method. Furthermore, contrary to the model of Freling et al., multiple train compositions need to be able to be stabled at the same track. Besides the LIFO rules, the TAP model also needs to adhere to that the available length of each stabling track must not be exceeded by the total length of all compositions stabled at the respective track.

SD stabling demands and plans creation:

The second step of the model is to create a set of SD stabling plans to which the BDu stabling plan can be compared to. Given the generated BDu stabling demand, a Monte Carlo simulation will be used to generate this set of SD stabling plans. Monte Carlo simulation is often used to estimate robustness of an initial plan when under the influence of stochasticity (Mira, Andrade, & Castilho Gomes, 2020) (van den Broek, 2022), and it has been argued in Section 2.3 that Monte Carlo simulation is the better option to assess the robustness compared to analytical methods. The simulation will first run the BDu stabling demand through a Stochasticity Algorithm, which subjects this demand to the stochastic events determined earlier, to generate an instance of SD stabling demand. The algorithm loops through the whole stabling demand and decides if and where the stochastic events occur, based on the probability that each of these events occur. The simulation will run the BDu stabling demand through this algorithm in parallel multiple times to obtain a set of SD stabling demands.
These SD stabling demands are then subsequently fed into the same TAP model as in the first step to generate a set of SD stabling plans. It is deemed acceptable to generate these SD plans without taking the BDu plan in mind, as it is assumed that the BDu stabling plan generated by the TAP model is the optimal solution to the BDu stabling demand, and since the SD stabling demand is a variation to the BDu stabling demand and therefore is very similar to it, the SD stabling plan should already look very similar to the BDu stabling plan as well.

Comparing the BDu and SD stabling plans:

The final step of the model is to compare the BDu stabling plan and the generated set of SD stabling plans to estimate the robustness of the BDu plan. This is done by comparing the BDu stabling plan to the SD plans to investigate how much they differ from each other, in other words, how many changes are required to the BDu stabling plan to turn into each of the SD plans? Before being able to determine how different two plans are from each other, first an algorithm needs to be created which is able to calculate and express the total difference between two stabling plans in an objective manner. With this algorithm, the BDu and SD stabling plans can then be compared to each other and subsequently, the robustness of the BDu stabling plan can be estimated by investigating whether the generated BDu stabling plan is indeed the best blueprint for the creation of SD stabling plans in the later planning phase, or that possibly some other stabling plan would serve better as a blueprint.

Figure 3.1 below gives a graphical representation of the model scheme for the Robustness Assessment Model.

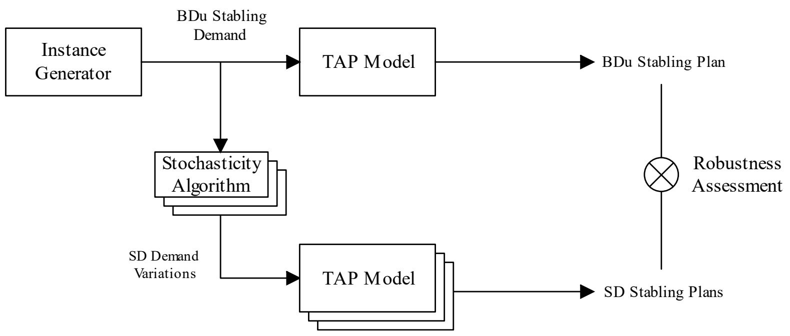


Figure 3.1: Robustness Assessment Model scheme.

Besides, also a sensitivity analysis will be performed to investigate the influence of each of the stochastic variables in the Stochasticity Algorithm on the robustness estimation. The sensitivity analysis will follow the one-factor-at-a-time method, in which only one of the variables will have a changed value at a time. The resulting key metrics of the robustness estimation will then be compared to the results of the model with the regular stochastic variable values to investigate how much the results have changed. For each stochastic variable, a high- and a low-scenario, in which the affected variable has an increased and decreased value respectively, is used. Comparing the results of both these scenarios for each of the stochastic variables will give an insight in which stochastic events will have a bigger impact on the robustness estimation metrics when their probability is either increased or decreased.

How do the case study stabling plans perform on robustness?

The model will be used on three case study locations, each with three capacity utilisation scenarios. This is first of all to see how the model performs across different stabling yards, and second of all to investigate if and how the robustness metrics change when the stabling demand becomes larger and therefore fills more of the stabling yard. Furthermore, a basic comparison will be made between stabling plans with a high robustness and stabling plans with low robustness to investigate whether possible patterns arise in these stabling plans which could act as a predictor of the robustness of a stabling plan, instead of having to run the Robustness Assessment Model for every BDu stabling plan to obtain an estimate of its robustness.

These sub research questions will together be able to answer the main research question of this research, which reads: *“How can the robustness of a stabling plan to stochastic events be defined and assessed?”*. With the first part of the sub questions, a well formulated definition can be given to the robustness of a stabling plan to stochastic events, which in this research will be focussed on changes in the stabling demand itself, as this has not yet been thoroughly researched. With this definition and the research on the operations at NS stabling yards, it is possible to create a model which is able to estimate the defined robustness for a given initial stabling plan. The sensitivity analysis of the model will then give further insights in which of these stochastic events influence the robustness of a stabling plan the most.

4 Stabling Yard Operations

In this chapter the operations of stabling yards will be discussed. First of all, an introduction to a function of a stabling yard and its distinct types is given. Thereafter, the several types of rolling stock units NS operate are discussed, followed by the different services being offered at stabling yards. This is followed by an explanation of the planning process NS adopts for the general timetabling as well as the planning of stabling yards. From these processes, the critical factors or processes which require alterations to the stabling plan are derived, which in turn leads to the definition of stabling plan robustness.

4.1 Stabling Yards

When a public transport operator does not operate all their services without change in frequency or total interruption, its vehicles need a place to be stored until they are required again to fulfil the timetable. Between peak hours, operators may decide to lower the service frequency to reduce costs, lowering the number of vehicles required for operation. When operators decide to seize operations during the night altogether, all vehicles are removed from service and need storage until the next morning when operations start again. The locations where rail vehicles are stored are called shunt or stabling yards. These yards offer a centralised point where a large number of trains can be safely stored, something which stations cannot do, both in capacity as well as the possible accessibility to the station of unauthorised persons, increasing the risk of vandalism on the valuable rolling stock. The stabling yards are also much more space efficient for storing trains due to the absence of wide passenger platforms. Finally, the stabling yards offer an isolated location where trains can be cleaned, serviced repaired, split or combined, and shunted without interfering with the regular services on the network.

Stabling yards come in all different sorts of sizes and configurations, tailored to the requirements for this specific yard. However, the general layout of these stabling yards can be categorised into two categories: shuffleboard and carousel layouts. Although the carousel layout stabling yards are out of scope in this research, it is important to understand the differences between the two different layouts.
The shuffleboard stabling yard layout features dead-end stabling tracks which can only be accessed and exited via the same side. The trains stored on these stabling tracks can be visualised as a stack, as they enter and exit the track under the LIFO principle, where the first arriving train will exit the track as the last, as it gets blocked in between the end of the track and the other trains stored at the same track that arrived at a later point in time. An example of a shuffleboard stabling yard is Amersfoort Bokkeduinen, shown in Figure 4.1.

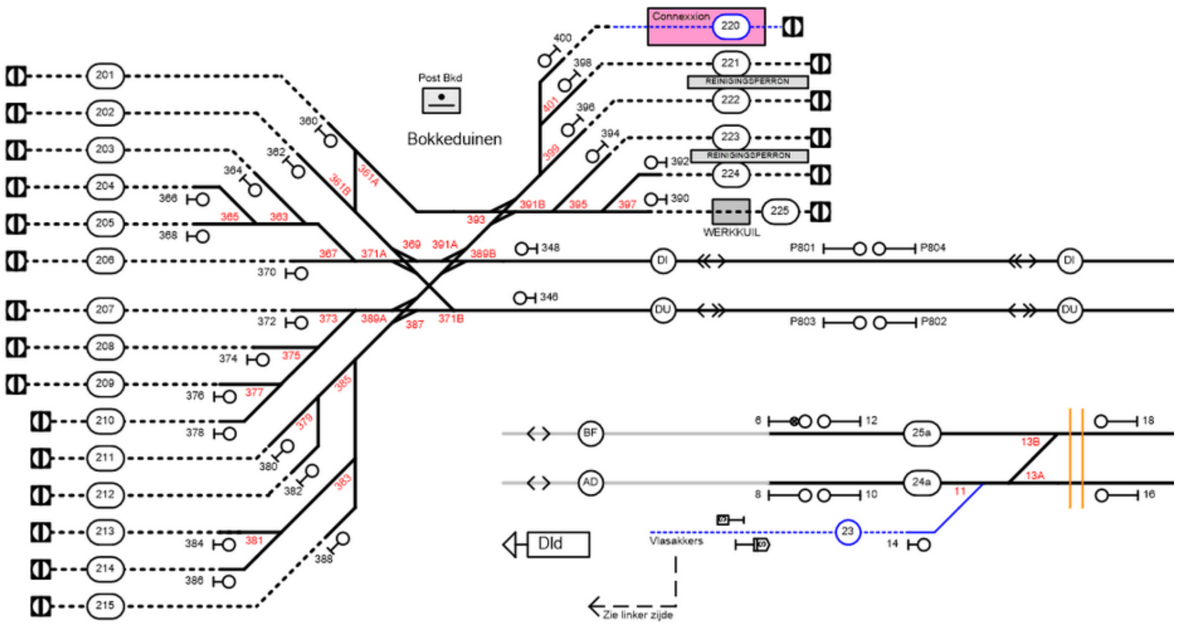


Figure 4.1: Stabling yard with shuffleboard layout at Amersfoort Bokkeduinen, retrieved from Sporenplan.nl (SporenplanOnline).

Carousel layouts feature through-tracks, where trains are able exit and enter the stabling yard from both sides. With this layout, two modes of operations are possible: FIFO and free operation. Under FIFO operations, trains enter the stabling track from one side and leave via the other, essentially forming a queue at each track, where the train that entered the stabling track first will leave first as well. Free operation is essentially a combination of LIFO and FIFO, where trains are able to enter and exit from either side of the stabling tracks. An example of a stabling yard with a carousel layout is Zutphen, shown in Figure 4.2.

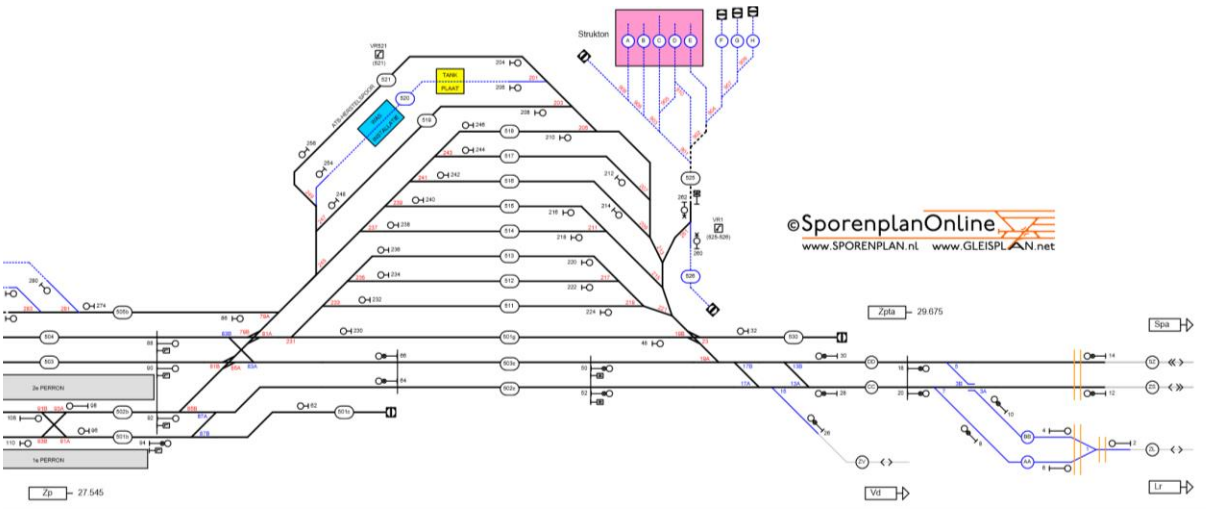


Figure 4.2: Stabling yard with carousel layout at Zutphen, retrieved from Sporenplan.nl (SporenplanOnline).

As mentioned, stabling yards are able to offer a number of services. First off, compositions are able to be split up here or combined. This splitting and combining generally takes place at the start of the night, as soon as the trains arrive at the stabling yard, so that the trains are stabled for the night in the right compositions ready for departure the next morning. Trains are also able to be cleaned at the stabling yard, both internally and externally. Cleaning staff will clean the inside of the train and removing waste from the wastebins on special cleaning platforms. The outside of the train can be washed in special train washing installations. Overnight, the technical components of the trains can also be periodically inspected. These inspections can be categorised in two types, A- and B-inspections, with the latter being a global inspection occurring frequently, and the former being a more extensive inspection on critical components but performed less frequent. Some larger stabling yard also have a technical centre which is able to perform small forms of maintenance when required.

At the stabling yards, two methods of servicing can be used. The first method is low servicing, where trains are serviced at their stabling location, meaning they do not need to be shunted to different servicing locations. The second method is carousel servicing, where trains are shunted between stabling tracks to specific servicing tracks when required. When trains in a stabling yard need to move to another track that requires movements in two directions, for instance to undergo servicing or cleaning, trains need to perform a so-called saw move, at which the train first needs to move to an opposing track, where the driver needs to transfer control from one side of the train to the other and walk to the other side of the train, and then drive the train to the destination track, which is illustrated in Figure 4.3. As the transfer of control from one cabin to another is a complex procedure and walking from one end of the train to another takes a long time, especially for longer compositions, the saw move operation is very time-consuming.

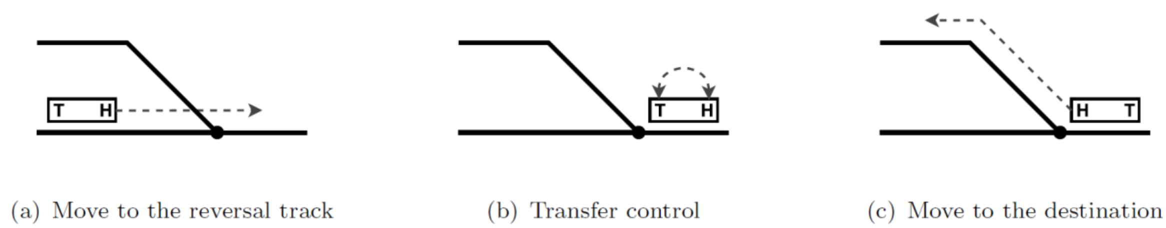


Figure 4.3: Illustration of saw move of a train (van den Broek, 2016).

4.2 Rolling Stock

Public transport networks form hierarchical levels inside them, either naturally due to, for instance, route preference of travellers or higher attractiveness of certain origin or destination nodes, or by design (van Nes, 2002). For example, a public transport network could be split up into two levels; a high and low network level, as illustrated in Figure 4.4, with the high-level network generally having a lower density and higher speed compared to the low-level network. Although the figure only illustrates a bi-level hierarchical network, the network can consist of more than two hierarchical levels (van Nes, 2002). The use of hierarchical levels in a public transport network enables network operators to offer both high accessibility and competitive travel times, whereas a single-level network has to make a trade-off between the stop density and the average operating speed.

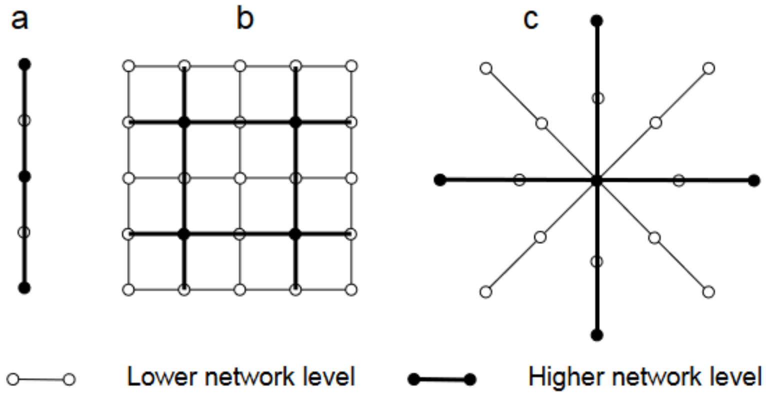


Figure 4.4: Examples of bi-level hierarchical networks in linear (a), grid (b), and radial form (c) (van Nes, 2002).

The NS operate such a bi-level hierarchical network, with long-distance services with a lower stop density and a higher operational speed and short-distance, regional services with a lower operational speed but with a higher stop density, increasing accessibility of the network. However, as it is found that the presence of a transfer in the route applies a hefty penalty to the utility of the route choice (van Nes, 2002), the boundaries between these two levels are somewhat blurred, with long-distance services also stopping at smaller stations to reduce the number of transfers required for affected travellers on this section of track.

The NS classifies these types of services as InterCity (IC) and Sprinter (SPR) services respectively. For these two kinds of services, NS generally operates rolling stock specifically tailored to these respective service types, with the IC trains generally offering more seating possibilities to increase passenger comfort during long trips, whereas SPR trains are closer to the light rail spectrum with higher acceleration and deceleration performance to minimise the time loss of the larger stop density, as well has focussing more on optimisation for passenger capacity in the carriages by swapping out a portion of the seating possibilities for standing instead, as average in-vehicle times are lower in this hierarchy level, making standing in the train more justifiable.

Table 4.1 shows the four different train unit types and their variants which NS use to operate the IC services. Table 4.2 shows this for the three different SPR train unit types and variants.

Table 4.1: Current IC rolling stock units of NS (Treinen van NS) (NS Reisplanner).

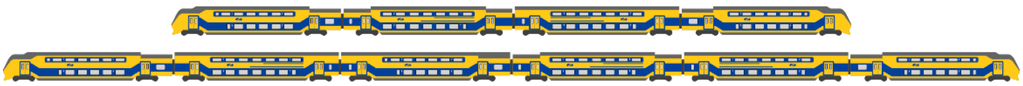
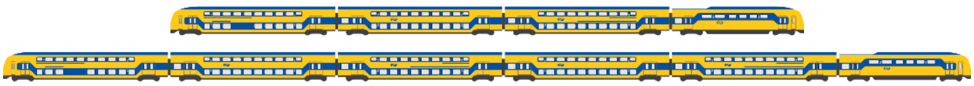
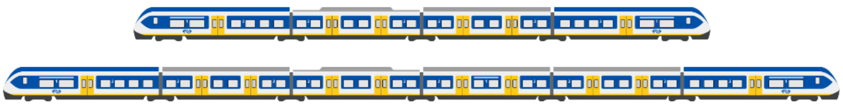
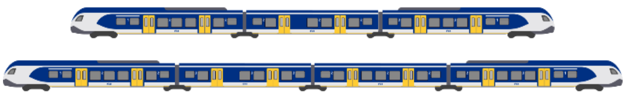
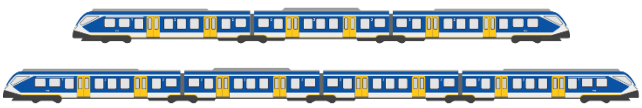
Train unit	Production	Amount	Variant	Carriages	Total Length [m]
VIRM	1991-2009	176	VIRM-IV	4	108.6
			VIRM-VI	6	162.1
					
ICMm	1991-1998	137	ICMm-III	3	80.6
			ICMm-IV	4	107.1
					
DDZ	1991-1998[1] 2012-2014[2]	50	DDZ-IV	4	101.8
			DDZ-VI	6	154
					
ICD	2014-2016[3]	77 ICR + 31 TRAXX	-	7/9	26.4 (ICR)[4] 18.9 (TRAXX)
					

Table 4.2: SPR rolling stock units of NS (Treinen van NS) (NS Reisplanner).

Train unit	Built	Amount	Variant	Carriages	Total Length [m]
SLT	2009-2012	131	SLT-IV	4	69.4
			SLT-VI	6	100.5
					
FLIRT	2015-2017	58	FLIRT-III	3	63.2
			FLIRT-IV	4	80.7
					
SNG	2017-2019	206	SNG-III	3	59.6
			SNG-IV	4	75.8
					

[1] Original DD-AR carriages
[2] Modernisation into DDZ units with motor carriage mDDM
[3] Modernisation
[4] Per carriage

4.3 Planning Process

Before being able to define the robustness of a stabling plan and build an assessment method of said robustness, an understanding of the planning process for creating a timetable at NS is crucial. A rail network timetable first of all contains the schedule of each line, which contains the running times between stops and dwell times at each stop. This is expanded by the planned train paths, both on open tracks between station nodes, as well as at the station nodes themselves at the interlocking areas in order to find conflict-free paths for each train. By having a timetable in place, both the customers and operators have an overview which services are running and at which times. The network services become predictable as a consequence, making trip planning easier from a demand perspective, and crew and rolling stock planning from a supply perspective.

The Infrastructure Manager (IM), which is ProRail for the Dutch railway network, is the final responsible party for the feasibility and execution of the complete timetable of the Dutch railway network, including both the passenger services and freight services. However, ProRail does not create the timetable from the ground up. Rather, every Railway Undertaking (RU) hands in their proposed timetables for their services to ProRail ahead of time. ProRail then has to ensure that the minimal requires safety margin between trains is respected by modifying the proposals of the RUs when necessary, such as changing departure or dwell times, or cancelling a service altogether, deeming it infeasible to operate on the network. The RUs then have the option to either accept the proposed changes by the IM or propose a new timetable themselves for review until an acceptable solution has been agreed upon.

The methods for creating a timetable for public transport operations are not set in stone, with each operator having found or created their own preferred planning method. Every year in December, the NS start to operate the newly designed timetable. In this subsection, the general timetabling method of NS is discussed.

The creation of a timetable spans multiple design phases, rather than just a single design phase in which the complete, functional timetable is created. A visualisation of how these planning phases are connected is shown in Figure 1.4. After the conduction of preliminary studies has been finalised, a Base Hour Pattern (BHP, Dutch: *Basis Uur Patroon*) is created around one and a half years prior to the eventual start of execution of the new timetable. At this planning stage, a standard hour of operation is decided upon, which will form the basis for the finalised timetable. With a periodic timetable, finding a feasible schedule is called the Periodic Event Scheduling Problem (Liebchen & Möhring, 2007). However, contrary to the network timetable, shunting movements to and from a stabling yard are rarely periodic. Therefore, these movements are not considered for the creation of a BHP.

After a BHP has been created, the timetable is expanded to span entire operating days, which is defined as the timetable of a Standard Day (BD, Dutch: *Basis Dag*), where in essence the BHP is copied from the starting time of operations until the end of operations at the end of this day. The first and final hours of operation may look different than the BHP due to the start-up and shut-down processes of each service. Furthermore, during peak hours, service intervals may vary compared to moments outside of peak hours. A BD timetable is created for every day of the week, taking into account that the daily timetables need to be compliant with each other in the transition from one day to another. The BD timetable is created around a year prior to execution of the new timetable, which is then being updated to a BD-update (BDu) timetable once every few months, which serves as a blueprint the next planning stage, as well as a feasibility check for the planned train movements. The goal of these updates is to further specify the timetable to the seasonal fluctuations in demand. After these updates have gone through, the timetable is ready for generic operation.

However, public transport operations are never ‘standard’. Rather, they are subject to continuous changes, the majority of which cannot be planned for beforehand. Examples of such changes are for instance large events such as concerts or sporting events, which change the transport demand in comparison to the projected demand in the BD(u) timetable. As a consequence, service frequency or rolling stock length might have to be increased or decreased to cater to this newly projected demand. Also, planned maintenance on the infrastructure can cause the BD(u) timetable to become infeasible, requiring rerouting or cancellation of trains. As these events are generally unknown during the creation of the BD timetables, it is impossible to plan for in advance. For these reasons, the *Specifieke Dag* (SD, Specific Day) timetable is created a few weeks before the specific day of operation. At this point in time, it is generally known if and when such events take place, and what the consequences are to the operation of the train services. The SD timetable takes the BDu timetable as a blueprint and is then subsequently altered to satisfy the feasibility requirements for that specific day of operation. The deadline for finalising this SD planning differs for the network schedule and the node planning, with the former to be handed over to ProRail four weeks prior to operation, whereas the node planning needs to be finalised 56 hours prior to the day of operation. After this deadline, the timetable ends up in a ‘black box’, as ProRail then locks the timetable in place. This phase marks the boundary between planning and operation, as after this point, changes in the schedule can only be made at the day of execution itself.

At the start of the day, the SD planning is adjusted to fit the newest available information, such as final rolling stock and crew availability as a result of rolling stock breakdown or non-attendance of train staff. Furthermore, planned infrastructural maintenance might have encountered delays, running past the planned maintenance duration, and therefore resulting in required rescheduling actions. During the day, train traffic controllers need to be wary for any type of disturbances or disruptions, such as delays due to calamities, or infrastructural or rolling stock breakdown. The final schedule that is actually being used, is defined as the Realisation. When no disturbances occur, no changes are necessary to the adjusted SD planning. However, when these disturbances do occur, the schedule has to be adapted on the go. This is done by the *Regionaal Besturingscentrum* (RBC, Regional Control Centre). For severe or somewhat frequently occurring disruptions, such as the blockage of a link between two stations, special disruption strategies have been determined. These plans indicate what rescheduling steps should be taken during a disruption to ensure the best service performance reasonably possible. Examples of such measures are the rerouting or short turning of services, or the use of buses between the two disconnected stations to serve as a replacement for the closed railway tracks.

4.4 Planning Uncertainties

As mentioned in Section 4.3, the creation of a timetable starts long beforehand. However, the major drawback of planning ahead is the lack of some of the necessary information in each of these planning steps. When part of the required information for a planning phase in unavailable, assumptions have to be made regarding this missing information. As time then progresses, it might become evident that the previously made assumptions were incorrect and as a result making the designed planning unusable. Consequentially, the plan needs to be adapted to fit the newly available information. However, the problem of a lack of necessary information plays up again; each planning phase requires assumptions which could be 'broken' in the following phase as new information becomes available. Even during operations itself, the possibility of disruptions in the realisation of the timetable means that even last-minute changes could be required. It is therefore crucial for good timetable planning to understand which information is known or unknown, when the unknown information becomes available, and how this influences the feasibility of the timetable planning.

Starting with the creation of the BHP, information regarding the network and the line plans are known, such as the general demand, running and dwell times, as well as the proposed frequency settings of each line and infrastructural constraints. This information allows the creation of the BHP. However, as the BHP is created very early in the process, there is still some useful information lacking resulting in the assumption that travel demand is fixed and does not change over time, while in reality this is not the case due to peak hours. Furthermore, the circulation of the rolling stock is out of scope in this phase, however in later phases this becomes important in planning the start and end points of the individual schedules of the rolling stock.
In the next phase this information is now known, which allows the creation of the BD schedule. However, when the BD schedule is created, maintenance to the infrastructure is not taken into account, as at that point in time this information is not available yet. Another assumption being made is that the travel demand does not differ throughout the months, while in reality this does change significantly. For example, demand is lower during the summer months due to holidays for both students and commuters. For these reasons, the BDu schedules are created, which do respect these changes in seasonal demand, as well as infrastructural maintenance that affects operations for a longer duration than a few days. However, short-term (un)planned maintenance is still unknown at this stage, as well as most of the large-scale events such as concerts or sporting events resulting in an increase of demand across affected routes, requiring more frequent or longer trains.
These matters could only become known just days before the specific day of operation. The SD planning phase allows planning staff to adjust the BDu schedule to fit the requirements and operational constraints on specific days of operation, therefore taking into account maintenance works as well as other planned events which may require alterations to the schedule such as rerouting of train services or cater to changes in expected demand. As mentioned in the previous section, the deadline for handing in the finalised SD schedule to the IM, ProRail, is 56 hours in advance for node schedules and four weeks for the network schedule between these nodes. Whenever any problems influencing the feasibility of the SD plan occur after this deadline NS is forced to sit and wait until the day of operation itself before it is able to adjust the plan again to fit the most recent available operational information. Information that can still be unknown in the SD planning phase is the availability of crew and rolling stock. Crew might call in sick at the start of the day, or train units might have broken down, making them unable

to run their planned service. This might require composition changes or cancellation of trains altogether. Furthermore, while the planned train types and compositions are determined beforehand, the specific train units of the planned operations can still vary or be unknown altogether. As a consequence, maintenance or inspection requirements of the planned compositions can still be unknown. Inspections and maintenance can take a significant amount of time, risking that the time between the scheduled arrival and departure to a stabling yard might not be sufficient to cover the required services for the specific train unit or that the planned stabling location might be unreachable due to the execution of said inspections. Furthermore, there is a risk that problems arise during these inspections, requiring the train unit in question to be removed from active service.

