Master thesis

Matching capacity with the predicted workload at the emergency call centre of ProRail

ProRail B.V.

Author:
H.B. Van Dijk (Arne)

Supervisors University of Twente:
Dr. E. Topan
Dr. Ir. E.A. Lalla

Supervisors ProRail:
A. Rosenboom
E. Hueting

10/05/2022

ProRail

UNIVERSITY OF TWENTE.

Colophon

Student
H.B. Van Dijk (Arne)
Industrial Engineering and Management
Planning, Logistics and Management specialization
Faculty of behavioral, management and social sciences

University
University of Twente
Drienerlolaan 5
7522 NB Enschede
The Netherlands

External Organization
ProRail BV
Moreelsepark 3
3511 EP Utrecht

Date of Publishing
10-5-2022

Preface

This thesis marks the end of my time as a student Industrial Engineering & Management at the University of Twente. I would like to thank several persons who contributed to this thesis.

First, I want to thank my company supervisors Anne Rosenboom and Edward Hueting, with whom I met many times for this research. I would like to thank both for always making time to meet me and discuss the latest changes in this research. They were always willing to help and quickly replied to any of my questions. In our meetings there was always time to discuss non-work-related matters, which I greatly appreciated. Furthermore, I want to thank all other colleagues at the MKS, with whom I enjoyed working. A special thanks goes to Saskia Wevers, who made it possible for me to execute this assignment at ProRail. She has also increased the quality of this research with her critical questions and comments.

Furthermore, I would like to thank my supervisors of the University of Twente: Engin Topan and Eduardo Lalla. I enjoyed working with Engin from the beginning. He always had good constructive feedback while he remained positive about the progress of this research. Furthermore, he always responded to my questions quickly and extensively. Engin has put a lot of time and effort in his guidance and that is something that I really appreciate. Eduardo has been an excellent second supervisor. His advice guided my research into the right direction at a crucial point. Besides, it was always enjoyable meeting with both of you.

I hope you enjoy reading this report.

Arne van Dijk
Genemuiden, May 2022

Management Summary

Problem Context

This research is conducted at the Meldkamer Spoor (MKS) of ProRail in The Netherlands. The MKS is an emergency call centre that serves as the central figure for communication of rail-related incidents. The management of the MKS often faces disruptions in the capacity. These disruptions are both short-term, for example due to sickness and long-term, for example due to retirements. Furthermore, the workload at the department is volatile and responsive. This means that the department cannot decide exactly when and how much work is carried out since this depends on the incidents that occur. Currently, the management lacks insight about the allocation of the workload over time (e.g., different hours in a day, different days in a week, different months in a year) and therefore there is no solid quantitative base to support decisions concerning capacity deployment. ProRail aspires to obtain more insight into the characteristics of the workload to better match the capacity and the workload at the MKS. The match between the capacity and the workload is evaluated by three KPIs: the undercapacity, overcapacity and service level. The aim is to minimize both the undercapacity (workload > capacity) and the overcapacity (workload < capacity), while meeting a predefined service level (percentage of hours where the capacity is greater than the service level). The main question that this research aims to answer is:

"How can ProRail estimate the workload at the MKS and deploy the necessary capacity in the most efficient and effective way and thus reduce both undercapacity and overcapacity while meeting the service level"

Methodology

In this research, the workload is estimated for every hour from 1-1-2018 to 30-9-2021, based on historical data about the workload factors. Frequencies for the tasks executed are already known. Durations of telephone calls are known as well, because these are registered. However, task times for administrative tasks are not known and therefore a work sampling study is conducted to determine these task times. This study is based on the method of Groover (2013). Then, the frequencies of tasks and the task times were combined in this study to estimate the workload for every hour in the aforementioned period. This sample with hours and their workload estimates was used as input for the scheduling model.

An MILP model is formulated using the elastic set-covering model, based on the article of Eveborn & Rönnqvist (2004). This model determines when shifts should start and how many shifts should start to meet the service level, while minimizing both under- and overcapacity. The under- and overcapacity are penalized with different weights. The model decides when and how much capacity is deployed. This way, the model effectively builds a schedule. Several model extensions were introduced to better match the model with the requirements, by including a service level constraint and constraints that ensure that different days have similar schedules. With this model 24 experiments are run that provide the model with different levels of freedom. For the experiments, the data is split up in a training and a test sample. Here, a schedule is created based on the training sample and evaluated on the test sample. The training sample and test sample are equal in size and seasonal factors, but they are non-overlapping. Furthermore, the data is split up in two seasons. The first season is the summer, where the average workload is higher. This season runs from June – September. The second season contains the data of the rest of the year, thus from October – May. This second season is called the winter season in this research.

Additionally, an analysis is executed that provides a practical roadmap for the decision maker about how capacity can best be deployed in case of disruptions. In this analysis, the optimal start times for regular shifts and day shifts are compared and the impact of omitting specific shifts from the schedule is reviewed.

Results

The work sampling study showed that only approximately 25% of the time was spent on the core dispatcher tasks, which comprise of the communication and administration of rail-related incidents and infrastructure malfunctions. Another 25% of the time was spent on other activities, such as shift evaluations and ancillary tasks and around 50% of the time was idle time. This relatively high portion of idle time indicates that there is room to improve the schedule while attaining the service level.

The analysis of the workload estimation showed that there is a positive trend in the workload, meaning that the average workload has grown over the past four years. Furthermore, it became clear that there are strong seasonal influences throughout the day, week, and year. Based on these results, we decided to create separate schedules for the summer and the winter period.

The results from the experiments with the scheduling model showed that it is possible to attain a 98% service level with on average less capacity than currently is deployed. For the summer, the model deployed a capacity of 57 shifts in the experiments where the schedule must meet practical requirements that are like those of the existing schedule, while only 51 shifts were deployed for the winter schedule. This is a significant difference compared to the regular schedule, where 56 shifts are deployed each week in both seasons. The model shifted the most capacity away from the Saturdays and Sundays.

Conclusion, limitations & recommendations

Our major finding in this research is that ProRail can reduce the overcapacity at the MKS by deploying less shifts per week than the company currently does. In particular during the winter, the available capacity can be decreased significantly. This can be particularly done by shifting the most capacity away from the Saturday and Sundays. The first place to start when decreasing the capacity is during the weekend, but during the winter, the capacity can also be decreased on most weekdays.

The model is sensitive to changes in the service level and the inflation factor that is used to inflate standard task times to normal task times. This inflation factor ensures that there is some time reserved for personal time, fatigue, and delay. This sensitivity to changes in these factors decreases the robustness of the model solutions.

For both the summer and the winter sample, one schedule was obtained from the experiments that performed better than the others. These experiments are recommended to ProRail, since they ensure that the required service level is attained throughout the year, while reducing the average capacity that is deployed. During the summer, 57 shifts are deployed and during the winter, 51 shifts are deployed. Since the winter period is two times as long as the summer period, this equates to a reduction of 3 shifts per year, which is 0.667 FTE. On the next page, the recommended summer and winter schedule are portrayed.

Figure 1: Recommended winter schedule

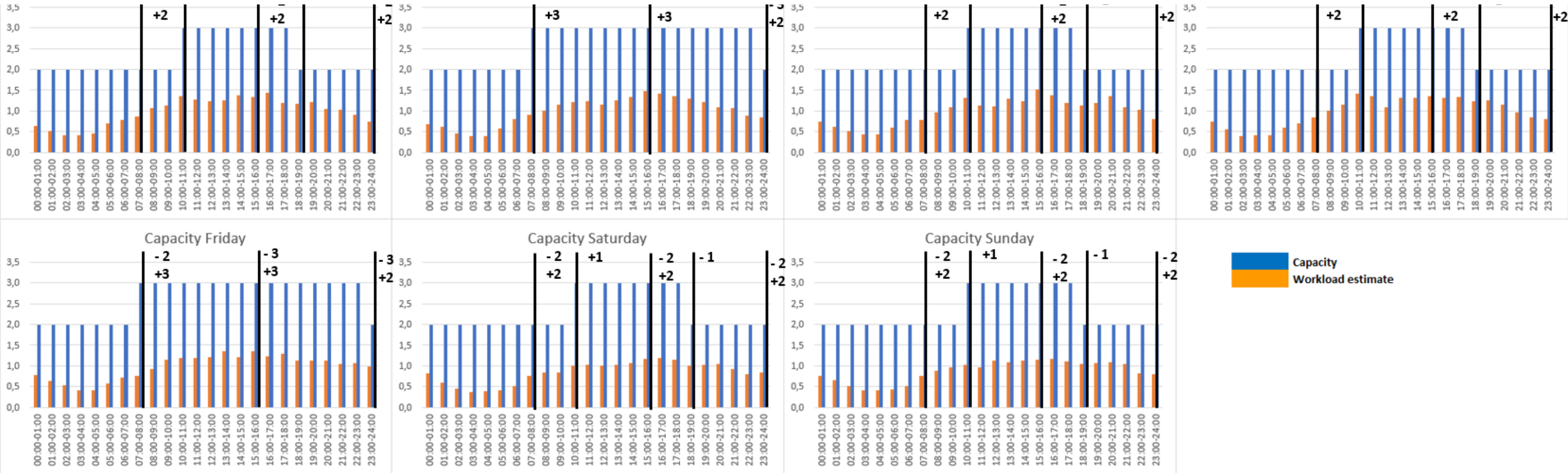


Figure 2: Recommended summer schedule

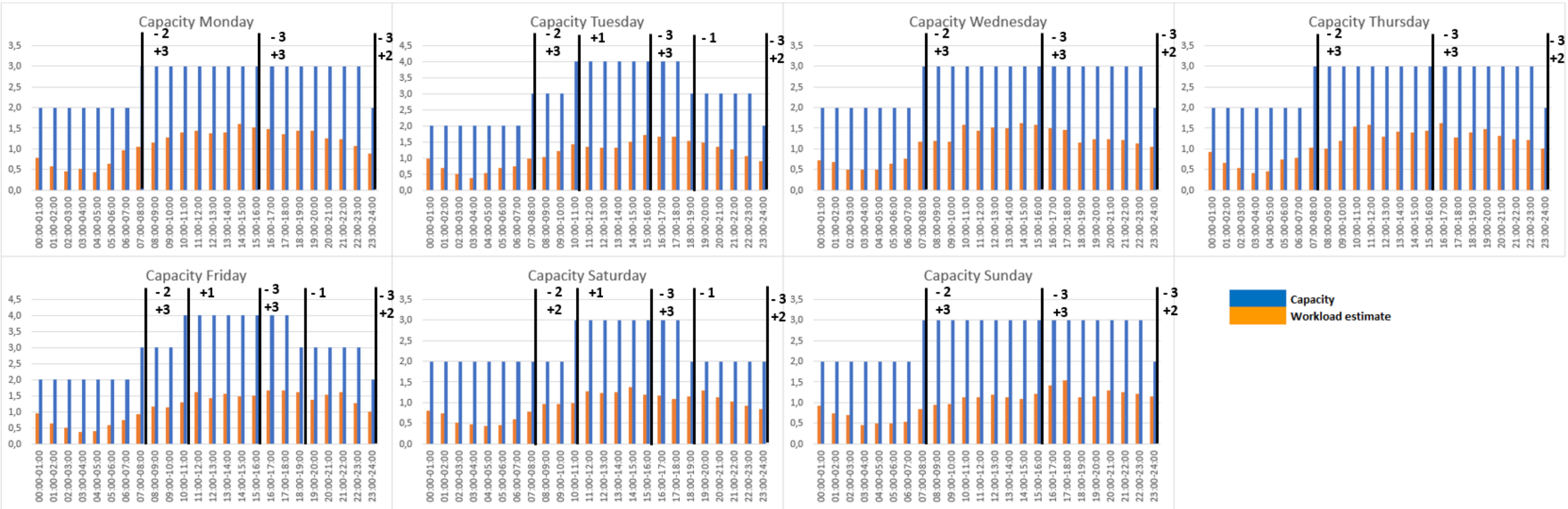


Table of contents

Colophon ii
Preface iii
Management Summary iv
- Problem Context iv
- Methodology iv
- Results v
- Conclusion, limitations & recommendations v

List of Figures xii
List of tables xiii
List of Abbreviations xiv
1 Introduction 1
- 1.1 About ProRail 1
- 1.2 About MKS 1
- 1.3 Problem description 2
  - 1.3.1 Assignment 2
  - 1.3.2 Problem context 2
  - 1.3.3 Planning and scheduling levels 3
  - 1.3.4 Problem cluster 3
  - 1.3.5 Core problem 4
- 1.4 Research approach 4
  - 1.4.1 Research objective 4
  - 1.4.2 Research scope 4
  - 1.4.3 Research questions 5
  - 1.4.4 Deliverables 7

2 Current system analysis 8
- 2.1 Handling incidents 8
  - 2.1.1 Intake 8
  - 2.1.2 Processing 9
  - 2.1.3 Follow-up notifications 10
- 2.2 Types of communication 11
  - 2.2.1 Telephone communication 11
  - 2.2.2 Transceiver communication 11
  - 2.2.3 Communication via digital forms 12
- 2.3 Incident analysis 13
  - 2.3.1 Frequency of incidents 13

2.3.2 Categorization of incidents 16
2.3.3 Workload per incident type 16
2.4 Schedule 17
2.4.1 Shift start and end times 18
2.5 Performance of the current schedule 18
2.5.1 Current KPIs 18
2.5.2 Proposed KPIs 18
2.6 Problems and changes in current structure 19
2.6.1 Limitations of current structure 20
2.7 Results from survey MKS 20
2.7.1 Workload experienced by dispatchers 20
2.7.2 Checking assumptions 20
2.7.3 Providing missing knowledge 21
2.7.4 Determining workload per incident 21
2.7.5 Factors contributing to a high workload 21
2.7.6 Solutions for lowering the work pressure 21
2.8 Summary 22
3 Literature review 24
3.1 Workload studies at ProRail 24
3.2 Workload estimation methods 24
3.2.1 Estimation based on expert knowledge 25
3.2.2 Historical records 25
3.2.3 Work measurement techniques 25
3.2.4 Direct workload measures 27
3.3 Stochasticity 27
3.4 Scheduling 27
3.4.1 Staffing Problem 28
3.4.2 Shift scheduling 28
3.4.3 Shift rostering 28
3.5 Function differentiation 29
3.6 Summary of the literature review 29
3.6.1 Conclusions of previous workload studies at ProRail 29
3.6.2 Conclusions of workload estimation methods 30
3.6.3 Conclusions of modelling stochasticity 30
3.6.4 Conclusions of scheduling 30
3.6.5 Conclusions of function differentiation 31

4 Solution design 32
4.1 Methodology 32
4.2 Workload estimation model 33
4.2.1 Size of the time buckets 33
4.2.2 Categories for different tasks 34
4.2.3 Mathematical formulation of workload estimation model 35
4.3 Determining task times RVOs and incidents 36
4.3.1 Study objective 37
4.3.2 Subjects 37
4.3.3 Output measures 37
4.3.4 Activity categories 37
4.3.5 Study design 37
4.3.6 Identify observers 39
4.3.7 Final steps of the work sampling study 39
4.4 Scheduling model 39
4.4.1 Model introduction 39
4.4.2 Mathematical model formulation 40
4.4.3 Explanation constraints 41
4.4.4 Extending the weekly schedule to all periods 42
4.4.5 Training data and test data 43
4.4.6 Model assumptions 43
4.5 Model extensions 44
4.5.1 Linked start times 44
4.5.2 Fixed total number of start times 45
4.5.3 Selecting one day shift 45
4.5.4 Fixed total capacity 46
4.5.5 Determining capacity for part of the week 47
4.6 Solution space and solving approach 47
4.7 Conclusions 47
5 Experiments 49
5.1 Results work sampling study 49
5.1.1 Observations 49
5.1.2 Measured proportions 49
5.1.3 Analysis work sampling study 50
5.1.4 Adjustment of task times 50
5.1.5 Transformation from normal times to standard times 52

5.1.6 Role of other activities in the model 52
5.2 Data preparation 53
5.2.1 Outlier detection 54
5.2.2 Trend analysis 54
5.2.3 Seasonality 56
5.3 Experiments 59
5.3.1 Experimental setup 59
5.3.2 Results of the experiments 61
5.3.3 Service level per week 62
5.3.4 Analysis of results 64
5.3.5 Modelling decisions 66
5.4 Recommended schedules 70
5.4.1 Comparison in performance 70
5.4.2 FTE reduction 70
5.4.3 Schedule explanation 71
5.5 Experiments for practical decision making 74
5.5.1 Optimal Start times 74
5.5.2 Optimal day shifts 75
5.5.3 Impact of omitting shifts 76
5.5.4 Break opportunities 79
5.6 Sensitivity Analysis 82
5.6.1 Impact of the service level 82
5.6.2 Impact of the inflation factor 83
5.6.3 Impact of weights for under- and overcapacity 83
5.6.4 Impact of processing RVOs during the night 84
5.7 Conclusions 85
6 Conclusion & Discussion 87
6.1 Conclusions 87
6.2 Discussion 88
6.2.1 Validity 88
6.2.2 Interpreting the results 89
6.2.3 Limitations 89
6.2.4 Suggestions for further research 90
6.3 Recommendations 91
7 References 93
Appendices 96

Appendix A 96
Appendix B 99
Appendix C 100
Appendix D 104
Appendix E 109
Appendix F 110
Appendix G 114

List of Figures

Figure 1: Recommended winter schedule ... vi
Figure 2: Recommended summer schedule ... vi
Figure 3: Problem cluster explanation of this research for ProRail ... 4
Figure 4: Flowchart work process MKS ... 8
Figure 5: Average number of telephone calls/hour (Jan. ‘18 – Nov. ‘21, source: MobiRail) ... 11
Figure 6: transceiver voice requests to MKS/hour (July ‘21 – Nov. ’21, source: SpoorWeb) ... 12
Figure 7: Number of RVOs created per hour (January 2018 – November 2021) derived from SAP ... 13
Figure 8: Incidents per month (July ‘17 – August ’21, source: SpoorWeb) ... 14
Figure 9: Histogram of incidents/day, with outlier bins (Jan. ‘18 – Nov. ’21, source: Spoorweb) ... 14
Figure 10: Average number of incidents per hour (Jan. ‘18 – Nov. ’21, source: SpoorWeb) ... 15
Figure 11: Boxplots of incidents per hour, excl. outliers (Jan. ‘18 – Nov. ’21, source: SpoorWeb) ... 15
Figure 12: Percentage of incidents per cause ... 16
Figure 13: Number of log lines per incident type (January 2018 – November 2021) ... 17
Figure 14: Average service level per hour (January 2018 – November 2021) ... 19
Figure 15: Classification of methods to determine time standards (Groover, 2013) ... 25
Figure 16: Flowchart diagram of the research methodology ... 33
Figure 17: Pie chart of results work sampling study ... 50
Figure 18: Proportion of time spend per category (including details about other activities) ... 53
Figure 19: Bar chart of the workload growth per workload factor (hours extra work per day) ... 55
Figure 20: Workload estimate per month and trend (January ‘18 – November ‘21) ... 56
Figure 21: Workload estimate per month de-trended (January '18 - November '21) ... 56
Figure 22: Hour seasonality of workload components (normalized) ... 57
Figure 23: Day seasonality of workload components (normalized) ... 58
Figure 24: Month seasonality of workload components (normalized) ... 58
Figure 25: Box whisker plots of weekly SL realizations exp. 1 – 6 (Summer 1) ... 62
Figure 26: Box whisker plots of weekly SL realizations exp. 7 – 12 (Summer 2) ... 63
Figure 27: Box whisker plots of weekly SL realizations exp. 13 – 18 (Winter 1) ... 63
Figure 28: Box whisker plots of weekly SL realizations exp. 19 - 24 (Winter 2) ... 64
Figure 29:Average capacity assigned per day by the scheduling model (summer) ... 66
Figure 30:Average capacity assigned per day by the scheduling model (winter) ... 66
Figure 31: Average capacity assigned per hour by the scheduling model (summer) ... 66
Figure 32: Average capacity assigned per hour by the scheduling model (winter) ... 66
Figure 33: Average workload estimate and capacity deployment regular schedule ... 68
Figure 34: Average workload estimate and capacity deployment result of experiment 17 (winter 1) 69
Figure 35: Capacity allocation for the recommended summer schedule ... 72
Figure 36: Capacity allocation for the recommended winter schedule ... 73
Figure 37: Illustration of the use of a day shift when a drop in capacity cannot be covered ... 75
Figure 38: Impact of replacing two shifts by a day shift per day (summer) ... 79
Figure 39: Impact of replacing two shifts by a day shift per day (winter) ... 79
Figure 40: Impact of removing two shifts (per day, summer) ... 79
Figure 41: Impact of removing two shifts (per day, winter) ... 79
Figure 42: Copy of the recommended summer schedule ... 91
Figure 43: Copy of the recommended winter schedule ... 92
Figure 44: Bar Charts of Q1-4 of questionnaire MKS ... 104
Figure 45: Bar charts of Q5-7 & 13 of questionnaire MKS ... 105

List of tables

Table 1: Research methodology and application 5
Table 2: Actions per incident category 9
Table 3: Example of determining the workload per hour 34
Table 4: Categories and corresponding numbers used in work sampling study 37
Table 5: Parameters chosen for work sampling study 38
Table 6: Observations required for each category 38
Table 7: Segment of the values of a_{ip} 42
Table 8: Measured proportion per workload category 49
Table 9: Calculation of task times 51
Table 10: Coefficient of variation of seasonal factors per workload category 57
Table 11: Characteristics per season used in the experiments 59
Table 12: Set-up and input all experiments 60
Table 13: Results of all 24 experiments 61
Table 14: Instances where a capacity of three workers is needed for sample W.1 67
Table 15: Performance current schedule over the year 70
Table 16: Performance of proposed schedules over the year 70
Table 17: Parameters for cost calculation 71
Table 18: Calculations for reduction in FTE and costs 71
Table 19: Results experiments for practical decisions (optimal Start Times) 74
Table 20:Results experiments for practical decisions (optimal day shifts) 75
Table 21: Comparison between moving capacity to a day shift and keeping it the same 76
Table 22: Results experiments for practical decisions (impact of omitting shifts) 77
Table 23: Results of experiments for evaluating break opportunities 80
Table 24: Cost savings resulting from break opportunities (lower bound) 81
Table 25: Cost savings resulting from break opportunities (upper bound) 81
Table 26: Experiments sensitivity analysis (Service level) 82
Table 27: Experiments sensitivity analysis (Inflation factor) 83
Table 28: Experiments sensitivity analysis (Weights for under- and overcapacity) 83
Table 29: Experiments sensitivity analysis (processing RVOs during the night) 84
Table 30: TIS-matrix (ProRail, 2020) 96
Table 31: Percentage of incidents per TIS-code 99
Table 32: Categorized answers to Q8 of questionnaire MKS 105
Table 33: Categorized answers to Q9 of questionnaire MKS 106
Table 34: Categorized answers to Q10 of questionnaire MKS 106
Table 35: Categorized answers to Q11 of questionnaire MKS 107
Table 36: Categorized answers to Q12 of questionnaire MKS 107
Table 37: Average workload per incident-label from questionnaire MKS 108
Table 38: Allowed start times scenario 1 109
Table 39: Schedules generated in all 24 experiments 110
Table 40: Schedules generated in sensitivity analysis experiments 114

List of Abbreviations

In the rail branch, it is very common to use abbreviations. These abbreviations might confuse the reader; therefore, each abbreviation is elaborated the first time it is used. Furthermore, the reader can use the list underneath as a reference.