At the day of operation, these possible issues have become apparent, enabling the control centre to adjust the SD plan where needed to incorporate the changes in rolling stock and crew availability. However, daily operations are never shy of disturbances and disruptions requiring further last-minute alterations to the schedule up until the moment of realisation itself. The realisation of the schedule is the final phase in the planning process and features a loop between unknown and known disturbances and disruptions, with each newly arising disruption becoming known information for the control centre to adjust the schedule accordingly while waiting for the next disruption to occur. Figure 4.5 visualises this stream of information knowledge across the timetable planning and realisation phases.

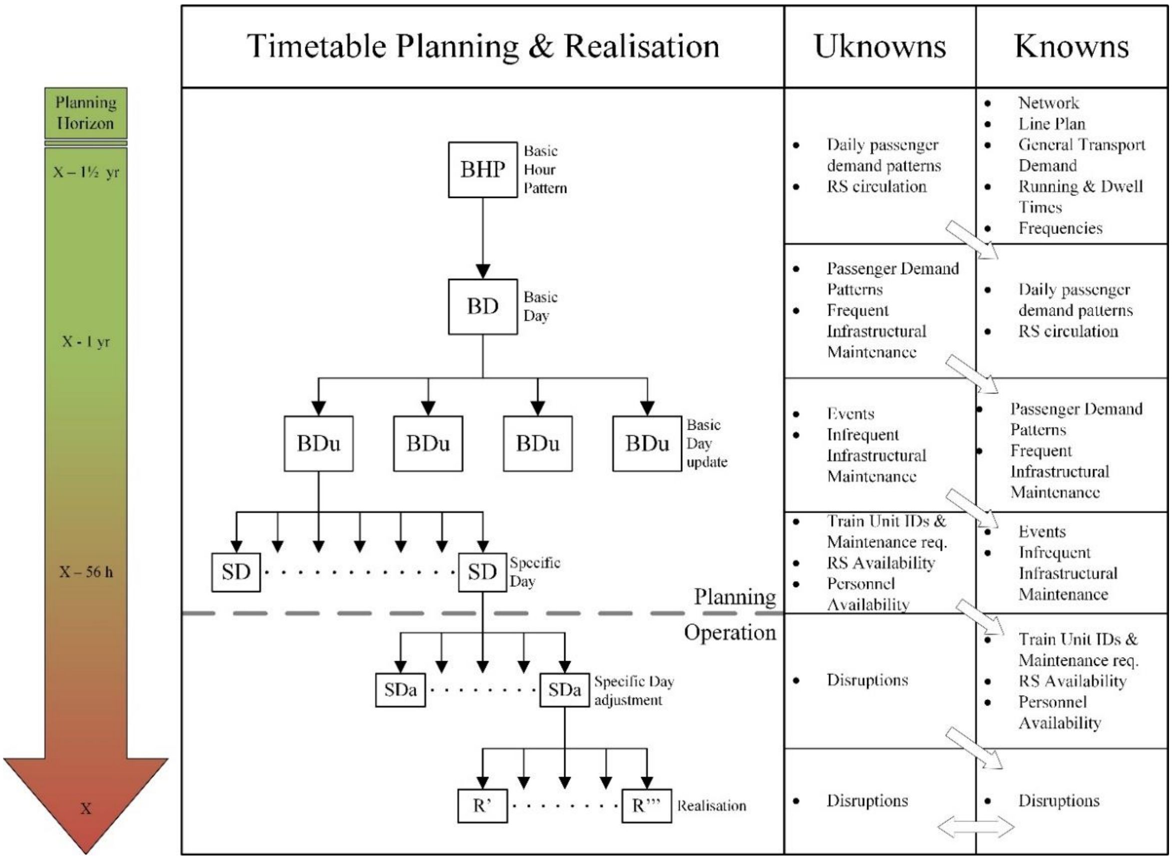


Figure 4.5: Information knowledge stream during planning phases.

4.5 Stabling Plan Robustness

As mentioned in the literature review, a well thought out definition of the robustness of a stabling plan in the earlier planning phase is critical for creating an assessment method for said robustness. It is therefore important to identify the uncertainties the stabling plan should be robust against in the BDu planning phase.

The cycle of unknown and known information during the different planning stages also affects the planning at and around stabling yards and their overarching node. Movements to and from the stabling tracks need to be carefully planned, as these movements should interfere with the already or still ongoing operations as little as possible, which may result in viable time windows for certain movements being very small. When time windows are tight, the movements are more susceptible to variability in events such as a change in arrival or departure time, resulting in necessary adjustments to the planned movements. This can further cause secondary inviabilities of other movements, as the arrival and departure order of train compositions at a stabling yard is of great importance, especially on shuffleboard layout stabling yards where shunting flexibility is limited the most.

The high variability in the node planning process, more specifically the movements to and from stabling yards, causes the BDu node plan to become infeasible very quickly due to a large number of conflicts. This has resulted in SD node planning staff to start their movement planning from scratch rather than using the BDu planning as a blueprint as intended. The stochasticity of events affecting the feasibility of a stabling plan do not stop there, though. After the SD plan is handed in to the IM, a multitude of things can happen which will require an adjustment of the stabling plan. The most common influential factors of variability between the design of the initial stabling plan and the realisation of the plan are:

- *<u>Change in infrastructural availability:</u>*
  It can occur that not all tracks at a stabling yard are operational. When the availability status of a track changes after the design of the initial stabling plan, changes in the stabling plan might be required. Reasons that certain stabling tracks become unavailable can be of technical nature in the case of (unexpected) maintenance, or of an operational nature, when certain stabling tracks are being reserved for trains out of the scope of the stabling plan, such as international trains or trains which are not yet in operation, such as the ICNG trains.

- *<u>Out of service trains present at stabling yard</u>*:
  Before the first train arrives at the end of regular operations, trains might already be present at the stabling yard and are not used in the scheduled operations. The presence, type, and location of these out of service trains are hard to predict in the BDu phase as at that point in time the presence, train type and variation, and stabling track allocation is generally unknown and varies often. The presence of out of service trains influences the track allocation of arriving and departing trains, as the available length at affected stabling tracks is reduced, requiring modifications to the stabling plan when different out of service trains are present than estimated in the BDu phase.

- *<u>Change in arrival/departure time:</u>*
  Besides changes in the rolling stock itself, it is also possible that compositions will have their arrival or departure time changed during the planning process. Note that these

changes in arrival or departure time are not delays in the operational phase, but mere rescheduling actions in the planning phase. These rescheduling actions could result in a BDu stabling plan becoming infeasible due to arising conflicts with arrival and departure order at a stabling track. When this is the case, a new stabling plan therefore would have to be created.

- *Change in train unit variant:*
  The arriving and departing train compositions in the SD phase might differ from the planned arriving and departing train compositions in the BDu phase. For example, due to infrastructural or operational issues such as rerouting services due to the unexpected closure of a route or crew shortages, arriving trains can differ from their planned arrival composition. As a result, the matched departures at the stabling yard will also suffer from the change in arriving composition, as these departing compositions are now different as planned as well. Besides operational problems rooting from these changes in compositions, the planned stabling plan might also encounter issues as a consequence of these changes. For instance, when a planned departure composition turns out to be of a longer variant than expected, for instance a SLT-VI instead of a SLT-IV, there is a chance it might not fit on its planned stabling track anymore, forcing the SD node planner to make changes to the stabling plan to find a suitable stabling location for this changed train.

- *Change in train unit type:*
  Another event is a change of train type altogether. Although this type of event is the least common out of all mentioned events, its effects can be significant. The first problem of this event is the change in composition length. For instance, an arriving composition which is planned to consist of SNG units now consists of SLT units, the length of the composition will be different than the planned length. In the case of a longer than planned composition, the problem could arise that the available room available at its intended stabling track is insufficient, requiring changes in the stabling plan. A second problem arises in the matching of arriving and departing trains. Considering that two different train unit types cannot be coupled, a change in arriving train unit types might result in infeasible departure compositions, which will require changes in the matching of arriving and departing units, and in the worst-case scenario departing with composition shorter than intended due to the lack of the required train unit types and variations.

- *Change in number of train units in composition:*
  A final possible change in actual arrivals compared to the planned arrivals which influences the scheduled departures for the next morning are changes in the number of train units an arriving or departing composition consists of. It can occur that in later planning stages, the number of train units in an arriving composition can be different than planned in the earlier BDu phase. As a consequence, this also influences the lengths of the departure compositions. From a stabling planning perspective, this could mean that there are more or less train units arriving and departing than planned earlier, which can result in the BDu stabling plan becoming infeasible and requiring alterations.

- *Unexpected maintenance requirements rolling stock:*
  As mentioned in the previous section, when designing the BDu plan, the exact trains executing the services are still unknown and therefore the requirements regarding inspection and maintenance as well. The inspections and maintenance can take a significant amount of time, increasing the risk that there is insufficient time between the scheduled arrival and departure of the affected train unit to have the requested inspections or maintenance performed. An accompanying risk to this is that the intended stabling location of the train unit in question is unreachable by the time the train has finished servicing, requiring alterations to the stabling plan. Furthermore, it is entirely possible that problems arise during inspection or maintenance which require the affected train unit to be pulled from service altogether.

- *Rolling stock breakdown:*
  Even at the point of startup of the operations in the morning, problems may still arise. One of these problems is the breakdown of a train during startup at the stabling yard. This can be extremely problematic in a shuffleboard layout stabling yard, where a broken-down train could block the exit of multiple other trains behind it. It will take some time before the broken-down train can be towed from the yard, so rescheduling actions are required.

- *Crew unavailability:*
  Besides unexpected breakdown of rolling stock, crew unavailability may also affect the feasibility of a stabling plan. This crew unavailability can stem from systemic shortage or crew calling in sick for the day. This might result in the plan becoming infeasible as there are not enough drivers to perform all shunting or service staff to perform the servicing actions in time. As a consequence, the initially designed stabling plan needs to be altered such that the plan can be executed in time, for instance by not performing cleaning to some of the trains, reducing the number of required crew and shunting time.

As said, all these events can have an impact on the feasibility of the initially designed stabling plan. With that many possible events which can impact the feasibility of a stabling plan, it means it is nearly impossible for an initial stabling plan to remain feasible until the realisation of the plan, as the input information and constraints to the stabling problem continuously changes. Therefore, the goal of assessing the robustness of a stabling plan should not be to find a stabling plan which stays feasible without any changes the longest, but rather to find a plan which is able to remain feasible in the most reasonable scenarios with on average the least number of changes required. This is in reasoning is similar to the recoverable robustness concept discussed by Morga (2013).
Regarding the stochastic events influencing the viability of a stabling discussed in this section, these events can occur in two timeframes: between the BDu and SD phase, and between the SD phase and the operational phase. The focus of the model that assesses the robustness of an initial stabling plan lies in the former timeframe, as the BDu and SD phases both are part of the planning phase. At this planning stage, the process of the stabling plan creation can still be controlled and adjusted before certain of these stochastic events occur in a proactive manner. Meanwhile, it is much harder to ‘protect’ the initial plan for the stochastic events occurring in

the second mentioned timeframe due to the ‘black box’ of ProRail, meaning that planning staff can only react to these events as they happen.
As the model focusses on the timeframe between the BDu and SD phase, the model should only incorporate the stochastic events which occur relatively frequently and to which the initial stabling plan is able to de made more robust to the consequences of these events. The stochastic events which comply with these characteristics and will therefore be incorporated in the model to assess the robustness of an initial stabling plan are:

- Presence of out of service trains at stabling yard.
- Change in arrival/departure time.
- Change in train unit variant.
- Change in train unit type.
- Change in number of train units in composition.

The stochastic events mentioned earlier in this section that are not incorporated in the model primarily did not make it into the model due to the fact that these events can generally only be acted upon in a reactive manner, rather than a proactive due to these events generally occurring after the node plan is handed over to the IM for the operational phase. Because this research is focussed on the proactive side of robustness, these events have therefore been left out of the model. Furthermore, regarding the changes in infrastructural availability, this event has also been left out due to the fact that these types of events do not occur regularly or in the same fashion across different stabling yard, such as reserving tracks for international trains, and are therefore hard to predict beforehand. Additionally, it is believed that the stochastic event of presence of out of service trains at the stabling yard already add some variability in infrastructural availability for the generation of the stabling plan, and that therefore the addition of such similar infrastructural variability would not add much value to the model.

With the events which will be incorporated into the model decided upon, as well as the exact scope of the model and the goals it is aiming to achieve, a definition can be construed for the robustness of a stabling plan in the planning phase. Given that an ideal designed stabling plan does not necessarily have the highest probability of remaining viable without any changes, but should rather be able to recover from stochastic events with little effort, the robustness of a stabling plan can be defined as the following:

“The robustness of an initial stabling plan is defined as the effectiveness at which the stabling plan is able to cope with the stochasticity of events during the later planning phase with the least number of changes compared to other stabling plans.”

As discussed in Chapter 3, the robustness of an initial stabling plan generated for the BDu stabling demand is estimated by generating a large number of SD stabling demands based on the BDu stabling demand when subjected to the stochastic events discussed in this section. These SD stabling demands will in turn result in generated SD stabling plans. Before the robustness of a BDu stabling plan can be estimated, the generated SD plans need to be filtered to only contain viable SD stabling plans. In this thesis, a stabling plan is viable if and only if all trains are stabled in the stabling plan. The percentage of viable stabling plans of all generated stabling plans is called the viability. The viability does not give an indication of the robustness of the BDu stabling plan, but rather indicates whether the stabling yard has sufficient remaining infrastructure to offer the flexibility to change the stabling plan when required. The robustness of the BDu stabling plan is then estimated using two metrics:

- Offset
- Spread

The offset of the BDu stabling plan is defined as how different the BDu plan on average is compared to the viable SD plans, with a lower offset value indicating that the BDu plan is on average very similar to the viable SD stabling plans.
The spread is defined very similarly to the offset in the way that it looks at the differences between plans. However with the spread, each viable SD plan is compared to all other viable SD plans. The average differences of each SD plan compared to the other plans result in a spread distribution rather than a singular value. The value of the offset is then compared to the spread distribution to estimate the robustness of the BDu stabling plan.

Table 4.3: Planning uncertainties and whether these are incorporated in the model.

Planning uncertainty	**Included in model?**
Change in infrastructural availability	No
Out of service trains present at stabling yard	Yes
Change in arrival/departure time	Yes
Change in train unit variant	Yes
Change in train unit type	Yes
Change in number of train units in composition	Yes
Unexpected maintenance requirements rolling stock	No
Rolling stock breakdown	No
Crew unavailability	No

4.6 Final Remarks

Concluding this chapter, two research questions can be answered. The first of these questions is:

What are the planned operations at (NS) stabling yards?

Based on earlier theses at NS regarding stabling yards, as well as discussing with and shadowing colleagues, a lot of insights have been gained into the different types of stabling yards and their operations. Besides parking being the main function of the stabling yard, the stabling yard can also serve as a location to split and combine compositions, perform internal and external cleaning, perform routine inspections, and in some cases perform simple maintenance to the trains.

The second research question which can now be answered is:

What are the biggest uncertainties in the initial planning stages regarding the feasibility of the stabling plan and how can the robustness of an initial stabling plan be defined?

The robustness of an initial stabling plan is defined as:

"The effectiveness at which the stabling plan is able to cope with the stochasticity of events during the later planning phase with the least number of changes compared to other stabling plans."

This definition of robustness is very similar to recoverable robustness as discussed by Morga (2013) in the way that it focusses not on the pure 'survivability' of a plan or system, but rather how capable it is able to adapt to changes in input data or changes in constraints.
The stochastic events which are most influential to the viability of a stabling plan are:

- Change of infrastructural availability.
- Presence of out of service trains at stabling yard.
- Change in arrival/departure time.
- Change in train unit variant.
- Change in train unit type.
- Change in number of train units in composition.
- Rolling stock breakdown.
- Crew unavailability.

5 Robustness Assessment Model

Now the robustness of an initial stabling plan has been defined, the models for assessing the robustness of a BDu stabling plan are discussed in this chapter. Together with the decided stochastic events in the previous section, this chapter will provide an answer to the fifth research question:

How can a stabling plan be assessed on robustness against stochastic events?

Using the model scheme discussed in Chapter 3 and shown again in Figure 5.1, three main sections can be identified. In the first section, a BDu stabling demand A will be generated by Instance Generator and is subsequently fed into the TAP model, which will generate a BDu stabling plan for this generated stabling demand. The second section of the model is the Monte Carlo simulation portion of the model. The simulation starts by feeding the generated BDu stabling demand into the Stochasticity Algorithm a large number of times, generating the desired number of SD stabling demand variations to simulate the changes in stabling demand in the SD phase compared to the earlier BDu phase. The SD demand variations are then all fed into the TAP model to create the SD stabling plans. These will then be compared to the BDu stabling plan in the third section of the model. The robustness assessment performed in the third section will finally return the values of the stabling plan robustness parameters discussed in Section 4.5.

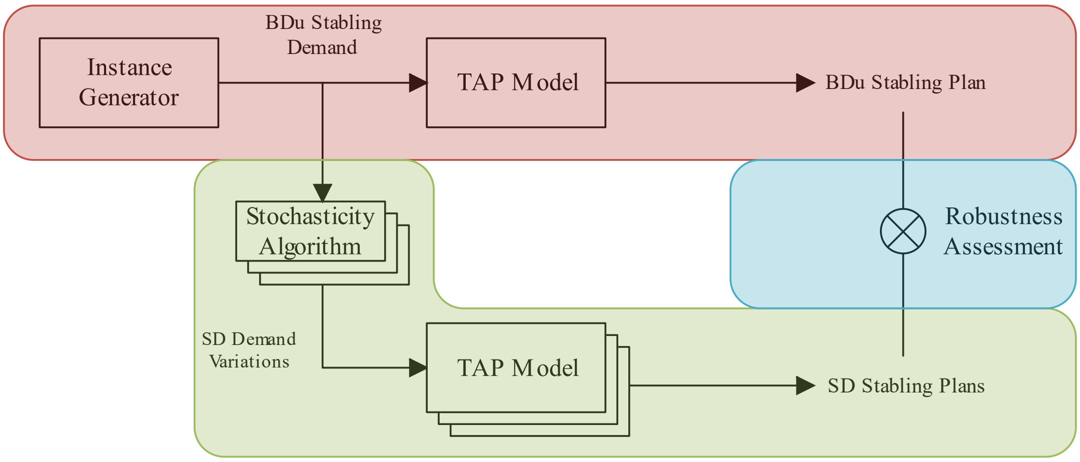


Figure 5.1: Model scheme for assessing and improving BDu stabling plan robustness, with the red section showing the process of creating a BDu stabling plan, the green section showing the Monte Carlo simulation for generating the SD plans, and finally the robustness assessment of the BDu plan in blue.

In this chapter first of all the required input data is discussed in Section 5.1. The Instance Generator created by NS is discussed in Section 5.2, which is then followed by the TAP model in Section 5.3. The Stochasticity Algorithm that generates variants to the initial stabling demand is explained in Section 5.4, with the robustness assessment being discussed in Section 5.5. Finally, the limitations of the models are given in Section 5.6.

5.1 Input Data

Before any models can be run, input data is required. For the Robustness Assessment Model (RAM), this data can be categorised into four sections: trains, stabling yard, minimum duration of stay, and stochastic probabilities.

The train-related data has already been discussed in Section 4.2. However, in this research not all train types are used in the model, rather only the types VIRM, ICM, SLT and SNG are used for running the RAM. Every incorporation of different train type will make it increasingly more complex for the Instance Generator to find a viable solution to the generation and matching of train units. This is also in line with the consideration that it is highly unlikely that all different train types are present at a single stabling yard.
An addition to the data is the maximum number of train units a composition can consist of for each train type. These values have been chosen to be equal to the largest compositions generally found in daily operation, and not necessarily the technically largest possible composition sizes. Table 5.1 gives an overview of this train-related data.

Table 5.1: Required train-related data for RAM.

Type	Service	Max. units	Variant	Amount	Length [m]
VIRM	IC	2	VIRM-4	98	108.6
			VIRM-6	80	162.1
ICM	IC	3	ICM-3	87	80.6
			ICM-4	50	107.1
SLT	SPR	2	SLT-4	69	69.4
			SLT-6	62	100.5
SNG	SPR	2	SNG-3	118	59.6
			SNG-4	88	75.8

Regarding the stabling yards which are used in the RAM, only two pieces of information are required. Firstly, the track identification numbers at which trains are allowed to be stabled are required. Secondly, for all these stabling tracks the available stabling length needs to be retrieved. This information will be introduced together with the case study stabling yards in Chapter 6.
For matching scheduled arrivals and departures, the minimum duration of stay needs to be determined. This duration is set to two hours in the Instance Generator and determines how long a train needs to be stabled at a track before it is allowed to depart again, which is done to ensure there is time for internal cleaning of the train and any required routine inspections. These operations are, however, not modelled in the RAM itself.
Finally, the probabilities for each stochastic event discussed in Section 4.5 needs to be retrieved. NS has recently conducted research in finding the variations in BDu and SD node plans and analysing how often these variations occur. Based on the findings in this analysis, the probabilities for each stochastic event have been estimated and are given in Table 5.2.

Table 5.2: Estimated stochastic event probabilities.

Event	Probability
Change in train unit variant	0.0767
Change in train unit type	0.0127
Addition/removal of train unit	0.0709
Change in arrival/departure time	0.0437

Whenever a train unit is subjected to a change in arrival or departure time, the change in the affected scheduled time is based on a normal distribution, with the average and standard deviation being extracted from the same BDu-SD variation analysis by NS. The average and standard deviation of changes in scheduled arrival or departure times found in the data used by NS for the BDu-SD variation analysis, are estimated to be 53 seconds and 4046 seconds respectively. This is a rather strange value for the standard deviation, as this would indicate that there is an extremely large spread in the distribution of the changes in the scheduled arrival or departure times. However, for the remainder of this research, these values extracted from the dataset provided by NS are assumed to be correct.
Besides changes in the stabling demand, the RAM also randomly generates out of service trains already present at the stabling yard. The aforementioned analysis could not be used to estimate these figures and moreover, these figures vary from stabling yard to stabling yard. Therefore, these values have to be assumed. An average of three out of service train units present at the stabling yard with a standard deviation of one unit has been deemed to be a reasonable estimate.

5.2 Instance Generator

The first step of the RAM is to load in an initial stabling demand, which is in this case from the BDu phase. This can either be done by loading in stabling demand from a real-life scenario, or by generating stabling demand via the Instance Generator (IG). The IG has been developed at NS by the department of R&D Hub Logistics with the goal to generate representative arrivals and departures for a stabling yard considering both realistic arrival and departure compositions, as well as realistic arrival and departure times. For this research, the decision was made to use the IG, as this allows more flexibility in creating multiple variations of stabling demand within a short amount of time as compared to the real-life scenarios.
When using the IG, the user checks the settings of the model, such as the number of desired units and whether or not the arriving and departing train units need to be matched inside the generator and changes them when desired. For this research, the IG is set up to not only generate the arrivals and departures, but also to match the arriving trains to the scheduled departures, whereas the number of units will vary for the different case study scenarios.

Given the number of desired units, the IG uses three main pieces of information to generate realistic arrival and departure schedules, simultaneously determining the compositions and their arrival and departure times:

- Train unit presence ratios.
- Composition presence ratios.
- Arrival and departure time distributions.

The IG then creates a realistic stabling demand in three steps. The first step is determining the arrival and departure compositions. The IG does this by using the train unit and composition presence ratios. The presence ratio of a train unit indicates the probability that a generated train unit is of a specific type compared to all possible train types. Similarly, the composition presence ratios determine how frequent a specific composition occurs in the stabling demand. The IG tries to generate a stabling demand which gets as close as possible to these ratios but does not see these as hard constraints. The unit and composition presence ratios can be found in Appendix A. The values of the ratios are set such that for both the arrivals and departures the sum of the products of the number of trains of a specific train unit in a composition and the ratio of this composition is equal to the presence ratio of the train unit as a whole. For example, the VIRM-4 train unit has a train unit presence ratio of 40376. In the arriving VIRM compositions, it shows that these composition ratios indeed add up to this value: $1 * 29682 + 2 * 3558 + 1 * 3578 = 40376$. The ratios between the possible compositions have already been empirically determined for the VIRM and SLT train units, whereas the ratios for the ICM and SNG train units have been manually added and estimated in this research, taking into account that the sum product of the number of train units in each possible composition and their composition ratios is equal to the overarching train unit presence ratio. Although these estimated ratios may not reflect the real-life ratios as well as for the VIRM and SLT train units, the realism of the estimated ratios are deemed sufficient enough for this research, as the main goal of this research is not to create the most realistic model of the IG. Furthermore, the composition ratios of the VIRM and SLT units have been specifically tuned using measurements across the Netherlands. In reality, different stabling yards will have different composition ratios and also train unit presence ratios due to the varying utilisation of the different train unit types and composition on routes across the country. However, as there is no data regarding the differences in these ratios for different stabling yards, it is assumed that the ratios of all stabling yards are equal to the nation-wide average ratios.

Then, the IG generates the arrival and departure times of the arrival and departure compositions respectively, based on the arrival and departure time distributions, which are visualised in Figure 5.2. Sequentially, the IG then matches the arrival and departure compositions, taking into account a minimum duration of stay of two hours.

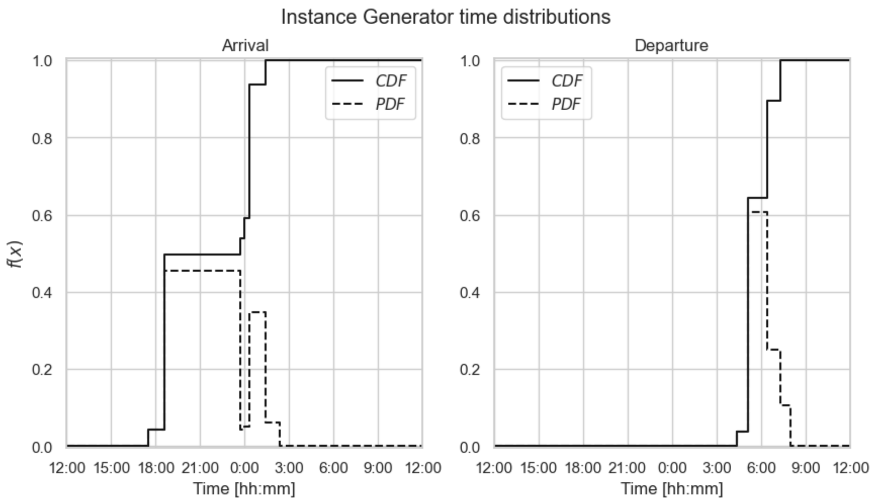


Figure 5.2: Train unit arrival (left) and departure (right) time distributions.

Finally, the IG assigns required services to the arrival compositions which would need to be performed before departure in the morning. However, as servicing is not taken into account for this research, these scheduled services will be disregarded for the continuation of the RAM. Figure 5.3 visualises the steps of the IG and its input and output.

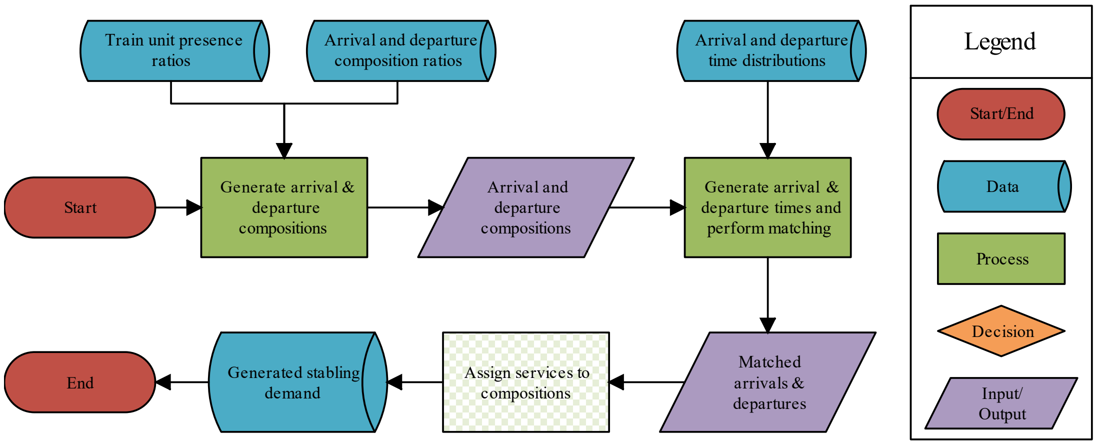


Figure 5.3: Instance Generator model steps.

When the IG has found a feasible solution for the stabling demand generation and matching of arriving and departing train units, the model returns the stabling demand in a table of arriving compositions and a table of departure compositions. For the RAM, these are then also used to create a table of the scheduled arrival and departure time of each train unit in the demand. An example of a generated stabling demand in the form of individual train units as well as in the form of departure compositions are given in Table 5.3 and Table 5.4 respectively.

Table 5.3: Example of train units generated by the IG.

id	Type	Variant	Length	Arrival time	Departure time
9403	VIRM	4	108.6	18:53:37	06:40:37
2403	SLT	4	69.4	20:26:05	06:34:37
2602	SLT	6	100.5	20:50:03	06:03:32
2603	SLT	6	100.5	22:10:34	05:57:32
8603	VIRM	6	162.1	22:16:34	05:32:30
2402	SLT	4	69.4	00:21:13	06:03:32
2601	SLT	6	100.5	00:54:08	05:20:29
2401	SLT	4	69.4	00:54:08	05:44:30
9402	VIRM	4	108.6	01:00:09	05:14:29
8602	VIRM	6	162.1	01:52:39	05:26:30
8601	VIRM	6	162.1	01:58:40	05:38:30
9401	VIRM	4	108.6	02:16:53	07:33:28

Table 5.4: Example departure compositions from generated stabling demand.

id	Type	Variant	Length	First Arrival	Last Arrival	Departure
9403	VIRM	4	108.6	18:53:37	18:53:37	06:40:37
2403	SLT	4	69.4	20:26:05	20:26:05	06:34:37
2603	SLT	6	100.5	22:10:34	22:10:34	05:57:32
8603	VIRM	6	162.1	22:16:34	22:16:34	05:32:30
2402 + 2602	SLT	4 + 6	169.9	20:50:03	00:21:13	06:03:32
2601	SLT	6	100.5	00:54:08	00:54:08	05:20:29
2401	SLT	4	69.4	00:54:08	00:54:08	05:44:30
9402	VIRM	4	108.6	01:00:09	01:00:09	05:14:29
8602	VIRM	6	162.1	01:52:39	01:52:39	05:26:30
8601	VIRM	6	162.1	01:58:40	01:58:40	05:38:30
9401	VIRM	4	108.6	02:16:53	02:16:53	07:33:28

5.3 TAP Model

After generating the initial stabling demand using the IG, the next step is to generate a stabling plan based on this initial demand. This is done using an optimisation model which solves the TAP. The linear solver is provided by the software package FICO Xpress and is called within the Python environment. The goal of the TAP is to find a feasible stabling plan given a stabling demand in the form of departure compositions C and a set of stabling tracks T by allocating a stabling track for each composition. This is done by two decision variables. The first variable determines whether a composition is stabled on a specific track, and the second indicates whether a composition could not be stabled. The model takes the arrival times, departure times, and the composition length from the set of departure compositions, as well as the track lengths for each of the stabling tracks and the distribution of the number of out of service trains present as input data. Table 5.5 shows how all the data and variables are represented in the model.

Table 5.5: TAP model sets, input, and decision variables.

Sets	
C	Set of departure compositions
T	Set of stabling tracks
Input	
M	Very large number (Big M)
t_c^{fa}	First arrival time of composition $c \in C$
t_c^{la}	Last arrival time of composition $c \in C$
t_c^{d}	Departure time of composition $c \in C$
l_c	Length of composition $c \in C$
L_t	Stabling length available a track $t \in T$
Decision variables	
$x_{c,t}$	$\begin{cases} 1 & \text{if composition } c \text{ is stabled at track } t; \\ 0 & \text{otherwise.} \end{cases}$
y_c	$\begin{cases} 1 & \text{if composition } c \text{ is not stabled;} \\ 0 & \text{otherwise.} \end{cases}$

When importing the data, the model has the functionality to add out of service train units to the stabling yard, which will affect the available stabling length on each of the stabling tracks. When this functionality is enabled in the model, first the total number of out of service train units is determined by randomly sampling from a normal distribution with a mean of three and standard deviation of one as discussed in Section 5.1, and is then rounded to the nearest integer number. When this random sample turns out to be negative, the number of out of service trains present at the stabling yard is assumed as none. In the case this number is larger than zero and therefore there are out of service train units present at the stabling yard, the model randomly assigns a stabling track for each unit, determines the train unit type and variant based on the train unit presence ratios, and subsequently removes the length of these train units off the available stabling length at the affected stabling tracks. These adjusted stabling lengths are then fed back into the TAP model.

As mentioned, the objective of the TAP model is to find a feasible stabling plan for a given stabling yard and stabling demand. A stabling plan is determined feasible when all the departure compositions have been stabled. However, when the model fails to find a stabling plan which allows all departure compositions to be stabled, it aims to minimise the sum of the lengths of the compositions which could not be stabled. This way, the model prefers stabling a longer composition (e.g., a VIRM-6 + VIRM-6) and decide to not stable to smaller compositions (e.g., two separate SLT-4 compositions) rather than the other way around. This line of reasoning is represented in the objective function of the TAP model in (5.1). The complete TAP model is then formulated as followed:

$$\text{Minimise} \quad \sum_{c \in C} y_c \cdot l_c \tag{5.1}$$

$$\text{Subject to} \quad \sum_{t \in T} x_{c,t} + y_c = 1, \qquad \forall c \in C \tag{5.2}$$

$$x_{c,t} \cdot t_c^{la} \leq x_{i,t} \cdot t_i^{fa} + \left(1 - x_{i,t}\right) \cdot M, \quad \forall t \in T, \;\; c \in \{1, \dots, C-1\}, \;\; i \in \{c+1, \dots, C\} \tag{5.3}$$

$$t_c^{d} \cdot x_{c,t} + \left(1 - x_{c,t}\right) \cdot M \geq t_i^{d} \cdot x_{i,t}, \quad \forall t \in T, \;\; c \in \{1, \dots, C-1\}, \;\; i \in \{c+1, \dots, C\} \tag{5.4}$$

$$\sum_{c \in C} x_{c,t} \cdot l_c \leq L_t, \quad \forall t \in T \tag{5.5}$$

$$x_{c,t} \in \{0,1\} \quad \forall\, c \in C, t \in T \tag{5.6}$$

$$y_c \in \{0,1\} \quad \forall\, c \in C \tag{5.7}$$

In the TAP model, the stabling constraint (5.2) ensures that a departure composition c can only be stabled at one track at most or not at all. The arrival constraint (5.3)(5.4) states that the last arrival time of composition c which is stabled at track t must be earlier than the first arrival of all other arriving compositions i on track t thereafter. The LIFO constraint (5.4) ensures that the stabled departure compositions adhere to the LIFO entry and exit order in similar fashion to the arrival constraint. The track length constraint (5.5) then ensures that the lengths of the departure compositions l_c stabled at track t do not exceed its stabling length L_t. Finally, (5.6) and (5.7) ensure that the two decision variables of the TAP model are binary variables.

As the stabling demand generated by the IG never results in a situation where a composition has departed before the last composition has arrived, there is no need for a time-dimension in this model to check the length constraint over the course of the night. However, the TAP model can be altered by adding such a time-dimension such that it is able to account for the case a composition departs before another composition arrives, freeing up more space on the respective stabling track. Furthermore, the TAP model does not consider constraints regarding the operational aspects of the stabling yard infrastructure, such movement restrictions for specific track sections. The model further assumes that every train has a conflict-free route to and from the stabling track on arrival and departure respectively, and that there is plenty sufficient time and crew to perform the shunting and servicing tasks. Therefore, it does not determine the routing of trains and scheduling of crew and shunting or servicing tasks, which greatly simplifies.

When the TAP model has found an optimal solution, the model returns the solution values of the decision variables $x_{c,t}$ and y_c. The outcome of these variables is then used to create a table of the stabling plan. This table shows for each composition, similarly to Table 5.4, the train unit ids, the type and variant, and the arrival and departure times. The first and last arrival times are combined in one single column to form the arrival time window of the departure composition. Using the results of the TAP and the stabling yard used in the model, the model determines the stabling track each composition is stabled at, as well as the arrival order of the compositions per track, with the first arriving composition being assigned order number one. The steps of the TAP model as discussed here are visualised in Figure 5.4. Given the departure compositions from Table 5.4 and the example stabling yard parameters in Table 5.6 below, a stabling plan returned by the TAP model is shown in Table 5.7.

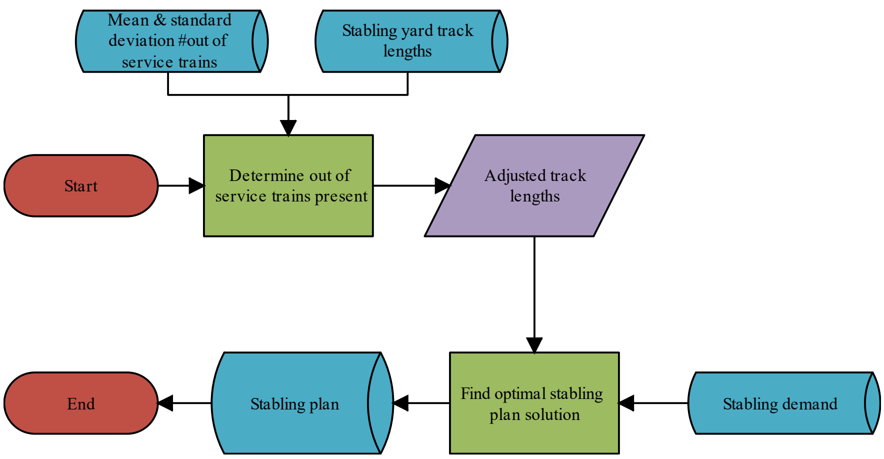


Figure 5.4: TAP model steps.

Table 5.6: Example stabling yard parameters.

Track	1	2	3	4	5	6	7
Length	380	381	311	364	376	338	293

Table 5.7: Stabling plan for generated example stabling demand.

Composition	Track	Order	id	Type	Variant	Length	Arrival Time	Departure Time
OOS 3	1	1	OOS 3	VIRM	4	108.6	-	-
OOS 4	1	2	OOS 4	VIRM	4	108.6	-	-
4	2	1	8603	VIRM	6	162.1	22:16:34	05:32:30
OOS 2	3	1	OOS 2	VIRM	6	162.1	-	-
1	3	2	9403	VIRM	4	108.6	18:53:37	06:40:37
OOS 1	4	1	OOS 1	SLT	4	69.4	-	-
2	4	2	2403	SLT	4	69.4	20:26:05	06:34:37
6	4	3	2601	SLT	6	100.5	00:54:08	05:20:29
8	4	4	9402	VIRM	4	108.6	01:00:09	05:14:29
5	5	1	2402 + 2602	SLT	4 + 6	169.9	20:50:03 - 00:21:13	06:03:32
9	5	2	8602	VIRM	6	162.1	01:52:39	05:26:30
3	6	1	2603	SLT	6	100.5	22:10:34	05:57:32
7	6	2	2401	SLT	4	69.4	00:54:08	05:44:30
10	6	3	8601	VIRM	6	162.1	01:58:40	05:38:30
11	7	1	9401	VIRM	4	108.6	02:16:53	07:33:28

5.4 Stochasticity Algorithm

After generating the BDu stabling demand by the IG and the resulting stabling plan using the TAP model, the next step is to generate a set of SD stabling plans. This is done by performing a Monte Carlo simulation which consists of two steps:

1. Generating a set of SD stabling demand based on the BDu stabling demand.
2. Using the aforementioned TAP model to generate SD stabling plans.

The first step of the Monte Carlo simulation is performed by the Stochasticity Algorithm (SA). For each iteration, the SA is called which takes as input the BDu stabling demand, the probability of occurrence for each stochastic event from Section 5.1, and data of all the different train types. The BDu stabling demand used as input in the SA is in the form of the list of train units, with their type, variant, length, arrival, and departure time, such as in Table 5.3. This format of train unit stabling demand has been chosen instead of the departure compositions format from Table 5.4, as this allows the possibility to alter the individual train units, whereas the latter format only allows changes in the composition as a whole. For generating the stochastic events occurring between the BDu and SD planning phases, the SA uses data of all the different train types found in Table 5.1.

Section 4.5 discussed the most common influential events which can jeopardise the feasibility of a BDu stabling plan in the SD phase. As a reminder, these were the following:

- Out of service trains present at stabling yard.
- Change in arrival/departure time.
- Change in train unit variant.
- Change in train unit type.
- Change in number of train units in composition.

As the generation of Out of Service trains present at the stabling yard is being done in the TAP model, the remaining events will be generated in the SA in this section, meaning that the SA solely focusses on the stochastic events in the stabling demand itself, rather than infrastructural stochastic events, such as the presence of Out of Service trains. The event generation of the SA is divided into two sections: individual train unit-specific events, and departure composition-specific events. Between these two sections, the BDu stabling demand in the form of individual train units are then converted into the form of departure compositions. The individual train unit-specific events are considered the change in arrival- or departure time, and changes in train unit type or variant, with the change in the number of train units in a composition being a departure composition-specific event. For each train unit for the former type of events, and for each composition for the latter type of event, the SA first determines whether or not each event will occur for each unit or departure composition. The probability of each event from occurring on a specific unit or departure composition has been given in Section 5.1. It is assumed all these events are independent of each other and can occur simultaneously on the same unit, with the exception of the change in train unit type or variant. Of these two stochastic events, only one of these events at most can occur for each train unit.

The first event discussed is the change in arrival and departure time. As given in Section 5.1, the probability for a train unit to have this happen is 0.0437. Although a change in scheduled arrival or departure time influences the whole composition and not just the single train unit, these changes are performed in the train unit-specific section of the SA. This is due to the chosen format in departure compositions showing only the first and last scheduled arrival time of the train units in the composition, rather than the arrival time of each individual train unit. In the case of compositions of more than two train units in length, this means that arrival times of other units in the composition get lost, making it impossible to determine what the arrival compositions were and to change the arrival time of the whole arrival compositions.
The algorithms for changing the arrival and departure time, Algorithm 1 and Algorithm 2 can be found in Appendix B, as well as all other algorithms discussed in this section. The former algorithm takes care of possible changes in arrival time, while the latter does the same for the scheduled departure times, and work in the same way. The algorithms first decide which train units will encounter a change in arrival or departure time by using a Bernoulli distribution which has only two outcomes: success (1) or failure (0), with the probability of success extracted from the stochastic event probabilities from Section 5.1. When the Bernoulli trial results in a success outcome for a train unit, the arrival or delay will be changed for that specific unit using the outcome of a normal distribution trial with a mean and standard deviation of 53 seconds and 4046 seconds respectively. As the creation of the departure compositions in the format of Table 5.4 considers train units with the same arrival or departure time to be of the same composition on arrival or departure respectively, the algorithms check whether the change in arrival or departure time conflicts with the scheduled arrival or departure time of

another train unit. As infrastructural constraints at the stabling yard are not taken account in this research, the minimum headway between arrivals or departures can be set to an arbitrary number, in this case 10 seconds. When the changed time is conflicting with other arrival or departure times, a new sample is taken from the normal distribution until the changed time does not result in any conflicts. When the train unit in question is coupled with other train units on arrival and its arrival time is being changed, the arrival time of the coupled train units will be changed correspondingly as well. The same will happen for changes in departure time for coupled train units at departure.

The next possible stochastic event occurring to a train unit of a BDu stabling demand, is the change of the train unit variant, with a probability of 0.0767. The algorithm for this event, Algorithm 3, starts by using a Bernoulli trial to determine which train units will have a change in train unit variant. When a train unit is determined to have a change in variant, the algorithm first retrieves the train unit type in order to get the list of variants of this type. By removing the current variant of the train unit, a list of candidate variants is created. Out of this list, a new variant is randomly chosen. As in this research each train type only has two variants, there is only one candidate for the variant change. However, for train types with more than two variants, the new variant is randomly chosen with the probability for each variant being based on the number of trains the operator has of this variant. When the new variant is chosen, the variant of the specific train unit and its length are changed in the stabling demand accordingly.

It is also possible that the type of a train unit gets changed, which can have significant effects on the stabling demand and the resulting stabling plan. This is due to the fact that this change could not only affect the train unit in question, but also different train units in the stabling demand when these train units are coupled to this unit on arrival or departure, as coupled train units need to be of the same type. In practice, there are a number of ways to deal with the change of type of a train unit, such as changing the matching of arrivals and departures, or taking the affected train unit out of service after arrival. However, as in this stage of the RAM the matching has already taken place in the IG, it is assumed that the matching is fixed, and it therefore is not possible to redo the matching process. As a result, when one train unit gets a unit type change event, all train units which directly or indirectly are coupled to the train unit in question need to change to this type as well. Since this would mean multiple train units could change type in one instance, the input probability for determining whether a train unit is an ‘instigator’ for a unit type change is not equal to the probability of a train unit experiencing a unit type change as discussed in Section 5.1. Rather this has to be a value tuned in a way such that the outcome of this event reflects this probability instead. The probability value has been tuned by executing the SA on a set of test stabling demands and noting down the percentage of train units having had a type change occurred. The probability value in the SA was then changed such that the outgoing percentage of train unit type changes matches the probability given in Table 5.2. The tuned probability for a type change came out to be equal to 0.0112.

With this tuned probability, the algorithm first decides for each train unit if it will be an ‘instigator’ for a unit type change. For each unit that will change its type, the algorithm performs three main steps: determining the new type, determining all units that require this unit type change, and finally performing the unit type change for all affected train units. For determining the new train type, the algorithm uses the current type of the train unit in question and its service type (IC or SPR) to retrieve a candidate list of new train unit types which match the service type of the original train unit type. One of these candidates is then randomly chosen, with the probability of each candidate type being chosen being based on the number of trains of each train type NS owns. The next step of the algorithm is determining which train units in the stabling demand are directly or indirectly connected to each other on arrival or departure. Train units are directly connected with each other when they are coupled on arrival or departure, whereas train units are indirectly connected with each other when these two units have both at one point been coupled to a specific different train unit. For instance, train units A and C are indirectly connected when train unit A is coupled to train unit B on arrival, and train unit C is coupled with train unit B on departure. This ensures that after a type change all train units are not coupled with a train unit of a different type. Finally, when all train units requiring the unit type change are identified, the new variant for each unit is randomly chosen based on the number of trains of each variant available. The type, variant, and length are then adjusted for all affected train units in the stabling demand accordingly.

After performing the train unit type changes, all train unit-specific stochastic events have been completed. The table of the modified stabling demand in the form of train units will now be converted to a table of departure compositions such as in Table 5.4, which allows for performing the final stochastic event: addition or removal of a train unit from a departure composition. For each departure composition, Algorithm 5 will first decide which compositions will be subjected to this stochastic event based on the probability for an addition or removal of a train unit from a departure composition given in Section 5.1, which was equal to . For each of the departure compositions which will have an addition or removal of a train unit, the algorithm first checks the current number of train units in the composition. If the composition only consists of a single train unit, a train unit can only be added, as for this research it is assumed that no composition 'disappears' from the stabling demand, as this allows for an easier comparison of the change in stabling location of a specific departure composition in the BDu phase and a generated SD phase. Whenever the current number of units in a composition equals the maximum number of units in a composition for the specific train unit type of the departure composition, the back train unit will get removed. When the current number of train units does not equal to one nor the maximum number of train units allowed, the algorithm will randomly decide whether a unit will be added or removed.
When a decision has been made for each affected departure composition whether an addition or removal of a train unit will take place, the algorithm will carry out the required actions. In the case of an addition, the model randomly adds a variant of the correct train unit type to the composition, with the probability for each variant being based on the number of trains of this specific variant NS owns. Where this extra comes from is not modelled in the RAM, but it is assumed that this extra unit has either arrived on its own outside of the generated BDu stabling demand or has arrived coupled with the train unit it is coupled with in the departure composition. The length of the composition is then subsequently updated with the length of the extra train unit. In the case of a removal, the algorithm removes the last train unit of the departure composition and updates the composition length accordingly. In this case, it is assumed that a removal of a unit in the departure composition is a result of the cancellation of an arriving train unit the evening before, or an arriving composition consisting of one fewer train unit.

With the final stochastic event finished, the SA returns the adjusted stabling demand in the form of the departure compositions, which act as a representation of the stabling demand in the SD phase. These adjusted stabling compositions can then be used by the previously discussed

TAP model to generate the SD stabling plan. The SA and TAP model can then be used together in the Monte Carlo simulation to generate a large number of SD stabling plans which can then subsequently be used to assess the performance of the original BDu plan when under the influence of the modelled stochastic events. All the steps the SA takes to generate a sample SD stabling demand based on a BDu stabling demand is given in Figure 5.5.

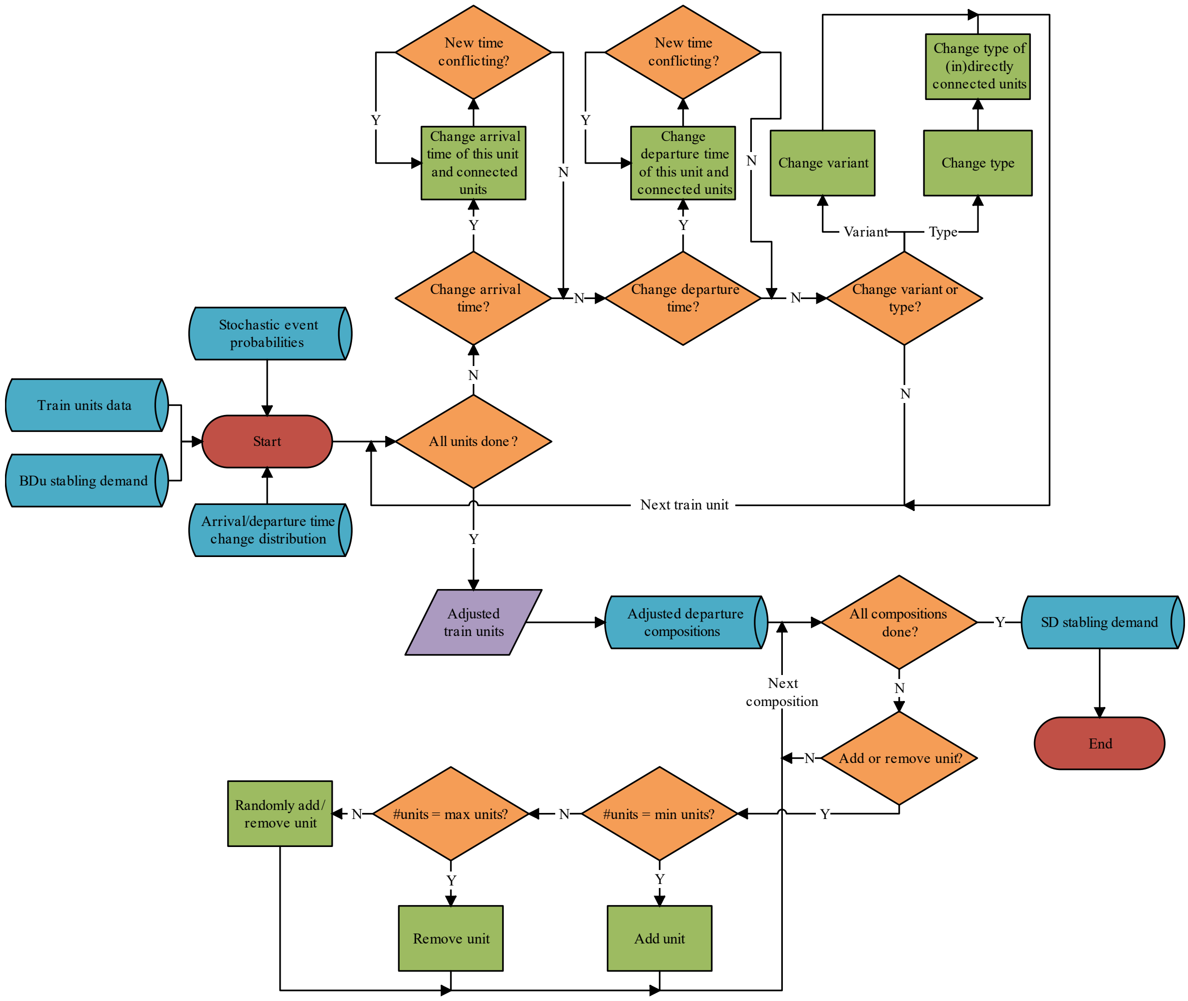


Figure 5.5: Stochasticity Algorithm steps.

5.5 Assessment Model

5.5.1 Main Calculations

The IG, TAP model, and SA together are able to create both a BDu stabling plan based on a generated stabling demand, as well as generate a desired number of SD stabling plans based on adjusted stabling demands subjected to stochastic events. With both the BDu and SD plans generated, the robustness assessment of the BDu plan can now be performed. As discussed in Section 4.5, the three key metrics used to estimate the robustness of a stabling plan were:

- Viability
- Offset
- Spread

Given a viable BDu stabling plan and a set of n generated SD plans at a specific stabling yard, the assessment model performs a number of steps. First off, the model decides which of the SD stabling plans can be deemed viable, which in this case is whenever all departure compositions are being stabled. The percentage of the n generated SD plans which have been deemed viable is the value of the viability metric. A high viability value would indicate that the stabling yard has sufficient remaining infrastructure to cope well with the stochastic events occurring in the original stabling demand, as it is able to find a feasible stabling plan for the majority of the SD stabling demand. This value therefore does not say anything about the robustness of the BDu stabling plan, but rather indicates the robustness of the stabling demand itself and the ability to find a feasible solution to the TAP of this stabling demand at the stabling yard in question.

The model then continues with extracting the stabling coordinates (track, stabling order) from the BDu plan and all viable SD plans, which will be used to calculate the BDu offset, and SD spread distribution. As in this research the SA can only modify the departure compositions, and not add or remove compositions as a whole, this allows for an easier comparison between the stabling locations of a departure composition in both stabling plans, as now it is not required to determine the coordinates or the penalty value of a departure composition that does not exist in one of the two stabling plans.