AKI: Afhandeling Kleine Incidenten
An AKI-dossier is the administrative processing of specific incidents without structural delay. With these incidents, ICB-teams do not have to go on site and no ALs are alarmed. The incident receives the TIS-code 1.1, so that a prognosis can be issued.

AL: Algemeen Leider
An 'Algemeen Leider' (general leader) is responsible for the of a team of incident emergency responders (see ICB). An AL is alarmed for incidents with a TIS-code of 1.1 and higher (except for AKI-dossiers). An AL is responsible for the coordination of an incident.

CaTo: Cameratoezicht
CaTo actively monitors cameras at crossings (and other locations) where many incidents happen in order to assist other departments and suicides.

CHI: Coördinator Herstel Infra
The CHI department is responsible for assigning contractors to repair defects and ensuring that the contracts are followed accurately. They are responsible for resolving infra defects.

DVL: Decentrale verkeersleider
A DVL is responsible for the logistics of a specific section of the rail network.

EMTP: Eerste Mens Ter Plaatse
The EMTP is the first person of ProRail ICB to arrive at the location of the incident.

ICB: Incidentenbestrijding
ICB is the department that is responsible for resolving the incidents. ICB-teams are teams that resolve incidents in the field. ICB-teams consist of incident emergency responders. A regular team is built up as follows: one squad-leader, one driver and two regular team members.

MKS: Meldkamer Spoor
The MKS is the department of ProRail that is the central figure in the communication about incidents. In this research, we call the employees at this department dispatchers. They take in calls and inform the relevant stakeholders in each situation.

OBI: Operationeel Besturingscentrum Infra
The OBI department monitors and operates the high-voltage installations, such as the overhead line equipment.

OCCR: Operationeel Controle Centrum Rail
The OCCR is the operational control centre rail. It is the location where all operational departments of ProRail are based. Furthermore, employees who lead the operations of NS are present at this location as well to ensure a smooth cooperation.

OVD-I: Officier Van Dienst Incidentenbestrijding

An OVD-I is the operational manager of the ICB-department. He or she is ultimately responsible for all incidents throughout the Netherlands. When there is a disagreement within IB, the OVD-I takes the final decision.

PuVo: Publieksvoorlichting
The PuVo-department receives non-urgent calls from regular citizens with questions and remarks about the rail. These calls come from the so-called PuVo-line. After office hours an during weekends, the line is forwarded to the MKS.

RVO: Rapport van onregelmatigheid
An RVO is a digital form that is used to report a defect and to provide instructions to a contractor. In an RVO the priority for the contractor is indicated as well.

TIS: Treinincident Scenario
The system of TIS-codes defines the type of incident and the work instruction that corresponds with this type of incident. The software used at the MKS automatically generates certain tasks based on the TIS-code that is entered. An explanation per TIS-code is provided in Table 30*.*

TRDL: Treindienstleider
A TRDL is a rail traffic controller (American: train dispatcher) and is responsible for the safety of train traffic in his or her domain. The TRDL communicates with train drivers, and actively guides train travel in special situations.

1 Introduction

In this chapter, the goal of this research, the core problem and the research design are introduced. Prior to all this, the company where the research is conducted, ProRail, and the department where this research is focused on, the control room rail (MKS), are introduced.

1.1 About ProRail

ProRail is a railway managing company in the Netherlands. They are responsible for the maintenance, renewal, expansion, and safety of the Dutch railway network. They also arrange all train traffic and build and manage stations (ProRail, 2021).

The Netherlands has one of the world's busiest railway networks. However, due to the covid-19 crisis, the number of travellers on the Dutch rail network plummeted. The average number of travellers from April 2020 onwards was only about 30% of that during the same period in 2019 (ProRail, 2020). In the year before the covid-19 crisis made its impact, every day 1 million people travelled by train and 100.000 tons of goods were transported over the 7000 kilometres of track (ProRail, 2019).

ProRail is convinced that it will see further mobility growth after the corona crisis. Based on, among other things, the number of requested train rides for 2021, long-term plans by train carriers, general sustainability efforts (shift from plane and road to train) and recent insights from Statistics Netherlands, the company still foresees major growth on the rail in the coming years (ProRail, 2020).

The construction and maintenance of the rail infrastructure are not performed by ProRail itself, but by different rail contractors depending on the region. Railway operators for passengers and goods pay a fee to ProRail for the use of railways (ProRail, 2019).

The Dutch railway system connects nearly all major cities in the Netherlands. Most of the tracks, in particular in the Randstad area, are used intensely. Therefore, a disruption in the tour of one train, or on one track can impact the journey of many travellers.

1.2 About MKS

The first point of contact when there is a (potential) disruption, is the Meldkamer Spoor (MKS) at the operational control centre rail (OCCR). The MKS is the emergency call centre of ProRail. It is housed at the OCCR, a large department of ProRail in Utrecht, where stakeholders concerned with the nation-wide operational functioning of the rail network are located. The MKS handles emergency calls (alarms) and calls for other cases that might impact train traffic at an operational level. In this research, employees at the MKS are called dispatchers.

Incidents or dangerous situations at the Dutch railways are reported to the MKS. When the situation needs to be registered, a form is created that generates an open report in SpoorWeb. This method ensures that relevant stakeholders are informed about the incident. If necessary, emergency services, such as the police, fire brigade or ProRail's own emergency responders (the ICB- department) are alerted.

Incidents are registered in Spoorweb, a software application that is further described later in this research. Each incident receives a TIS-code that indicates which follow-up actions must be taken. An explanation of each TIS-code is provided in Appendix A. Furthermore, when a contractor needs to fix an infrastructure-related problem, then an RVO is created by the dispatcher. This is a form in which the problem is administered and in which the contractor receives the instructions he or she needs.

Besides telephone communication, the MKS communicates via transceivers as well. A transceiver is a walkie-talkie on a fixed location. With the transceiver, a dispatcher can communicate conveniently with ProRail's emergency responders that resolve the incidents throughout the Netherlands. More information about the work of dispatchers at the MKS is provided in Section 2.1.

1.3 Problem description

In this section, the problem raised by ProRail is elaborated. The problem context is described textually, and graphically by means of a problem cluster. Furthermore, the core problem is described, and the goal of this research is stated.

1.3.1 Assignment

The workload at the MKS has a volatile nature. It is not uncommon to have half an hour without notifications and consequently receive so many calls that all phone lines are occupied. Furthermore, it is crucial that the MKS is always reachable, since their work impacts the safety of other stakeholder, such as train drivers and rail personnel. Therefore, it is important that peak demand can be met. Currently, the MKS has sometimes issues with meeting peak demand, while the department is overstaffed at other moments. The problem raised by the company is:

"Matching capacity and workload at the MKS during the day has proven to be challenging."

ProRail wishes to gain insight into the structure of the workload, to determine recurring elements and ultimately devise a layout for the long-term schedule. In this schedule, it should be determined when capacity is deployed and how much capacity is deployed at these times. They wish to find a structure of shifts that can be used in general and a scheduling model that optimizes the deployment of staff within this structure for an average week. Long-term in this project means roughly a year. Furthermore, ProRail wishes to flatten the peaks in the workload and gain insight into the optimal staffing in various situations, including situations where the MKS has more tasks and responsibilities.

1.3.2 Problem context

Employees at the MKS have indicated that they often encounter a workload that they experience as unpleasant. Two-thirds of the employees indicated that they experienced this once per week or more, with 22% indicating that they experience this multiple times per shift. Since recruiting and training staff takes time and could be outpaced by the departure of elderly employees, the management of the MKS wants to deploy the existing capacity more efficiently. In this situation, that means that capacity is deployed in such a way that the number of times that dispatchers encounter a workload that is too high is minimized.

The capacity at the MKS is generally structured in a way that there are three dispatchers working during the morning shift, three during the evening shift and two during the night. Furthermore, there is a relief shift on weekdays which means that from 11.00 to 19.00 there is one more dispatcher present. In Section 2.4, more information about the structure of the schedule is provided. In November 2021, 22 full-time employees are employed at the MKS.

Work at the MKS is primarily reactive, the demand for work (workload) per shift is not known upfront. Currently, management of the MKS lacks insight in the distribution of the workload over the days and weeks. The combination of these two factors makes it difficult to properly match the workload to the capacity. Furthermore, it is not clear whether the current schedule is near-optimal or not. Also, ProRail currently has no tool to forecast the workload. Only the rail-focused weather forecast provides some insights into the chances of more incidents occurring. When a weather-code has been issued, for example a substantial chance on high temperatures, the staff at the MKS know that issues related to this weather code might arise during the day.

In conclusion, to match the capacity and the workload better, one needs to discover whether there are seasonal patterns in the workload. Based on the information derived from a workload analysis, an improved long-term schedule should be created. This schedule should provide information about how much capacity is deployed at each moment in the planning horizon. *The optimal schedule should determine when and how much capacity should be deployed, where the capacity is minimized, while the service level is met.* Here, the service level is defined as the percentage of hours where the capacity is sufficient to meet the demand. Furthermore, overcapacity (capacity > workload) and undercapacity (capacity < workload) are kept to a minimum by penalizing both. The knowledge problems encountered in this research are stated in Section 1.4.3, with an explanation about the method that is used to solve the problem.

1.3.3 Planning and scheduling levels

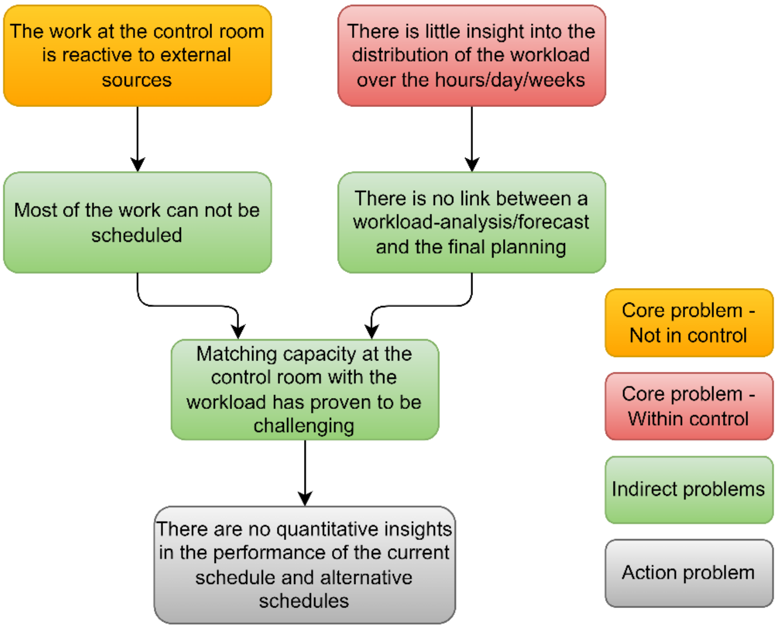
The schedule at the MKS has evolved over the years. The department has grown in responsibility and number of employees. This growth influenced the structure of the schedule. In the schedule, three different levels can be distinguished. Per levels, decision-makers and other influences are described.

- **Tactical**: On a tactical level, the structure of the shifts (e.g., when they start and end, and when there are breaks), the number of FTEs and the division of roles and tasks are decided. The tactical scheduling level is focused on the long term. Other factors that influence the tactical schedule are the CAO (collective work agreement) and national laws. For example, by determining the increment in salary for night shifts, a minimum number of hours between two shifts etc. The tactical schedule decisions are made by the team manager of the MKS.
- **Offline operational:** The offline operational schedule concerns assigning employees to the shifts. This schedule is medium-term. Employees are assigned to shifts almost a year in advance, yet regularly there are changes in the occupation of shifts weeks or days before the shift is planned. Each department at the OCCR has a dedicated planner.
- **Online operational:** The online operational schedule concerns how different tasks are divided over the dispatchers during the shift. When there is a dispatcher in training, he or she handles as much incidents as possible (further described in Section 2.1.1). Furthermore, dispatchers who process the intake of a specific incident, generally process follow-up calls of this incident as well. The dispatchers in operation are together responsible for the online operational schedule.

1.3.4 Problem cluster

In the figure underneath, the relations between the different problems in this research are depicted graphically.

Figure 1: Problem cluster explanation of this research for ProRail



1.3.5 Core problem

The fact that most of the work at the MKS is reactive to external sources cannot be changed. Reactive in this context means that work at the MKS is the result of actions of external stakeholders who cannot be influenced. Therefore, the focus is on the other core problem: to improve the knowledge about the workload. Hereafter, this research focuses on the action problem, by creating a prescriptive model which helps the decision makers by deciding how the capacity should be deployed. The different problems are stated in the problem cluster (see Figure 1). In Section 1.4.3, the research questions that are linked to each subproblem are stated, in combination with the approach that is used to solve the core problem.

1.4 Research approach

In this section, the approach to solve the core problem and achieve the research goal is explained. First, the research objective is defined, and the practical and scientific contributions are outlined. Next, the scope of this research is defined. Hereafter, the sub-questions that break down the core problem are elaborated. In the final section of the research approach, the deliverables are defined.

1.4.1 Research objective

The main objective of this research is to create a prescriptive model that helps the decision maker to deploy capacity more efficiently. In this situation, efficiently means that the capacity is deployed in such a way that the service level is met. This objective could be achieved by improving the schedule in such a way that more dispatchers are present at moments that the workload peaks, or by flattening the peaks in the workload by moving tasks to later moments.

This research aims to help ProRail solve the practical problem of deploying capacity efficiently. While doing that, this research aims to generate knowledge about how an organization can deploy its capacity efficiently in a situation where the demand is not known upfront.

1.4.2 Research scope

This research can be split up in different stages: First, an analysis of the workload is carried out and the workload is estimated for periods in the past based on historical data and expert knowledge. The workload estimation is used to create a (near-) optimal schedule. In this schedule, the capacity is a parameter, not a decision variable. The decision variables are the moments when a shift start and how much capacity is allocated to each shift.

Determining start times of shifts and assigning capacity to each shift are tasks for the scheduling model that are within the scope. However, assigning individual employees to different shifts (rostering) falls outside the scope of this research. To illustrate this with an example, the scheduling model could decide that a capacity of three is required on Monday from 15:00 to 23:00, though it does not decide which employees would work during this shift. That is the task of the planner. The implementation of a new schedule falls outside the scope of this research as well. Furthermore, this research focuses on the rail network in the whole of The Netherlands, since the MKS processes calls for every region in The Netherlands.

1.4.3 Research questions

To solve the core problem and subproblems and achieve the research goal, several research questions are answered. The process to achieve this goal is threefold. First a workload estimation model is created that produces the input for the scheduling model. Hereafter, the scheduling model is created. Finally, the results that are produced by the scheduling model are analysed and validated. Each chapter focuses on one research question. The different research questions are drafted to answer the main research question:

"How can ProRail estimate the workload at the MKS and deploy the necessary capacity in the most efficient and effective way and thus reduce both undercapacity and overcapacity while meeting the service level"

The structure of these questions is based on the first five phases of the managerial problems solving method (MPSM) (Heerkens & Van Winden, 2012). The last two phases: implementing the solution and evaluating the performance of the solution in reality, fall outside the scope of this research. Table 1 provides an overview of this method and its application in this research.

Table 1: Research methodology and application

Methodology (MPSM)		**Research Application**		
Phase	Description	Research Question	Chapter	Description
1	Problem definition	-	1	Introduction
2	Approach	-	1	Introduction
3	Analysing the problem	1	2	Current system analysis
4	Develop alternative models	2 3	3 4	Literature review Solution design
5	Select model and evaluate performance	4	5	Results

1. ***What are the characteristics of the workload and schedule at the MKS?***
   Workload breakdown:
   - Which tasks does the work of dispatcher on a regular day comprise?
   - How are the regular tasks distributed over the day and week?
   - Which incident types can we distinguish and with which frequency do they occur?
   - How much work stems from frequent tasks?
   - When and how often does the workload peak?

Schedule breakdown:

- What does the current schedule look like?
- What influence do weather warnings have on the schedule?

Performance:

- Which problems are encountered in the current schedule?
- How does ProRail currently measure the performance of the schedule?
- Which KPIs could serve as indicators for scheduling performance?

The first research question aims to analyse the current situation at the MKS. First, the structure of the work is analysed. Here, a general overview of the intake process of incidents is provided. Furthermore, an analysis of the different incident types is given.

Hereafter, an analysis of the schedule is provided, and the influence of weather notices is researched. Finally, the performance of the current schedule is analysed. Different KPIs are discussed, even as the bottlenecks of the schedule and how the current schedule functions.

The interviews with dispatchers at the MKS are used to gain more insights into the problems at the department. Furthermore, it is used to provide more information about the workload that results from specific actions and incidents. Thirdly, the interviews could provide possible solutions for reducing the moments at which the MKS cannot meet the demand in terms of workload.

2. ***What is known in the literature about estimating and scheduling reactive work?***
   Data collection
   - Which methods are suggested in scientific literature to estimate the workload of specific tasks?

   Schedule creation and validation
   - Which mathematical models are known that can be used to create a fixed schedule based on an uncertain input for an emergency call centre?
   - Which methods can be used to test the performance of a proposed schedule?

   Function differentiation
   - What is known in the literature about function differentiation at (emergency) call centres?

The second research question aims at finding the relevant literature for this study. There are several challenges in this research, for which scientific literature provides a roadmap. The first challenge concerns finding a good method to estimate the average time that is needed to process RVOs and incidents. In order to do so, previous studies at ProRail that were focused on the workload are reviewed, as well as more general studies that focus on estimating the workload. Furthermore, a literature study must provide answers to the question about how one can create a schedule while the input contains a lot of uncertainty. Lastly, studies that investigate possibilities around function differentiation are reviewed.

3. ***How can we create and validate the workload estimation and scheduling models?***
   - How can we estimate the workload in the past based on historical data?
   - How can we validate the workload predictions?
   - How can we generate and evaluate different schedules based on different criteria and preferences?

When suitable methods to gather the necessary data, estimate the workload and create a schedule are found in literature, the process of implementing and testing these methods starts. First a model that translates data from incidents, telephone calls, transceiver calls and RVOs into an accurate estimate of the workload is created and validated.

Hereafter, a scheduling model is created that minimizes the capacity while meeting the service level constraint. The schedule is based on estimates of the historical workload. The variables in this model are the number of employees per shift and the start- times of shifts.

4. ***What solutions are proposed for different problem scenarios by the scheduling models?***
    - Are all restrictions and desired features included in schedules generated by the scheduling model?
    - What is the performance of the schedule on the selected KPIs?

The fourth research question is focused on the performance of the developed models. The answers to the sub-questions in this chapter form the results of this research. The performance of the scheduling model is examined. Different scenarios are compared, and these results form the basis for the conclusions and recommendations from this research. Large data samples are used for experiments with the start- and end-times of shifts. The schedules are validated using cross-validation.

5. ***How should the results of the performance be interpreted and what consequences would the implementation of the models have?***
    - What conclusions and recommendations stem from the results?
    - What are the consequences of the models for stakeholders (e.g., planners and dispatchers) if the models are implemented?
    - What are the limitations of this research?
    - How does this research contribute to the scientific body of knowledge?

The fifth and final research question aims at interpreting the results. In the chapter that corresponds to this research question, the conclusion, discussion, and recommendations are stated. Here the interests of other stakeholders are included, and the limitations of the research are discussed.

1.4.4 Deliverables

This research intends to deliver the following results:

- An extensive quantitative and qualitative analysis of the work activities of dispatchers at the MKS.
- A model that estimates the perceived workload per dispatcher based on the number of incidents, type of incidents, telephone and transceiver calls and RVOs in a specified period in the past.
- A model that determines the optimal start- and end -times per shift and number of dispatchers per shift that minimizes the capacity needed per week to meet a predefined service level.

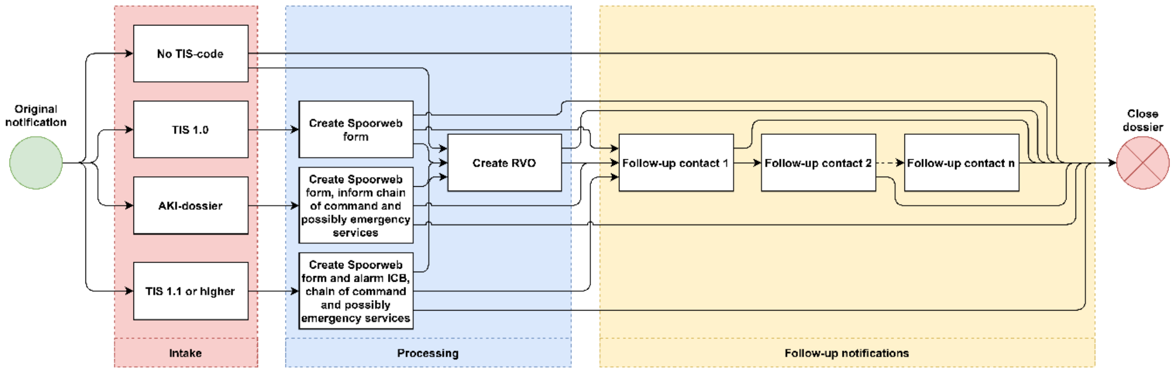
2 Current system analysis

In this chapter, the current situation at the MKS at ProRail is explained in detail. In Section 2.1, the main tasks of the dispatchers are elaborated. Each factor that has a significant contribution to the total amount of work is clarified. In Section 2.2, the different types of communication at the MKS and the seasonal patterns for the different types are explained. In Section 2.3, an analysis of incidents is provided. In Section 2.4, the structure of the schedule is analysed, as well as the influence of weather warnings. In the Section 2.5, the performance of the current schedule and different KPIs are analysed. In Section 2.6, the problems and changes in the current structure are explained. In Section 2.7, a qualitative analysis of the survey that was conducted at the MKS is provided. The chapter ends with a summary.

2.1 Handling incidents

In this section, the general method that is used to process notifications/incidents is explained. The process can broadly be divided into three phases: the intake, processing, and handling follow-up notifications. Each of these four phases is briefly described in the next sections. Furthermore, a graphical summary of the process is depicted in Figure 2.

Figure 2: Flowchart work process MKS



2.1.1 Intake

The intake of an incident concerns the telephone conversation in which a specific incident is made known to the MKS for the first time and the tasks immediately flowing from it. Currently, all dispatchers at the MKS are eligible to take the call. However, when there is a dispatcher available, who is still in the training phase (a relatively new dispatcher), then this person will take the call, while a more experienced dispatcher listens in on the conversation. The goal of the intake is to obtain all relevant information as quickly as possible and determine the essential follow-up actions.

TIS 1.1 or higher: During the intake, the dispatcher determines which type of incident is being reported. The first decision a dispatcher needs to make is whether he or she needs to alarm or not. In general, when the incident hinders train traffic and is expected to cause delay of 5 minutes or more for at least 30 minutes, or travellers are not free in their movement (e.g., when a train is stranded, and passengers cannot leave the train) then the incident receives a TIS-code of 1.1 or higher. Examples of incidents with TIS 1.1 and higher can be found in Appendix A.

AKI-dossier: When an incident does not meet the requirements to be labelled as TIS 1.1 or higher, but the incident has one of the following five causes, then other stakeholders would like to be informed and receive a prognosis and therefore, the incident will be labelled as TIS 1.1.

- Defect materials

- Hindrance caused by emergency services (police or fire brigade)
- Hindrance caused by the behaviour of travellers or personnel
- Hindrance due to the health of travellers or personnel
- Hindrance due to a calamity outside of The Netherlands

It is currently programmed in Spoorweb that a prognosis can only be given if an incident is labelled with a TIS-code of 1.1 or higher. Therefore, when a prognosis is desired, but the incident is from an administrative point of view a TIS 1.0 incident, dispatchers label the incident with TIS 1.1. And that is the case for incident with the five causes mentioned above. An example of an AKI-dossiers is when a defect train is stranded at the platform (so personnel and travellers can move in and out freely). For AKI-dossiers, AL's and ICB-teams normally do not move to the site of the incident.

TIS 1.0: When there is an incident or (possible) disruption of the train service that does not meet the requirements for TIS 1.1 or higher and does not have one of the five causes of AKI-dossiers, then the incident is labelled as TIS 1.0. An example of an incident or (possible) disruption that is labelled TIS 1.0 is a signal malfunction.

No TIS-code: When there is a situation reported that does not impact the train service, the incident does not receive a TIS-code and is thus not registered in Spoorweb. An example of such a situation is when there is a switch defect that has no logistical impact, for which only an RVO is created.

Almost all incidents are reported via a telephone call, however there are two exceptions: In some cases, CaTo detects a dangerous or odd situation and reports this to the MKS directly. They can report it face-to-face, because their office is located next to the office of the MKS. The second exception is when a member of an ICB-team or AL accidently encounters a dangerous or odd situation. ICB-teams are teams that resolve incidents in the field. ICB-teams consist of incident emergency responders. In this case, an incident can be reported via transceiver; the primary mode of communication between the MKS and IB'ers and ALs.

2.1.2 Processing

After the intake, one can separate specific cases for which standard tasks must be completed. These tasks depend on the type of incident and the role of the reporter. Which task needs to be undertaken in which situation is depicted in Figure 2 and is summarized in Table 2. The most important tasks are explained below:

Table 2: Actions per incident category

Category	Certain actions	Possible actions
No TIS-code	n.a.	-Create RVO
TIS 1.0	-Registration in Spoorweb	-Create RVO
AKI-dossier	-Registration in Spoorweb -Inform chain of relevant stakeholders and provide a prognosis	-Create RVO -Contact emergency services
TIS 1.1 and higher	-Registration in Spoorweb -Inform chain of relevant stakeholders and provide a prognosis -Alarm ICB	-Create RVO -Contact emergency services

Registration in Spoorweb: This is generally the first action that is taken by a dispatcher. Usually, a dispatcher starts with creating a form in Spoorweb during the intake. He or she records the location, time, type of incident and the reporter, among other things.

Inform chain of relevant stakeholders and provide prognosis: When an incident is registered in Spoorweb and labelled with a TIS-code of 1.1 or higher, then the chain of relevant stakeholder is automatically informed, and prognosis is automatically provided. A prognosis can be altered by an AL.

Alarm ICB: When an incident is registered in Spoorweb and labelled with a TIS-code of 1.1 or higher, then the department that resolves incidents throughout the Netherlands (ICB) is alarmed. The AL that is the closest to the location of the incident is automatically informed and instructed to move to the site. Furthermore, the person within the ICB-department that is the closest to the location of the incident (the EMTP) is alarmed as well, so he or she can arrive quickly at the site of the incident. The AL contacts the MKS that he or she will take over the coordination of the incident. During this conversation, the MKS provides the AL with all the necessary information.

Creating an RVO: When there is an infrastructure defect, a constructor needs to move to the site to fix the problem. To instruct the constructor, an RVO-form is created by a dispatcher that contains information about the problem.

Contacting emergency services: In case of acute danger, or an investigation from the police is required (e.g., for a suicide), emergency services are contacted. This is for example the case when there is a fire, or a suicidal person close to the track. Since the emergency services are located at more stations and they have more personnel, they can generally arrive more quickly at the site than staff of ProRail.

Creating a PuVo-form: The PuVo-line is a telephone line on which civilians can call ProRail about a wide variety of issues. During the day this line is answered by the PuVo-department and only some issues are forwarded to the MKS. When someone calls via the PuVo-line outside the office hours of the PuVo-department, a dispatcher must create and fill in a PuVo-form. This form contains a description of the situation that was addressed by the person who called. This form is used to inform the PuVo-department, who will resolve the issue on the next workday.

2.1.3 Follow-up notifications

During the period that the incident is active, which is the period starting with the intake until the dossier is closed, the incident is monitored by all dispatchers. During this period, dispatchers regularly check multiple things:

- Will the prognosis expire soon?
- Are there any developments in the situation that were not reported yet: e.g., the train is moved?
- Did other stakeholders in the process respond to their call-ups (ALs, emergency services etc.)
- Have the emergency services arrived on the site?

During the incident, the dispatchers and AL communicate about the actions that are taken by both parties. This is the case for incidents with a TIS-code of 1.1 or higher that are not AKI-dossiers. For AKI-dossiers, the MKS informs ICB in SpoorWeb as well by logging information.

Next to these active tasks, there are several reactive tasks, such as communicating with the ALs, members of ICB-teams and the rail traffic control to exchange information. ICB-teams are groups of emergency responders that resolve incidents in the field. The information that the MKS receives is logged into SpoorWeb. This way, other stakeholders can access this information as well.

On average 7.66 dossiers were active on any given moment between July 2017 and august 2021. Around 1.22% of the time, there was not a single active dossier. These periods occurred mostly during the night.

The MKS also closes every SpoorWeb dossier. For incidents without structural delay, the MKS closes a dossier when they consider all related tasks as completed. For incidents with structural delay, the ICB-team must first indicate that they have completed their work and the central traffic control (VLC) must give a green light as well, before the dossier can be closed.

2.2 Types of communication

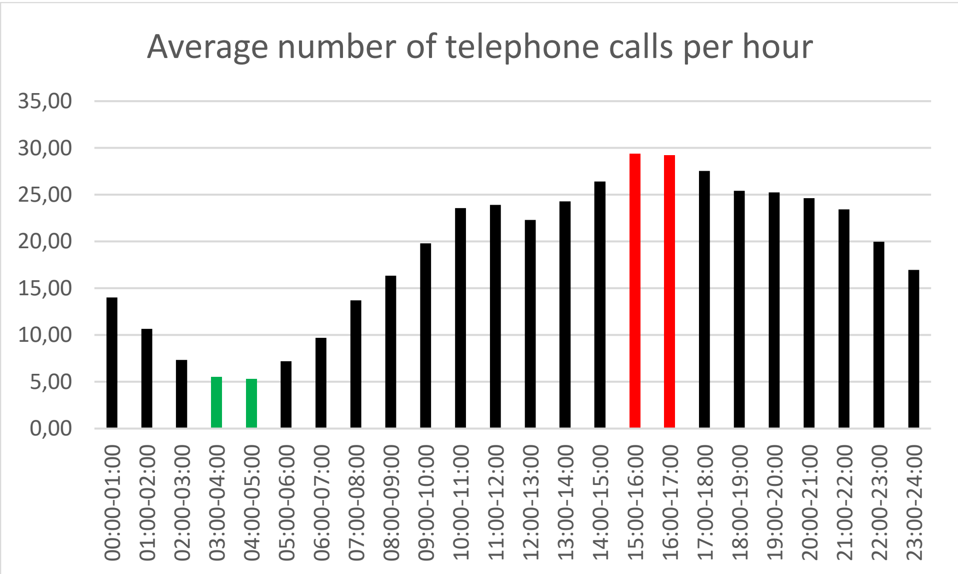
At the MKS the work that dispatcher carry out is very diverse. As mentioned in Section 1.2, the MKS is the first point of contact for incidents on and around the rail network in The Netherlands. Therefore, the dispatchers are communicating with various stakeholders, e.g., with emergency services, TRDLs, ALs or incident emergency teams of ProRail. Since dispatchers often need to provide information to other stakeholders rapidly, most communication is done via telephone. In special cases, communication goes face-to-face as well. This is only possible with departments that are located at the OCCR as well, like camera surveillance and CHI. The CHI coordinates the restoration of infrastructure problems. The communication with ALs and ICB-teams can be done by telephone as well as transceiver.

2.2.1 Telephone communication

Most of the communication at the MKS goes via telephone. The calls come from several different reporters, like emergency services and ProRail and NS employees that are responsible for the train traffic. The intake of an incident is almost always done via telephone.

In Figure 3, the average number of telephone calls per hour between January 2018 and November 2021 is depicted. From the figure it becomes clear that the high peaks in total number of telephone calls lies between 15:00 and 17:00. During the morning there are less phone calls than during the afternoon and during the night there are less phone calls than during the morning or afternoon.

Figure 3: Average number of telephone calls/hour (Jan. ‘18 – Nov. ‘21, source: MobiRail)

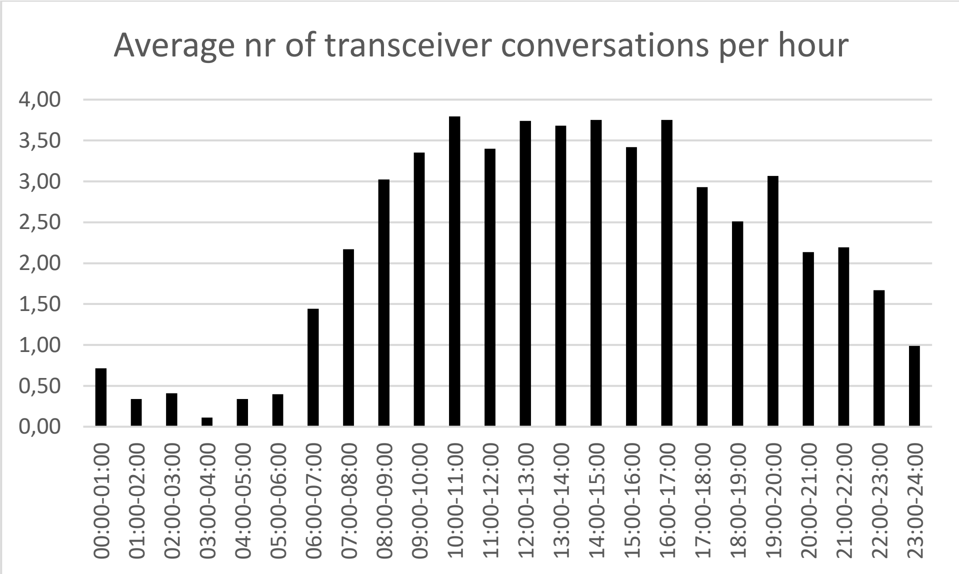


2.2.2 Transceiver communication

Since June ’21, the MKS can communicate with members of all ICB-teams and ALs throughout the Netherlands via transceiver. Dispatchers at the MKS can communicate directly with the ALs and members of incident emergency teams via the transceiver. Via the transceiver, communication in the form of one-to-one and one-to-many are possible. This second form is an advantage in comparison to general telephone communication. This allows for effective communication in situations where multiple employees need the same information.

From Figure 4 it is clear that most of the communication via the transceiver takes place during the day. This graph shows more fluctuations than those of the distribution of telephone calls or incidents over the day. This is probably due to the limited amount of data. What stands out is that, in contrast to the telephone data, the morning is a bit busier than the afternoon. Furthermore, the night is very quiet compared to the day.

Figure 4: transceiver voice requests to MKS/hour (July '21 – Nov. '21, source: SpoorWeb)



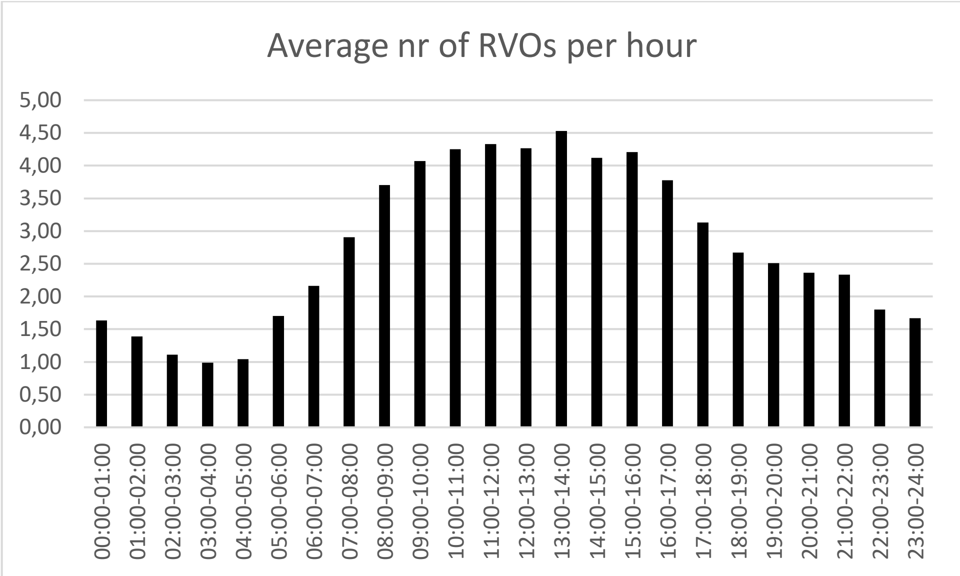
2.2.3 Communication via digital forms

In many situations, an incident or notification that is received at the MKS is the result of an infrastructure defect. Before a contractor is sent to the relevant location, an RVO is created. RVOs are reports that contain the basic information about a defect that is needed by a contractor before he or she visits the location where the repairment is needed. The CHI and OBI create RVOs as well, where each department creates RVOs that are related to their expertise. The OBI department monitors and operates the high-voltage installations, such as the overhead line equipment.

The historical data of RVOs is saved in SAP and contains, among other things, the time the RVO was created, a description of the problem and the priority it has for the constructor. Furthermore, each RVO has a unique identifier. This unique identifier is an 8-digit number, that can start with a 5, a 7 or an 8. RVO's that start with a 5 are created by TRDLs, with a 7 by dispatchers at the MKS and with an 8 they are either created by a dispatcher at the MKS, by the CHI or by OBI. The unique identifier does not contain more information. Currently it cannot be derived from historical data which department created the RVO for each RVO that starts with an 8. Further analysis should provide an estimate of the number of RVOs that are created by the MKS of the total number of RVOs that start with the number 8.

In Figure 5, the average number of RVOs per hour where the unique identifier starts with a 7 or an 8 between January 2018 and November 2021 is depicted. From this figure it becomes clear that most RVOs are created in the morning and the early afternoon. Since the CHI-department currently only works during office hours, they do not create RVOs during the night. Therefore, when one only looks at the RVOs created by the MKS, the number of RVOs created during the night shift is high compared with the number created during the day shifts. One possible reason for this is that a lot of the construction work at the track is performed during the night, since it then has less effect on the train service compared with construction work during the day.