As mentioned in Section 4.5, the offset and spread are determined using the difference between two stabling plans. However, before these calculations can commence, first a method to measure the difference between two stabling plans needs to be proposed. How different a stabling plan is compared to any other plan can depend on a lot of factors and is very subjective, as one person might find a certain track change of one departure composition ‘worse’ worse than another person. Also, the type and length of a departure composition might have an influence on how impactful a change of location for this composition might seem to a person. In the end, one could add countless factors and special cases to determine the difference between two plans. This would lead to a very precise calculation method for this difference, reducing the instances where two subjectively different changes in a stabling plan can result in the same difference score. However, the more factors are added to a model, the less explainable its results are, as it becomes increasingly harder to interpret how each of these factors contributes to the final plan difference score, especially when the weight of these factors are heavily subjective. Furthermore, the more extensive the method is, the higher the runtime will be, which is detrimental for a model which generates a large number of stabling plans which need to be compared. Therefore, a trade-off has to be made regarding the explainability of the method which determines the difference between two stabling plans, and its precision. For this research, three metrics are proposed which will compose the method which calculates how different two stabling plans are:

1. Track changes
2. Order changes
3. Neighbour changes

These three metrics are found to be a good trade-off regarding the explainability and the precision of the results. The coordinate system of stabling track and arrival order provides an easy and objective method for determining the location of a departure composition in a stabling plan and how this differs from any other stabling plan. As mentioned, these coordinates indicate the location of a departure composition, but it does not say anything about what this composition consists of. This does not pose any problems for determining the differences between two stabling plans, as for this research the difference between the plans is only related to the locations of the departure compositions in their entirety. The contents of each departure compositions and the possible occurred stochastic events which led to changes in the contents of a departure compositions are not considered to be changes in the stabling plan and are therefore not relevant for this calculation.
With the track and order coordinates indicating where a departure composition itself is located and how this differs from the location of this composition in another stabling plan, how a stabling plan 'looks' different compared to another is also due to how the location of a departure composition changes with regard to the compositions around it, its neighbours. The track, order, and neighbour changes together will give insight in how different two stabling plans are to each other from the perspective of the individual departure compositions as well as from the perspective of the plan for the stabling yard as a whole.

The calculation for the difference between two stabling plans, A and B, can be split into two main parts: a track- and composition-perspective. The first two metrics are part of the track-perspective calculations. For each track t, the calculation model checks which compositions are stabled at this track in the plan A. The model then retrieves the stabling tracks for these compositions in plan B to check whether each composition stabled at track t in plan A has moved and how many tracks they have moved across. Instead of adding each track move individually to the difference score, the model rather adds only the unique track changes to the difference score once. That means that two compositions from track t that move to the same track result in the same addition to the difference score as when only one composition from track t moves to this other track. By using only the unique track changes instead of the individual track changes of each composition per original stabling track, the model favours compositions from the same track to stay together as much as possible over minimising the number of individual track changes, increasing recognisability between the two plans.
Figure 5.6 shows a simple stabling plan consisting of three tracks and nine departure compositions labelled A to I on the left. In the middle stabling plan, two single compositions are swapped, with composition C moving to track 2 and composition F moving to track 1. As each of these compositions move over one single track, the value of both of these track changes is equal to one, resulting in the total difference score being increased by two. The stabling plan on the right has two groups of compositions swapped. The compositions B and C from track 1 are moved to track 2, and as they move together from their original track to the same new track,

this equates to one single track change equal the number of tracks the compositions have moved over, which is in this case one. Similarly, compositions E and F move from track 2 to track 1 together, which is therefore also a single track change with a value of one, resulting in the total difference score of this stabling plan compared to the standard stabling plan equalling two.

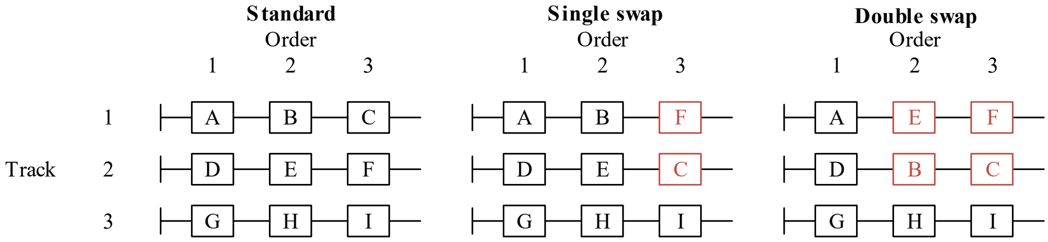


Figure 5.6: Example track changes with the same score.

For the changes in arrival order of a composition at its stabling track between the two stabling plans, the same line of reasoning is used. For each stabling track t in plan A, the model checks for each group of compositions that have moved to the same track in plan B the unique order changes and adds them to the difference score. However, there is one exception to determining the unique order changes and adding them to the score. The last arriving composition on a stabling track can have a different order value, based on the number of compositions already arrived at the track. This could mean that a composition that is the final arriving composition at its track in both plan A and plan B, but with different order values, is considered to have occurred an order change and is therefore added to the difference score although it is still the final arriving composition to its stabling track. Therefore, the model makes a special exception to these occurrences and does not take the composition in question into account for determining the unique order changes.

Figure 5.7 shows another stabling plan variant from the standard stabling plan from Figure 5.6. Here, the track changes of compositions A and B together, and composition H on its own, result in a track change score of four being added to the difference score. For the calculation of the order change score, one track change group is investigated at a time. The first track change group consists of compositions A and B, which had an original arrival order of 1 and 2 respectively. In the new stabling plan, the order of composition A has increased by one to an order of 2. If composition B would have also had its order increased by one to a value of 3, these two order changes would be seen as a single unique order change. However, composition B has its order value increased by two instead, which means that the order changes of compositions A and B are seen as two separate changes, with the first change resulting in an addition to the difference score of one, and the second resulting in an addition of two. With the order of composition H being reduced by one, the total order change score is $1 + 2 + 1 = 4$.

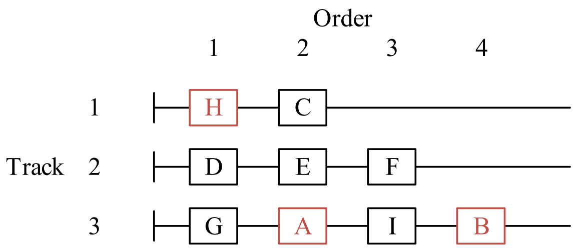


Figure 5.7: Example order changes.

Finally, the model assesses the difference between two stabling plans from a composition-perspective as well, in the form of changes in 'neighbours'. Each composition can have up to two neighbours, one in front of itself and one behind. The model checks for each departure composition what its neighbours are in plans A and B. The number of changes in neighbours for each departure composition is halved and then added to the difference score. The reason the number of changes in neighbours is halved, is that a specific neighbour change will arise twice in the model: once from the perspective of one composition, and once from the perspective of the changed neighbour in question, resulting in the specific change being added twice.

Continuing with the example from Figure 5.7, composition A had composition B as neighbour in the original stabling plan. In the new stabling plan, composition A is neighboured by compositions G and I, which are two neighbour changes. On the other hand, composition H was originally neighboured by compositions G and I, but is now neighboured by composition C on one side and no composition on the other, which are therefore also two neighbour changes. Determining these changes for each composition results in a total neighbour change score of $(2 + 2 + 1 + 1 + 2 + 2)/2 = 5$. With the track change score, order change score, and now the neighbour change score calculated, the total difference score is equal to $4 + 4 + 5 = 13$.

The total difference score after all tracks and compositions have been analysed is then divided by the number of departure compositions in the stabling demand, which allows the ability to easier investigate the average number of changes per departure composition. The steps for the calculation of the difference score are visualised in Figure 5.8.

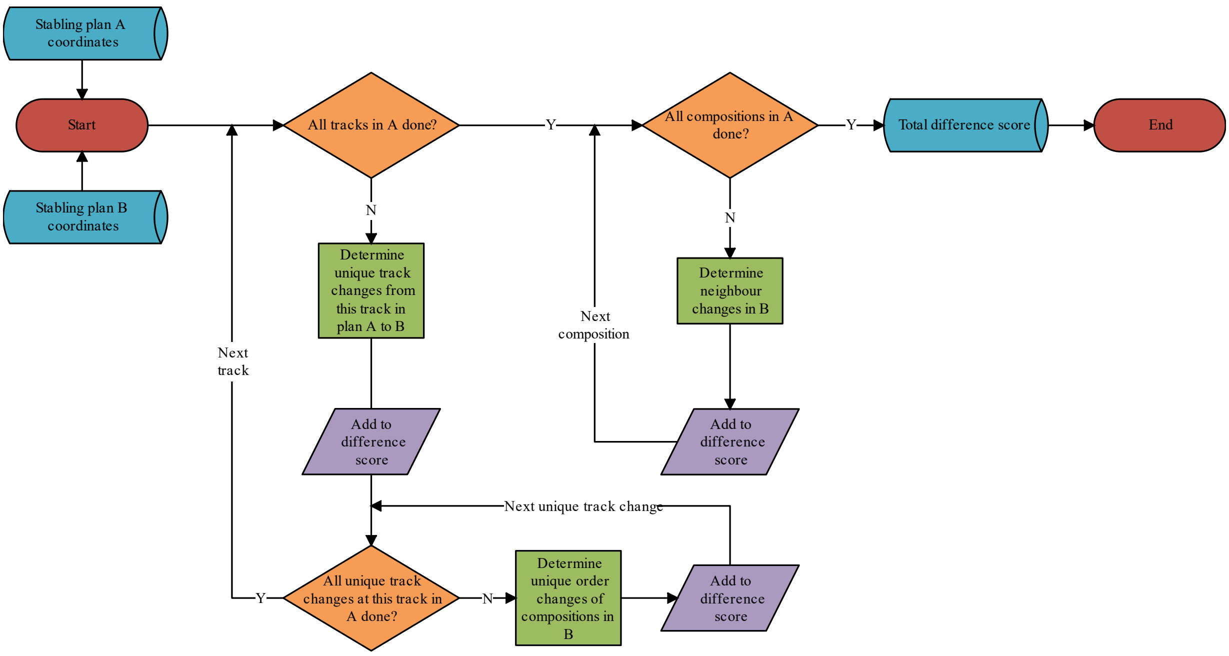


Figure 5.8: Stabling plan difference score calculation steps.

For the calculation of the offset, the BDu plan is compared to all viable SD plans to obtain a distribution of these difference scores. Instead of simply taking the mean of these scores once to obtain the value of the offset, the offset is estimated using bootstrapping. Bootstrapping is a random resampling method which enables obtaining a set of offset estimations, which can be used for determining the confidence interval of the offset value of the BDu stabling plan. By performing 25,000 bootstrap samples from the list of difference scores of the BDu stabling plan and the viable SD stabling plans, a distribution of the offset estimations is obtained. From this distribution, the mean and the 95% confidence interval are determined. Figure 5.9 shows a distribution of the bootstrapped offset estimations, together with the mean estimate and the 95% confidence interval.

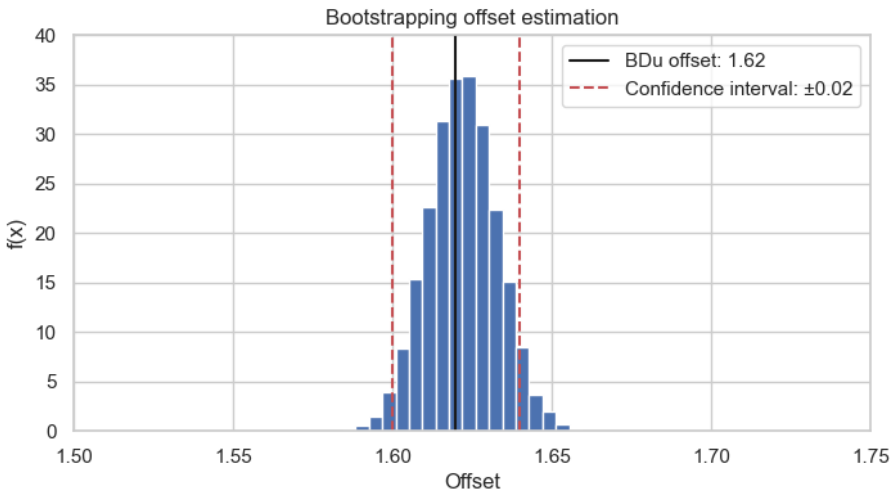


Figure 5.9: Bootstrapped offset estimation distribution.

An important condition for performing bootstrapping, is that the original sample must be a good representation of the population distribution, as a badly distributed original sample will never result in bootstrap samples reflecting the population. When testing the RAM, it was found that a set of 5000 generated SD stabling plans from the Monte Carlo simulation was sufficient to obtain a smooth and well-formed offset estimation distribution. Furthermore, this distribution remained consistent when running the RAM multiple times, indicating that the chosen sample size for the Monte Carlo simulation is sufficient and a good representation, as if this would not have been the case, the spread distribution would significantly differ between test runs of the RAM. Figure 5.10 shows the distribution of the difference scores of the BDu stabling plan compared to the viable SD plans across three runs of generating the SD plans based on the same BDu stabling demand and stabling plan. The distributions of these three individual runs are very similar, indicating that the sample of the SD plans represent the population of possible SD plans very well.

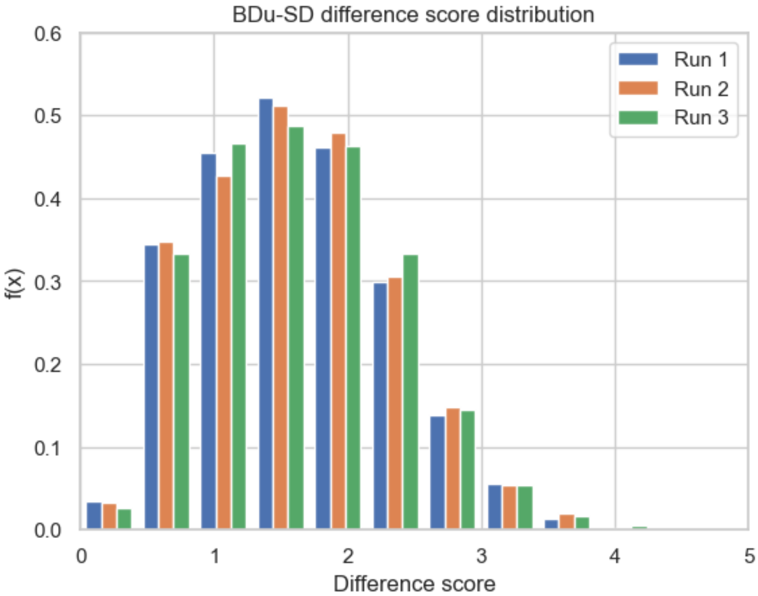


Figure 5.10: Distribution of the difference scores between BDu and SD plans across three runs.

The spread distribution is the result of comparing all SD plans to each other. With n_{viable} viable SD plans, this results in a matrix of size $(n_{viable} \times n_{viable})$ with the difference scores of all SD comparisons. As a comparison between plans A and B results in the same difference score as a comparison between plans B and A, only the upper triangle of this matrix needs to be calculated and can then me mirrored along the diagonal to obtain the difference scores of all SD pairs. The spread distribution then consists of the means of each row, excluding the difference score of the comparison of an SD plan with itself from the calculation of the means. The calculation for the average spread for a viable SD stabling plan i compared to other viable SD plans j, is shown in (5.8).

$$\overline{spread_{\iota}} = \frac{\sum_{j \neq i \in n_{viable}} D(SD_i, SD_j)}{n_{viable} - 1} \tag{5.8}$$

This value gives an indication of how similar viable SD stabling plan i is compared to all the others when this SD stabling plan would have been generated as the initial, BDu stabling plan. A lower mean value of this SD plan compared to the offset would then indicate that this SD plan requires less changes to obtain the optimal stabling plan for the other SD stabling demands compared to the original BDu stabling plan.

The offset confidence interval and the spread distribution can then be used together to analyse the robustness of the BDu stabling plan. The values of the spread distribution entries and the offset all say something about the performance of the SD plans and BDu plan respectively when subjected to the stochastic events from the SA. A lower value would indicate that on average less changes to the initial stabling plan are required to generate a feasible solution for each of the adjusted stabling demands generated by the SA. Therefore, when one of the SD stabling plans has a lower average difference score than the offset value of the BDu plan, this means that this specific SD plan is more robust to the stochastic events generated by the SA than the BDu plan is, as on average it needs less changes to the stabling plan to become feasible again. As a result, this SD plan has a higher reusability compared to the BDu plan, while ideally the BDu plan has the highest reusability, as this would indicate that the initially generated BDu plan is indeed the better stabling plan for both the initial stabling demand, as well as the possible stochastic changes this stabling demand is subjected to. The robustness of the BDu stabling plan is estimated by determining the percentage of SD plans which have a higher average difference score than the BDu offset value, as these SD plans perform worse than the BDu stabling plan. As the offset is estimated in a 95% confidence interval, the robustness estimation of the BDu plan will therefore also be in the form of a 95% confidence interval. The steps taken to obtain the robustness estimation of a BDu stabling plan generated by the used TAP model are visualised in Figure 5.11.

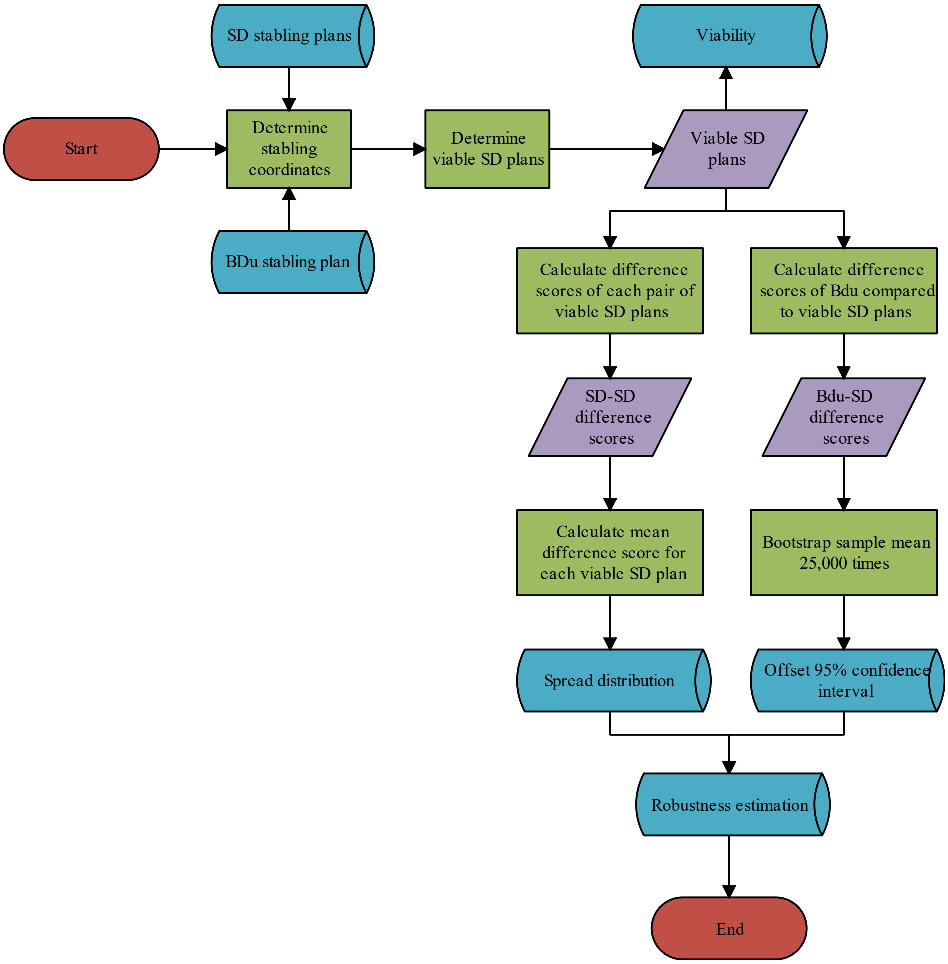


Figure 5.11: Robustness estimation steps.

5.5.2 Sensitivity Analysis

As the RAM uses stochastic events to estimate the robustness of a BDu plan, it is important to not only estimate the robustness of BDu stabling plans in specific scenarios, but also investigate what slight changes in the probabilities of each stochastic event do to the robustness estimation. Performing such a sensitivity analysis can give insight in how all of these stochastic events and their probabilities influence the robustness estimation.
The sensitivity analysis for the RAM is performed following the one-at-a-time principle, where the value of only one variable is altered, and the rest are kept the same. This is then repeated

until the model has gone through each variable. For the RAM in this research, the four stochastic variables which will be altered in the sensitivity analysis are:

- Probability of variant change
- Probability of type change
- Probability of addition or removal of train unit from composition
- Probability of change in arrival or departure time

For each of these four variables, two variations are defined for the sensitivity analysis, resulting in a total of eight variations which need to be carried out by the RAM. The first variation for each of the variables is a 10% increase in the probability value, with the second variation being a 10% decrease. Given the generated BDu stabling plan in the 'normal' scenario, the sensitivity analysis model inserts the variation of the probability values into the SA to generate new SD stabling demand. The RAM then continues as normal to calculate the viability, offset, and spread distribution. The results of the 'normal' RAM are then compared to the RAM results of each variation of the sensitivity analysis and the percentual change of the viability, offset, and final robustness estimation is calculated. These percentual changes will then give the insights in how sensitive the RAM to small changes in the probabilities of the individual stochastic events in stabling demand. Figure 5.12 visualises how these sensitivity scenarios enable a comparison of how the viability, offset, and robustness estimation change when the four stochastic variables have a higher or lower probability of occurrence.

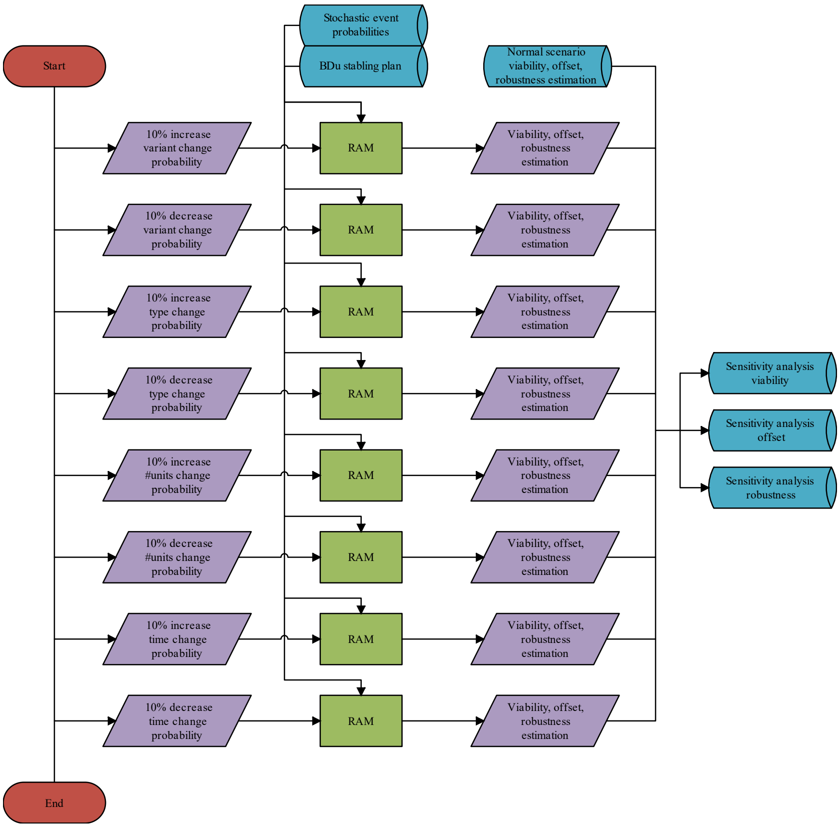


Figure 5.12: Sensitivity analysis steps.

5.6 Model Limitations

Although the RAM is quite extensive, the model does have some limitations as a result of simplifications and scope reductions. First of all is a simplified definition of the viability of a stabling plan in the RAM. For this research, a stabling plan is only viable if all departure compositions from the stabling demand could be stabled. This has been chosen such that the comparison of the location coordinates of a departure composition in two different plans is easier. In the same manner, the SA is unable to remove departure compositions or add them completely, while in reality this happens regularly. However, as these additions or removals of departure compositions from the stabling plan would make a comparison of stabling locations much harder, since then a penalty for the addition or removal of a specific departure composition would have had to be defined.

A further simplification to the model has been made to the TAP model. The model only seeks to minimise the total length of the not-stabled departure compositions. As a result, the solution space with equal objective function values can be quite large, especially when the number of train units is far below the capacity of the stabling yard. As a consequence, it is very much possible that the generated BDu stabling plan could not have the highest estimated robustness out of all possible BDu stabling plans. The TAP model could be extended by incorporating specific stabling preferences, such as minimising the number of different train types on the same track or minimising the walking time of drivers.

From an operational perspective, certain scope reductions have been made as well. For a start, the stabling plan is only a portion of the whole node planning, which also shows the train movements itself, to and from the stabling yard but also around the station(s) of the node in question. However, the RAM only looks at the stabling plan of the stabling yard itself. Aiming to optimise the robustness of an initial stabling plan should be done with a healthy dose of caution, as an optimisation of the stabling plan does not necessarily lead to an optimal or even only feasible node planning, as some train movements might be causing conflicts due to this optimised stabling plan.

This immediately leads to the next scope reductions. The stabling yards used in this research are of the shuffleboard layout, where trains enter and exit the stabling tracks in a LIFO principle. However, stabling yards with a carousel layout can have trains entering and exiting the stabling tracks from both sides. The RAM model would need to be adjusted to be able to work with these layouts as well. The TAP model, as well as the coordinate system for determining the stabling locations of the departure compositions, would need to take into account the entry and exit directions of the compositions and individual train units and the order value of the coordinate system might need to be slightly redefined to take into account these entry and exit directions. Furthermore, the RAM only considers the low servicing method, where all routine servicing of the trains is done on the stabling tracks itself. This means the only two movements of a train arriving at a shuffleboard stabling yard with low servicing are the arrival to the stabling track and the departure from the stabling track, excluding shunting movements as a result of a composition split, which are not modelled in the RAM with the duration of the split being incorporated into the arrival time of the composition in question, and the assumption that there is always available room to perform this composition split. When using carousel servicing, movements to and from the servicing tracks need to be modelled. These movements should take into account allowed routes at the stabling yard, check for conflicting routes and calculate the duration of the movement. The current RAM, however, does not model any of these movements and would therefore need to be adjusted to incorporate the planning of train movements when desired. Moreover, also staff scheduling is not incorporated in the RAM, but rather the assumption is made that there is plenty of staff available to perform every shunting or servicing task.

Finally, the stochastic events in the RAM only occur after the matching of arriving and departing units has taken place, while in reality changes to the stabling demand can already happen prior to this phase. As a result, the SA might yield stabling demands which are not fully in line with realistic stabling demands. This could become especially apparent under a type change of a train unit. In the RAM, all directly or indirectly connected train units in the stabling demand will also change their type to ensure no train unit is connected to a train unit of a different type. In reality, other rescheduling actions would be performed when one train changes its type, such as leaving only the affected train unit at the stabling yard and reducing departing train compositions when required.

These limitations to the RAM are deemed acceptable, as a trade-off has to be made regarding the complexity of the model and its runtime, with the current state of the RAM found to be a good balance between the two. Additionally, as this research is the first to attempt to both define and measure stabling plan robustness in this manner, it is argued that a strong foundation in a simpler environment, which can later be extended upon, is much more valuable than a 'shallower' model which does incorporate a portion of these limitations.

6 Robustness Assessment

In this chapter, the RAM is used to assess the robustness of BDu stabling plans of varying capacity utilisation rates across three case study locations. The three case study locations, as well as the three capacity utilisation scenarios per location, yielding nine scenarios in total, are given in Section 6.1. The results of the RAM are discussed in Section 6.2, with the viability scores of all nine scenarios discussed in Subsection 6.2.1, the offset, spread, and robustness of the scenarios across the three case study locations in Subsections 6.2.2, 6.2.3, and 6.2.4 respectively, and the results of the sensitivity analysis in Subsection 6.2.6. Finally, a reflection on the results can be found in Section 6.3.