Figure 5: Number of RVOs created per hour (January 2018 – November 2021) derived from SAP



Furthermore, the number of Spoorweb forms that are created per hour is analysed in Section 2.3. Here, graphs with the distribution of the number of incidents (measured in number of Spoorweb forms) over the day and week are provided.

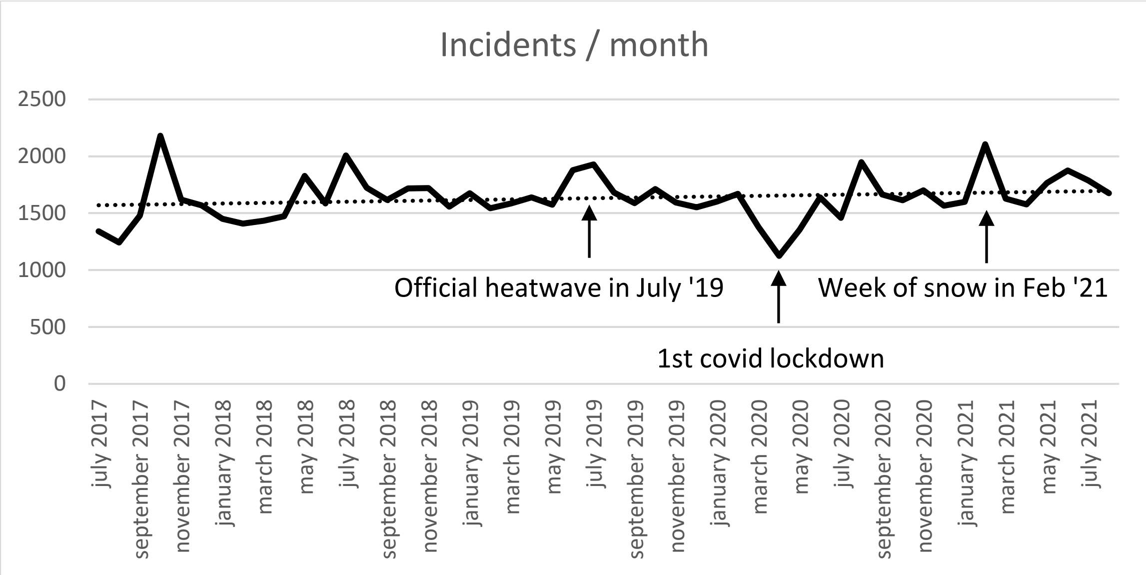
2.3 Incident analysis

In this section, a general analysis of the frequency, duration, number of loglines and causes of different incident types is provided. Furthermore, a description of AKI-dossiers is given.

2.3.1 Frequency of incidents

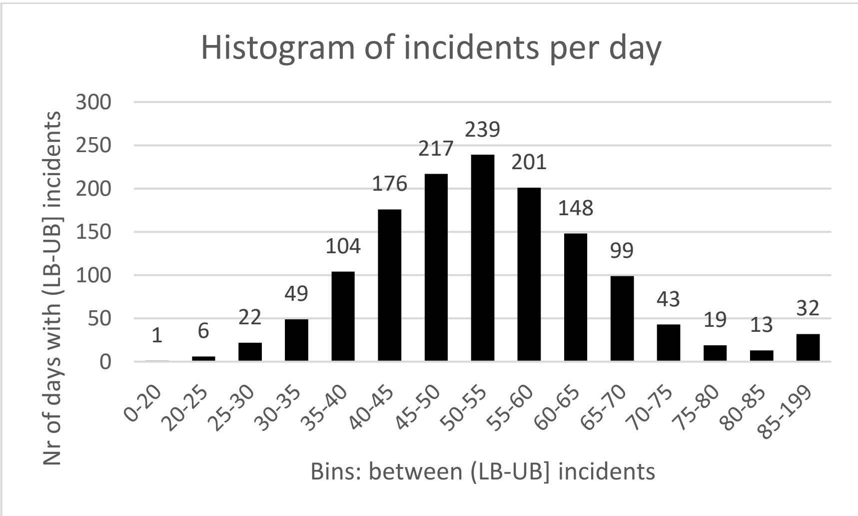
The number of incidents per day has been relatively stable over the past four years. A dip in the number of incidents due to the lockdowns of the covid-19 crisis is visible and peaks from extreme weather events as well. In Figure 6, the number of incidents per month and the underlying trend are shown. All incidents (thus also TIS 1.0) are included. In this figure, a clear drop in the number of incidents is visible from the first covid-19 related lockdown. Furthermore, the incidents are aggregated per month since a graph from the incidents per day over the last four years fluctuates heavily, which makes it harder to read. Due to these monthly aggregations, the graph becomes smoother, however single-day anomalies are not visible anymore.

Figure 6: Incidents per month (July ‘17 – August ’21, source: SpoorWeb)



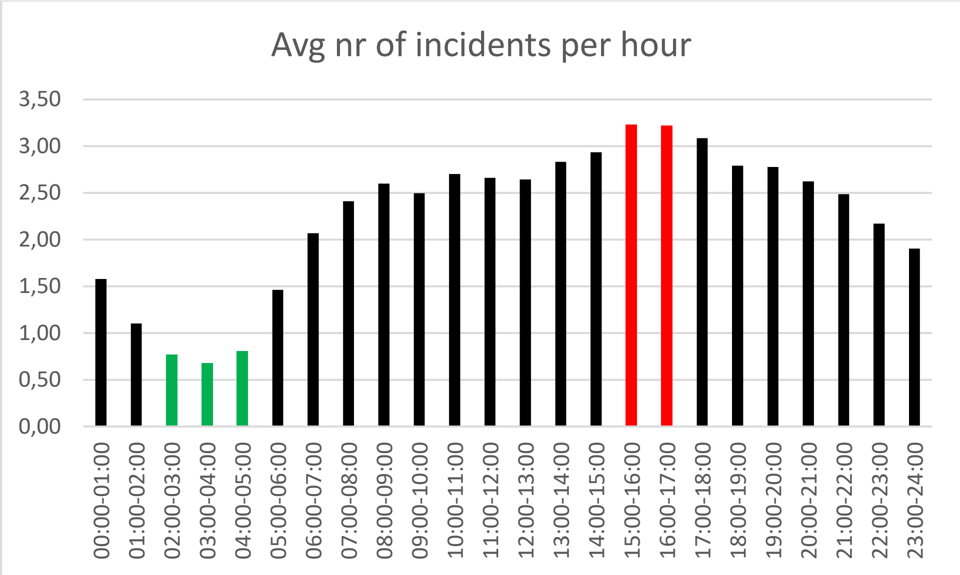
In Figure 7, a histogram of the number of incidents per day is depicted. For this histogram, there are two large bins (0-20) and (85-199) for the extreme values (below and above 1.5 times the IQR). From this histogram it seems that the average number of incidents per day is symmetrically distributed, with a mean of 53.60 and a standard deviation of 15.36. The coefficient of variation is thus 15.36/53.6=0.29, which is significantly lower than that of the standard Normal Distribution (1). This shows that variation in the number of incidents per day is relatively low.

Figure 7: Histogram of incidents/day, with outlier bins (Jan. ‘18 – Nov. ’21, source: Spoorweb)



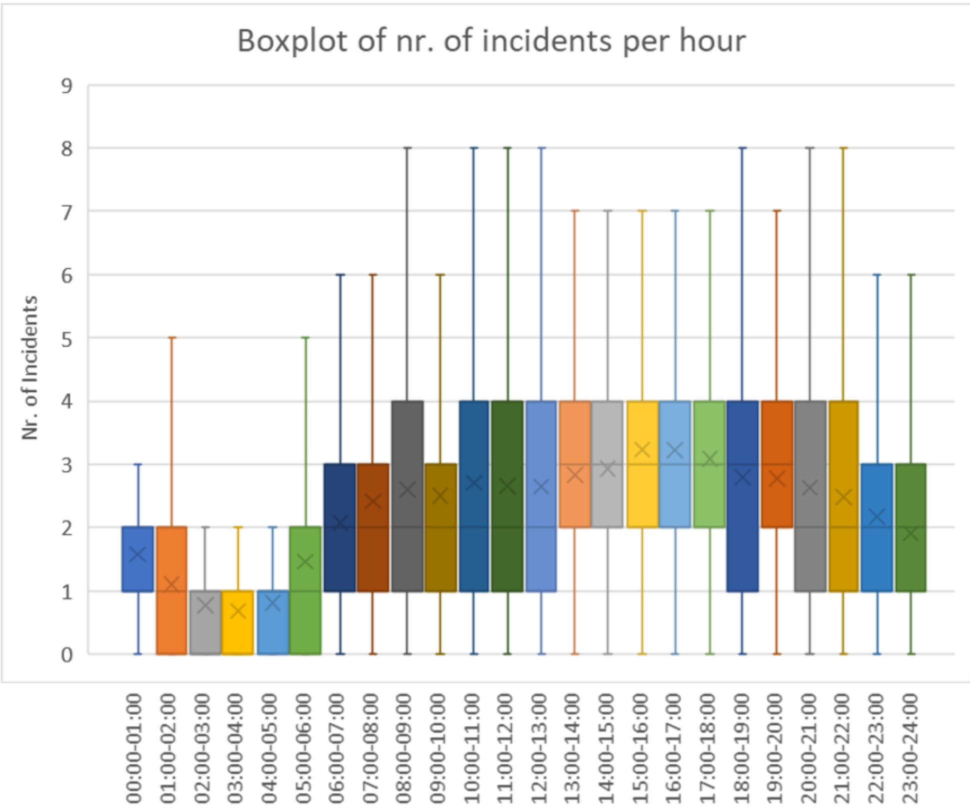
In Figure 8, the average number of incidents per hour is depicted. From this figure, it becomes clear that during the night, there are the least incident and that the most incidents happen in the afternoon. The average number of incidents per hour increases from 4:00 until 9:00. Then it stays relatively stable until 21:00 with a peak between 15:00 to 17:00. The average number of incidents per hour decreases from 21:00 until 4:00. This graph shows that the demand is the highest between 10:00 and 21:00 and especially between 15:00 and 17:00. During the night it is quieter. What stands out is that this graph has a lot of similarity with the number of telephone calls per hour.

Figure 8: Average number of incidents per hour (Jan. '18 – Nov. '21, source: SpoorWeb)



In Figure 9, the number of incidents per hour is shown in a box-whisker graph. In this figure, the box represents how many incidents there were in a specific hour on 50% of all days. The X represents the average number of incidents in that hour. The ⊦ shaped line represents the 50% of periods with the most (above the box) and least (below the box) number of incidents. What stands out in this figure is that the average number of incidents per hour was the highest in the afternoon, but the variance is lower in this period than in the morning.

Figure 9: Boxplots of incidents per hour, excl. outliers (Jan. '18 – Nov. '21, source: SpoorWeb)



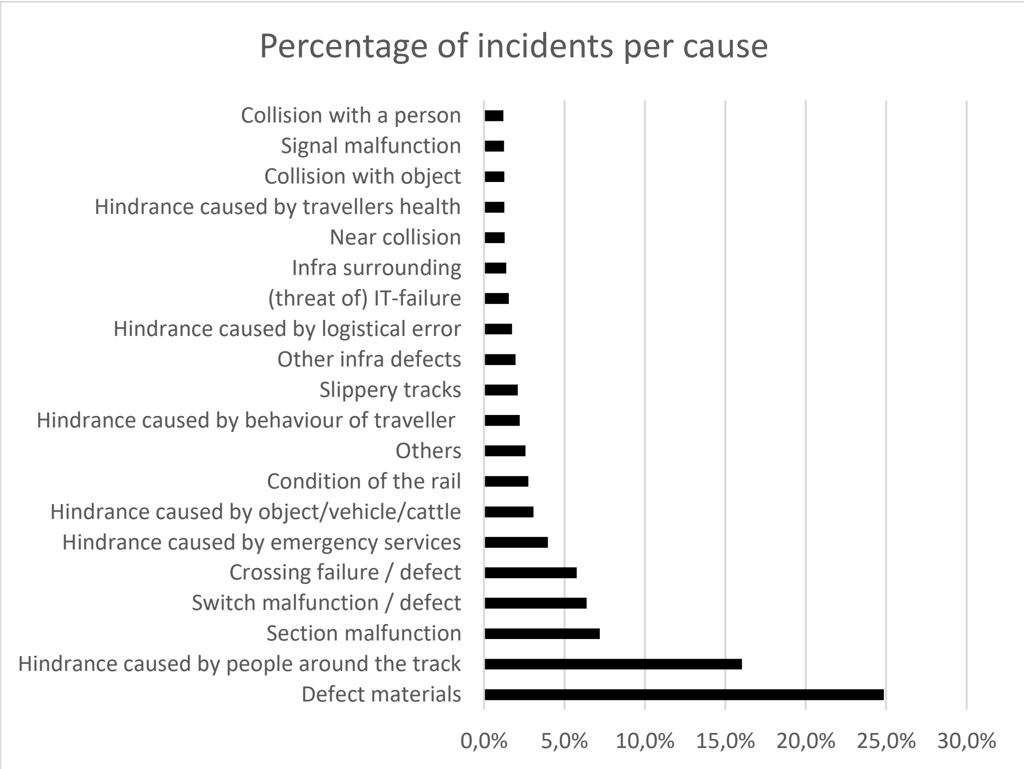
2.3.2 Categorization of incidents

All registered incidents are categorized in two ways: with the TIS-code and with the incident-label. This TIS-code indicates the work instruction that needs to be used and is therefore more solution-oriented than explanatory. When we look at how common each TIS-code is, it becomes clear TIS 1.0 is much more common than all other incident-types. More than 70% of all incidents receive the TIS-code 1.0. Furthermore, from the 21 different TIS-codes, 15 are used less than 1% of the time. There is also a lot of variation between different incidents with TIS 1.0. Therefore, the TIS-code is not a suitable indicator for determining the impact of an incident on the workload. An overview of the percentage of incidents per TIS-code is provided in Appendix .

The second method that is used for categorizing incidents is by their incident label. For the incident-label, one of 84 descriptions is used to explain the cause of the incident. In Figure 10, the 20 most occurring incident-labels and the percentage of incidents that correspond with it are depicted.

These categorizations can be used for splitting up incidents in groups based on their respective workload. More information about the workload per incident is provided in Section 2.3.3.

Figure 10: Percentage of incidents per cause



2.3.3 Workload per incident type

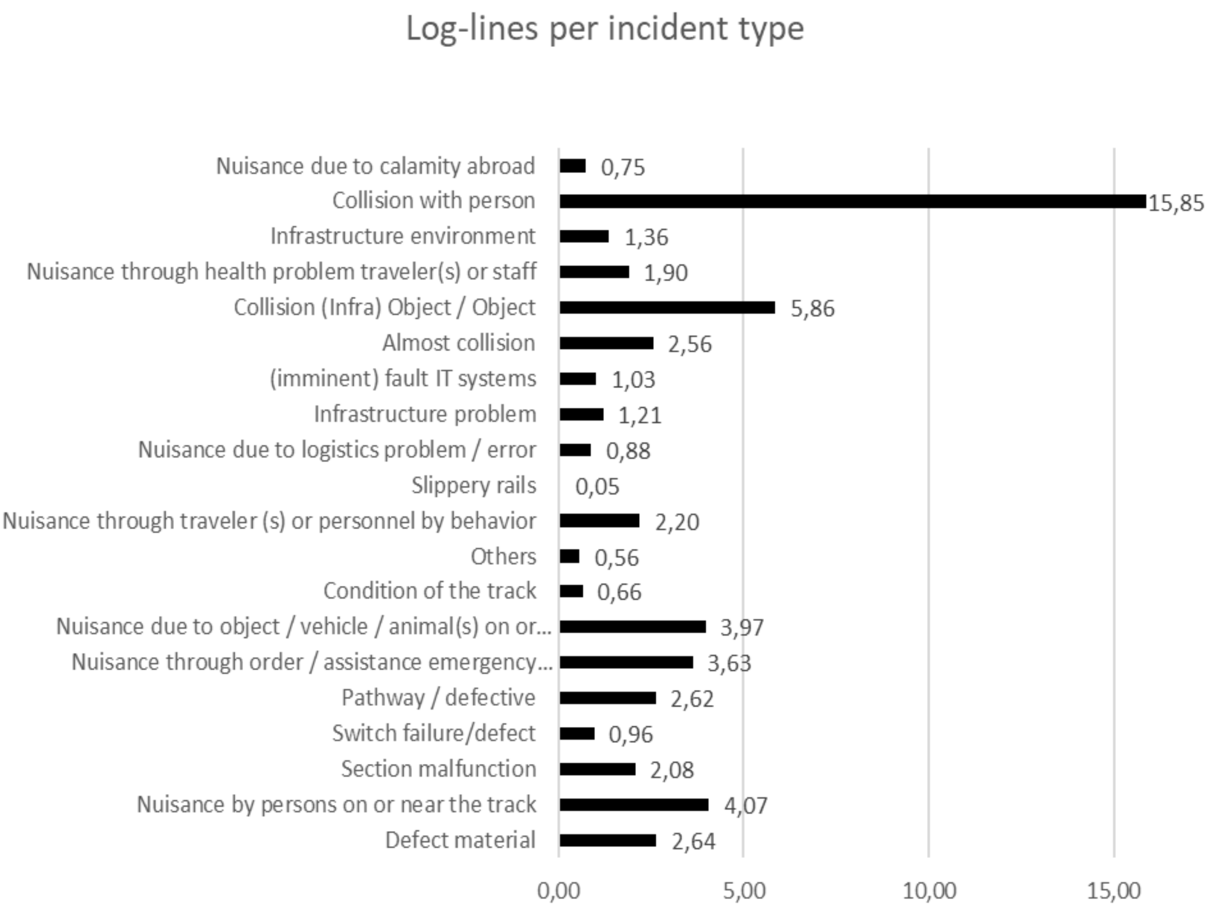
Not every incident is the same in terms of impact on the workload at the MKS. One incident requires more follow-up tasks than other incidents. Each incident is categorized with an incident label and a TIS-code. Furthermore, a log is kept where several stakeholders log their actions. For small incidents, only a couple of actions are logged before the dossier is closed. For more impactful incidents, more actions need to be taken to solve the problem and therefore more actions are logged. The number of loglines logged by the MKS of an incident is thus an indicator for the workload of that incident. Other stakeholders also make loggings in SpoorWeb dossiers, but we do not take these logged lines in consideration, since these might not have resulted in any work for the MKS.

Every incident starts with an intake and ends when a dispatcher closes the dossier. In general, larger incidents take more time to resolve than smaller incidents and require more attention from a dispatcher as well. Therefore, the time that a dossier is active is an indicator for the workload that stems from that dossier. However, some dossiers require little attention, but are open for a relatively long time. For example, a dossier for a switch failure might be open for multiple hours before the incident is closed, but relatively little work resulted from the incident. Therefore, the time a dossier has been active will not be used to determine the workload per incident type.

Since a logline by the MKS directly indicates that the MKS received or provided information, the number of lines that are logged is a more reliable indicator of the workload than the time an incident was active. Therefore, the number of loglines is used as key indicator for the workload per incident in this research.

In Figure 11, the average number of log lines, logged by dispatchers, per incident type is depicted. Only the twenty most common types of incidents were used, the incidents are ranked from least to most frequent in the figure, with 'Nuisance due to calamity abroad' being the 20th most frequent and 'Defect material' being the most frequent type of incident. From this graph it becomes clear that collisions result in much more logged actions than other incidents. Furthermore, it is visible that some incidents lead to almost no loglines from the MKS.

Figure 11: Number of log lines per incident type (January 2018 – November 2021)



2.4 Schedule

In this section, the structure of the schedule is clarified. Furthermore, the different scheduling levels are explained and the influences of weather notices on the schedule are analysed.

2.4.1 Shift start and end times

In the current schedule, every day there are three dispatchers present during the early shift, which is from 7:00 until 15:00. There are also three dispatchers present during the late shift, from 15:00 until 23:00. During the night, there are only two dispatchers present. This shift starts at 23:00 and ends at 7:00. Furthermore, from Monday to Friday there is one dispatcher on a so-called relief shift. This shift starts at 11:00 and ends at 19:00. The moments the shifts start, and end are like those of other departments at the OCCR and originate from the moments that the train service start and end.

The relief shift ensures that employees on the early shift, as well as employees on the late shift can take a break. The three dispatchers on the early shift take a break alternately between 11.00 and 13.00 and the three dispatchers in the late shift take a break alternately between 17.00 and 19.00. The four hours in between these break moments can be used for portfolio tasks. Furthermore, when there are scheduling problems, these relief shifts are the first shifts that are rescheduled. In case of such a rescheduling, a dispatcher that was originally scheduled to work a relief shift, will then replace someone in a regular shift. When no relief shift is scheduled, which is the case during the weekends and intermittently on weekdays, then dispatchers officially do not have a break. As a substitute they receive a financial compensation for the breaks they missed.

2.5 Performance of the current schedule

To measure a possible improvement in performance, one must first define how the performance of the schedule is measured. Currently, there is one KPI in place that is actively measured at the MKS. This KPI is described in the following section. Furthermore, several possible KPIs are suggested.

2.5.1 Current KPIs

As of January 2022, there is one key performance indicator (KPI) in place in the ICB-department for which the MKS is the main responsible actor. This KPI is that in case there is an incident of TIS 1.1 or higher (and it does not fall in the AKI-category), the AL should be alarmed within 5 minutes after the incident became known to ProRail.

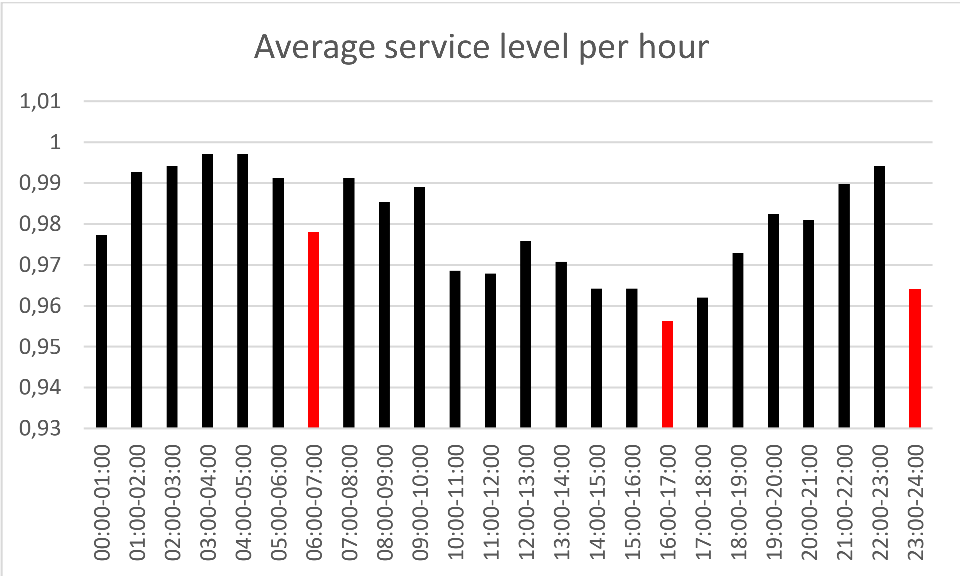
Since the goal of this research is to better match the available capacity to the workload, a KPI should be determined to measure this match. The existing KPI that an AL must be alarmed within 5 minutes after the dispatcher is notified about the incident, aims at measuring the speed with which the first stage of an incident is handled. Currently, this KPI is considered as an indirect KPI for 'the average number of understaffed hours at the MKS'. However, it is not a reliable indicator. A workload that is too high (for example all dispatchers are already occupied with other phone calls) could be the reason that the goal of alarming within 5 minutes is not met. However, it does not say much about the workload over the day or entire shift of the dispatcher. Besides, there might be other reasons for failing to achieve this goal and therefore it is not a good KPI for what we want to measure in this research. Therefore, this KPI will not be used in this research. Also, there are intakes for specific incidents where it is normally not possible to meet the 5-minute threshold, since a substantial number of actions must be completed. This is for example the case when hazardous substances are involved. An intake that is longer than 5 minutes for such an incident is the standard and does not imply that a dispatcher has provided bad work.

2.5.2 Proposed KPIs

Since this research aims at lowering the work pressure for dispatchers at the MKS, the main KPI in this research is: **the service level expressed as the percentage of shifts where the capacity is sufficient**. This KPI is currently not actively measured. We defined that an hour is understaffed (capacity is not sufficient) when the time needed to properly handle all incidents is more than the capacity (number of available workers * 60 minutes) that is available during a certain period of time. The time needed to handle all incidents is estimated with the workload estimation model of Section 4.2

Figure 12, the average service level per hour over the years 2018 – 2021 is depicted. From this figure it becomes clear that situations where the capacity does not meet the demand happen the most in the afternoon. Furthermore, it stands out that the service level is low at the beginning of the night and the end of the night. This is logical, since the workload per hour decreases smoothly during the evening and into the night, while the capacity directly drops with 1/3rd after 23:00. Especially at the beginning of the night there are on average much more times that the capacity is not sufficient. During the morning, the workload increases steadily, but the capacity does not increase until 7:00. As expected, the number of understaffed hours between 6:00 and 7:00 is high.

Figure 12: Average service level per hour (January 2018 – November 2021)



A second KPI that was proposed by the management of the MKS is the number of times that an emergency call (telephone call on the alarm line) is missed. Dispatchers are instructed to terminate their activities and directly handle the emergency call. Therefore, such a call could only be missed in extreme circumstances. From January '18 – November '21, a total of 419 emergency calls were missed. Per day this number equals 419/1341 ≈ 0.312, or roughly once in three days. However, only 26.01% of these calls rang for 7 seconds or more. Therefore, one can conclude that dispatchers often did not get the chance to answer the call. One can doubt the urgency of these emergency calls, given that they were discontinued so quickly. The 26.01% of the calls that rang for 7 seconds or more make up a total of 109 missed emergency calls over a period of almost 4 years. This number is too low to draw reliable conclusions when it comes to seasonal effects and therefore, missed emergency calls is not used as a KPI in this thesis.

2.6 Problems and changes in current structure

To process incidents as effective as possible, the management of the MKS actively tries to improve the MKS in several ways. One of them is by looking into the capacity allocation, which is where this research is focused on. Another way is by looking into the division of tasks at the MKS. A project intended to restructure the division of labour is currently underway. The reason for instigating this project is explained in Section 2.6.1.

2.6.1 Limitations of current structure

When a new incident is reported to the MKS, the dispatcher that answers the call assesses the urgency of the incident and directly acts. In most cases, a report in SpoorWeb is created in which the incident and its main characteristics are summarized. This method ensures that all relevant stakeholders are informed about the incident. Also, other dispatchers at the MKS are informed this way and are therefore able to handle new calls about this specific incident. This is useful when another call about this disruption comes in, while the recipient of the first call of the situation is actively handling another incident. Since all dispatchers are updated and informed about every incident, each dispatcher can take this incoming call, which makes the team very flexible. When a caller that is connected to a certain incident calls the MKS and the dispatcher who did the intake is occupied, other dispatchers can take the call. This way a caller hardly ever must wait.

However, this flexibility comes at a cost. Dispatchers need to be updated about multiple incidents, either by receiving information directly from their colleagues, or by listening in on their conversations on the phone, or by looking at the information that is logged in SpoorWeb about the incidents. All these methods require additional work and thus increase the workload. Furthermore, since there is no differentiation between roles at the MKS and all dispatchers perform the same tasks, it can occur that a relatively inexperienced dispatcher handles a complex incident. This could be a suboptimal situation if the incident is handled in a less effective way. It also might be less efficient compared to a situation where every dispatcher handles tasks that align with the qualities and experience of him or her.

2.7 Results from survey MKS

In November 2021, a survey was conducted at the MKS with the following goals:

- Provide insights about how often workers experience a workload that is too high
- Confirm or disprove existing assumptions
- Provide missing information about the workload that stems from specific tasks
- Provide insight about what would be the best method to determine the impact of the workload of different incidents
- Provide insights about which factors contribute the most to exceedances in the maximum workload
- Provide insights about possible solutions for lowering the work pressure at the MKS

The questionnaire used in this survey was in Dutch. The answers have been translated and for the open questions, the answers have been categorized. A qualitative summary of the survey is provided in this section. The (categorized) answers to all questions are provided in Appendix D.

2.7.1 Workload experienced by dispatchers

As explained in Section 1.3.2, the survey showed that dispatchers frequently encounter a workload that they experience as unpleasant. Two-thirds of the employees indicated that they experienced this once per week or more, with 22% indicating that they experience this multiple times per shift. Furthermore, 77.8% of the respondents indicated that they do not have enough time for working on their portfolio tasks. However, it should be noted that portfolio tasks are less important than regular work at the MKS.

2.7.2 Checking assumptions

Concerning the confirmation of assumptions that are important in this research: 61.1% of the respondents indicated that the workload that they experienced had a strong impact on the pace with which they executed tasks. 27.8% indicated that it had little impact on the pace, while 11.1% indicated that it had no impact on the pace with which they executed tasks. The answers were more divided regarding the influence of the workload on the quality of their work: 33.3% indicated that the workload that they experienced had a strong impact on the quality of their work, 38.9% indicated that it had little impact, 22.2% indicated that it had no impact and 1 person (5.6% of respondents) indicated that he or she did not know the answer. These answers show that the perceived workload has a strong impact of the work pace and some impact on the quality of work.

2.7.3 Providing missing knowledge

There were three questions that were used to fill in missing knowledge about the workload that stems from specific tasks. The question "*How long does it take you to do the intake of an incident?"* resulted in a variety of answers. However, the highest answer was 5 minutes, with 27.8% of the respondents indicating 5 minutes as upper bound for the time it took to do the intake of an incident. The question *"How long does it take you to process an RVO"* Also resulted in a variety of answers, but on average the estimations were slightly lower than for the incidents. Only one person indicated a maximum of 5 minutes, while three persons indicated a maximum of 4 minutes and the rest of the answers provided even a lower estimate. The question *"How long does it take you to fill in a form related to the calls from PuVo, DVP and Storing Publiek?"* had the most differences between the answers, with four people indicating that it only took them 1 minute and four people indicating that it took them 5 minutes (or more), with the rest of the answers in between.

The averages of the answers were also determined. If people had responded with one value, then this value was used for the average. If people gave a lower an upper bound, then the value in between these bounds was used. The average time dispatchers answered to spend on an intake was 2.79 minutes and 2.93 minutes for processing an RVO. The full lists of answers are stated in Table 34 and Table 35.

2.7.4 Determining workload per incident

Question 13 was focused on finding out what the best method would be to determine the workload that resulted of specific incidents. Here, the largest group of respondents (50%) indicated that the number of loglines would be the best indicator. Furthermore, 27.8% choose for the incident-labels, 11.1% for the TIS-code and another 11.1% for the time that the dossier of an incident was active as best indicator for the work that stems from specific incident. Therefore, one can conclude that the number of loglines of an incident is likely the best indicator for the work that an incident has generated for the dispatchers.

2.7.5 Factors contributing to a high workload

Regarding the factors that contribute the most to exceedances in the maximum workload, 55.6% of the respondents indicated that a high work pressure was often the result of multiple (large) incidents happening simultaneously. A train colliding with a person, or situations with suicidal people on or around the track were mentioned explicitly as types of incidents that often cause a high work pressure by 4 respondents. Furthermore, 27.7% of the respondents indicated that situations when other dispatchers work on portfolio tasks during a regular shift often cause a high work pressure for other dispatchers.

2.7.6 Solutions for lowering the work pressure

The respondents were asked what they thought what the best solution would be for lowering the workload. Since it was an open question, the answers had to be categorized. In 33.3% of the answers, respondents indicated that more capacity (as in more personnel) would be the solution. Other solutions that were provided were to work only during the night shifts on portfolio tasks, to change the way the intake process works and to improve SpoorWeb (the software application in which the incidents are registered).

2.8 Summary

This section summarises the scheduling problem that Chapter 1 identifies and Chapter 2 analyses.

Currently, there is little insight in the distribution of the workload and whether there are seasonal factors or trends present in the data. Furthermore, most of the work at the MKS is reactive to external sources (incidents). These two factors combined result in the situation that the management of the MKS questions whether the current schedule makes sense.

The goal of this research is to determine whether the current schedule is in line with the perceived workload and whether and how the schedule could be improved. The scheduling problems concern the long-term (tactical) schedule. The objective of the schedule is to use the available capacity as effective as possible, by minimizing the number of times that the maximum workload is exceeded. The main variables are the start- and end-times of the shifts and the number of dispatchers that are working per shift. Furthermore, different divisions of roles and tasks and developments herein are included in the analysis.

In Chapter 0, the different components of the work at the MKS are explained and analysed. The following becomes clear from the analyses:

- The number of telephone calls, transceiver conversations and incidents per hour follow a similar distribution. During the night is it quiet, with a sharp increase in the morning (between 05:00 and 07:00), a peak in the afternoon (between 15:00 and 17:00) and a steady decrease from the peak in the afternoon until the middle of the night.
- The number of RVOs per hour follows a different distribution than that of the factors above. This distribution has effectively two peaks: a peak in the morning and a lower peak in the night. Early in the morning (between 05:00 and 07:00) and in the evening (between 06:00 and 01:00) it is relatively quiet.
- The number of incidents per shift follows the following trend: During each of the day shifts there are on average twice the number of incidents as during the night shift. For every day, the afternoon shift has slightly more incidents on average than the morning shift.
- The number of incidents per day is relatively stable around an average of 53.6 registered incidents per day with a standard deviation of 15.4. The coefficient of variation is therefore 0.29, which is relatively low. The outliers in the number of incidents per day are often caused by extreme weather.
- Extreme weather has a major impact on the workload
- In terms of TIS-codes, by far the most incidents are categorized as TIS 1.0 (72.40%). This makes this indicator unsuitable for determining a weight per incident.
- The number of loglines is an indicator for the workload an incident generated for the MKS. Multiplying the number of loglines with the frequency of an incident type showed that the categories 'defect material' and 'hindrance caused by people around the track' are by far the most impactful categories. The category 'Collision with a person' is the third most impactful category, even though this category is only the ten most frequent type of incident. This shows that these incidents generate a lot of work when they happen.
- The main KPI in this research to measure the quality of a schedule is the number of workload exceedances per shift and the average idle time per dispatcher per hour.

- The MKS is going through a restructuring, with changes roles and responsibilities. This project might influence the practical applicability of this research and therefore these developments must be followed closely.
- It is difficult to distillate an optimal schedule directly from the different figures about the workload factors. Not all factors follow the same seasonality patterns and not all factors have the same impact on the workload for dispatchers. In the modelling part, the strong seasonal patterns of the different factors are exploited

3 Literature review

The aim of this literature review is to answer the following research question:

"What is known in literature about estimating and scheduling reactive work at an (emergency) call centre?"

This research question has been split up in four different parts. Section 3.2 summarizes several studies into the workload of other departments at ProRail. Furthermore, it provides insights into how missing data might be gathered and transformed into a workload estimate. Section 3.3 describes how one can model stochasticity and explains how simulation can be used to create a robust schedule. Section 3.4 explains the three different stages of staffing and scheduling in a call centre. Section 3.5 provides insights into the differences between call centres with and without function differentiation. Lastly, Section 3.6 provides a summary of the literature review.

3.1 Workload studies at ProRail

In the past, several studies into the workload of train rail controllers (TRDLs) of ProRail have been conducted. Since TRDLs also perform reactive work and work for the same organisation as the dispatchers at the MKS, these studies could provide insights in how one can measure the workload at the MKS.

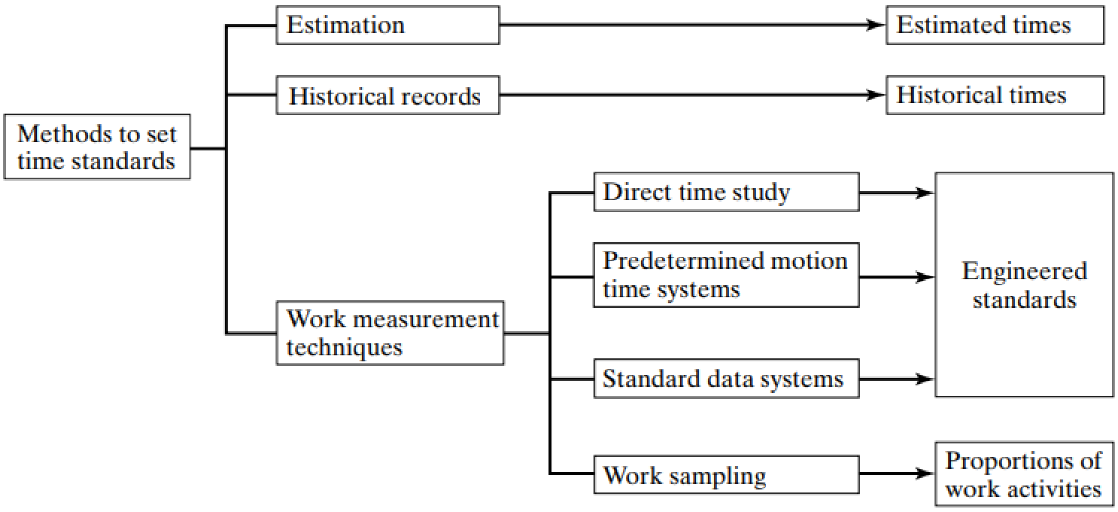
Inspectie verkeer en waterstaat, the Dutch department for inspection of traffic and water services, investigated the workload of TRDLs in 2005 (Inspectie Verkeer en Waterstaat, 2005). In their research, they first conduct a literature study, to gain insight in the tasks of TRDLs and in what is already known about the workload amongst them. Hereafter, a combination of interviews, a questionnaire and analysis of incidents and logged information was used to estimate the workload (Inspectie Verkeer en Waterstaat, 2005).

This study provides a starting point for how one can translate historical data into a workload estimate. However, the goal of the previously mentioned study was not to create a prescriptive model that could be used to decide how one could deploy capacity. Since the goal is different from this study, the approach cannot be taken over one-to-one.

3.2 Workload estimation methods

To estimate the workload at a given moment, qualitative and quantitative methods are used. The goal is to estimate the workload in the past based on historical data and estimate for each hour how many minutes of work had to be carried out. Therefore, this study is not focused on the cognitive workload, where cognitive workload is defined as the level of mental resources required of a person at any one time (Human Reliability, 2021). There are many different methods available for measuring the workload related to specific tasks or setting time standards for tasks. In the book of (Groover, 2013) called 'Work Systems: The Methods, Measurement and Management of Work', multiple methods are proposed. They can be split up in estimation, historical records, and work measurement techniques. In the following sections, every category is explained.

Figure 13: Classification of methods to determine time standards (Groover, 2013)



3.2.1 Estimation based on expert knowledge

Estimation is a technique in which a person familiar with the jobs performed in the department is asked to judge how much time should be allowed for the given task (Groover, 2013). Because this method depends on the estimator's judgment, it is the least accurate of the techniques for determining time standards. When multiple subjects are asked to judge to estimate the time that should be allowed for the same task, the estimate becomes stronger.

In the survey conducted at the MKS, workers were asked to judge how much time should be allowed for doing the intake of an incident and for creating an RVO (see Appendix C).

3.2.2 Historical records

The actual times and production quantities from records of previous identical or similar job orders are used to determine the time standards. Historical records are an improvement over expert estimates because they represent actual times for amounts of work completed (Groover, 2013). However, historical records might not always be available.

In this research, historical records are available for the phone calls. The time that a telephone line is ringing is known, even as the time that workers are on the phone.

3.2.3 Work measurement techniques

To measure the workload during a given period, the workload related to specific tasks that were carried out in that period must be estimated. Furthermore, when a substantial part of the time can be categorized as idle time, which is often the case in work environments with reactive work, then it is favourable to have an estimate of the time spent in this category as well. Since this study aims at matching the available capacity to the workload, it is important that the workload estimate is of high quality. Therefore, it is important to have a workload estimate that is of high quality. A common way to do so is by a time and motion study.

Time and motion study (also referred to as motion and time study, the terms are used interchangeably) is the scientific study of the conservation of human resources in the search for the most efficient method of doing a task. It consists of a wide variety of procedures for determining the amount of time required, under certain standard conditions of measurement, for tasks involving some human activity (Harper & Mousa, 2013).

Groover (2013) introduces a set of four techniques that are concerned with the evaluation of a task in terms of the time that should be allowed for an average human worker to perform that task: (1) direct time study, (2) predetermined motion time systems, (3) standard data systems, and (4) work sampling, an alternative work measurement technique in which statistical measures are determined about how workers allocate their time among multiple activities. Among the four techniques, work sampling should be differentiated from the other three. The primary purpose of work sampling is to determine proportions of time spent in various categories of work activity using randomized observations of the subjects in interest. On the other hand, the principal purpose of the other three techniques is to establish standard times (Groover, 2013).

The time it takes to execute a task might differ a lot, since different tasks in the same category might have specific features that make one task much harder than another task. For example, creating an RVO for a leaking cargo rail wagon is much more worker than creating one for a switch failure. Since these times might differ a lot, it is not interesting to establish standard times. Therefore, the workload estimate must be validated with a work sampling study.