6.1 Case Study Locations

The three case study locations, featuring a shuffleboard layout and using the low servicing method, which have been chosen for this research are Carthesiusweg (ctw), Nijmegen (nm), and the shuffleboard stabling yard at Zwolle (zl). The schematic overview of these stabling yards can be found in Appendix C.
Carthesiusweg is a stabling yard northwest of Utrecht Central Station and is the largest stabling yard in terms of capacity of the three case study locations. The stabling yard consists of 22 tracks in the shuffleboard layout, with the track numbers between 251 and 272. However, not all of these tracks are used for stabling for the scenarios at ctw. The tracks which are not being used as stabling tracks in these scenarios is based on the night of August 2nd to 3rd 2020. At this night, tracks 251 and 262 are being used for stabling the upcoming ICNG trains, which are not part of this research, as these are not yet in full operation. Track 263 is also not used as a stabling track, as this track is used for maintenance on the Dutch train protection systems of trains. Finally, track 272 is unavailable due to the presence of an aerial work platform. Table 6.1 shows the active stabling tracks and their available stabling length in metres.

Table 6.1: Carthesiusweg stabling track lengths.

Track	**252**	**253**	**254**	**255**	**256**	**257**	**258**	**259**	**260**	**261**	**264**	**265**	**266**	**267**	**268**	**269**	**270**	**271**
Length [m]	244	235	247	238	252	304	271	259	325	378	230	169	170	187	154	155	165	169

Nijmegen is the smallest location of the three with thirteen stabling tracks labelled 1R-13R. Tracks 7R and 8R are not electrified and is used for stabling trains of the local railway operator, track 9R is unreachable, and 13R is not used as stabling track at the night of the 2nd to 3rd of August 2022. Table 6.2 shows the available stabling lengths of the stabling tracks at Nijmegen.

Table 6.2: Nijmegen stabling track lengths.

Track	1R	2R	3R	4R	5R	6R	10R	11R	12R
Length [m]	337	337	229	229	271	331	276	261	263

The final stabling yard location are the shuffleboard stabling tracks at the southwest side of Zwolle Station. There is also a carousel layout stabling yard at the east of Zwolle Station, but as this research only focusses on shuffleboard layouts, this yard is left out of the scenarios at Zwolle. The stabling tracks at the shuffleboard stabling yard of Zwolle are labelled 403 to 414 and 430 to 433. Tracks 403 to 408 are not electrified and used by local railway operators, meaning these tracks are not taken into account by the RAM. Furthermore, at the night of June 9th to 10th 2021 tracks 409, 410, and 433 are not used as stabling tracks. That leaves the remaining tracks to be used for stabling, which are given in Table 6.3 below along with their available stabling lengths.

Table 6.3: Zwolle shuffleboard stabling track parameters.

Track	411	412	413	414	430	431	432
Length [m]	371	373	374	375	422	456	456

At each of the three locations, three scenarios will be run. Each of these scenarios will have a different number of train units in the stabling demand, relative to the total capacity of the stabling yard. The three scenarios will run with a number of train units in the stabling demand of around 40%, 60%, and 80% of the capacity of the stabling yard.
The NS have already defined the capacity of each stabling yard in the Netherlands in form of *bakken* (carriages). This unit of measure is equal to the standard length of a carriage, which is equal to 27.2 metres. The capacity of each stabling track has then been calculated by dividing the effective stabling length by this length and rounding down to an integer value. Newer generation of trains generally have shorter carriages than this length, meaning that their length in standard carriages is lower than their actual number of carriages. For the VIRM and ICM train units, their length in standard carriages is equal to their number of carriages. For the SLT-4 and SLT-6 train units, their length in standard carriages is 2.8 and 3.9 respectively, whereas for the SNG-3 and SNG-4 train units the length in standard carriages is 2.3 and 2.9 respectively. Taken into account the number of trains NS owns of each train unit, the average length of a train measured in standard carriage length is given in (6.1).

$$\bar{L} = \frac{98*4+80*6+87*3+50*4+69*2.8+62*3.9+118*2.3+88*2.9}{98+80+87+50+69+62+118+88} \approx 3.52 \tag{6.1}$$

With the stabling capacity of the locations in standard carriages, the average train unit length in standard carriages, a “cutting loss of 7%” as defined by NS, and the average number of out of service trains present at the stabling yard, the required number of train units that need to be generated in the stabling demand for each scenario can now be determined. The stabling tracks used for this research at ctw have a combined capacity of 123 standard carriages, which is equal to 34 train units rounded down. Reducing this value by the 7% cutting loss results in an effective stabling capacity of 31 train units rounded down. Finally, to obtain the required number of train units in the stabling demand for each scenario, this capacity is reduced by the average number of out of service train units, which is three, and then multiplied by the corresponding capacity utilisation percentage of the respective scenarios. This leads to a stabling demand of 11, 16, and 22 train units respectively. The results of these calculations as well as for the other two stabling locations can be found in Table 6.4. The stabling demand for all nine scenarios is given in Appendix D.

Table 6.4: Stabling demand sizes per location and scenario.

Location	Total capacity	Scenario 1	Scenario 2	Scenario 3
Ctw	31	11	16	22
Nm	24	8	12	16
Zl	26	9	13	18

6.2 Assessment Results

In this section, the results of the RAM are discussed. With each of the three locations having three scenarios determined based on the capacity utilisation of the stabling yard in question, nine runs of the RAM have been performed. The results of these nine runs are divided into four subsections. Subsection 6.2.1 will discuss the viability of the three locations across the three scenarios. The following three subsections thereafter will discuss the further results of the RAM scenarios of Carthesiusweg, Nijmegen, and Zwolle respectively. Finally, Subsection 6.2.6 will discuss the results of the sensitivity analysis of the RAM. This sensitivity analysis has been obtained by running a ‘low’ and ‘high’ variant of the probability of each of the four stochastic events to which the stabling demand can be subjected to. This results in eight RAM runs per location, per scenario, or 72 runs in total.

6.2.1 Viability

Figure 6.1 below shows how the viability of the stabling plans of the SD stabling demand changes for each location when the utilisation rate of the capacity of the respective stabling yard is increased. What is immediately noticeable is that the viability is very high in all scenarios across the three locations, staying well above a value of 99 percent. At Zwolle, the viability even stayed at 100% in both the first and second scenarios, whereas Carthesiusweg and Nijmegen never reach a viability of 100% and are also decreasing more compared to Zwolle when the capacity utilisation increases.

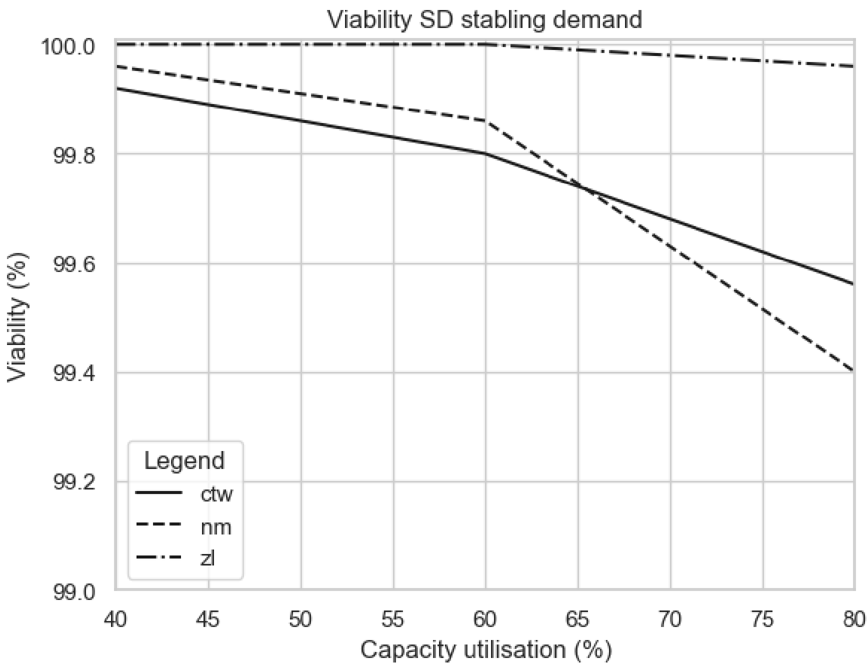


Figure 6.1: Viability of SD stabling demand.

The high viability across all locations and scenarios indicate that the case study stabling yards remain to have sufficient infrastructural leeway to be able to cope with the stochastic changes the stabling demand can be subjected to. This high viability is not surprising when the capacity utilisation is very low, but it seems that even with 80% capacity utilisation the viability stays very high. Although the viability in the third scenario is still very high for each location, the decrease in viability seems to worsen when the capacity utilisation increases further. It is expected that the viability eventually will hit a capacity utilisation limit at which the stabling yard is occupied with trains to the extent it does not have any spare room to cope with changes in the stabling demand, which will heavily impact the viability of the SD stabling demand. This is due to the fact that the higher the capacity utilisation of the BDu stabling demand is, the smaller the solution space is for finding a feasible stabling plan, up until there is only one single possible solution which is able to stable all demand. Whenever this is the case, the probability is higher that a change in stabling demand will make this final solution infeasible as well, without opening up the solution space to more possible stabling plans.

6.2.2 Carthesiusweg

The first location of which the RAM results are discussed is the stabling yard of Carthesiusweg in Utrecht. Figure 6.2 below shows the offset-spread plot of all three capacity utilisation scenarios. In these plots, a number of things can be analysed. First of all, the viability of the SD stabling demand, discussed in the previous subsection, is given above the respective scenario plot. Within the plot itself, the spread distribution is shown. As mentioned in Section 5.5, this spread distribution is a histogram of the average difference between one of the viable SD stabling plans and all other viable SD stabling plans. This gives an indication in how well each of these SD stabling plans are able to cope with the changes in stabling demand. Further visible in each plot is the BDu-SD offset of the respective scenario, together with the confidence interval of this offset estimation. The upper and lower bound of this offset are then used to calculate the percentage of SD plans which on average perform worse than these bounds. These percentages then make up the robustness estimation of the BDu stabling plan.

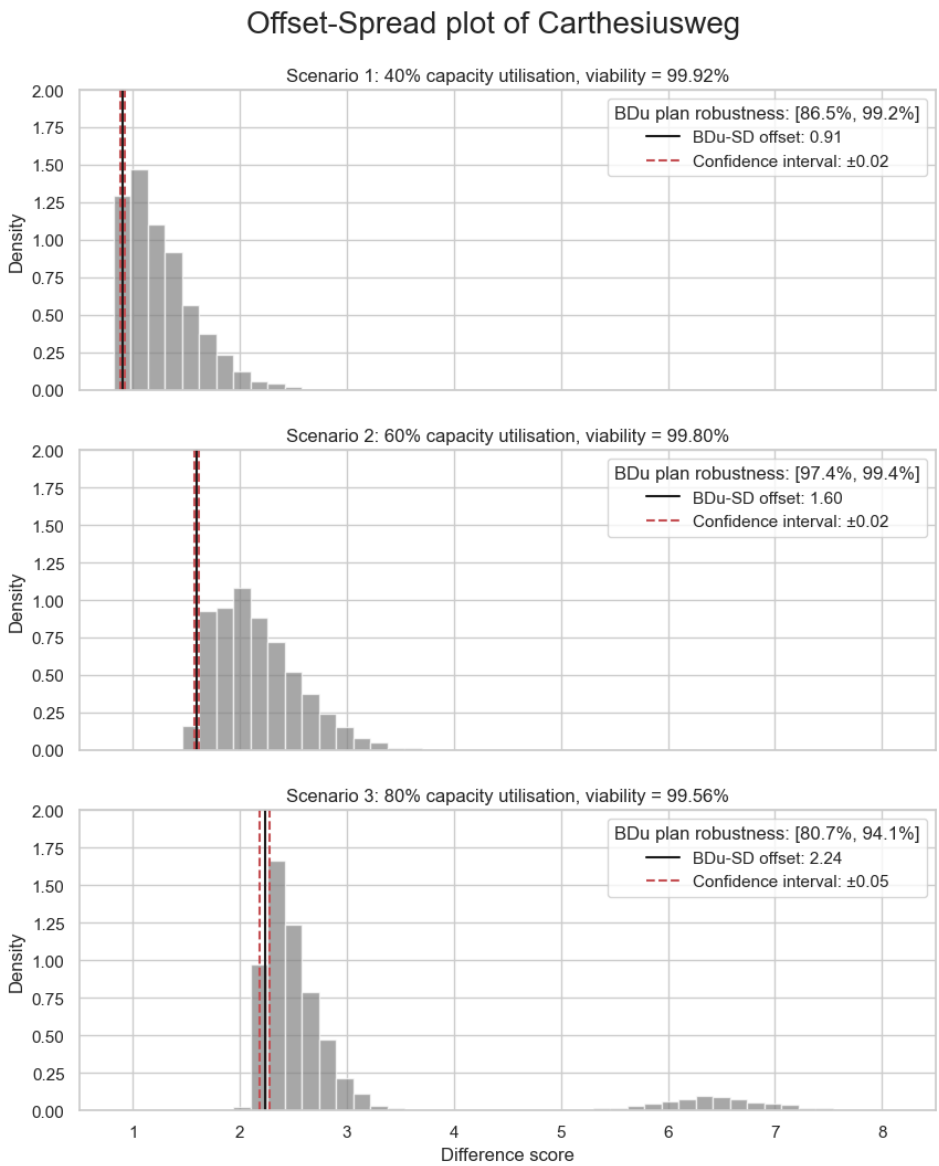


Figure 6.2: Offset-Spread plot of the scenarios at Carthesiusweg.

What first of all is noticeable of these results, is that the BDu plan robustness interval can be quite large, even though the confidence interval of the offset estimation is rather narrow. This is especially apparent in the first scenario, where a confidence interval of only ± 0.02 results in a robustness interval with a width of 12.7%. This larger window is, although not ideal, hard to avoid, especially when a large portion of the spread values lie inside the confidence interval of the offset. In these cases, a small difference between the upper and lower bound of the offset could still result in a substantial difference between the robustness estimation of these bounds.

What is further noticeable is that the average offset value increases with every scenario. While the average offset estimation in the first scenario is equal to 0.91 average changes per departure composition, this value increases to 1.60 in the second scenario and 2.24 in the third. As the offset value is the average change score per departure composition, this increase indicates that when the capacity utilisation increases, on average relatively more changes need to be made to make the stabling plan feasible again in the SD-phase when the stabling demand has changed. This is not only visible in the offset, but also in the spread distribution, as it starts to move towards the right side of the figure, indicating that also the generated SD stabling plans need to perform relatively more changes on average to make the stabling plan feasible for all other SD stabling demand.
Regarding the shape of the spread distributions, the shape of the distributions in the first and second scenario seem very similar, however the third scenario shows something rather unexpected. The spread distribution of the third scenario shows a second peak between a difference score of 5.5 and 7.5. It is unclear as to why the density of the spread distribution behaves this way in this scenario, especially since this phenomenon does not show in the other two scenarios at this location.

The offset estimation of the BDu stabling plan and the spread distribution of the SD plans lead to an estimated BDu stabling plan robustness of $[86.5\%, 99.2\%]$ in the first scenario, which means that, with 95% confidence, the generated BDu stabling plan outperforms at least 86.5% of the generated SD stabling plans, based on the average number of required changes per departure composition, with an upper bound of 99.2%. For the second scenario, this robustness interval is equal to $[97.4\%, 99.4\%]$ and the BDu stabling plan of the third scenario has a robustness estimation of $[80.7\%, 94.1\%]$. Although intuitively one might expect the robustness estimation of the BDu stabling plan to decrease when capacity utilisation increases, this is not necessarily the case. This is because the robustness estimation is a measure of how well the TAP model is able to generate a stabling plan for a BDu stabling demand that can more efficiently cope with small changes in this stabling demand compared to the SD stabling plans based on the changed stabling demand variations. With the robustness estimations of the three scenarios being relatively high, this indicates that the TAP model is able to immediately generate a stabling plan which is more versatile compared to the later generated SD stabling plans, as it is able to deal with the stochasticity of the stabling demand with on average less changes to the original plan.

6.2.3 Nijmegen

The second location at which the RAM has been performed is the stabling yard of Nijmegen. This stabling yard is a lot smaller in terms of capacity than the previous location of Carthesiusweg, meaning that the number train units in the stabling demand at each of the scenarios is lower, as discussed in Section 6.1.

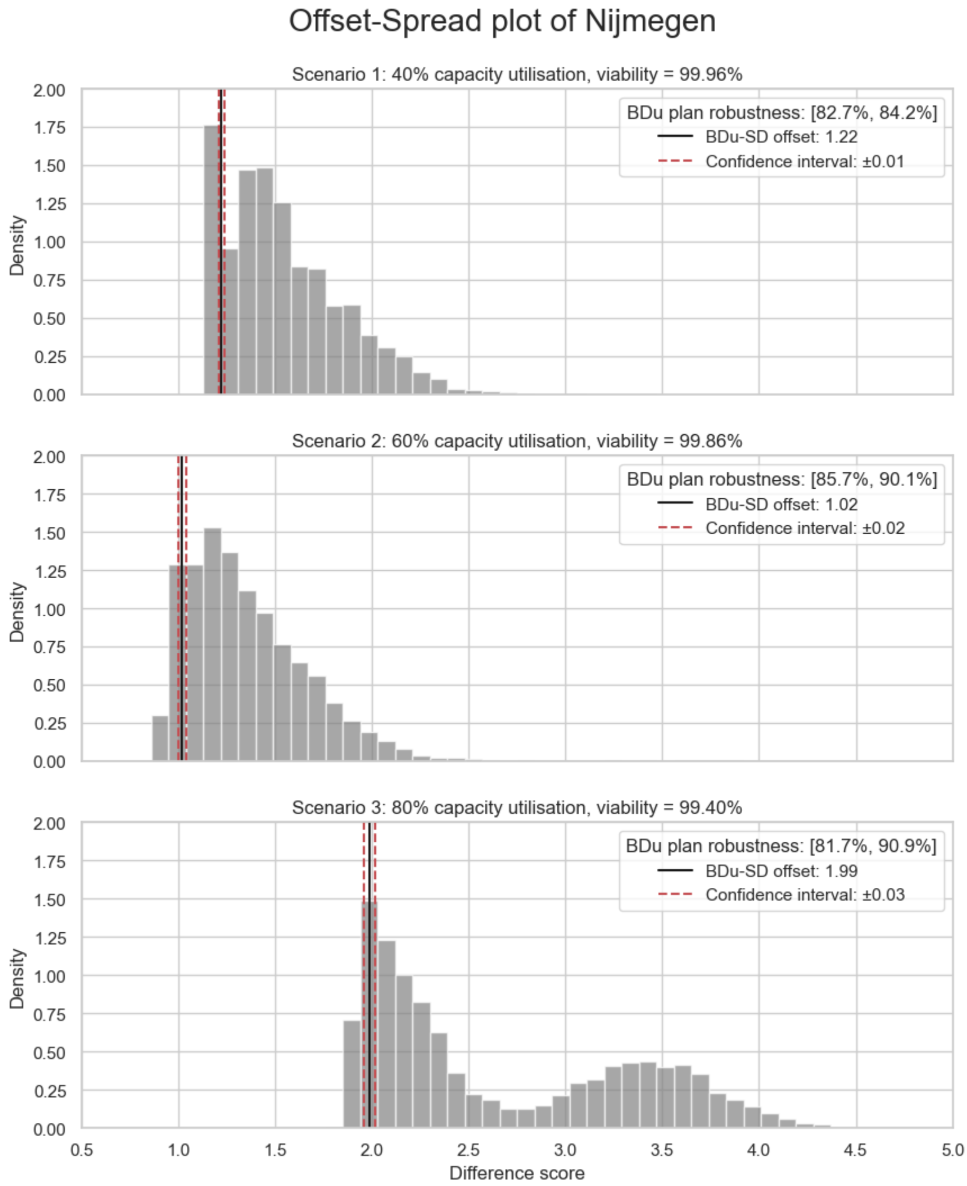


Figure 6.3: Offset-Spread plot of the scenarios at Nijmegen.

The results of the RAM at Nijmegen seem fairly similar to the results of the Carthesiusweg scenarios. The shape of the spread distributions in the first and second scenario show the same skewness towards the left side, similar to Carthesiusweg, and the third scenario also has the second peak, albeit more profound in this case.

In the first scenario, the offset estimation has a slightly surprising value, as with 1.22 it is higher than the offset of the second scenario, which is equal to 1.02. With less capacity utilisation, it would be expected that, due to the higher remaining capacity, the stabling plan would require less changes on average to the stabling plan to cope with the changes in the stabling demand. However, in this scenario this is not the case, as also the SD plans in the spread distribution do not seem to achieve significantly lower average difference scores. Between the second and third scenario, the offset estimation does behave as expected, with the third scenario having a higher offset of 1.99.

The BDu stabling plan robustness estimations are quite high in all three scenarios, with the first scenario having a very narrow robustness estimation interval of $[82.7\%, 84.2\%]$. The SD plans which have a lower average difference score are only slightly lower, with the 'optimal' SD plan having an average score of 1.14, which is only 0.08 lower than the BDu stabling plan. The second scenario has a BDu stabling plan robustness of $[85.7\%, 90.1\%]$, which means the BDu plan generated in this second scenario performs slightly better than the BDu stabling plan of the first scenario when compared to their generated SD stabling plans. Finally, the third scenario, as mentioned, sees a large increase in the offset estimation and also the SD plans seem to struggle a lot more with the 80% capacity utilisation scenario compared to the 60% capacity utilisation in the second scenario. Although the upper bound of the robustness estimation is very similar to the upper bound of the robustness estimation in the second scenario, its lower bound has gone down compared to the second scenario, leading to a robustness estimation interval of $[81.7\%, 90.9\%]$. As mentioned earlier, the second peak in the spread distribution is very pronounced in this scenario at Nijmegen, with a valley between 2.5 and 3.0 average difference score, before the density rises up significantly again, reaching a peak at an average difference score of around 3.5.

6.2.4 Zwolle

The final location used for the RAM is Zwolle, more specifically, the shuffleboard layout stabling yard southwest of Zwolle Station. Figure 6.4 below shows the offset-spread plots of the three capacity utilisation scenarios.
First noticeable is the very high viability in all three scenarios, with the first two scenarios having a viability of 100%, and the third scenario only being slightly lower at 99.96%. Compared to Carthesiusweg and Nijmegen, this is quite a bit higher. One explanation for these higher viability values is that the shuffleboard stabling yard of Zwolle generally has longer stabling tracks compared to the other two locations. As a consequence of this, the stabling yard does not suffer as much when a composition becomes longer due to stochastic events, as the relative increase in length of the affected composition compared to the total length of the stabling track is smaller. Another, less satisfying, explanation could have been that the three scenarios at Zwolle got 'lucky' with the generated stabling demand from the SA, however when rerunning the RAM for these scenarios led to the same high performance on the viability, indicating that this explanation is less likely.

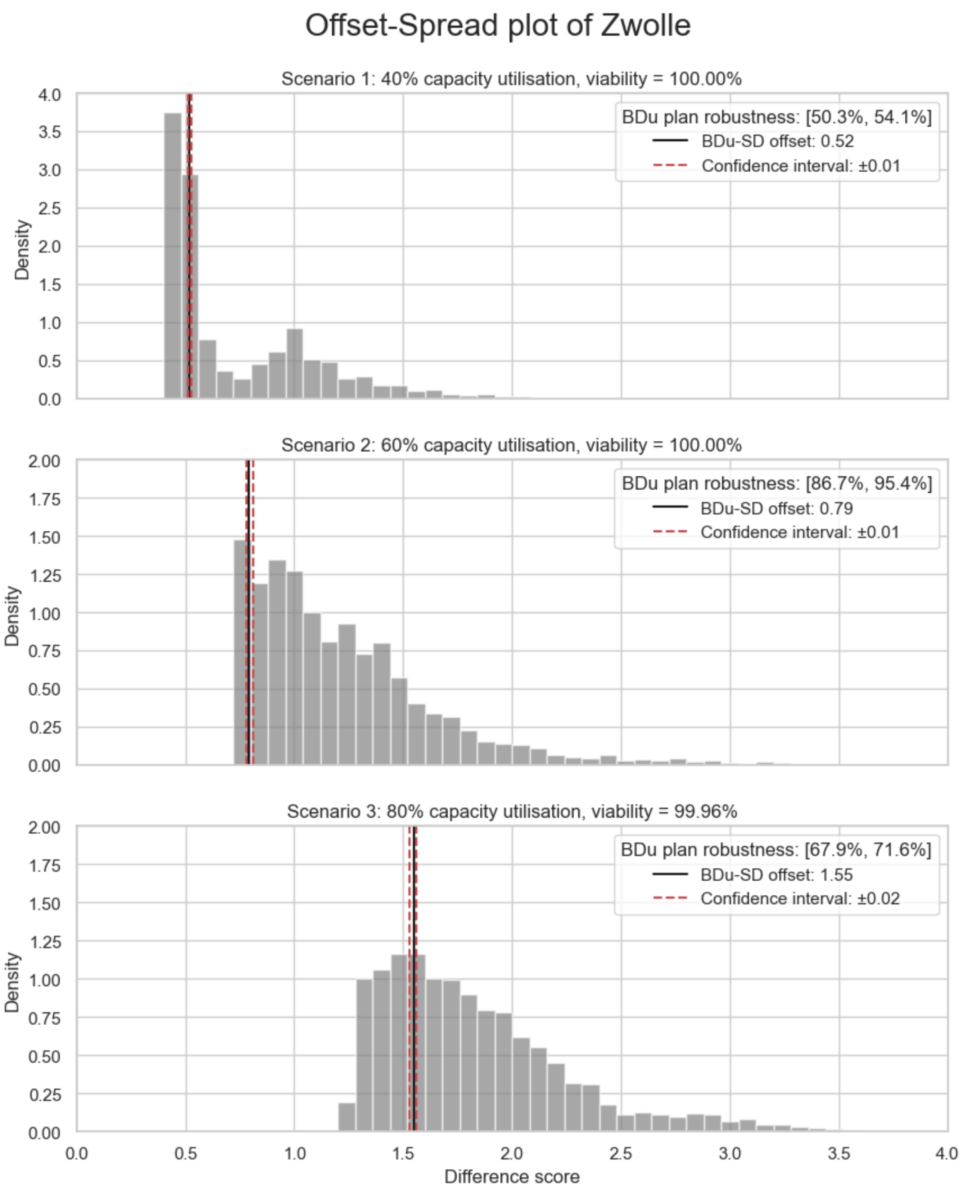


Figure 6.4: Offset-Spread plot of the scenarios at Zwolle.

The offset values for each scenario are systematically lower than at the other two locations, with an estimated offset of 0.52 in the first scenario, 0.79 in the second scenario, and 1.55 in the third scenario. These lower offset values compared to the other two locations are likely due to the longer length of the stabling tracks as mentioned, as well as the fact that this location has the least number of stabling tracks of all, meaning that the penalty of track changes generally has a lesser contribution to the total difference score.

The spread distributions in these three scenarios generally show the same behaviour as at the other two locations – with the exception of the first scenario at Nijmegen – in how the whole distribution moves towards the right when the capacity utilisation is increased. The first scenario at Zwolle seems to show a second peak, just like the third scenario at both Carthesiusweg and Nijmegen. It is unclear whether this is due to the randomness of the SA, or that this double-peaked distribution is not limited to the third scenario only. The third scenario at Zwolle does not have a distinct second peak, but it does show a sharp decrease in the slope of the distribution after an average difference score of 2.5.

Although the offset at the first scenario is very far to the left side of the distribution, the robustness estimation is relatively low, coming in at an interval of $[50.3\%, 54.1\%]$, as there are a lot of SD stabling plans which have a slightly lower average difference score. Even though these SD plans have only a slightly lower score, the fact that such a large amount of SD stabling demands result in an on average better performing SD plans means that the TAP model was unable to find a better stabling plan using the initial stabling demand. The BDu stabling plan in the second scenario performs much better, with a robustness estimation of $[86.7\%, 95.4\%]$. The BDu stabling plan in the final scenario performs somewhat better than the BDu plan in the first scenario but is still well away from the robustness at the second scenario, with a robustness estimation interval of $[67.9\%, 71.6\%]$.