Work sampling is an appropriate method for determining the proportion of the time spend on specific tasks. Work sampling is desirable when sufficient time is available (e.g., several weeks), there are multiple subjects, and the tasks are nonrepetitive but categorizable. All these factors apply to the case in this research. Another important advantage of work sampling is the convenience for the subjects. Being a subject in a work sampling study tends to be less demanding than in a direct time study, since the observations are made quickly at random times rather than over a long continuous period (Groover, 2013).

In a work sampling, one or more researchers measure the activity of subjects at random moments. Randomizing the sampling times ensures that the subjects do not know when a sample will be taken. If human subjects could anticipate when the work sampling observer were coming, they might be inclined to adjust their behaviour in response. This would bias the estimates of activity category proportions. As more samples are taken, the researcher can estimate the true proportion of the workload that stems from specific tasks with more certainty.

For each work category k, the proportion of work that falls in this category p_k can be estimated by dividing the number of times that work from this category is observed by the total number of observations. The statistical base of this approach is the binomial distribution. However, with many observations n, the binomial distribution can be approximated with the normal distribution. The estimated proportion for each category $\hat{p}_k$ will approximate the true proportion p_k.

Groover (2013) proposes the following 9-step approach for a work sampling study.

1. Define the objective(s) of the study
2. Define the subjects to be studied
3. Define the output measure(s)
4. Define the activity categories
5. Design the study
6. Identify the observers who will do the sampling
7. Announce the study
8. Make the observations
9. After completing the study, analyse and present the results

The proportions measured in the work sampling study can be used to either validate or disprove the workload estimate that is based on historical data. The proportions measured in the work sampling study should match the estimates from the workload estimate.

3.2.4 Direct workload measures

Since we are interested in moments when the workload is too high, another option for validating the workload estimate is by directly measuring the perceived workload. To create a sample of the workload that is experienced per hour, one can let the subjects indicate per hour how high the workload was that they perceived.

With this approach, several categories are created upfront that represent different levels of the workload. For example, one could work with three categories, where the first category represents quiet hours, the second represents regular hours and the third represents busy hours. For a predefined period of time, employees indicate the workload they perceived, and this sample can be compared with the workload estimate. The workload estimate can be split up in the same categories, based on certain thresholds. For example, the 20% of the hours with the highest workload estimates can be grouped under the category busy, the 20% of the hours with the lowest workload estimates can be grouped under the category quiet and the rest will fall under the category regular hours.

If the correlation between the categorized workload estimates and the workload measures is high, then the workload estimate would be a good predictor for the perceived workload.

3.3 Stochasticity

There are essentially two ways to include stochasticity in a scheduling problem: introducing stochasticity in the forecasting stage (e.g., by using quantile forecasting) or by introducing stochasticity in the mathematical model (e.g., by introducing probabilistic constraints in the mathematical model).

One way to solve a stochastic optimization problem is by using sample average approximation. In this technique, the expected objective function of the stochastic problem is approximated by a sample average estimate derived from a random sample. The resulting sample average approximating problem is then solved by deterministic optimization techniques. The process is repeated with different samples to obtain candidate solutions along with statistical estimates of their optimality gaps (Verweij, Ahmed, Kleywegt, Nemhauser, & Shapiro, 2003). Monte Carlo simulation is used to generate the different samples.

When it comes to determining the correct number of staff in each time interval, queueing models and simulation models may be used. Queuing models are elegant and may give analytical results but in general, many real-world simplifications need to be made. Simulation can take many practical factors into account, but these may be very computationally expensive solutions. Sometimes, queuing models and simulation are combined to obtain ideal staff requirements (Ernst, Jiang, Krishnamoorthy, & Sier, 2004).

3.4 Scheduling

In general, creating a schedule for a call centre consists of three steps. In the first step, the number of workers needed per period is determined based on certain goals and restrictions (e.g., costs and/or service levels). In the second step, the results of the staffing problem serve as input for the shift scheduling problem. Here, the restrictions for shifts are considered and an optimal collection of shifts to be worked is determined. In the third and final step, the rostering problem combines shifts into rosters and provides the actual matching between employees and rosters (Aksin, Armony, & Mehrotra, 2009). In conclusion, the staffing problem is concerned with the capacity that is needed/desirable for each period of time. The scheduling problem is concerned with determining the optimal set of shifts; when the shifts start, when they end and how many workers should be assigned to each shift. In the scheduling problem, information that is gained in the staffing problem is used as input. In the rostering problem, real workers are assigned to certain shifts. This rostering problem is out of the scope of this assignment.

3.4.1 Staffing Problem

In literature, there are various articles concerning the staffing of call centres. However, the traditional approach to call centre resource deployment decisions is to attempt to build an agent schedule that minimizes costs while achieving some customer waiting time distribution objectives (Aksin, Armony, & Mehrotra, 2009). Examples of articles that focus on minimizing the staffing cost while attaining specific service levels with different methods or restrictions are (Atlason, Epelman, & Henderson, 2008), (Cezik & L'Ecuyer, 2008), (Liao, Koole, van Delft, & Jouini, 2011), (Liao, van Delft, & Vial, Distributionally robust workforce scheduling in call centres with uncertain arrival rates, 2013) and (Gans, et al., 2015).

To evaluate the performance of an outcome of the staffing problem, one can use simulation models or analytic queuing models. Queueing models originally focused on staffing problems with homogenous workers. Simulations models are more apt to deal with multi-skill call centres. (Mehrotra & Fama, 2003).

3.4.2 Shift scheduling

The results of the staffing problem typically provide the input for the shift scheduling problem. As input for the problem, the desired capacity per time period is known. The shift scheduling problem determines the optimal collection of shifts. The traditional approach to the scheduling problem is to formulate and solve a mathematical program to identify a minimum cost schedule (Aksin, Armony, & Mehrotra, 2009). The traditional mathematical programming approach is based on the assumption that all agents are able to handle all in-coming calls. Avramidis, Chan, & L'Ecuyer (2009) have devoloped search methods that focus on producing agent schedules for a multi-skill call center.

3.4.3 Shift rostering

The third step in scheduling the staff at a call centre is to roster the shifts. This step combines shifts into rosters and matches employees to the shifts. For many organisations, as well as for ProRail, this step is performed by a planner. Over a longer period, every dispatcher has the same number of morning-, evening- and night shifts, except for older dispatchers who might opt for a schedule without night shifts. Dispatchers regularly exchange shifts to better match the roster with their personal agenda.

A different solution for matching workers to shifts is by using an auction-based approach. Workers can "bid" on specific shifts in order to match labour supply with labour demand (Aksin, Armony, & Mehrotra, 2009). The order of bidding can be based on factors such as seniority and previous quality of service delivered.

One study that investigated the effects of self-rostering for nurses in a hospital in Boston found that the number of change requests from the nurses decreased after the self-scheduling implementation. The number of sick calls remained steady, while the nurses' need for control and flexibility decreased gradually as the self-scheduling implementation progressed. Furthermore, self-scheduling was perceived by the nurses to give them more time to spend with their families and to provide what they felt was better patient care. Despite all these positive results, the trial eventually stopped due to practical reasons. Nurses would roster themselves for shifts that were already full and leave large blocks of time without a full roster. Furthermore, they sometimes signed up for consecutive days and nights, which is not allowed by regulation. When a planner would shift people around to solve these scheduling problems, the nurses became annoyed that their preferences were not honoured. Therefore, the nurse manager stopped the project altogether (Bailyn, Collins, & Song, 2007).

3.5 Function differentiation

Functions at (emergency) call centres can be split in two groups: specialists and generalists. Specialists are workers that are specialised in certain tasks. These workers are not capable of performing all different tasks, or at least not at the level of a generalist. A generalist can perform more tasks and is therefore more flexible. In the current situation at the MKS, all dispatchers have the same responsibilities (apart from side responsibilities). Furthermore, each worker can handle all tasks (every type of incident, creating RVOs, communicating by transceiver) and can therefore be considered generalists. Most of the early literature on staffing deals with these problems in settings with a single pool of homogeneous agents, while recent literature on staffing models focuses on multi-skill settings. (Aksin, Armony, & Mehrotra, 2009).

As mentioned in Section 3.4.2, there are several articles that focus on solving the problem of scheduling multi-skill call centres. Cezik and L'Ecuyer (2008) use linear programming and simulation to staff a multiskill call centre while minimizing the staffing costs. In this paper, heuristics are proposed to solve large-sized problems.

Since ProRail is working on a project that would change the current generalist approach to a situation with function differentiation, it is important to compare the advantages and disadvantages of the approaches. Van Buuren, Kommer, van der Mei and Bhulai (2017) compare call centre models with and without function differentiation for an emergency call centre in the Netherlands. In a situation with function differentiation, call takers handle the inbound requests and perform communication with the caller and dispatchers take care of the outbound calls. In a situation without function differentiation, generalists perform both tasks. The researchers use simulation models to compare the different call centre models on multiple KPIs. In their study they use three KPIs: the fraction of calls answered in 6 seconds, the total response time and the total costs for the policy. They experimented with 20 different scenarios where the incoming calls had a different arrival rate in each scenario. In 17 out of the 20 different scenarios, the researchers concluded that a policiy of function differentation performed better than one with solely generalists. However, this was heavily influenced by the KPI that focuses on costs, since the generalists were considerably costlier than call takers and dispatchers. In situations with an equal number of workers for the same arrival rate, the policy with solely generalists always performs better on all KPIs, except for the costs (van Buuren, Kommer, van der Mei, & Bhulai, 2017).

3.6 Summary of the literature review

The main goal of the literature review is to provide methods to map the workload and schedule dispatchers in such a way that a threshold for the maximum workload per dispatcher is exceeded as little as possible.

3.6.1 Conclusions of previous workload studies at ProRail

From the literature review it becomes clear that a range of methods can be used to determine the workload for specific time periods or tasks. Time and motion studies can be applied, although they require a lot of measuring time. Interviews, questionnaires, and logged information were used to estimate the workload studies at other departments at ProRail (Inspectie Verkeer en Waterstaat, 2005).

3.6.2 Conclusions of workload estimation methods

In this literature review, many methods to set time standards were discussed. Estimation based on expert knowledge can be used, since this knowledge is already documented in the results of the survey on the MKS. Historical records can be used to determine the time standards for telephone calls, since there is data available about the time and duration of all telephone calls during the last four years.

Estimation based on expert knowledge is the least accurate of the techniques for determining time standards (Groover, 2013). To obtain a more precise estimate, or to validate the estimate based on expert knowledge, work measurement techniques can be used. Of these techniques, work sampling is the most appropriate for this case since the tasks are not very repetitive. Another way to validate the estimate based on expert knowledge is by letting the subjects directly estimate the perceived workload, as discussed in Section 3.2.4.

Both methods have their own advantages and disadvantages. The work sampling method is a method that is well-known in literature. It is a statistically sound method of which it is clear how it should be implemented. However, it is a very time-consuming method for the researcher. Furthermore, the method intends to validate whether the estimate of the proportion of time that is spend by an employee on certain tasks is correct, this is not a direct estimate of when the workload is too high.

Comparing the experienced workload with the estimate focuses directly on the workload that dispatchers experience. An advantage of this method is that the researcher does not necessarily have to sample all data himself, since the dispatchers can indicate on paper or via an online tool how busy each hour was. However, there is more subjectivity involved in this method, since a ‘busy’ hour for one employee might be a ‘normal’ hour for another one and vice versa. Furthermore, this method results in less observations than the work sampling method. It is questionable whether several weeks of sampling would provide enough data to draw reliable conclusions. Furthermore, the quality of the sample relies on the willingness of the research population to cooperate. If some of the dispatchers are not interested in cooperating in this research, the sample becomes smaller and then an even longer period would be needed to draw reliable conclusions.

3.6.3 Conclusions of modelling stochasticity

Since the core problem of this study contains many uncertain elements (e.g., the differences in workload from hour to hour, differences per dispatcher in what is perceived as a workload that is too high etc.), stochasticity must be incorporated into the problem. To cope with this, several solutions are suggested in literature. Examples are sample average approximation (SAA), quantile forecasting or creating a simulation model.

3.6.4 Conclusions of scheduling

Scheduling a modern call centre could be seen as a three-stage problem. One should first determine the optimal staffing levels for each time period. Hereafter, the optimal set of shifts should be determined to match the capacity with a feasible roster. Finally, individual workers should be assigned to specific shifts (Aksin, Armony, & Mehrotra, 2009). Where workers are currently assigned to shifts by a planner, an alternative is to use self-scheduling. While self-scheduling has shown to lead to multiple improvements (i.e., less need for rescheduling, an improved sense of being in control and more flexibility), it might lead to new scheduling problems (i.e., lots of rescheduling due to infeasible rosters) ((Bailyn, Collins, & Song, 2007).

Furthermore, in this research the goal is to match the workload and capacity in such a way that the number of times that the workload per hour is exceeded is minimized.

3.6.5 Conclusions of function differentiation

Finally, the differences between call centres with and without function differentiation were investigated. From this analysis it became clear that the optimal situation depends on multiple factors, including the arrival rate and the objective of the problem owner. In the experiments ran by Van Buuren et al (2017), where the objective was to minimize costs, using function differentiation was most often the optimal choice. However, when the goal is to maximize the service level with a fixed capacity, a pool of homogeneous workers is expected to perform better.

4 Solution design

In this chapter, the models that are the core of this research are described. The research question that is central in this chapter is the following:

How can we create and validate the workload estimation and scheduling models?

The way each model is built, the goal of each model and the assumptions that are ingrained in each model are explained. In Section 4.2, the model that estimates the workload for periods in the past based on historical data is described. Secondly, the categories in which the workload is split up are explained. Thirdly, a mathematical formulation of the workload estimation model is provided. In Section 4.3, the work sampling study that is used to determine the task times for resolving incidents and RVOs is explained. In Section 4.4, a mathematical formulation of the scheduling model is provided and explained. In Section 4.5, several model extensions are provided and in Section 4.6, the solving approach is elaborated. Finally, the chapter ends with the conclusions in Section 4.7.

4.1 Methodology

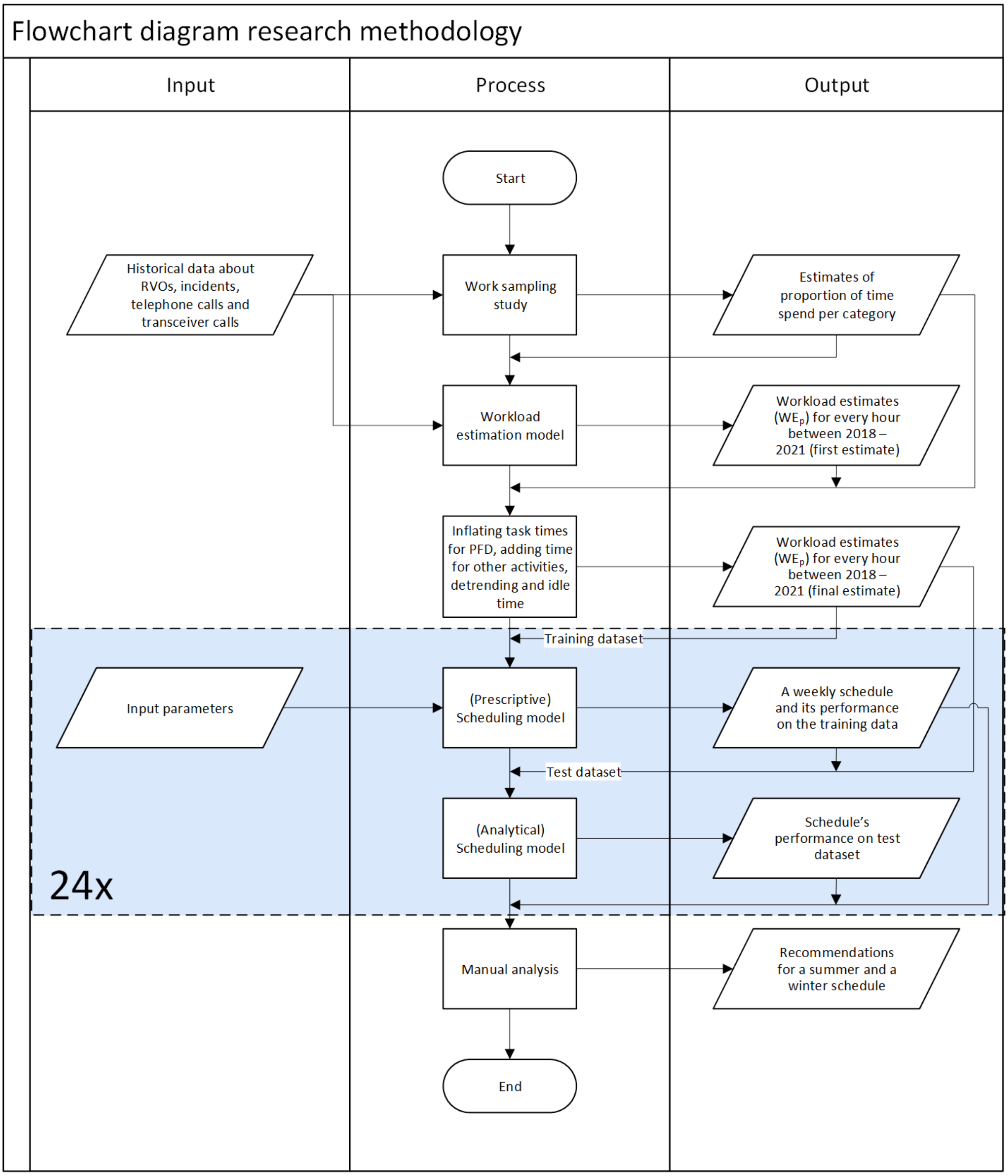
In this section, it is explained how the workload estimation model and the scheduling model fit into the overall methodology. A flowchart of the methodology is provided, and this flowchart is explained in words.

In Figure 14, the research methodology is explained by means of a flowchart. To provide recommendations for the schedule(s) that the MKS should use, the following methodology is used: First, a work sampling study is conducted, to determine the task times that correspond to incidents and RVOs. This work sampling study is explained in Section 4.3 and the results of this work sampling study are explained in Section 5.1. The historical data about the different work categories that were discerned in Chapter 2 and the task times that are obtained from the work sampling study serve as input for the workload estimation model. This model is explained in Section 4.2 and the output of this model is explained in Section 5.2. The output of the workload estimation model is modified to include the trend, idle time and time for personal things, fatigue, and delay. This process of data preparation is described in Sections 5.1 and 5.2. After the modifications, the final values for the workload estimate are obtained.

The dataset containing the final workload estimate is split into four datasets: two for the summer and two for the rest of the year. For both seasons, which we call summer and winter in this research, the two datasets act alternately as training set and test set.

The scheduling model is explained in Section 4.4. This model can be used prescriptively (to generate an optimal schedule) as well as analytically (to evaluate the performance of a given schedule). For each experiment, a training dataset and the set of input parameters are provided to the scheduling model as input, and it provides an optimal schedule as output. Thereafter, a test dataset, which has no overlap with the training dataset is used as input for the scheduling model. The schedule that was created by the scheduling model with the training dataset is evaluated on the test dataset. In total, 6 experiments are executed for all 4 datasets, resulting in 6*4=24 experiments. In the end, an optimal schedule is determined for the winter as well as for the summer in a manual analysis. This means that the analysis is not based exclusively on the quantitative outcomes, but there is also some qualitative analysis involved.

Figure 14: Flowchart diagram of the research methodology



4.2 Workload estimation model

In the following sections, the workload estimation model is explained. First, the decision for the size of the time buckets is explained. Hereafter, the different categories that are used are explained.

4.2.1 Size of the time buckets

With the workload estimation model, the workload has been estimated for every hour between January 1st 2018 and September 30 2021. Time buckets of one hour have been used. With this size, there was enough data available for each time-bucket to create a reasonable estimate, while also ensuring that the workload is split up in enough periods to provide a multitude of options for the scheduling model. With smaller time buckets, the data available per period is very limited and therefore the quality per estimate would be worse. With bigger time buckets, peaks in the workload might be missed, since they are spread out over a larger period.

Furthermore, the time buckets should be in line with the decision variables for the scheduling model. If the time buckets are large, the number of decision variables becomes small. This reduction of the solution space might result in a situation where many good solutions are omitted. With small time buckets, the opposite problem occurs: the solution space becomes too extensive. Time buckets of one hour each are deemed appropriate in this research.