6.2.5 Best and Worst Stabling Plans Analysis

Besides analysing the robustness estimation for the initially generated BDu stabling plans originating from the stabling demand scenarios, it is also insightful to analyse how these BDu stabling plans differ from the best and worst performing SD stabling plan.
Comparing two stabling plans and determining why one plan performs better on robustness compared to the other is an extremely complex task, as it is very hard to predict how each change plays out in the stochastic environment. However, in this subsection a small step is taken into this direction by comparing the stabling plans from three scenarios. As it is easier to compare two extremes, the estimated best performing SD stabling plan is compared to the estimated worst performing SD stabling plan with regard to stabling plan robustness. Note that as these SD stabling plans are created with changed stabling demand, the train units in each of the departure compositions might be different due to variant or type changes, or the addition or removal of a train unit. Furthermore, the scheduled arrival or departure time of the departure composition might have been changed as well. As discussed in Section 5.5, this does not influence the comparison of two stabling plans, as this is done based on the stabling track and arrival order of each of the departure compositions.

The first scenario which is being analysed is Scenario 1 at Carthesiusweg, of which its best and worst performing stabling plans are found in Appendix E.1. Comparing these two stabling plans, two main things are noticed. First of all, has two used stabling tracks, namely tracks 253 and 265, which are almost packed to the brink, with only a few metres left to spare. The present departure composition at track 253 is a SLT-4, which means that a variant change would immediately lead to a required change in the stabling plan, as there is not enough room for a SLT-6 train unit here. The same goes for track 265, where a SNG-3 train unit that changes in a SNG-4 unit will also require changes to the stabling plan. Having a significant portion of the used stabling tracks being close to maximum capacity, the probability of stochastic events resulting in required changes to the stabling plan increases, which negatively affects the robustness of the plan. On the other hand, the best performing plan suffers far less from this issue, with the most packed used stabling track, which is track 269, still having around 20 metres of track to spare. At this track, a SNG 3+4 composition is present, and given the fact that 2 SNG units is the maximum number allowed in a composition in this thesis and that the difference in length between a SNG-3 and SNG-4 train unit is less than the remaining room at the stabling track, this track is not in danger of requiring changes due to stochastic events. Overall, the best performing stabling plan has on average a lot more room to spare on the used stabling tracks relatively to the worst performing stabling plan. As a result, the best plan generally has more margins to cope with the stochastic events. These factors have likely contributed to the fact that the best performing SD stabling plan had an estimated offset of 0.88, whereas the worst performing plan had an estimated offset of 3.62.

The next scenario is the second scenario at Nijmegen, of which the best and worst stabling plans, which have an estimated offset of 0.93 and 4.76 respectively, are found in Appendix E.2. The first difference between these two stabling plans is that in the best plan, every track is used for stabling, whereas in the worst plan track 1R is not used for stabling. This results in relatively less room available in the latter plan to cope with stochastic events without requiring changes to the stabling plan. The best SD plan is able to alleviate a bit of the pressure on the other stabling tracks by using track 1R for stabling as well. Furthermore, the best performing SD plan actually has one extra out of service train unit at the stabling yard, which tips the scales even more towards this plan, as it is able to leave on average more room on each track relative to their total lengths, while having more train units to deal with. Looking at the individual stabling tracks, the worst performing SD plan has a bottleneck at track 3R, which stables both a VIRM-4 unit and a SLT-6 unit and only has around 20 metres of empty stabling track remaining. This extra length is insufficient when another train unit is added to any of the two departure compositions at this track, or even when the VIRM-4 unit encounters a variant change to the VIRM-6 train unit, which increases the risk of required stabling plan changes. The best SD plan suffers less from these issues, with tracks 3R, 5R, and 1R having the least margin. However, track 3R stables an ICM 3+4 composition and has enough room to undergo a variant change which would result in an ICM 4+4 composition without the need for changes to the stabling plan, resulting in the only threat being an addition of a third unit to the composition. Track 5R stables two separate VIRM-4 units and has sufficient room for one of these units to become a VIRM-6 train unit. Extra train units added, or two VIRM-6 train units will however require plan changes. Finally, track 10R has 38 metres to spare and stables a VIRM-6 unit and a SNG-4 unit. As these two units are both the largest variant, only a train unit addition to either would result in required stabling plan changes.

The final scenario is the second scenario at Zwolle. Its best and worst SD stabling plans are found in 0 and the estimated offset are 0.77 and 3.75 respectively. With the very long but few stabling tracks at Zwolle, it is inevitable in higher capacity utilisation scenarios that a much larger number of departure compositions need to be stabled at the same track compared to Carthesiusweg and Nijmegen, which generally have shorter stabling tracks. As a result, the risk increases for each stabling track to not have sufficient remaining room to cope with the possible stochastic events the departure compositions are subjected to, since more departure compositions on the same track increase the probability that one or more stochastic events occur at this stabling track. In the worst performing SD stabling plan, stabling track 432 stables two separate SNG-4 units, an ICM-3 unit, a SNG-3 unit, and a VIRM-6 train unit. This is right up to the maximum capacity of the track, as it only has 2 metres of room left. This means that any increase in length of one of these departure compositions due to either a change in variant or addition of a train unit to the composition, will immediately result in required changes to the stabling plan. Comparing this to the most bottleneck stabling track at the best performing SD stabling plan, track 414, the risk of required stabling plan changes is lower here. Track 414 has one SNG-3 unit, one ICM-4 unit, one SLT-6 unit, and one ICM-3 unit stabled, and has around 27 metres of spare room. This is enough room for one of the two train units which are of the shorter variant, the SNG-3 and the ICM-3 unit, to become its longer variant without the need for changes in the stabling plan. The other two departure compositions are train units which are already their longest variant, so a variant changer for any of them would result in an increase in remaining room at the stabling track. The only further risk here is an addition of a train unit to one of the departure compositions. Furthermore, the worst performing plan has on average a lower percentage remaining room on its stabling tracks compared to the best plan, indicating that is has less leeway to cope with stochastic events across the stabling yard. However, it has to be noted that the worst performing stabling plan has six out of service train units present in its stabling plan compared to the three in the best performing stabling plan, which means the average percentage remaining room of both of these plans could in a more general case be more similar.

From the analysis of the best and worst stabling plans for these three scenarios, some interesting patterns are starting to show up. First of all, it seems like evenly spreading the departure compositions across the whole stabling yard is quite beneficial for the robustness of a stabling plan, as not using available stabling tracks means that the other stabling tracks will become busier, and each added departure composition to a stabling track is an increased risk in there not being sufficient room for these departure compositions when stochastic events which increase the length of a composition have occurred. Furthermore, the most bottlenecked tracks seem to be an important predictor of the robustness of a stabling plan, as these tracks will be the first to crumble under the effects of the stochastic events. Special attention regarding this is required for the differing probabilities of occurrence for each of the stochastic events, as a stabling track which would only succumb to a stochastic event which has a lower probability of occurring is in less of a threat than a stabling track which might have more leeway in it but could succumb to a stochastic event with a higher probability of occurrence. Also, the number of departure compositions on the same stabling track likely has an effect on the robustness of a stabling plan, as each extra composition on the stabling track will increase the probability that at least one of these compositions is subjected to a stochastic event.

6.2.6 Sensitivity Analysis

Besides running the RAM at the three locations using the three defined capacity utilisation scenarios to obtain the estimated robustness of the initial stabling plan of each of the nine scenarios in total, it is also important to investigate the sensitivity of the RAM itself. The sensitivity of the model indicates how much the results of the model in question change when one of the model variables has a different value. This is done by performing a one-at-a-time sensitivity analysis, where one of the four stochastic variables is either increased or decreased. These four stochastic variables are the probabilities of a variant change (“Variant”), type change (“Type”), addition or removal of train unit (“Unit”) and change in arrival or departure time (“Time”), as discussed in Section 5.5. To get the clearest picture of the sensitivity of the RAM, the sensitivity analysis will be run eight times, a 10% increase (+) and a 10% decrease (-) compared to the normal value for each variable, for each of the nine regular scenarios, leading to 72 runs in total. The viability, average estimated offset, and estimated robustness based on this average offset of each of these eight sensitivity runs are then compared to the values of the respective regular run. From these values, a percentage change between the values of each sensitivity run and the normal run is then calculated. For each variable change and each metric, the average percentage change of the nine scenarios is taken. These results are shown in Table 6.5. The individual results of the 72 total sensitivity analysis runs can be found in Appendix F.

Table 6.5: Sensitivity analysis results.

Metric	Change compared to normal results [%]							
	Variant+	**Variant-**	**Type+**	**Type-**	**Unit+**	**Unit-**	**Time+**	**Time-**
Viability	0.02	-0.01	0.02	-0.02	-0.06	0.01	0.02	0.00
Offset	0.23	0.15	0.12	-0.11	1.74	-1.72	2.66	-1.71
Robustness	0.99	1.39	0.58	0.52	1.62	-0.37	1.78	-1.22

Looking at the average change of the viability compared to the normal scenarios, an increase or decrease in the probability of any of the four stochastic events occurring does not seem to have any significant effect on the viability. The ‘largest’ influencer is an increase in the probability of the addition or removal of a train unit to a departure composition, resulting in a decrease of just 0.06%, which is negligible compared to the 10% increase of the respective variable. In fact, the small changes for each of the variables are more likely to be the cause of the randomness of the SA when generating new stabling demand.

For the estimation of the offset, generally the same story applies. An increase or decrease in the chance of a variant or type change are very small compared to the respective increase or decrease and therefore these changes are more likely to be due to the randomness of the model than due to the changes in event probability. The changes at the other two variables seem to have a bigger effect on the offset. The offset on average increases by 1.74% and decreases by 1.72% when the probability of the addition or removal of a train unit to the departure composition is increased or decreased respectively. These effects can be explained due to the fact that in a majority of the cases a train unit will be added rather than removed, as generally departure compositions consist of only one train unit in the generated BDu stabling demand, and that these additions have a significant effect on the composition length, therefore increasing the probability that this composition does not have sufficient room at its original stabling track and needs to be stabled at a different track, resulting in a higher difference score for this case. Inversely, a decrease in the probability of the addition or removal of a train unit will generally result in shorter trains, creating more buffer space on each stabling track which in turn means less changes to the stabling plan could be required.

Changes in the probability of a change in arrival or departure time have similar effects, with an increase in this probability results in an average offset increase of 2.66% and a decrease in probability in an average offset decrease of 1.71%. Whenever a composition gets its arrival or departure time changed, there is a risk that it arrives at or departs from its stabling track in a different order than planned, therefore requiring changes in the stabling plan, which explains the observed changes in the average offset.

Although changes in these two variables are significantly larger than the first two variables, they still are around an order of magnitude smaller than the change in the variable value, indicating that also these last two variables do not have a large impact on the offset value.

One might expect that the robustness would change inversely of the change in offset, however, a lower offset value does not necessarily mean a higher robustness. The viability and offset are calculated based on the BDu stabling demand and stabling plan and the SD counterparts, with only the latter being affected by changes in the probability of stochastic events from occurring. However, the robustness is estimated using the spread distribution, which is the result of determining the difference between each SD stabling plan with each other SD stabling plan. As a result, a change in the probability of an event could have both negative and positive outcomes when it comes to comparing two SD stabling plans. For instance, a higher probability of the addition or removal of a train unit to a departure composition would generally lead to a higher offset, as the BDu plan remains unchanged, and an SD plan now have one lengthened composition requiring stabling plan changes. However, another SD plan now has a higher probability of having this event occurring to its departure composition as well, which might mean that this stabling plan needs to be changed the same way as the first SD plan, lowering their differences, which on a large scale results in the spread distribution moving more to the left of the offset, lowering the robustness. Inversely, it could happen that the increase in probability would rather increase the differences between two SD plans instead of cancelling each other out in the aforementioned example. Therefore, the results of the offset changes are not necessarily directly connected to the changes in robustness.
The changes in robustness seen in Table 6.5 show that changes in the probability of the last two variables are at the same order of magnitude as for the offset changes and of the same sign, with an increase in offset also having an increase in robustness and vice versa. Changes in the probability of a variant change seem to have more effect on the robustness compared to the offset, whereas changes in unit type are still very little. This could be explained by the fact that a change in variant results in a changed length, although not by the same significance of the addition or removal of a complete train unit, which can result in required changes to the stabling plan. On the other hand, type changes generally have less effect on the composition length, as the lengths of the small and large train unit variants of one train type are very similar to those of the other train type with the same service type, therefore resulting in a lower probability of required changes to the stabling plan.

Overall, the changes in robustness are all quite small compared to the 10% change in probability of the variables, indicating that the model output does not change heavily when these probabilities have a change in value.

6.3 Reflection on Results

The RAM has shown to give valuable insights regarding the robustness of an initial stabling plan by comparing it to a set of generated stabling plans of a later phase.

The spread distributions generated by the RAM in the chosen scenarios generally seem well-formed and smooth, indicating that the chosen number of required SD stabling plan generations are sufficient, as a too low number of SD plans would result in a much more jagged histogram. To increase the resolution of the spread distribution even more, the number of SD plans could be increased. However, this increase in resolution likely does not outweigh the increased runtime of the RAM, as the spread distribution already has a good resolution, and increasing this resolution further would most likely not lead to significantly different or better results.

The analysis of the best and worst performing stabling plans in the previous section did show some patterns which could have an impact on the robustness of a stabling plan. These patterns were all very much infrastructure related. However, also the headways between composition arriving at or departing from the same stabling track have been globally analysed, but these did not show any apparent pattern which could give away how well a stabling plan performs on robustness Even though one would except that a plan with very small headways is much more vulnerable to changes in the scheduled arrival or departure time, these patterns of small headways in worse performing plans and larger headways in better performing plans did not show. Rather, the distribution of larger and smaller headways in a plan seemed rather random and not related to the offset of a stabling plan, which leads to the robustness estimation of the plan.

However, the RAM is not perfect. First of all, the elephant in the room needs to be addressed: the runtime of the model. The runtime of the model is pretty long, especially when it is desired to incorporate the RAM into a larger model, such as the HIP software from NS, with the intent to run the RAM repeatedly in quick succession, as it would significantly increase the computational workload of this node planning software, only to assess the performance of just a small portion of this planning. The runtime of the model at Carthesiusweg ranged from 25 minutes for the first scenario to 35 minutes for the third scenario. For Nijmegen and Zwolle, the runtime ranged between 15 and 25 minutes. Of this runtime, the generation of the BDu stabling plan and the SD stabling demand and stabling plans only took between 30 and 90 seconds, depending on the number of train units in the stabling demand. The rest of the runtime is therefore solely due to the assessment part of the model, more specifically, the calculation of the spread distribution. As it needs to compare each viable SD stabling plan to each other viable SD stabling plan, the number of calculations increases quadratically with the number of viable SD plans. Optimisations to the runtime of this portion of the calculation has already been made by only calculating the upper triangle of the comparison matrix and then mirroring the results along the diagonal, as the difference score between plans A and B is the same as the difference score between plans B and A, which as a result already has halved the calculation time of the spread distribution. However, the runtime is still suboptimal. The number of SD plans which are generated by the RAM could be halved to 2500 to reduce the runtime of the RAM by around 75%. However, reducing the number of generated SD plans will result in less reliable results for viability, offset, spread distribution and therefore BDu stabling plan robustness.

A second shortcoming of the RAM is its simple TAP model, more specifically its objective function. The current objective function only consists of minimising the total length of the departure compositions which are not stabled. As a result, there can be a multitude of stabling plans which have the same optimal objective function value, namely zero. It is not clear whether the linear solver the TAP model uses in this research, FICO Xpress, chooses the first optimal solution it has found, or that it randomly selects one from a pool of optimal solutions. However, in any case this could mean that the TAP model is not consistent in its output, especially at scenarios with a lower capacity utilisation, as in these scenarios the solution space is the largest. Conversely, a scenario with higher capacity utilisation will have a much smaller solution space, as there are more constraints and less room to shuffle trains around.
As a result of this simpler TAP model, the offset estimation and therefore the estimated robustness of the BDu stabling plan returned by the TAP model can vary significantly when

rerunning the RAM and a different BDu stabling plan is returned. This is most noticeable in the robustness estimation of the first scenario, which is generally on the lower side, as a larger BDu stabling plan solution space increases the probability that the chosen BDu stabling plan is not one of the better performing plans.
However, this does not mean that the RAM itself is therefore not performing as intended, in fact, it is performing exactly as intended. The TAP model is only an auxiliary model to the RAM, just as the IG is, to create the BDu stabling plan, whereas the main goal of the RAM is to assess this BDu plan. How 'well' this BDu stabling plan creation is working is of lesser importance, as the output of the RAM might be dependent on the chosen BDu stabling plan, how the RAM itself works does not. Given that the RAM is able to consistently return very similar spread distributions for the same scenario every time, it indicates that the RAM is working as intended in illustrating the possible performance of other, SD stabling plans.

Ideally, the RAM could be used to not only assess the generated BDu stabling plan it receives, but also assess the performance of the TAP model itself in the way it finds the optimal stabling plan in the BDu phase. Further analysis into which patterns in the stabling plans, emerging from the RAM, distinguish a robust stabling plan from a less robust stabling could be very beneficial to improving the TAP model connected to the RAM. With these improvements, a stabling plan generated in the BDu phase could be better prepared for the changes in stabling demand occurring between the BDu and SD phase, requiring the least number of changes to the initial stabling plan to remain feasible.

7 Conclusions and Recommendations

This chapter concludes this thesis, with the conclusions to the research questions being discussed in Section 7.1, a summarising discussion about this research in Section 7.2, and finally the recommendations for further research in Section 7.3.

7.1 Conclusion

The main research question to this thesis read:

How can the robustness of a stabling plan to stochastic events be defined and assessed?

In order to properly be able to answer this main research question, six sub research questions have been formed. These sub questions have been researched and the answers to these questions are summarised here. The first reads:

What knowledge is currently available in literature regarding stabling yard operations and, if applicable, how do they take stochasticity of such operations into account?

This research question has been answered by conducting a literature study in Chapter 2 across two themes: robustness in transport, and the Train Unit Shunting Problem (TUSP) and stochasticity. Research in robustness in railway schedules has generally been focussed on operational robustness when subjected to stochastic events during the moment of operation itself, such as delays. However, research in robustness between two planning phases to changes in information affecting the schedule creation, has been severely lacking. Furthermore, stabling yards have been out of scope in the research in robustness in railway schedules, and therefore a detailed definition of stabling plan robustness and its influential factors could not be found. The problem of planning of operations at a stabling yard, called TUSP, has been heavily researched the past two decades, but the vast majority of these models are still deterministic. In some models, some form of stochasticity has been implemented, however stochasticity in servicing needs and arriving and departing train compositions has been lacking.

The second research question, given below, has also been answered in Chapter 2.

Which methods can be used to assess robustness?

This research question has been answered by the third theme in the literature study, which focusses on the possible assessment methods for robustness of a system. For this assessment, two methods are possible: an analytical model or a simulation model, more specifically the Monte Carlo simulation. The vast majority of researchers prefer this simulation over the analytical method, as Monte Carlo simulation is more efficient in complex systems. Therefore, it has been concluded that the MC simulation is the best method to assess the robustness of an initial stabling plan.

The third research question, which is stated below, has been answered in Section 4.1.

What are the planned operations at (NS) stabling yards?

This research question has been answered in Section 4.1. A stabling yard generally comes in one of two layouts, which are a shuffleboard layout where trains can only access and exit the tracks from one side, and a carousel layout, where trains can enter and exit at both sides. Besides splitting and combining train compositions and parking them overnight, a stabling yard can also offer additional services, such as cleaning and routine inspections, and in some cases also small maintenance works. These services can be performed in two ways: low servicing and carousel servicing. At the first method, trains are being serviced at their stabling track, whereas in the second method trains have to be shunted to special service tracks instead.

How can the robustness of an initial stabling plan be defined and what are the biggest uncertainties in the initial planning stages regarding the feasibility of the stabling plan?

The fourth research question, stated above, is answered in Section 4.5, following the information in Section 4.3, and consists of two parts. First of all, the robustness of an initial stabling plan is defined as *"the effectiveness at which the stabling plan is able to cope with the stochasticity of events during the later planning phase with the least number of changes compared to other stabling plans"*.
This definition focusses not on the pure 'survivability' of a plan or system, but rather how capable it is in adapting to changes in input data or changes in constraints, as it is nearly impossible for a plan to remain feasible in its original form when the stabling demand changes. The planning uncertainties which are found to be most influential to the variability between an initial stabling plan and the realisation of this plan are:

- Change of infrastructural availability.
- Presence of out of service trains at stabling yard.
- Change in arrival/departure time.
- Change in train unit variant.
- Change in train unit type.
- Change in number of train units in composition.
- Rolling stock breakdown.
- Crew unavailability.

How can a stabling plan be assessed on robustness against stochastic events?

For this research question, the created Robustness Assessment Model (RAM) is presented in Chapter 5. In this thesis, scope of the RAM has been reduced to a shuffleboard stabling yard with the low servicing method. Of the planning uncertainties discussed in the answer to the previous research question, the following uncertainties have been incorporated into the RAM:

- Presence of out of service trains at stabling yard.
- Change in arrival/departure time.
- Change in train unit variant.
- Change in train unit type.
- Change in number of train units in composition.

The RAM then estimates the robustness of an initial stabling plan, which in this case is from the Basic Day update (BDu) phase with regard to the stochastic events in the next phase, the Specific Day (SD) phase, in three steps.
First, a BDu stabling demand and stabling plan needs to be generated. The demand is realistically generated by extending the Instance Generator (IG) created by NS. The stabling plan is created by solving the Track Assignment Problem (TAP) using a Mixed Integer Programming model.
Second, a Monte Carlo simulation has been created that generates a set of SD stabling plans by running the BDu through an algorithm which alters this stabling demand based on the chosen stochastic events.
Finally, the RAM estimates the robustness of the BDu stabling plan by calculating the offset, which is equal to the average difference between the BDu plan and the SD plans, and a spread distribution, which is the average difference of each SD plan compared to every other SD plan. The offset is estimated in a 95% confidence interval by using bootstrapping over the original sample of differences between the BDu and SD plans, therefore also resulting in a 95% confidence interval for the robustness estimation.

The final research question of this thesis is:

How do the case study stabling plans perform on robustness?

This research question discusses the performance of the initial, BDu generated stabling plan, with regard to its robustness to the stochastic events incorporated in the model, on the three locations using three scenarios. After filtering the SD plans for viable SD plans, the RAM assesses the BDu stabling demand and stabling plan on the offset, which is the average number of changes per departure composition between the BDu stabling plan and all viable SD stabling plans, and a robustness estimation, which is the percentage of SD plans which perform worse on average difference with the other SD plans compared to the offset value of the BDu stabling plan. Table 7.1 shows a summarised view of the results of the RAM for all three case study locations and the three capacity utilisation scenarios, which are 40%, 60%, and 80% of the theoretical maximum capacity. The offset value in the table is the average offset estimation, whereas the robustness estimation is given as a 95% confidence interval.

Table 7.1: RAM normal results summary.

Location	**Utrecht \Carthesiusweg**			**Nijmegen**			**Zwolle**		
Used capacity	**40%**	**60%**	**80%**	**40%**	**60%**	**80%**	**40%**	**60%**	**80%**
Viability [%]	99.92%	99.80%	99.56%	99.96%	99.86%	99.40%	100%	100%	99.96%
Offset	0.91	1.60	2.24	1.22	1.02	1.99	0.52	0.79	1.55
Robustness interval	86.5% 99.2%	97.4% 99.4%	80.7% 94.1%	82.7% 84.2%	85.7% 90.1%	81.7% 90.9%	50.3% 54.1%	86.7% 95.4%	67.9% 71.6%

An analysis of the best and worst performing SD stabling plans of three scenarios originating from the respective spread distributions of these scenarios showed that evenly spreading the departure compositions across all available stabling tracks is beneficial for the robustness of said stabling plan. Furthermore, analysing the most bottlenecked stabling tracks in a stabling plan could give indications as to how well or how bad a stabling plan would perform on robustness. No pattern could be found, however, between the headways of compositions arriving at or departing from the same stabling track.
A sensitivity analysis has been performed, where each of the stochastic variables are changed one at a time, to investigate how the estimation results change. The percentage of SD plans deemed viable did not change significantly when stochastic variables were changed. For the offset estimation, changes in the number of train units in a composition or changes in arrival or departure time had a small effect on the results. The changes in the results for the robustness estimation were also relatively small. Judging by the results of the sensitivity analysis, it can be concluded that the RAM is very stable and does not yield significantly different results when probabilities of stochastic events are altered slightly.

Given these answers to the research questions, the majority of the research gaps defined in this thesis have been filled. A well-argued definition of the robustness of an initial stabling plan has been given, and also an assessment model has been created which is able to judge whether the creation process of an initial stabling plan is able to provide robust solutions in these early planning stages. The addition of stochasticity in the stabling demand in the form of changes in the rolling stock also provides a contribution to the research in the TUSP, as this has not yet been incorporated into the models of other researchers. However, as this thesis has focussed on the transition from the BDu to the SD phase, stochasticity in service requirements of trains have not been considered. Therefore, there is still knowledge to be gained for TUSP researchers in that regard.

7.2 Discussion

It has become clear at the start of this thesis that research in the robustness of an initial stabling plan was lacking. This has posed a problem for NS due to the fact that changes in information, more specifically the stabling demand, between the BDu and SD planning phases can lead and has led to major changes in the stabling plan being required to make the initially designed stabling plan feasible again. The goal of this research has been to first of all define the robustness of an initial stabling plan, as well as to create a method which assesses how efficiently an initial stabling plan is able to deal with these changes in information.
Given a scenario with stabling demand and a stabling yard, the RAM is able to create a BDu stabling plan, generate SD stabling demand based on the BDu stabling demand and the influential stochastic events, generate SD stabling plans, and finally assess the performance of the BDu stabling plan based on its differences with the generated SD stabling plans.
The BDu stabling plans of the three scenarios at the three case study locations generally performed very well on robustness, with the upper bound of the robustness estimation often exceeding 90%. Some did perform a little worse due to the generated BDu plan being suboptimal.

However, the results of the RAM are still very insightful, as it both shows the ability of the TAP model to generate robust initial stabling plans and the possible performance it could achieve by analysing the spread distribution. The generally good results of the RAM across the scenarios show that the TAP model used for the RAM is able to generally find an initial stabling plan which is close to the optimal offset value, and therefore robustness, as the SD plans with the lowest average difference score compared to the other SD plans are generally only a fraction lower than the estimated offset of the initial stabling plan. One unexpected result was the Nijmegen 40% capacity utilisation scenario. Across all other scenarios, a pattern was visible in how both the offset and the lowest average difference score in the spread distribution would increase when the capacity utilisation is higher. However, in the Nijmegen 40%-scenario, both these values were actually higher than their counterparts in the Nijmegen 60%-scenario. Rerunning this specific scenario a number of times did not result in a different outcome, indicating that the results of this scenario are not due to ‘bad luck’ in the randomness of the RAM. Why this scenario has a worse performance is therefore still unclear.

The results of this thesis can be very insightful for NS, but also other rail operators who suffer from the consequences of changes in scheduled stabling demand after an initial stabling plan has been created. First of all, defining the robustness of an initial stabling plan has opened the door into investigating how certain factors affect this robustness. Continuing from this definition, the created RAM can be used in multiple ways. The RAM could be used to either assess a set of candidate initial stabling plans to investigate which of these plans performs the best on robustness, or the RAM could be used to assess the ability of the TAP model creating these initial plans to generate robust stabling plans immediately. Furthermore, the RAM could also be used on a set of pre-defined scenarios, such as been done in this thesis, to find patterns in stabling plans which are able to predict or influence the robustness of the stabling plan, such as the spread of departure compositions across all available stabling tracks and the most bottlenecked stabling tracks discussed in Subsection 6.2.5.

The RAM is not perfect, however. First of all, the scope of the model has been narrowed down to only shuffleboard layout stabling yards using a low servicing method. Secondly, operational simplifications have been made by assuming there is enough infrastructure and staff to perform the splitting and shunting of trains when they arrive or depart, as the staff and movement planning is not part of the RAM. Furthermore, the SA of the RAM takes place after the matching of arrivals and departures, as this happens in the IG. The effect of changes in the stabling demand in the RAM could therefore be slightly overestimated, as in reality the matching of arrivals and departures could be redone if this would make finding a new stabling plan easier. Also, the SA is not able to add or remove a departure composition completely, as the model needs the location of the departure composition in both stabling plans it is comparing with each other. Regarding the input data of the RAM, all values are reasonable, with one exception. The mean and standard deviation of the change in arrival or departure time seem a bit unnatural, with a relatively low mean and a standard deviation two orders of magnitude larger. However, the assumption has been made that the delivered data these values were extracted from are correct. The two main limitations of the RAM came to light when running the case study scenarios, which are the simplicity of the TAP model and the runtime of the RAM.

The simple objective function of the TAP meant that, especially in lower capacity utilisation scenarios, multiple initial stabling plans existed with the same optimal objective function value. As a result, rerunning the same scenario would often result in a significantly different offset estimation due to the linear solver used in the TAP picking a different initial stabling plan than before, and therefore also resulting in a significantly different robustness estimation. This does not nullify the results of the RAM, as it still works consistently as intended since the spread distribution does remain consistent due to the distribution being less affected by these small inconsistencies as they even out when generating a large number of SD plans.

The other major limitation of the RAM is, as mentioned, its runtime. A complete run with 5000 SD plans takes between 15 and 35 minutes, depending on the number of train units in the stabling demand, meaning that running the nine regular scenarios and the 72 sensitivity analysis scenarios took around 30 hours in total. The larger runtime of the RAM would be especially problematic when the desire is to frequently assess the robustness of generated initial stabling plans. However, the runtime could be found less problematic when using the RAM to assess the performance of the TAP model in generating robust stabling plans, as ideally when this TAP model is optimised, the RAM does not need to be used continuously anymore. Additionally, the RAM could be used to find patterns in the stabling plans predicting or influencing the robustness of the stabling plan, which could then in turn be incorporated into the TAP model.

7.3 Recommendations

This thesis is not the end of research into stabling plan robustness, rather it acts as the gateway to further developments. In this section, recommendations are given as to how the Robustness Assessment Model (RAM) could be used in its current state, as well as recommendations for further research into assessing and optimising for stabling plan robustness.

In its current form, the RAM can be used in two ways. The first method is to use the RAM to judge a set of candidate stabling plans to determine which of these stabling plans can achieve the highest robustness when subjected to stochastic events between the BDu and SD phase. A second method for operating the RAM is to use the RAM to assess the ability of a TAP model to return robust stabling plans as well as optimise the TAP model further by analysing patterns in stabling plans which could affect its robustness.

Furthermore, the HIP package of NS takes a long time to create a shunt and stabling plan due to its use of Local Search. This is especially problematic when a plan needs to be created again when information changes. Ideally, HIP could first call on a database of earlier developed plans and use one of those plans which had a similar stabling demand as starting point. The RAM is able to aid in determining which plans should be put in the database, as ideally the database consists of plans which have a high robustness, since these are able to cope more efficiently with changes in stabling demand.

Besides NS with the HIP package, other railway operators are also able to use the RAM when they want to investigate how robust their initial stabling yard planning processes are when stabling demand changes at a later phase. Optimising this process can lead to an increase in planning efficiency due to the lower number of required man-hours to adjust a plan at a later phase.

For recommendations for further developments, first of all, it would be very fruitful to extend the RAM to not only cover the simplest combination of stabling yard layout and servicing method, which is the shuffleboard layout with low servicing as presented in this thesis, but also the carousel layout and the carousel servicing method. For the carousel layout implementation, the entry and exit side to the stabling track needs to be taken into account, whereas that is of no concern in the shuffleboard layout as there is only one way of entering and exiting the stabling track. Furthermore, the order determination of the coordinate system needs to be redesigned. One method could be to split up the order coordinate into a distinction between the compositions leaving in direction A and compositions leaving in direction B, or change to a different coordinate value altogether, such as the location of the front or centre of the composition on the stabling track in metres compared to a defined reference point on the track. For implementing carousel servicing, this would mean that a train has multiple arrival and departure times, possibly on different stabling tracks as well. This would require a change to the TAP model such that it is able to plan every movement, rather than just a single stabling location for each train. This would significantly increase the complexity of the model, however, as the runtime of the RAM is 95% due to calculating the spread distribution, it is estimated that this would most likely not result in an extreme increase in total runtime of the RAM.

Incorporating the planning of train movements at the stabling yard nicely leads to the next recommendation, which is further research in assessing the robustness of not only the stabling plan, but the whole node plan. Analysing the robustness of planned train movements in interlocking areas could for instance give insights in which movements or movement

combinations generally need rescheduling in later planning stages.
Additionally, it would also be beneficial to incorporate the possibility of departure compositions being added to or removed from the plan between the planning stages. The current version of the RAM does not support this, as it needs a composition which exists in the first phase to exist in the second phase as well, regardless of what the composition consists of, such that it is able to compare the stabling location of this composition in both plans. Incorporating the addition or removal of a composition to the stabling demand would require an implementation of some sort of penalty to the difference score for each addition or removal, of which the value of would also need to be determined.

Moving from possible extensions to the RAM, it would also be insightful to also investigate which factors are influential to the robustness of a stabling or node plan in other planning horizons, for instance from the SD phase right up until the operational phase. It might be found here that certain factors which were influential in one phase, are of lesser influence on the robustness of a plan in another phase. Furthermore, it could also be interesting to research how the infrastructure affects the robustness of a stabling or node plan, for example when certain tracks or switches are unavailable.

The current method to determine the difference between two stabling plans is not extremely sophisticated and rather aims at simplicity and explainability. Because of this, it is rather common that two different sorts of changes between two stabling plans are given the same difference score, where a person might find that one specific change is more different than the other. Making such a model which determines the differences between two plans is always a trade-off between explainability and realism, as a too sophisticated model might make it very hard to understand how the model has reached a certain difference score for the comparison of two stabling plans. It would therefore be very interesting to investigate which changes to a stabling plan are seen as more severe by human planners, be it from visual standpoint or how complex it is to create the rest of the plan around these changes.

8 Bibliography

Abbink, E., van den Berg, B., Kroon, L., & Salomon, M. (2004). Allocation of railway rolling stock for passenger trains. *Transportation Science, 38*, 33-41.

Bao, S. (2018). *A Robust Solution to Train Shunting Using Decision Trees.* Delft University of Technology. Retrieved from MSc thesis. https://repository.tudelft.nl/islandora/object/uuid%3A9acb3bd1-1ffb-4ebb-b886-f10c7291b101

Beerthuizen, E. (2018). *Optimising train parking and shunting at NS service yards.* Delft University of Technology. Retrieved from MSc thesis. https://repository.tudelft.nl/islandora/object/uuid%3Ad5c93539-f182-4d4d-931b-c5356727c827

Borecka, J. (2021). *Routing Optimization for the Train Unit Shunting Problem in a Multi-Agent Deep Reinforcement Learning Framework.* Delft University of Technology. Retrieved from MSc thesis. https://repository.tudelft.nl/islandora/object/uuid%3Ab08fc160-934f-4000-8b09-3308adce3b6d

Centraal Bureau voor de Statistiek. (2022, July 8). *Verkeersdeelname en deelnaame openbaar vervoer; persoonskenmerken*. (CBS) Retrieved November 17, 2022, from https://opendata.cbs.nl/#/CBS/nl/dataset/84707NED/table

Chang, S., & Wang, Z. (2021). Simulation Analysis of the Exit Capacity of depot of Urban Rail Transit. *Journal of Physics: Coference Series. 1972*, p. 012096. IOP Publishing.

Cicerone, S., D'Angelo, G., Di Stefano, G., Frigioni, D., & Navarra, A. (2007). Robust Algorithms and Price of Robustness in Shunting Problems. *7th Workshop on Algorithmic Approaches for Transportation Modeling, Optimization, and Systems*, (pp. 175-190). Sevilla.

Eurostat. (2022, July 13). *Modal split of passenger transport*. (Eurostat) Retrieved November 17, 2022, from https://ec.europa.eu/eurostat/databrowser/view/t2020_rk310/default/table?lang=en

Freling, R., Lentink, R., Kroon, L., & Huisman, D. (2005). Shunting of Passenger Train Units in a Railway Station. *Transportation Science, 39*(2), 261-272.

Ge, L., Voß, S., & Xie, L. (2022). Robustness and disturbances in public transport. *Public Transport, 14*, 191-261.

Gilg, B., Klug, T., Martienssen, R., Paat, J., Schlechte, T., Schulz, C., . . . Tesch, A. (2018). Conflict-free railway track assignment at depots. *Journal of Rail Transport Planning & Management, 8*, 16-28.

Goverde, R. (2007). Railway timetable stability analysis using max-plus system theory. *Transportation Research Part B: Methodological, 41*(2), 179-201.

Goverde, R., Heidergott, B., & Merlet, G. (2009). Railway timetable stability analysis using stochastic max-plus linear systems. *Proceedings 3rd International Seminar on Railway Operations Modelling and Analysis* (pp. 1-18). Zurich: Swiss Federal Institute of Technology.

Guimarães, V., Leal Junior, I., & da Silva, M. (2018). Evaluation the sustainability of urban passenger transprotation by Monte Carlo simulation. *Renewable and Sustainable Energy Reviews, 93*, 732-752.

Haahr, J., Pisinger, D., & Larsen, J. (2015). *Reactive Robustness and Integrated Approaches for Railway Optimization Problems.* DTU Management Engineering.

Huizingh, E. (2018). *Planning first-line services on NS service stations.* Univeristy of Twente. Retrieved from MSc thesis. https://essay.utwente.nl/74968/1/Thesis2018_E.H.R.Huizingh_IEM.pdf

Kroon, L., Schrijver, A., & Lentink, R. (2008). Shunting of Passenger Train Units: an Integrated Approach. *Transportation Science, 42*, 436-449.

Landex, A., & Jensen, L. (2013). Measures for track complexity and robustness of oepration at stations. *Journal of Rail Transport Planning & Management, 3*, 22-35.

Liebchen, C., & Möhring, R. (2007). The Modeling Power of the Periodic Event Scheduling Problem: Railway Tiemtables - and Beyond. In F. Geraets, L. Kroon, A. Schoebel, D. Wagner, & C. Zaroliagis, *Algorithmic Methods for Railway Optimization* (Vol. 4359). Berlin: Springer. doi:10.1007/978-3-540-74247-0_1

Liebchen, C., Lübbecke, M., Möhring, R., & Stiller, S. (2009). The Concept of Recoverable Robustness, Linear Programming Recovery, and Railway Applications. In *Robust and Online Large-Scale Optimization* (Vol. 5868, pp. 1-27). doi:10.1007/978-3-642-05465-5_1

Lusby, R., Larsen, J., & Bull, S. (2018). A survey on robustness in railway planning. *European Journal of Operational Research, 266*(1), 1-15. doi:10.1016/j.ejor.2017.07.044

Mira, L., Andrade, A., & Castilho Gomes, M. (2020). Maintenance scheduling within rolling stock planning in railway operations under uncertain maintenance durations. *Journal of Rail Transport Planning & Management, 14*, 100177.

Morga, L. (2013). *Robustness and Recoverability in Transport Logisctics.* Madrid: Universidad Politecnica de Madrid.

Norrbin, P. (2016). *Railway Infrastructure Robustness.* Lulea University of Technology.

NS. (2019). *Jaarverslag 2019*. (NS) Retrieved November 17, 2022, from https://2019.nsjaarverslag.nl/

NS. (n.d.). *NS Reisplanner*. (NS) Retrieved December 2023, from https://www.ns.nl/reisplanner/#/

NS. (n.d.). *Treinen van NS*. (NS) Retrieved November 18, 2022, from https://www.ns.nl/over-ns/treinen-van-ns

NS. (n.d.). *Verantwoordelijkheden*. (NS) Retrieved November 18, 2022, from https://www.ns.nl/over-ns/de-spoorsector/verantwoordelijkheden.html

Ozkan, O., & Kilic, S. (2019). A Monte Carlo Simulation for Reliability Estimation of Logistics and Supply Chain Networks. *IFAC-PapersOnLine, 52*(13).

Rusca, A., Popa, M., Rosca, E., Rosca, M., Dragu, V., & Rusca, F. (2016). Simulation model for port shunting yards. *IOP Conference Series: Materials Science and Engineering, 145*(8).

SporenplanOnline. (n.d.). (Sporenplan) Retrieved from http://sporenplan.nl/

Tahmasseby, S., van Oort, N., & van Nes, R. (2008). The role of infrastrcutures on public transport service reliability. *2008 First International Conference on Infrastructure Systems and Services: Building Networks for a Brighter Future (INFRA)*, (pp. 1-5).

UK Department for Business, Energy & Industrial Strategy. (2019). *Greenhouse gas reporting: conversion factors 2019*. (UK Government) Retrieved November 17, 2022, from https://www.gov.uk/government/publications/greenhouse-gas-reporting-conversion-factors-2019

van Cuilenborg, D. (2020). *Train Unit Shunting Problem, a Multi-Agent Pathfinding approach.* Delft University of Technology. Retrieved from MSc thesis. https://repository.tudelft.nl/islandora/object/uuid%3Ac9a0f41c-de6f-4ef4-8d94-6d0a89df3eec

van den Broek, R. (2016). *Train Shunting and Service Scheduling: an integrated local search approach.* Utrecht University. Retrieved from MSc thesis. https://studenttheses.uu.nl/bitstream/handle/20.500.12932/24118/ThesisRoelvandenBroek.pdf?sequence=2

van den Broek, R. (2022). *Towards a Robust Planning of Train Shunting and Servicing.* Utrecht: Utrecht University.

van den Broek, R., Hoogeveen, J., & van den Akker, J. (2018). How to Measure the Robustness of Shunting Plans. *18th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems (ATMOS 2018).* Dagstuhl.

van Hövell, M. (2019). *Increasing the Effectiveness of the Capacity Usage at Rolling Stock Service Locations.* Delft University of Technology. Retrieved from MSc thesis. https://repository.tudelft.nl/islandora/object/uuid%3A7d2b5493-9cd0-4197-991f-1937f007e6c9

van Leeuwen, A. (2021). *Optimization of the short-term labor and task scheduling in the maintenance of regional trains.* Delft University of Technology. Retrieved from MSc thesis. https://repository.tudelft.nl/islandora/object/uuid%3A7db77462-e6d7-4116-b00d-1c118f1de4c4

van Marsbergen, M. (2018). *Estimating the capacity of a stabling yard with the implementation of 100% servicing.* Delft University of Technology. Retrieved from MSc thesis. https://repository.tudelft.nl/islandora/object/uuid%3Ad78dd1fc-b42e-43f2-a3c7-6de37a777116 van Nes, R. (2002). *Design of multimodal transport networks: A hierarchical approach.* Delft University of Technology.

van Oort, N. (2011). *Service Reliability and Urban Public Transport Design.* Delft: TRAIL Research School.

Vromans, M., Dekker, R., & Kroon, L. (2006). Reliability and heterogeneity of railway services. *European Journal of Operational Research, 172*(2), 647-665.

Wang, J., Zhao, Y., Gronalt, M., Lin, B., & Wu, J. (2022, May). Synchronized Optimization for Service Scheduling, Train Parking and Routing at High-Speed Rail Maintenance Depot. *IEEE Transactions on Intelligent Transportation Systems, 23*(5), 4525-4540. doi:10.1109/TITS.2020.3045852

Wegener, M. (2004). Overview of land use transport models. *Journal of Discrete Algorithms, 5*.

Zieger, S., Weik, N., & Nießen, N. (2018). The influence of buffer time distributions in delay propagation modelling of railway networks. *Journal of Rail Transport Planning & Management, 8,* 220-232.

Zinser, M., Betz, T., Becker, M., Geilke, M., Terschlüsen, C., Kaluza, A., . . . Warg, J. (2019). PRISM: A Macroscopic Monte Carlo Railway Simulation. *The 12th World Congress on Railway Research.* Tokyo.

Appendices

Appendix A: Instance Generator Ratios

A.1: Presence ratio of each train unit.

Train unit	Amount	Presence ratio
VIRM-4	98	40376
VIRM-6	80	32960
ICM-3	87	35844
ICM-4	50	20600
SLT-4	69	28428
SLT-6	62	25544
SNG-3	118	48616
SNG-4	88	36256

A.2: VIRM composition ratios.

VIRM-4	VIRM-6	Arrival ratio	Departure ratio
1	0	29682	33239
0	1	26894	29281
2	0	3558	2546
1	1	3578	2045
0	2	832	405
Total	VIRM-4	40376	40376
	VIRM-6	32136	32136

A.3: ICM composition ratios.

ICM-3	ICM-4	Arrival ratio	Departure ratio
1	0	29396	28559
0	1	15473	15034
2	0	1708	1910
1	1	1463	1504
0	2	1316	1045
3	0	285	326
2	1	282	328
1	2	150	327
0	3	150	330
Total	VIRM-4	35844	35844
	VIRM-6	20600	20600

A.4: SLT composition ratios.

SLT-4	SLT-6	Arrival ratio	Departure ratio
1	0	16062	22784
0	1	17038	21636
2	0	3506	1581
1	1	5354	2482
0	2	1576	713
Total	VIRM-4	28428	28428
	VIRM-6	25544	25544

A.5: SNG composition ratios.

SNG-4	SNG-6	Arrival ratio	Departure ratio
1	0	27617	39033
0	1	24021	30633
2	0	6296	2843
1	1	8407	3897
0	2	1914	863
Total	VIRM-4	48616	48616
	VIRM-6	36256	36256

Appendix B: SA Algorithms

Algorithm 1: Change arrival time algorithm.

Data: BDu stabling demand A, event probabilities P, train data (type, service, max. units, variants, amount, length)
Result: Altered arrival time

for $i \in A$ **do**
  $changearrival = choice([0,1], p = [1 - P_{arrival}, P_{arrival}])$
  if $changearrival == 1$ **do**
    $deviation = round\big(normal(\mu = 53.1, \sigma = 4046)\big)$
    $indices = where(A[Arrival\ time] == A[Arrival\ time][i])$
    while $any(|A[Arrival\ time][i] + deviation - A[Arrival\ time]| < 10)$ **do**
      $deviation = round\big(normal(\mu = 53.1, \sigma = 4046)\big)$
    end
    for $j \in indices$ **do**
      $A[Arrival\ time][j] = A[Arrival\ time][j] + deviation$
    end
  end
end

Algorithm 2: Change departure time algorithm.

Data: BDu stabling demand A, event probabilities P, train data (type, service, max. units, variants, amount, length)
Result: Altered departure time

for $i \in A$ **do**
  $changedeparture = choice\big([0,1], p = \big[1 - P_{departure}, P_{departure}\big]\big)$
  if $changedeparture == 1$ **do**
    $deviation = round\big(normal(\mu = 53.1, \sigma = 4046)\big)$
    $indices = where(A[Departure\ time] == A[Departure\ time][i])$
    while $any(|A[Departure\ time][i] + deviation - A[Departure\ time]| < 10)$ **do**
      $deviation = round\big(normal(\mu = 53.1, \sigma = 4046)\big)$
    end
    for $j \in indices$ **do**
      $A[Departure\ time][j] = A[Departure\ time][j] + deviation$
    end
  end
end

Algorithm 3: Change train unit variant algorithm

Data: BDu stabling demand A, event probabilities P, train data (type, service, max. units, variants, amount, lengths)

Result: Altered train unit variants

for $i \in A$ **do**

  $changevariant = choice([0,1], p = [1 - P_{variant}, P_{variant}])$

  if $changevariant == 1$ **do**

    $currenttype = A[Type][i]$

    $currenvariant = A[Variant][i]$

    $V = variants[currenttype]$

    remove $currentvariant$ from V

    $p_{newvariant} = \left[\frac{amount[currenttype][j]}{\sum_k amount[currenttype][k]}\right], \qquad \forall(j,k) \in V$

    $newvariant = choice(V, p = p_{newvariant})$

    $A[Variant][i] = newvariant$

    $A[Length][i] = lengths[currenttype][newvariant]$

  end

end

Algorithm 4: Change in train unit type.

Data: BDu stabling demand A, event probabilities P, train data (type, service, max. units, variants, amount, length)

Result: Altered train unit types

$arrivals = A[Arrival\ time]$

$departures = A[Departure\ time]$

for $i \in A$ **do**

  $changetype = choice\big([0,1], p = [1 - P_{typetuned}, P_{typetuned}]\big)$

  if $changetype == 1$ **do**

    $currenttype = A[Type][i]$

    $othertypes = where(services == services[currenttype]$

    remove $currenttype$ from $othertypes$

    $p_{newtype} = \left[\frac{\sum_k amount[m][j]}{\sum_n \sum_k amount[n][k]}\right], \quad \forall(m,n) \in othertypes$

    $newtype = choice(othertypes, p = p_{newtype})$

    // Determine which units to change type of and change them

    $matchindices = where(arrivals == arrivals[i])$

    $tocheck = matchindices$

    $checked = []$

    while $tocheck \neq \emptyset$ **do**

      $connected_a = where(arrivals == arrivals[tocheck[0]]$

      For $j \in connected_a$ **do**

        if $j \notin tocheck$ **and** $j \notin checked$ **do**

          Append j to $matchindices$

          Append j to $tocheck$

        end

      end

      $connected_a = where(arrivals == arrivals[tocheck[0]]$

      For $j \in connected_a$ **do**

        if $j \notin tocheck$ **and** $j \notin checked$ **do**

          Append j to $matchindices$

          Append j to $tocheck$

        end

      end

      Append $tocheck[0]$ to $checked$

      Remove first index from $tocheck$

    end

    $p_{newvariant} = \left[\frac{amount[newtype][j]}{\sum_k amount[newtype][k]}\right], \ \forall(j,k) \in variants[newtype]$

    for $m \in matchindices$ **do**

      $newvariant = choice(variants[newtype], p = p_{newvariant})$

      $A[Type][m] = newtype$

      $A[Variant][m] = newvariant$

      $A[Length][m] = lengths[newtype][newvariant]$

    end

  end

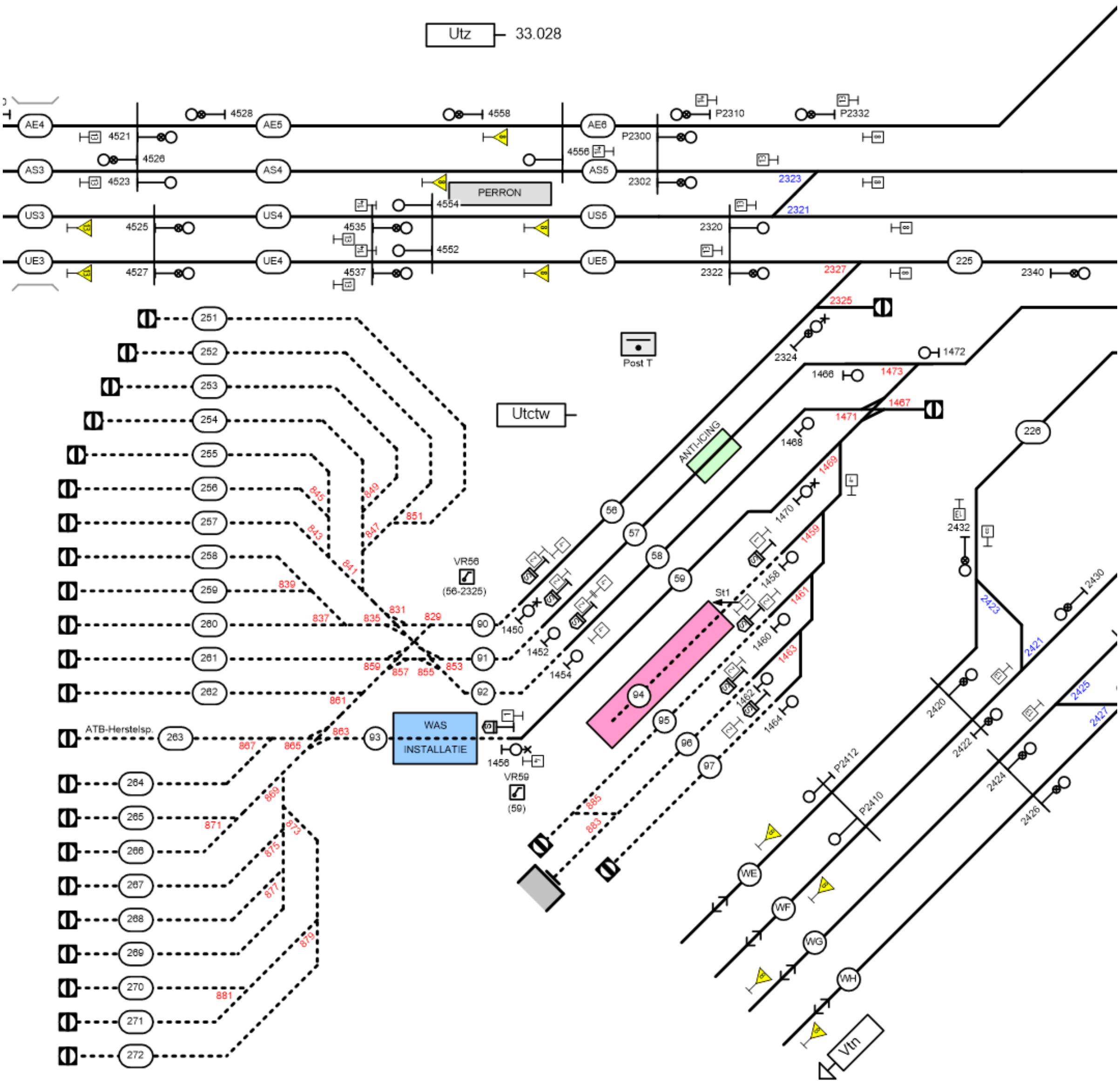
end

Algorithm 5: Addition or removal of train unit from departure composition.