Furthermore, the data is split into two seasons, a summer season, and a winter season. This division is made based on the results of the seasonality analysis. The summer season runs from June – September, while the winter season covers the rest of the year. The argumentation for this division is explained in Section 5.2.3.

4.2.2 Categories for different tasks

Based on the available data in the information systems of ProRail, the following categories of different tasks were selected for the workload estimation model:

- Resolving incidents
- Creating RVOs
- Telephone calls
- Transceiver calls

Each category has been explained in detail in Sections 2.1 and 2.2. To prevent double counting some of the work, the groups must be mutually exclusive. Therefore, the category 'resolving incidents' is defined as all tasks related to incidents (intake, logging, looking for information), that are executed when a dispatcher is not on the phone. For the category 'creating RVOs', the same holds: only the time spend on creating RVOs when the dispatcher is not on the phone is counted. To estimate the workload per hour, all four data sources that impact the workload are incorporated in one model. To provide more clarity, an example of how the workload is determined for one specific hour is provided in Table 3. The time used to multiply the different components with stems from the results of a work sampling study, which is explained letter. The hour used for the example took place on 26-9-2021 from 14:00 to 15:00.

Table 3: Example of determining the workload per hour

Component	Number (tally)	Time factor	Time impact
Incidents	5	Differs per incident	00:27:28
RVOs	4	00:02:29	00:07:27
Telephone calls	23	Differs per call	00:20:42
Transceiver calls	5	Differs per call	00:01:40
		Sum:	00:57:16

This calculation is performed for every hour every hour between January 1st 2018 and September 30th 2021. The time factors used differ per category, or even within the category. The following approaches were used:

- **Resolving incidents:** For incidents, the fixed time that is counted is determined in the work sampling study. Furthermore, 1 minute per logline logged by the MKS is added to determine the final time impact.
- **RVOs:** For RVOs, a fixed time factor is used, which is determined in the work sampling study.

- **Telephone calls:** For telephone calls, the duration of each call is known. For incoming calls, only the time on the phone is counted. For outgoing calls, 10 seconds are added to incorporate the time needed to search the right line and wait for the response. The time spent on telephone calls is validated in the work sampling study
- **Transceiver calls:** For transceiver calls, the duration of outgoing communication from the MKS is known. However, the duration of the incoming communication is not known and the duration of the periods between the incoming and outgoing communication is unknown as well. With active conversations, these short time frames between the outgoing and incoming messages cannot be used for other work and therefore this time must be considered as well. Therefore, the duration of the communication from the MKS is multiplied with a factor of 2.5 to compensate for the unknown duration from the incoming communication and time between the voice memos. According to C. Snels from Entropia (the company who provides the transceiver communication service), this is an adequate estimate. The time spent on transceiver calls is validated in the work sampling study.

4.2.3 Mathematical formulation of workload estimation model

The output of the workload estimation model serves as input for the scheduling model. To make the connection clearer, a mathematical formulation of the workload estimation model is provided in this section.

The Workload Estimation (WE) is determined for periods p=1, 2, …, P = 32,856. Where p=1 is the period on January 1st, 2018, from 00:00 to 01:00, and P = 32,856 is the period on September 30th, 2021, from 23:00 to 24:00. Each period p has a start time: ST_p and an end time: ET_p.

Each category has a Time Impact (TI) per period, which provides us with the following four categories: $TI_{incidents}$, TI_{RVO}, $TI_{telephone}$ and $TI_{transceiver}$. The sum of these categories for each period provides us with the workload estimate per period.

The Time Impact per period is determined differently for every category, as described in Section 4.2.2. For every member of every category the time of registration (T_creation) is known. Furthermore, for all incidents i, the $loglines_i$ are known. For every incoming telephone call, the duration of the call is known. For every transceiver call l, the duration of the outgoing communication is known. Note that the time impact is expressed in minutes, so the 1/6th that is added to outgoing telephone calls stands for 10 seconds. The information results in the following set of equations:

Indices:

Periods p = 1, 2, …, 32856.

Incidents I = 1, 2, …, 73697.

RVOs j = 1, 2, …, 91245

Telephone calls k = 1, 2, …, 618571

Transceiver calls l = 1, 2, …, 4694.

Parameters:

WE_p	A weighted estimated of the workload for every period p in minutes
$TI_{incident,\,p}$	The time impact of all incidents in period p
$TI_{RVO,\,p}$	The time impact of all RVOs in period p

$TI_{telephone,"I",p}$	The time impact of all incoming telephone calls in period p
$TI_{telephone,"O",p}$	The time impact of all outgoing telephone calls in period p
$TI_{telephone,p}$	The time impact of all telephone calls in period p
$TI_{transceiver,\ p}$	The time impact of all transceiver calls in period p
$FT_{incidents}$	The fixed time impact of incidents
FT_{RVOs}	The fixed time impact of RVOs

Calculation of parameters

$$WE_p = TI_{incident,p} + TI_{RVO,p} + TI_{telephone,p} + TI_{transceiver,p} \qquad \forall p \qquad (1)$$

$$TI_{incident,p} = \sum_i (FT_{incidents} + loglines_i) \qquad \forall p \ where \ ST_p \leq T\_creation_i < ET_p \qquad (2)$$

$$TI_{RVO,p} = \sum_j FT_{RVOs} \qquad \forall p \ where \ ST_p \leq T\_creation_j < ET_p \qquad (3)$$

$$TI_{telephone,"I",p} = \sum_k calling_k \qquad \forall p \ where \ ST_p \leq T\_creation_k < ET_p \qquad (4)$$

$$TI_{telephone,"O",p} = \sum_k ({}^1/_6 + calling_k) \qquad \forall p \ where \ ST_p \leq T\_creation_k < ET_p \qquad (5)$$

$$TI_{telephone,p} = TI_{telephone,incoming,p} + TI_{telephone,outgoing,p} \qquad \forall p \qquad (6)$$

$$TI_{transceiver,p} = \sum_l (2.5 * calling_l) \qquad \forall p \ where \ ST_p \leq T\_creation_l < ET_p \qquad (7)$$

4.3 Determining task times RVOs and incidents

To determine the task times of RVOs and incidents, a work sampling study is conducted. The results are validated using expert estimations from the survey. The work sampling study is the more reliable than expert estimations according to Groover (2013) and therefore, the work sampling study is leading, while the expert estimations are used for validation. The choice for a work sampling study is based on the following advantages of this method:

- Work sampling is desirable when there are multiple subjects, since these can be measured at the same time in a work sampling study (in contrast with other work measurement techniques). This is the case in this research, and it lowers the time that is needed to reach the required number of samples.
- Work sampling is desirable when tasks are nonrepetitive but categorizable, which applies to this case.
- Being a subject in a work sampling study is less demanding than in an observation study with one of the other workload estimation techniques.

- Besides obtaining an estimate for the proportion of time spent on specific tasks, one can derive an estimate for the standard time that it takes to execute certain tasks as well from this estimate.
- Compared to the method described in Section 3.2.4, work sampling is less subjective, takes less time and is well-known in scientific literature.

The design of the work sampling study is based on the 9 steps that were laid out by (Groover, 2013).

4.3.1 Study objective

The goal of this study is to validate or improve the workload estimate that is based on a combination of expert knowledge and historical records, by measuring the proportion of time spend on tasks in different predefined categories. In case there is a significant difference, then the results of the work sampling study are leading, since this method for setting time standards has a higher relative accuracy than estimation based on expert knowledge.

4.3.2 Subjects

The subjects in this study are the employees of the MKS. We assume that all employees are the same. The employees that are present differ per day and therefore in this study, we will measure the performance of worker 1 and worker 2 and … and worker K that are present at the time of measuring. To reduce the relative performance difference between workers, the work sampling study will only be conducted on days when there are no employees in training present.

4.3.3 Output measures

The output measures in this study are the proportions of time spend on tasks of each category and the average task times for resolving incidents and creating RVOs.

4.3.4 Activity categories

The categories in this case are ‘resolving incidents’, ‘creating RVOs’, ‘being in a telephone call’, ‘being in a transceiver call’ and ‘idle time’. Furthermore, there is one category called ‘other activities’, such as working on portfolio tasks, communicating with the OvD-I, checking e-mails etcetera. In the study design, the category ‘other activities’ is currently not considered, since there is currently no estimate available for the time spent in this category. The first four categories consist of reactive, urgent tasks. The fifth category: ‘other activities’, contains tasks that are not urgent, but useful and related to work. The sixth category: ‘idle time’ contains all the moments when subjects are not performing useful work for ProRail.

Table 4: Categories and corresponding numbers used in work sampling study

Category	Number
Resolving incidents	1
Creating RVOs	2
In a telephone call	3
In a transceiver call	4
Other activities	5
Idle time	6

4.3.5 Study design

In Table 5 and Table 6, the parameters that are used for the maximum deviation from the true proportion and the confidence interval (CI) are shown. Different maximum deviations from the actual value of the proportion are accepted, based on the size of the proportion. In Table 6, the observations required to validate the estimate of the proportion per category is depicted. The z-value comes from the normal distribution: the z-value is the outcome of the inverse of the normal distribution for $\alpha/2$, where alpha in this case is 1-0.95=0.05.

Table 5: Parameters chosen for work sampling study

Description	Value
CI	95%
Z-value of 95% CI	1,95996

The estimates $\hat{p}$ for the proportions of the time spent on work for each category stem from the workload estimate. The sum of time spend on a category is divided by the total time available in the same period. The estimated time spent on work of each category is calculated for the hours in which the observations are made. It would not make sense to use estimates that are (partly) based on for example hours in the nights, or Sundays and Mondays, while no measurements are taken during these hours or these days. The periods that are selected for the workload estimate resemble the hours that are observed. In Section 4.2.3, two parameters for the workload estimation were unknown: the fixed time impact for incidents ($FT_{incidents}$) and the fixed time impact for RVOs (FT_{RVOs}). To obtain an estimate of the proportion for every category, an estimate of these two parameters is needed. This is attained by averaging the answers given to Question 10 and Question 11 from the survey, which resulted in the following values: $FT_{incidents}$ = 2.79 (2 minutes and 47 seconds) and FT_{RVOs} = 2.34 (2 minutes and 21 seconds). The full lists of answers are stated in Table 34 and Table 35.

The CIHW is calculated by dividing the maximum deviation by the z-value (≈1.96). Only for the transceiver it is different: since the lower bound would be below 0, the CIHW is calculated by taking the difference between the UB ($\hat{p}$) + 0.01 (max deviation) and dividing it by 2. The number of observations required stems from the following formula from Groover (2013):

$$n = \frac{\hat{p}(1-\hat{p})}{\hat{\sigma}_p^2}$$

Where $\hat{\sigma}_p^2$ is the confidence interval half-width squared. The values obtained for n are stated in the fifth row of Table 6.

Table 6: Observations required for each category

	Incidents	RVOs	Telephone calls	Transceiver calls	Other activities + Idle time
Estimate of p ($\hat{p}$):	10.43%	2.63%	13.57%	0.65%	72.71%
Max deviation of true proportion	4pp	4pp	4pp	1pp	8pp
CIHW:	2.041%	2.041%	2.041%	0.743%	4.082%
Observations required:	266,56	177,59	251,26	87,57	132,96
Obs. Req. Rounded up:	267	178	252	88	133

Regarding the practical design of the study, several things are important. Every hour, four observations are made. With, on average, three workers present, a total of 4*3=12 observations are made per hour and 12*8=96 observations per shift. With the parameters in Table 5, a total of 267 observations are required for the category 'Incidents'. 267 divided by 96, rounded up to the nearest integer results in three days of measuring. After these three days, the proportions found will determine if more observations are needed and how many. The last two days can be used for this purpose.

A pseudo random number generator in Excel VBA is used for generating random numbers that determine the times on which the measurements are taken. The random number generator is initialized with a seed value, so that the same random numbers can be generated again. The numbers are stratified per hour, so an equal number of observations are done per hour (and thus also per day). The procedure draws four random numbers and sorts these from first to last time. It then checks if there is at least a 5-minute difference between every number and between the first number of this list and the last number of the previous set of 4 random numbers. If that is all true, the numbers are stored on the sheet and converted to actual times. If one of the requirements is not met, four new random numbers are drawn, and the process continues until all 8*4=32 sampling times for the shift are generated.

A procedure in VBA is started as soon as the observer starts measuring. The code ensures that a message (with audio) is displayed every time an observation should be done.

4.3.6 Identify observers

In this study, the researcher is the only observer.

4.3.7 Final steps of the work sampling study

The work sampling study has been announced on the information portal of the MKS in the second week of January '22. The observations are made on January 15, 24, 27, 28 and February 2. The results are analysed and presented in Chapter 5.

4.4 Scheduling model

In the following sections, the scheduling model is explained. The objective and constraints are first explained, and a mathematical model is provided. Hereafter, different model extensions are explained, and the solving approach is discussed.

In the paper of (Ernst, Jiang, Krishnamoorthy, & Sier, 2004), which concerns scheduling problems, the authors argue that the **set-covering model** is so general that many problems in staff scheduling, and rostering can be described in this unified format. Furthermore, they explain that the **elastic set-partitioning model** is a useful variation, since it allows both under and over coverage. In this research, we deal with work that is reactive to a volatile workload, and therefore allowing both under and over coverage is a logical choice.

4.4.1 Model introduction

The goal of the scheduling model proposed in this research is to determine the minimum capacity that is needed to meet a specific quality standard. This quality standard is defined as meeting a predefined percentage of periods where the capacity present per period p is higher than the estimated workload per period p. This percentage is called the service level. Simultaneously, we want to minimize both the overcapacity and undercapacity. This is achieved by penalizing these factors per time period p, where undercapacity (y_p^-) is penalized with a higher weight g than the overcapacity (y_p^+), which is penalized with a factor f. The formulation is based on a paper of Eveborn & Rönnqvist from 2004. However, working on specific tasks has been left out of consideration, and constraints are added that ensure that the intended service level is met. The model is based on the elastic set-partitioning model. In the paper, the branch-and-price method is used to obtain solutions (Eveborn & Rönnqvist, 2004).

The model determines the number of employees (x_i) that start in each period i. These periods (i) are the 168 periods of one hour in a week (7 days * 24 hours = 168 periods). The schedule repeats itself every week, which is the reason that the model only needs to determine the number of employees that start for every period i, instead of every period p. How the model ensures that the weekly schedule is extended to all periods is explained in Section 4.4.4. For every period, the amount of over- or undercapacity is measured with the variables y_p^+ and y_p^- and the number of understaffed hours is tracked with the variable w_p.

Furthermore, the parameters 'f' and 'g' determine the weights that are assigned to one hour of over- and undercapacity, respectively. The variable 'SL' stands for the service level. This parameter determines the maximum proportion of hours that may be understaffed. WE_p is the workload estimate for period p that results from the workload estimation model. The values of a_{ip} link the employees that start in period i (x_i) to the corresponding periods p. The parameter P is the total number of periods p that is used as input for the model and K is the number of weeks of data in the sample. This parameter K is also used in the objective function, to ensure that the weight of each factor that the model tries to minimize (the number of shifts per week, the undercapacity and the overcapacity) is not dependent on the size of the input sample. Finally, the parameters AST_i are used to restrict the start times that the model can use. The decision-maker can limit the allowed start times (AST_i) to ensure that a convenient schedule is obtained.

4.4.2 Mathematical model formulation

A model with the following variables is proposed:

$x_i = nr\ of\ employees\ starting\ in\ period\ i$

$y_p^+ = overcapacity\ in\ period\ p\ in\ hours$

$y_p^- = undercapacity\ in\ period\ p$ in hours

$$w_p = \begin{cases} 1, & if\ there\ is\ undercapacity\ in\ period\ p\ (y_p^- > 0) \\ 0, & otherwise \end{cases}$$

The following parameters are required for the model:

$f = the\ penalty\ given\ to\ one\ hour\ of\ overcapacity$

$g = the\ penalty\ given\ to\ one\ hour\ of\ undercapacity$

$SL = service\ level\ (\%\ of\ hours\ where\ capacity > Workoad\ estimate)$

$WE_p = Workload\ estimate\ for\ period\ p\ (estimate\ of\ needed\ capacity)$

$$a_{ip} = \begin{cases} 1, & if\ employees\ that\ start\ at\ time\ i\ work\ in\ period\ p \\ 0, & otherwise \end{cases}$$

$P = total\ number\ of\ time\ periods\ used\ in\ the\ sample\ for\ optimization$

$K = number\ of\ weeks\ of\ data\ in\ the\ sample\ \ (\frac{P}{I})$

$$AST_i = \begin{cases} 1, & if\ employees\ are\ allowed\ to\ start\ at\ time\ i \\ 0, & otherwise \end{cases}$$

Furthermore, we use the following index sets:

I: *set of periods in which an employee can start*

P: *set of time periods used for the optimization*

Goal: Determine the minimum capacity needed so that the capacity exceeds the estimated workload SL% of the time, while overcapacity and undercapacity are minimized.

The model can be stated as:

$$Min\ z = K * \sum_{i \in I} x_i + \sum_{p \in P} (f y_p^+ + g y_p^-)$$

s.t.

$$\sum_i a_{ip} x_i - y_p^+ + y_p^- = WE_p \qquad \forall p \qquad (8)$$

$$\sum_i a_{ip} x_i \geq 1 \qquad \forall p \qquad (9)$$

$$\sum_i a_{ip} x_i \leq 6 \qquad \forall p \qquad (10)$$

$$x_i \leq 6 * AST_i \qquad \forall i \qquad (11)$$

$$M w_p \geq y_p^- \qquad \forall p \qquad (12)$$

$$\sum_p w_p \leq P * (1 - SL) \qquad (13)$$

$$y_p^+, y_p^- \geq 0 \qquad \forall p \qquad (14)$$

$$x_i \geq 0, integer \qquad \forall i \qquad (15)$$

$$w_p \in \{0,1\} \qquad \forall p \qquad (16)$$

4.4.3 Explanation constraints

In the objective function, the sum of the workers (x_i) is minimized in combination with the penalties for overcapacity and undercapacity. Notice that the overcapacity and undercapacity are expressed as real positive number and they do not have to be integers. The weights f and g determine how much influence each of the three factors (personnel costs, cost of overcapacity and cost of undercapacity) have. The model can easily be extended by incurring personnel costs c or c_i, which could make the model more comprehensible. Furthermore, constraint (8) ensures that x_i is an integer greater than or equal to 0.

Constraint (8) ensures that the overcapacity and undercapacity receive the correct values. The factor $\sum_i a_{ip} x_i$ represents the capacity present in period p, since a_{ip} = 1 when a worker that started in period i is present in period p. Since a_{ip} is one of the parameters, the values for a_{ip} for each combination of i and p are known upfront. The capacity present in period p is made equal to the workload estimate for this period (WE_p) by either subtracting the overcapacity y_p^+, or by adding the undercapacity y_p^-. Constraint (7) ensures that y_p^+ and y_p^- are greater than or equal to 0.

Constraints (9) and (10) ensure that there is always at least one worker and at most six workers present at the MKS. Constraint (11) ensures that workers will only start in periods that are allowed by the decision maker. The decision maker might for example want to exclude start times that are during the night. The AST_i is multiplied with 6 to ensure that multiple employees can start at the same time, but no more than 6.

Constraints (12) ensures that the total number of periods with undercapacity is counted correctly. This number is needed to calculate the service level. The 'M' in this constraint represents a large number, which ensures that M * w_p is always greater than y_p^- in case w_p = 1. Constraint (16) ensures that w_p is equal to either 0 or 1.

Constraint (13) ensures that the service level is met, by setting the sum of all hours where there is undercapacity equal to one minus the service level multiplied with the total number of periods. When the sample contains for example 20 weeks, this means that there are 20 * 7 * 24 = 3360 periods. In case the service level is set to 98%, there may at most be 3360 * (1-0.98) = 67.2 periods with undercapacity. Therefore, the sum of w_p (periods with undercapacity) may not be greater than 67.