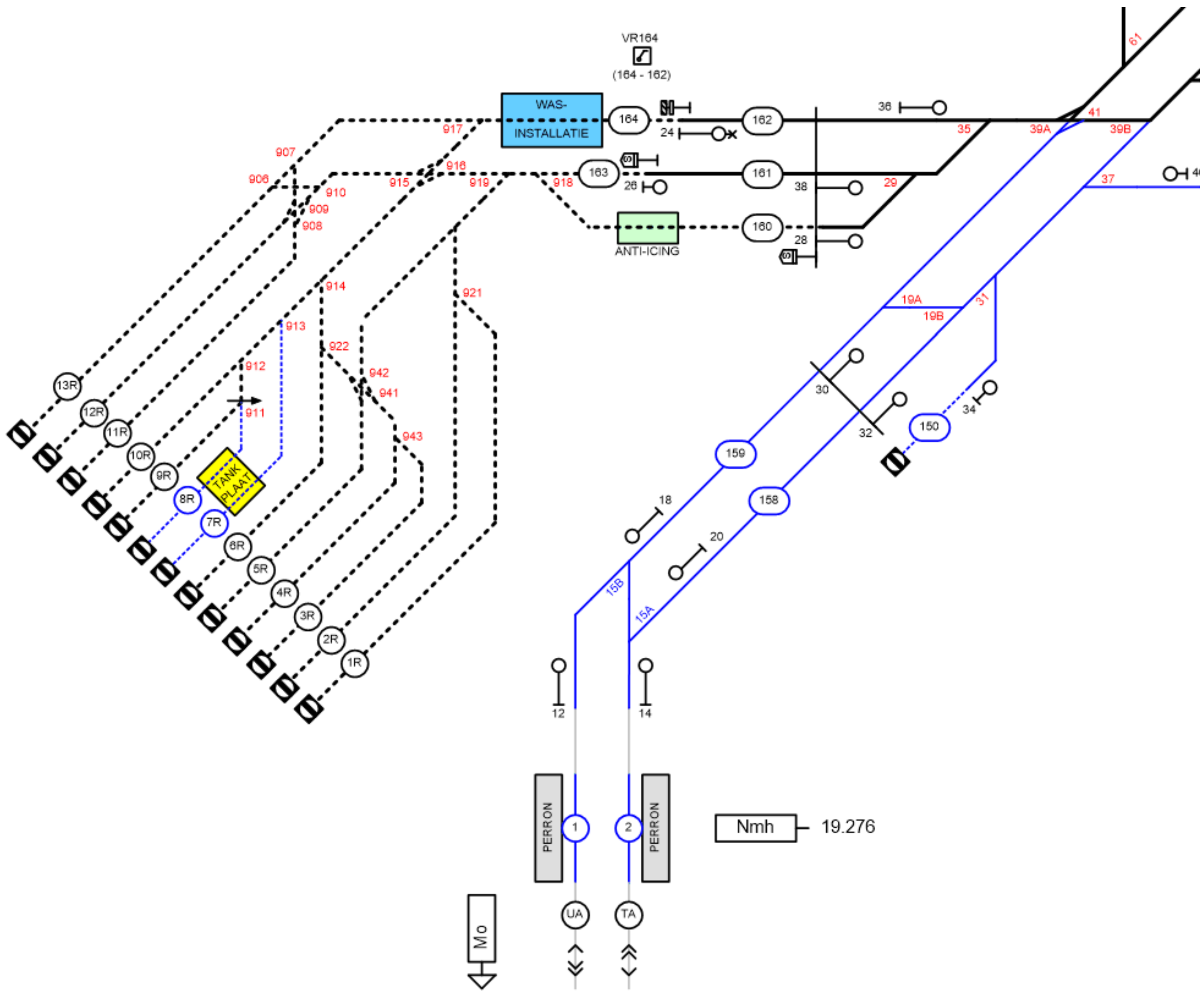
Data: BDu departure compositions C, event probabilities P, train data (type, service, max. units, variants, amount, length)
Result: Altered arrival time

for $i \in C$ **do**
  $addremove = choice([0,1], p = [1 - P_{addremove}, P_{addremove}])$
  if $addremove == 1$ **do**
    $currenttype = C[Type][i]$
    $currentunits = \#units\ currently\ in\ composition$
    if $currentunits == 1$ **do**
      $action = 1$
    else if $currentunits == maxunits[currenttype]$ **do**
      $action = 0$
    else do
      $action = choice([0,1])$
    end
    if $action == 1$ **do**
      $p_{newvariant} = \left[\frac{amount[currenttype][j]}{\sum_k amount[currenttype][k]}\right],$
      $\forall (j,k) \in variants[currenttype]$
      $newvariant = choice(variants[currenttype], p = p_{newvariant})$
      $C[Variant][i] = C[Variant][i] + newvariant$
      $C[Length][i] = C[Length][i] + lengths[currentype][newvariant]$
    end
    else do
      $removalunit = where(variants[currenttype] == C[Variant][i][-1])$
      Delete $removalunit$ from $C[Variant][i]$
      $C[Length][i] = C[Length][i] - lengths[currentype][removalunit]$
    end
  end
end

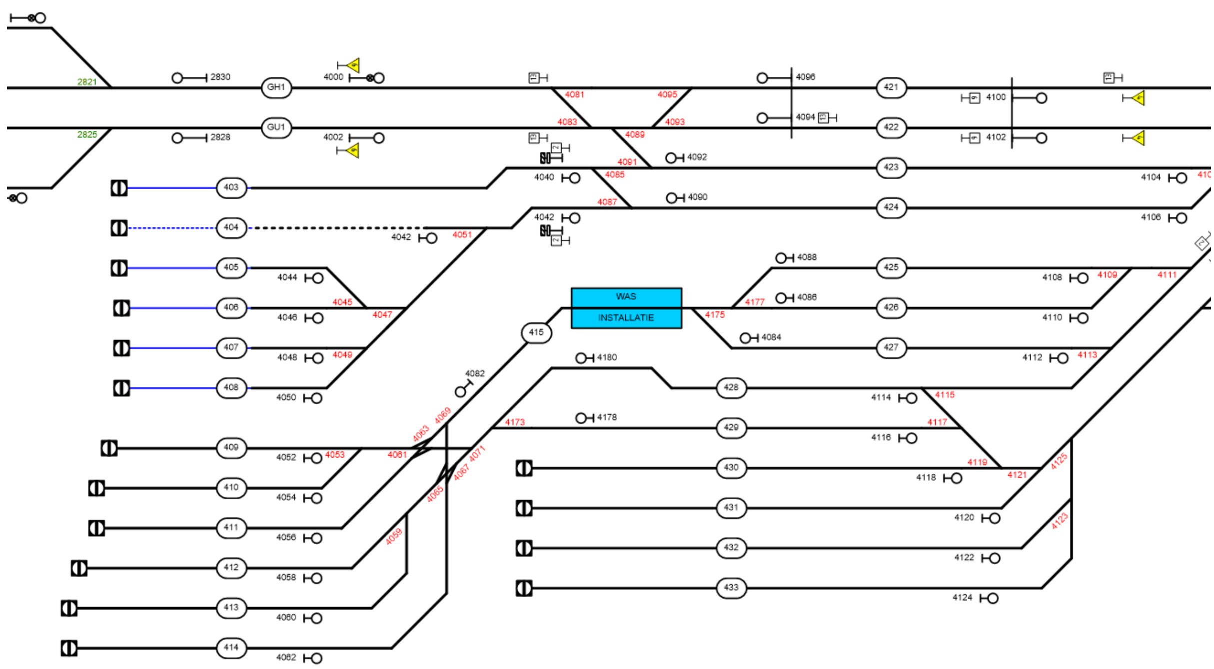
Appendix C: Case Study Locations



C.1: Carthesiusweg stabling yard (SporenplanOnline, sd).



C.2: Nijmegen stabling yard (SporenplanOnline, n.d.).



C.3: Zwolle shuffleboard stabling yard (SporenplanOnline, n.d.).

Appendix D: Case Study Stabling Demand

Appendix D.1: Carthesiusweg

D.1: Carthesiusweg scenario 1 stabling demand.

id	Type	Variant	Length [m]	Arrival time	Departure time
2601	SLT	6	100.5	18:16:08	05:29:31
4001	ICM	3	80.6	19:17:08	05:47:32
4201	ICM	4	107.1	19:32:38	05:15:50
9402	VIRM	4	108.6	21:23:36	05:41:32
9401	VIRM	4	108.6	21:47:12	05:55:43
8601	VIRM	6	162.1	23:39:52	07:41:29
2301	SNG	3	59.6	23:45:53	07:01:40
2702	SNG	4	75.8	23:45:53	06:34:09
2401	SLT	4	69.4	23:54:25	06:07:43
2701	SNG	4	75.8	01:15:17	05:35:31
2302	SNG	3	59.6	01:21:18	06:01:43

D.2: Carthesiusweg scenario 2 stabling demand.

id	Type	Variant	Length [m]	Arrival time	Departure time
2302	SNG	3	59.6	20:42:11	07:39:23
9402	VIRM	4	108.6	21:14:40	06:13:09
4002	ICM	3	80.6	21:55:25	06:19:09
2402	SLT	4	69.4	22:01:25	05:15:20
2602	SLT	6	100.5	22:01:25	05:56:02
2702	SNG	4	75.8	22:07:25	05:33:20
2301	SNG	3	59.6	22:07:25	07:45:24
2401	SLT	4	69.4	22:43:26	05:09:20
2601	SLT	6	100.5	22:49:56	06:59:19
8601	VIRM	6	162.1	23:42:47	05:03:20
2701	SNG	4	75.8	23:50:50	07:32:57
4001	ICM	3	80.6	00:27:22	05:27:20
4201	ICM	4	107.1	00:33:23	06:07:08
2303	SNG	3	59.6	00:41:42	05:21:20
8602	VIRM	6	162.1	00:47:42	05:39:20
9401	VIRM	4	108.6	01:16:19	07:05:20

D.3: Carthesiusweg scenario 3 stabling demand.

id	Type	Variant	Length [m]	Arrival time	Departure time
4002	ICM	3	80.6	17:37:46	07:01:42
8601	VIRM	6	162.1	17:45:02	05:41:09
4003	ICM	3	80.6	18:55:09	05:23:08
8603	VIRM	6	162.1	22:23:03	07:34:57
2701	SNG	4	75.8	22:52:26	07:07:42
2302	SNG	3	59.6	23:09:17	05:35:09
9403	VIRM	4	108.6	23:22:05	07:59:41
8602	VIRM	6	162.1	23:49:06	06:26:48
2402	SLT	4	69.4	23:55:06	07:22:57
4202	ICM	4	107.1	00:01:06	06:06:30
2304	SNG	3	59.6	00:07:06	05:35:09
2702	SNG	4	75.8	00:13:06	07:28:57
4201	ICM	4	107.1	00:19:06	05:29:08
2401	SLT	4	69.4	00:25:06	05:47:09
2602	SLT	6	100.5	00:31:06	06:38:34
2301	SNG	3	59.6	00:37:06	06:12:31
2601	SLT	6	100.5	00:43:07	05:17:08
4001	ICM	3	80.6	00:49:07	04:53:08
9401	VIRM	4	108.6	00:55:07	04:59:08
9402	VIRM	4	108.6	01:22:44	07:13:49
2703	SNG	4	75.8	01:43:11	04:24:07
2303	SNG	3	59.6	01:43:11	05:11:08

Appendix D.2: Nijmegen

D.4: Nijmegen scenario 1 stabling demand.

id	Type	Variant	Length [m]	Arrival time	Departure time
4001	ICM	3	80.6	18:54:25	06:05:51
2701	SNG	4	75.8	21:15:49	05:53:08
4201	ICM	4	107.1	22:00:47	05:20:24
2401	SLT	4	69.4	22:06:47	05:26:24
2601	SLT	6	100.5	23:44:19	05:38:24
8601	VIRM	6	162.1	00:20:13	06:15:35
9401	VIRM	4	108.6	00:48:15	05:14:23
2301	SNG	3	59.6	01:13:27	05:32:24

D.5: Nijmegen scenario 2 stabling demand.

id	Type	Variant	Length [m]	Arrival time	Departure time
4001	ICM	3	80.6	19:16:23	06:02:14
9402	VIRM	4	108.6	19:44:20	06:45:14
2302	SNG	3	59.6	20:47:45	06:55:20
2601	SLT	6	100.5	00:06:12	05:36:25
9401	VIRM	4	108.6	00:12:12	05:42:26
4002	ICM	3	80.6	00:29:27	06:39:13
8601	VIRM	6	162.1	00:48:54	07:53:46
2301	SNG	3	59.6	01:01:49	07:07:20
2702	SNG	4	75.8	01:01:49	05:53:31
2401	SLT	4	69.4	01:11:47	07:13:20
2701	SNG	4	75.8	01:17:48	05:16:56
4201	ICM	4	107.1	01:24:13	07:01:20

D.6: Nijmegen scenario 3 stabling demand.

id	Type	Variant	Length [m]	Arrival time	Departure time
2701	SNG	4	75.8	17:56:42	05:08:01
2301	SNG	3	59.6	17:56:42	06:14:08
4201	ICM	4	107.1	18:33:01	06:08:08
4001	ICM	3	80.6	19:52:49	06:02:07
8601	VIRM	6	162.1	20:29:59	05:44:07
4002	ICM	3	80.6	21:47:36	06:26:41
2401	SLT	4	69.4	22:07:15	04:35:02
2702	SNG	4	75.8	23:25:35	06:41:07
9401	VIRM	4	108.6	23:47:43	05:20:01
2302	SNG	3	59.6	00:11:29	06:20:41
9402	VIRM	4	108.6	00:17:29	05:56:07
2402	SLT	4	69.4	00:23:30	05:02:00
2602	SLT	6	100.5	00:23:30	06:48:32
8602	VIRM	6	162.1	01:01:21	05:50:07
2601	SLT	6	100.5	01:12:58	07:14:01
2303	SNG	3	59.6	01:24:55	05:14:01

Appendix D.3: Zwolle

D.7: Zwolle scenario 1 stabling demand.

id	Type	Variant	Length [m]	Arrival time	Departure time
8601	VIRM	6	162.1	20:51:02	05:14:28
2401	SLT	4	69.4	21:04:01	06:04:23
4201	ICM	4	107.1	21:29:32	06:59:55
4001	ICM	3	80.6	23:49:35	07:30:54
2301	SNG	3	59.6	23:57:25	05:38:44
2601	SLT	6	100.5	00:47:27	06:10:24
2302	SNG	3	59.6	00:53:27	07:58:57
9401	VIRM	4	108.6	01:07:38	05:07:16
2701	SNG	4	75.8	02:14:37	05:58:23

D.8: Zwolle scenario 2 stabling demand.

id	Type	Variant	Length [m]	Arrival time	Departure time
2601	SLT	6	100.5	17:34:46	05:29:40
2301	SNG	3	59.6	19:12:50	06:12:27
9401	VIRM	4	108.6	20:06:52	05:35:40
2702	SNG	4	75.8	21:19:35	06:58:15
9402	VIRM	4	108.6	21:31:09	05:11:40
4201	ICM	4	107.1	21:37:09	06:06:26
2401	SLT	4	69.4	21:47:04	05:23:40
4001	ICM	3	80.6	22:04:29	06:20:22
4002	ICM	3	80.6	23:25:17	05:17:40
2302	SNG	3	59.6	23:38:49	06:00:26
8602	VIRM	6	162.1	00:24:44	05:54:26
2701	SNG	4	75.8	00:34:33	06:26:23
8601	VIRM	6	162.1	01:16:17	07:08:35

D.9: Zwolle scenario 3 stabling demand.

id	Type	Variant	Length [m]	Arrival time	Departure time
9402	VIRM	4	108.6	17:50:15	07:57:55
2401	SLT	4	69.4	18:36:59	06:22:43
2702	SNG	4	75.8	19:50:53	05:28:21
4001	ICM	3	80.6	20:54:13	06:34:43
9401	VIRM	4	108.6	22:55:35	05:22:20
2303	SNG	3	59.6	23:28:17	06:52:43
2703	SNG	4	75.8	23:28:17	06:52:43
9403	VIRM	4	108.6	23:47:16	07:17:27
4002	ICM	3	80.6	23:53:16	06:10:43
2301	SNG	3	59.6	00:10:09	06:16:43
2602	SLT	6	100.5	00:16:09	05:46:04
2402	SLT	4	69.4	00:16:09	07:25:28
8602	VIRM	6	162.1	00:22:09	06:04:43
2601	SLT	6	100.5	00:46:40	05:16:20
4201	ICM	4	107.1	01:03:52	06:40:43
2302	SNG	3	59.6	01:10:05	06:28:43
2701	SNG	4	75.8	01:16:05	05:34:21
8601	VIRM	6	162.1	01:33:10	06:46:43

Appendix E: Best & Worst Stabling Plans

Appendix E.1: Carthesiusweg 40%-scenario

E.1: Best performing SD stabling plan Carthesiusweg scenario 1.

Composition	Track	Order	id	Type	Variant	Length	Arrival Time	Departure Time
1	261	1	2601	SLT	6	100.5	18:16:08	05:29:31
2	264	1	4001	ICM	3	80.6	19:17:08	05:47:32
3	264	2	4201	ICM	4	107.1	19:32:38	05:15:50
4	265	1	9402	VIRM	4	108.6	21:23:36	05:41:32
5	266	1	9401	VIRM	4	108.6	21:47:12	04:53:07
6	267	1	8601	VIRM	6	162.1	23:39:52	07:41:29
OOS 1	268	1	OOS 1	ICM	3	80.6	-	-
7	269	1	2702	SNG	3 + 4	135.4	23:45:53	06:34:09
8	270	1	2301	SNG	3	59.6	23:45:53	07:01:40
10	270	2	2701	SNG	4	75.8	01:15:17	05:35:31
9	271	1	2401	SLT	4	69.4	23:54:25	06:07:43
11	271	2	2302	SNG	3	59.6	01:21:18	06:01:43

E.2: Worst performing SD stabling plan Carthesiusweg scenario 1.

Composition	Track	Order	id	Type	Variant	Length	Arrival Time	Departure Time
OOS 1	253	1	OOS 1	VIRM	6	162.1	-	-
9	253	2	2401	SLT	4	69.4	23:57:38	06:07:43
4	261	1	9402ca	VIRM	4 + 6	270.7	21:23:36	05:41:32
2	264	1	4001ca	ICM	3 + 4	187.7	19:17:08	05:47:32
OOS 2	265	1	OOS 2	SLT	6	100.5	-	-
8	265	2	2301	SNG	3	59.6	23:45:53	07:01:40
1	266	1	2601	SLT	6	100.5	18:16:08	05:29:31
6	267	1	8601	VIRM	6	162.1	23:39:52	07:41:29
3	268	1	4201	ICM	4	107.1	19:32:38	05:15:50
5	269	1	9401	VIRM	4	108.6	21:47:12	05:55:43
10	270	1	2701	SNG	4	75.8	01:15:17	05:35:31
7	271	1	2702	SNG	4	75.8	23:45:53	06:34:09
11	271	2	2302v	SNG	4	75.8	01:21:18	06:01:43

Appendix E.2: Nijmegen 60%-scenario

E.3: Best performing SD stabling plan Nijmegen scenario 2.

Composition	Track	Order	id	Type	Variant	Length	Arrival Time	Departure Time
7	10R	1	8601	VIRM	6	162.1	00:48:54	07:53:46
8	10R	2	2702	SNG	4	75.8	01:01:49	05:53:31
9	11R	1	2301	SNG	3	59.6	01:01:49	07:07:20
11	11R	2	2701	SNG	3	59.6	01:17:48	05:20:13
10	12R	1	2401	SLT	4	69.4	01:11:47	07:13:20
12	12R	2	4201	ICM	4	107.1	01:24:13	07:01:20
OOS 4	1R	1	OOS 4	SNG	3	59.6	-	-
OOS 1	2R	1	OOS 1	SNG	4	75.8	-	-
OOS 2	2R	2	OOS 2	SLT	4	69.4	-	-
OOS 3	2R	3	OOS 3	ICM	3	80.6	-	-
1	3R	1	4001	ICM	3 + 4	187.7	19:16:23	06:02:14
4	4R	1	2601	SLT	4	69.4	00:06:12	05:36:25
2	5R	1	9402	VIRM	4	108.6	19:44:20	06:45:14
5	5R	2	9401	VIRM	4	108.6	00:12:12	05:42:26
3	6R	1	2302	SNG	3	59.6	20:47:45	06:55:20
6	6R	2	4002	ICM	3	80.6	00:29:27	06:39:13

E.4: Worst performing SD stabling plan Nijmegen scenario 2.

Composition	Track	Order	id	Type	Variant	Length	Arrival Time	Departure Time
OOS 1	10R	1	OOS 1	ICM	3	80.6	-	-
9	10R	2	2301	SNG	3	59.6	01:01:49	07:07:20
10	11R	1	2401	SLT	4	69.4	01:11:47	07:13:20
11	11R	2	2701	SNG	4	75.8	01:17:48	05:16:56
OOS 3	12R	1	OOS 3	SLT	6	100.5	-	-
12	12R	2	4201	ICM	4	107.1	01:24:13	07:01:20
1	2R	1	4001	ICM	3 + 3	161.2	19:16:23	06:02:14
2	3R	1	9402	VIRM	4	108.6	19:44:20	06:45:14
4	3R	2	2601	SLT	6	100.5	00:06:12	05:36:25
OOS 2	4R	1	OOS 2	SNG	3	59.6	-	-
5	4R	2	9401	VIRM	4	108.6	00:12:12	05:42:26
3	5R	1	2302	SNG	3	59.6	20:47:45	06:55:20
6	5R	2	4002	ICM	3	80.6	00:29:27	06:39:13
7	6R	1	8601	VIRM	6	162.1	00:48:54	07:53:46
8	6R	2	2702	SNG	4	75.8	01:01:49	05:53:31

Appendix E.3: Zwolle 60%-scenario

E.5: Best performing SD stabling plan Zwolle scenario 2.

Composition	Track	Order	id	Type	Variant	Length	Arrival Time	Departure Time
OOS 1	411	1	OOS 1	VIRM	4	108.6	-	-
OOS 2	411	2	OOS 2	SNG	4	75.8	-	-
1	412	1	2601	SLT	6	100.5	17:34:46	05:29:40
3	413	1	9401	VIRM	4	108.6	20:07:53	05:35:40
5	413	2	9402	VIRM	4	108.6	21:31:09	05:11:40
2	414	1	2301	SNG	3	59.6	19:12:50	06:12:27
6	414	2	4201	ICM	4	107.1	21:37:09	06:06:26
7	414	3	2401v	SLT	6	100.5	21:47:04	05:23:40
9	414	4	4002	ICM	3	80.6	23:25:17	05:17:40
8	430	1	4001	ICM	4	107.1	22:04:29	06:20:22
10	430	2	2302	SNG	3	59.6	23:38:49	06:00:26
11	430	3	8602	VIRM	6	162.1	00:24:44	05:54:26
4	431	1	2702	SNG	4	75.8	21:19:35	06:58:15
12	431	2	2701	SNG	4	75.8	00:34:33	06:26:23
OOS 3	432	1	OOS 3	SNG	4	75.8	-	-
13	432	2	8601	VIRM	6	162.1	01:16:17	07:08:35

E.6: Worst performing SD stabling plan Zwolle scenario 2.

Composition	Track	Order	id	Type	Variant	Length	Arrival Time	Departure Time
OOS 6	411	1	OOS 6	VIRM	4	108.6	-	-
4	411	2	2401	SLT	4	69.4	20:21:01	06:11:30
8	411	3	4201	ICM	4	107.1	22:16:16	06:06:26
13	412	1	8601	VIRM	6	162.1	01:16:17	07:08:35
OOS 1	413	1	OOS 1	VIRM	4	108.6	-	-
OOS 2	413	2	OOS 2	VIRM	4	108.6	-	-
6	413	3	9402	VIRM	4	108.6	21:31:09	05:45:10
OOS 5	414	1	OOS 5	SNG	3	59.6	-	-
12	414	2	2701	SNG	4	75.8	00:34:33	06:26:23
OOS 4	430	1	OOS 4	ICM	3	80.6	-	-
2	430	2	2301	SNG	3	59.6	19:12:50	06:12:27
3	430	3	9401	VIRM	4	108.6	20:06:52	05:35:40
1	431	1	2601	SLT	6	100.5	17:34:46	05:29:40
9	431	2	4002	ICM	3	80.6	23:25:17	05:17:40
OOS 3	432	1	OOS 3	SNG	4	75.8	-	-
5	432	2	2702	SNG	4	75.8	21:19:35	06:58:15
7	432	3	4001	ICM	3	80.6	22:04:29	06:20:22
10	432	4	2302	SNG	3	59.6	23:38:49	06:00:26
11	432	5	8602	VIRM	6	162.1	00:24:44	05:54:26

Appendix F: Sensitivity Analysis Results

F.1: Sensitivity analysis metric values.

Metric	Scenario	Standard	Variant+	Variant-	Type+	Type-	Unit+	Unit-	Time+	Time-
Viability	ctw40	99.9	99.9	99.9	99.9	99.9	99.9	100.0	99.9	100.0
	ctw60	99.8	99.9	99.8	99.8	99.9	99.7	99.8	99.8	99.9
	ctw80	99.6	99.5	99.7	99.8	99.5	99.6	99.7	99.7	99.7
	nm40	100.0	100.0	100.0	99.9	99.9	99.9	99.9	100.0	100.0
	nm60	99.9	99.9	100.0	99.9	100.0	99.9	99.9	99.9	99.9
	nm80	99.4	99.6	99.3	99.3	99.3	99.2	99.4	99.4	99.3
	zl40	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	zl60	100.0	99.9	99.9	100.0	99.9	99.9	99.9	100.0	99.9
	zl80	100.0	100.0	99.9	99.9	100.0	99.9	100.0	100.0	99.9
Offset	ctw40	0.91	0.90	0.89	0.89	0.90	0.93	0.88	0.92	0.90
	ctw60	1.60	1.62	1.58	1.61	1.59	1.62	1.58	1.63	1.57
	ctw80	2.24	2.25	2.22	2.23	2.23	2.28	2.16	2.22	2.22
	nm40	1.22	1.22	1.22	1.20	1.21	1.23	1.20	1.23	1.21
	nm60	1.02	1.03	1.03	1.03	1.02	1.04	1.00	1.05	1.01
	nm80	1.99	2.00	2.02	1.99	2.03	2.04	1.94	2.02	2.00
	zl40	0.52	0.52	0.52	0.53	0.52	0.53	0.52	0.56	0.50
	zl60	0.79	0.80	0.81	0.81	0.80	0.81	0.79	0.84	0.77
	zl80	1.55	1.53	1.56	1.54	1.53	1.56	1.53	1.59	1.49
BDu Robustness	ctw40	91.3	97.9	97.8	99.0	90.9	91.6	90.4	91.8	91.2
	ctw60	98.4	98.5	98.0	98.5	98.3	98.6	99.4	98.5	98.9
	ctw80	87.0	87.2	87.9	88.2	87.6	88.6	86.2	89.1	86.7
	nm40	84.1	83.2	83.3	84.3	83.2	84.3	83.2	85.0	83.0
	nm60	87.4	87.5	87.6	86.7	88.5	88.5	87.2	90.4	88.3
	nm80	86.2	86.4	86.2	85.9	87.0	87.8	85.8	88.6	85.9
	zl40	51.1	53.6	54.5	51.3	53.9	55.1	52.5	52.8	49.4
	zl60	91.4	90.0	89.6	90.0	91.5	90.7	90.3	92.2	88.1
	zl80	69.2	68.3	69.8	67.5	67.7	70.6	67.7	70.3	66.7

F.2: Sensitivity analysis percentage change in metric values.

Metric	Scenario	Variant+	Variant-	Type+	Type-	Unit+	Unit-	Time+	Time-
Viability	ctw40	0.0	0.0	0.0	-0.1	-0.1	0.0	0.0	0.0
	ctw60	0.1	0.0	0.0	0.1	-0.1	0.0	0.0	0.1
	ctw80	-0.1	0.1	0.2	-0.1	0.1	0.1	0.2	0.1
	nm40	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	nm60	0.1	0.1	0.1	0.1	0.0	0.1	0.0	0.1
	nm80	0.2	-0.1	-0.1	-0.1	-0.2	0.0	0.0	-0.1
	zl40	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	zl60	-0.1	-0.1	0.0	-0.1	-0.1	-0.1	0.0	-0.1
	zl80	0.0	-0.1	0.0	0.0	-0.1	0.0	0.0	-0.1
Offset	ctw40	-1.1	-2.2	-2.2	-1.1	2.2	-3.3	1.1	-1.1
	ctw60	1.3	-1.3	0.6	-0.6	1.3	-1.3	1.9	-1.9
	ctw80	0.5	-0.9	-0.5	-0.5	1.8	-3.6	-0.9	-0.9
	nm40	0.0	0.0	-1.6	-0.8	0.8	-1.6	0.8	-0.8
	nm60	1.0	1.0	1.0	0.0	2.0	-2.0	2.9	-1.0
	nm80	0.5	1.5	0.0	2.0	2.5	-2.5	1.5	0.5
	zl40	0.0	0.0	1.9	0.0	1.9	0.0	7.7	-3.9
	zl60	1.3	2.5	2.5	1.3	2.5	0.0	6.3	-2.5
	zl80	-1.3	0.7	-0.7	-1.3	0.7	-1.3	2.6	-3.9
BDu Robustness	ctw40	7.2	7.1	8.3	-0.5	0.3	-1.0	0.5	-0.2
	ctw60	0.1	-0.3	0.1	-0.1	0.2	1.0	0.1	0.5
	ctw80	0.3	1.0	1.4	0.7	1.8	-0.9	2.4	-0.3
	nm40	-1.0	-1.0	0.3	-1.1	0.2	-1.1	1.0	-1.3
	nm60	0.1	0.2	-0.8	1.3	1.3	-0.2	3.4	1.1
	nm80	0.2	0.0	-0.4	0.8	1.8	-0.6	2.7	-0.4
	zl40	4.9	6.7	0.5	5.6	7.9	2.9	3.4	-3.2
	zl60	-1.6	-2.0	-1.6	0.2	-0.8	-1.2	0.9	-3.6
	zl80	-1.3	0.8	-2.6	-2.2	1.9	-2.2	1.6	-3.6