4.4.4 Extending the weekly schedule to all periods

To get a feasible schedule, the values of a_{ip} must be filled correctly. The value is 1 when a worker that starts in period i works in period p. Table 7 underneath shows the values of a_{ip} for several i and p:

Table 7: Segment of the values of a_{ip}

i \ p	1	2	3	4	5	6	7	8	9	10	⋯	3360
1	1	1	1	1	1	1	1	1	0	0	⋯	0
2	0	1	1	1	1	1	1	1	1	0	⋯	0
3	0	0	1	1	1	1	1	1	1	1	⋯	0
4	0	0	0	1	1	1	1	1	1	1	⋯	0
5	0	0	0	0	1	1	1	1	1	1	⋯	0
6	0	0	0	0	0	1	1	1	1	1	⋯	0
7	0	0	0	0	0	0	1	1	1	1	⋯	0
8	0	0	0	0	0	0	0	1	1	1	⋯	0
9	0	0	0	0	0	0	0	0	1	1	⋯	0
10	0	0	0	0	0	0	0	0	0	1	⋯	0
:	:	:	:	:	:	:	:	:	:	:	⋰	:
158	0	0	0	0	0	0	0	0	0	0	⋯	0
159	0	0	0	0	0	0	0	0	0	0	⋯	0
160	0	0	0	0	0	0	0	0	0	0	⋯	0
161	0	0	0	0	0	0	0	0	0	0	⋯	1
162	1	0	0	0	0	0	0	0	0	0	⋯	1
163	1	1	0	0	0	0	0	0	0	0	⋯	1
164	1	1	1	0	0	0	0	0	0	0	⋯	1
165	1	1	1	1	0	0	0	0	0	0	⋯	1
166	1	1	1	1	1	0	0	0	0	0	⋯	1
167	1	1	1	1	1	1	0	0	0	0	⋯	1
168	1	1	1	1	1	1	1	0	0	0	⋯	1

Table 7 also shows the relationship between the first periods of the week and the last periods. There are 7 * 24 = 168 periods per week and someone who starts working in period i = 164, will work the last 5 periods of the week (164, 165, 166, 167 and 168) and the first 3 periods of the week (1, 2 and 3). The last 5 hours of the week corresponds with Sunday evening from 19:00 – 00:00 and the first 3 hours of the week correspond with Monday morning from 00:00 – 03:00. Since period 3360 is the last period of a 20-week sample it corresponds with Sunday evening from 23:00 – 00:00. Therefore, this period is served if someone starts at i = 161, 162, …, 168. In this example, dispatchers work for 8 hours straight, and breaks are not included. It results in a grid with a diagonal pattern, where for each row i, the number of 1's is equal to P/168*8 and for each column p, the number of 1's is equal to 8. Since for every period p there are 8 moments when dispatcher can start where he or she will be present in period p. For every period i where a dispatcher starts working, he or she will be present in 8 periods times the number of weeks, because the schedule repeats itself weekly. This does not have to be the same dispatcher.

A procedure fills all values for a_{ip} before the MILP is solved. The values are determined in the following way:

$\boldsymbol{If}\ p \bmod 169 \ \leq 7\ \boldsymbol{Then}$
  $\boldsymbol{If}\ (p \bmod 169 + 161 \leq i)\ \boldsymbol{or}\ (p \bmod 169 \ \geq i)\ \boldsymbol{Then}$
    $a_{ip} = 1$
  $\boldsymbol{Else}$
    $a_{ip} = 0$
  $\boldsymbol{End\ If}$
$\boldsymbol{ElseIf}\ (p \bmod 169 - 7 \leq i)\ \boldsymbol{and}\ (p \bmod 169 \ \geq i)\ \boldsymbol{Then}$
  $a_{ip} = 1$
$\boldsymbol{Else}$
  $a_{ip} = 0$
$\boldsymbol{End\ If}$

4.4.5 Training data and test data

Since the model explained in the previous sections treats the data as deterministic, a unique approach has been chosen to evaluate the performance of different schedules. For each experiment, the input dataset is split into a training dataset and a test dataset. The training dataset is used as input for the scheduling model. The scheduling model generates a schedule that is optimal for this training data. However, an optimal solution on the training data does not automatically guarantee an optimal solution for future workload realizations. To evaluate each schedule's performance on 'future realizations', the test dataset is used. If a schedule performs well on the test dataset, we can conclude that the schedule generated by the scheduling model is a robust schedule.

The training and test datasets are mutually exclusive, while containing an equal amount of data, with the same seasonality influences (the same number of datapoints from different weekdays and different months). This ensures that the schedule is tested fairly, since it is created for the same period as it is tested on, without having prior knowledge of the actual workload realizations.

4.4.6 Model assumptions

Several assumptions are made to simplify the model. The three most important assumptions are described in this section:

- **No breaks:** Breaks are not included in the model. In the current situation, dispatcher often do not have official breaks too.
- **No transfer of work:** When the workload is higher than the capacity, it is expected that dispatchers do not have enough time to fulfil all the tasks related to this workload. In reality, some of the work might be transferred to the subsequent hour. However, it would be complex to include this in the model in a realistic way and therefore it is assumed that the workload does not transfer from one hour to the next.
- **Work pace is constant:** It is assumed that work is executed with a constant pace. So, the pace does not depend on the workload. However, the task times are inflated to deal with some deviations, which is explained in Section 5.1.5.

4.5 Model extensions

The base model in Section 4.4 simultaneously minimizes the capacity (total number of workers needed), the overcapacity and the undercapacity. It does this while meeting certain constraints, such as a service level constraint and minimum and maximum capacity per hour constraints. However, several requirements that might result in a more convenient schedule are not included. In the following sections, model extensions that include these demands are explained. Examples of such requirements are for example: matching the start times of shifts in different days, fixing the total capacity upfront and determining the capacity only for a part of the week.

4.5.1 Linked start times

With the base model in Section 4.4, there are no constraints that schedules should look similar for different days of the week. However, it is convenient to have start times for shifts on different days of the week. For example, a schedule where every worker starts at either 7:00, 15:00 or 23:00 every day is more convenient than a schedule where workers start at 7:00, 15:00 or 23:00 on Monday, while they start on 6:00, 8:00, 14:00 and 23:00 on a Tuesday, another set of different start times on Wednesday, etc.

There are several options to cope with this inconvenience. For example, one can limit the allowed start times (AST_i) for every day to a few specific options. However, when you leave only three or four start times open per day, the solver would be very limited in its options. When you leave more options open (e.g., 6 – 8 per day), you could still end up in a situation where every day has different start times.

Another approach is to add a constraint that ensures that no employees start at time i when no employees started at the same time the day before (i-24). This provides the model more freedom than limiting the allowed start times to a few options, while ensuring that employees start at similar times every day. The following constraint is added in this case:

$$x_i \leq M * x_{i-24} \qquad for\ i = 25, 26, \ldots, 168 \qquad (17)$$

In constraint (17), the M is a number that ensures that more shifts are allowed to start at time i than at time i – 24. In this case, M = 6, since it is never allowed that more than 6 shifts start at the same time.

With constraint (17) in place, it is in theory still possible to have many different start times for shifts per day, but this will never result in a good solution. Since the constraint requires that capacity can only be deployed if capacity is deployed at the same time on the previous day, the model will result in a situation with either a limited number of start times (and low/regular total capacity) or with a higher number of start times and thus high capacity. Since high capacity is expensive and results in high overcapacity penalties, this is an option that will not be selected by the model. Therefore, the number of start times will remain limited, resulting a schedule that looks similar every day.

To better exploit the differences in demand between the weekdays and the weekend, it might be beneficial to allow different start times in the weekend compared to weekdays. One can easily adapt constraint (17) to this situation by changing for which values of i the constraint holds. If one would allow different start times for shifts in the weekend, the constraint would look the following:

$$x_i \leq M * x_{i-24} \qquad for\ i = 25, 26, \ldots, 120\ and\ for\ i = 145, 146, \ldots, 168 \tag{18}$$

Furthermore, it is also possible to use constraints (17) and (18) in combination with constraint (11), where a number of start times are manually included by setting the parameter AST_i to 0 for several i's, to ensure that no employees will start at those specific times.

4.5.2 Fixed total number of start times

In Section 4.5.1, it is explained how the start times might be linked to each other, which improves the practicality of the schedule. However, the model is still allowed to choose a relatively high number of start times. For example, if the model decides to deploy 56 shifts over the course of the week, then 8 shifts will start each day. The model can decide to let these shifts start at 8 different start times, which provides it with more flexibility for the subsequent days. This is more than twice as many start times as there are in the current schedule, and it reduces the clarity of the schedule.

To cope with this issue, we can extend the model with a new variable and two new constraints to ensure that the total number of start times remains limited. A new binary variable v_i is used to count the number of start times and a constraint is added that limits the number of start times v_i to a value of a new parameter: Total Start Times (TST). The following is added to the model of Section 4.4:

$$v_i = \begin{cases} 1 & if\ one\ or\ more\ workers\ start\ at\ time\ i\ (x_i \geq 1) \\ 0 & otherwise \end{cases}$$

$$Mv_i \geq x_i \qquad \forall i \tag{19}$$

$$\sum_i v_i \leq TST \tag{20}$$

$$v_i \in \{0,1\} \qquad \forall i \tag{21}$$

Constraint (19) ensures that v_i is 1 when one or more workers start at time i. M in this equation must be equal to the maximum number of workers that are allowed to start at one time, which is 6 in case the general model is followed. Constraint (20) ensures that no more than TST different start times are utilized. And finally, constraint (21) ensures that v_i is binary.

The model extension described in this section can be applied on its own, however it is sensible to use it in combination with the model extension of Section 4.5.2. Otherwise, the start times might still differ a lot every day, resulting in an impractical schedule.

4.5.3 Selecting one day shift

In some cases, having one day shift instead of a morning and an evening shift might be sufficient. Such a day shift would have overlap with both the morning and the evening shift and therefore it would normally start either at 9:00, 10:00, 11:00 or 12:00. With the constraints from Sections 4.5.1 and 4.5.2, we can limit the total number of start times and link them on different days, however there are some limitations with these constraints. For example, if constraint (17) is enforced, then it is for example not possible to have a day shift at 10:00 on Tuesday, no day shift at Wednesday and again a day shift at 10:00 on Thursday, since the model is only allowed to assign capacity to a specific start time if capacity was assigned to the same start time on the previous day. Since one might want to allow this option in reality, a different set of constraints is proposed. Furthermore, one support variable is introduced. The following set of constraints should be used in combination with the model extension of Section 4.5.2.

$$B_l = \begin{cases} 1 & \textit{if the day shift of set } l \textit{ is selected} \\ 0 & \textit{otherwise} \end{cases}$$

$$v_{10} + v_{34} + v_{58} + v_{82} + v_{106} + v_{130} + v_{154} \leq 7B_1 \quad (22)$$

$$v_{11} + v_{35} + v_{59} + v_{83} + v_{107} + v_{131} + v_{155} \leq 7B_2 \quad (23)$$

$$v_{12} + v_{36} + v_{60} + v_{84} + v_{108} + v_{132} + v_{156} \leq 7B_3 \quad (24)$$

$$v_{13} + v_{37} + v_{60} + v_{85} + v_{109} + v_{133} + v_{157} \leq 7B_4 \quad (25)$$

$$\sum_l B_l \leq 1 \quad (26)$$

$$B_l \in \{0,1\} \qquad \forall l \quad (27)$$

In the set of constraints above, the support variable B_l = 1 when the day shift of set l is selected. The four constraints (22) - (25) ensure that all v_i's that corresponds to one specific shift are either 0 (when the shift is not included in the schedule), or that the model is free to choose for each v_i whether it is 1 or 0. For example, when B_3=1, that means that the day shift of 11:00 is selected and therefore, the v_{12} is allowed to be 0 or 1, v_{36} is allowed to be 0 or 1, etcetera. On its turn, if v_{12} = 1, then the model is allowed to assign capacity to x_{12}, while this is not the case when v_{12} = 0.

When one of the B_l's = 1, constraint (26) ensures that no other B_l can be 1. This constraint can simply be relaxed to allow for more day shifts by setting the right-hand side to a larger integer of choice. Furthermore, constraint (27) ensures that all B_l's are binary.

4.5.4 Fixed total capacity

Another extension of the base model that is often practical is to fix the total capacity upfront. By restricting the total capacity upfront, the model distributes the capacity in such a way that it matches the workload the closest. This model extension is useful for planners that must deploy the capacity when the total capacity available is known. The following constraint is added to ensure that all the available capacity is deployed:

$$\sum_i x_i = TC \quad (28)$$

In constraint (28), TC stands for total capacity. Furthermore, the 'is equal'-sign can be relaxed to a 'is smaller than or equal to'-sign, that ensures that a maximum capacity is determined upfront. Note that setting a low TC in combination with a high service level (SL) might result in infeasible models.

4.5.5 Determining capacity for part of the week

A problem that management of the MKS often encounters is that capacity unexpectedly drops and that therefore, the schedule must be changed to cope with this drop in demand. In most cases, a couple of shifts (that one employee would work) are affected. Therefore, it is often not desirable to change the schedule for the whole week.

To ensure that the schedule is not changed for days that are not impacted by the drop in capacity, the decision variables x_i can be turned into parameters. Turning a part of the decision variables x_i into parameters automatically turns the portion of variables that is solely dependent on these decision variables into parameters as well. The constraints can be kept the same. It might be useful to set the total capacity to a fixed number upfront to ensure that all shifts that one wants to deploy are deployed.

4.6 Solution space and solving approach

In the current situation (Q1, 2022), at each of the 168 start times, between 0 and 6 shifts can be planned. Every week must have the same (base) schedule and every shift must start at the beginning of an hour. This means that the solution space has a size of $168^7 \approx 3.77*10^{15}$. Although only a small proportion of this number comprises the feasible solutions, it takes too long to review all of them. Furthermore, this is without considering the uncertainty in the input variables.

The problem can be classified as an MILP-problem. In the article of (Ernst, Jiang, Krishnamoorthy, & Sier, 2004), several methods are explained that are often used to find good solutions for scheduling and rostering problems. Constraint programming is mentioned as an effective approach for highly constrained problems, where a feasible solutions will suffice even if it is not optimal. Furthermore, it is explained that metaheuristics are often used to solve hard combinatorial optimisation problems. Metaheuristics are relatively robust, produce reasonably good feasible solutions in a limiting amount of running time, are relatively simple to implement and can deal with complex objectives. Because of these advantages it was initially decided to use metaheuristics for finding solutions. However, these metaheuristics only found small improvements in the objective. Therefore, a separate approach has been chosen.

The model is programmed in Excel and the OpenSolver add-in from the University of Auckland is used to set up the problem. Then, Gurobi, a commercial solver, is used to solve the MILP problem to optimality. With a few months of data, it takes the model around 10 minutes to solve it to optimality. With multiple years of data, the model takes a couple hours to solve it to optimality.

4.7 Conclusions

In chapter 4, the workload estimation model and the scheduling model are explained. With the workload estimation model, the workload is estimated for every period p in the past, based on data about incidents, telephone calls, RVOs and transceiver communication. The data available for the workload estimation covers all dates between 1-1-2018 and 30-9-2021. The estimate for every period p serves as input for the scheduling model. The workload that stems from specific telephone calls and transceiver conversations has been registered and is directly used for the workload estimate. The average workload that results from incidents and RVOs is determined in the work sampling study. The results of this study are presented in Section 5.1.

The scheduling model is formulated as an elastic set-partitioning model. This model closely follows the requirements of a schedule for the MKS, since it allows for both under- and overcapacity. The model assigns capacity to different shifts over the week in such a way that the capacity is minimized while the service level is met. Furthermore, undercapacity and overcapacity are both penalized with their respective weights. Additional constraints were introduced to ensure that the capacity per hour would keep between a minimum and maximum. The parameter a_{ip} that determines which shift(s) each period is covered is determined with a straightforward algorithm.

Several model extensions were added to allow for experiments with different levels of freedom for the model. The introduced extensions were:

- **Linked start times:** This model extension ensures that capacity can only be deployed at start times that are exactly 24 hours after preceding capacity deployments.
- **Fixed total number of start times:** This model extension ensures that the number of different times at which capacity is deployed does not exceed a certain threshold. This ensure that capacity is pooled, which keeps the schedule more practical.
- **Selecting one day shift:** This model extension ensures that only one repeating start time is chosen for the day/relief shift. It would not be practical if the day shift would start at different times every day and these constraints ensure that that is not the case.
- **Fixed total capacity:** This model extension ensures that a fixed number of dispatchers is deployed during the week.
- **Determining capacity for part of the week:** This model extension ensures that the capacity deployment is fixed for a certain part of the week and that the model will only decide the capacity deployment for the other part of the week.

The scheduling problem is solved using the OpenSolver software with Gurobi as solver engine. The solver finds optimal solution. However, it takes a relatively long time to solve one experiment.

5 Experiments

In this chapter, the setup and results of the experiments are explained. The experiments are set up in such a way that a variety of schedules is produced where each schedule is optimized with different requirements or with different input data. The research question that is central in this chapter is the following:

What solutions are proposed for different problem scenarios by the scheduling model?

In Section 5.1, the results of the work sampling study are explained. In Section 5.2, the process of preparing the data for the experiments is explained. In Section 5.3, the experiments and their results are explained. In Section 5.4, an additional set of experiments is introduced that are designed as support the planner when he or she must react to sudden capacity changes. Section 5.6 contains the sensitivity analysis, in which the impact of adjustments in the parameters are analysed. Finally, the chapter ends with the conclusions in Section 5.7.

5.1 Results work sampling study

The design of the work sampling study is described in Section 4.3. In the following sections, the total number of observations that were made in the study is calculated and the estimated proportion per category is explained. Furthermore, the average task times for incidents and RVOs are calculated, and these task times are inflated to normal times.

5.1.1 Observations

On the first observation day (January 15), one of the workers in the early shift had to go in quarantine due to close contact with someone who contracted Covid-19. Therefore, a last-minute change in the schedule translated to a situation where one of the workers from the late shift came earlier (he worked from 11:00 to 19:00). The researcher was present from 9:00 to 17:00 and therefore, there were two hours with less observations. On this day, a total of 4*2*2+4*3*6=88 observations were done (four measuring times per hour and two workers present during the first two hours and three during the last six hours). On the other three days the regular 96 (8 *4*3 = 96) observations were done. This results in a total number of 88+3*96=376 measurements.

5.1.2 Measured proportions

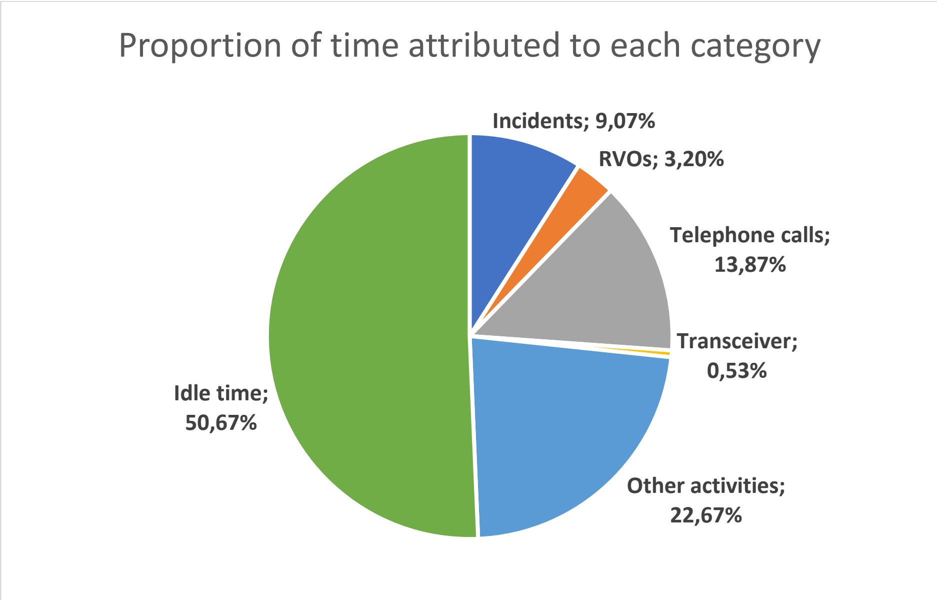
The proportions $\hat{p}$ of work that was spent on each category that were estimated with the historical records and the expert estimation are depicted in the second row of Table 8 ($\hat{p}$ HR+EE). Here, HR stands for 'Historical Records' and EE stands for 'Expert Estimate'. The proportions $\hat{p}$ that were measured on the four observation days in the work sampling study are depicted in the third row of Table 8 ($\hat{p}$ WS). The difference between the two methods is depicted in the fourth row.

Table 8: Measured proportion per workload category

	Incidents	RVOs	Telephone	Transceiver	Others	Idle time
$\hat{p}$ (HR + EE)	10.43%	2.63%	13.57%	0.65%	72.71%	
$\hat{p}$ (WS)	9.07%	3.20%	13.87%	0.53%	22.67%	50.67%
Difference	1.36pp	0.57pp	0.30pp	0.12pp	0.63pp	

Table 8 can be read in the following way. From the estimate based on expert knowledge (in the survey), an estimate for the average time spent on an intake, logging information, and creating an RVO was obtained. Equations (2) - (7) and the historical records (number of incidents, loglines and RVOs during the observed hours) were used to calculate the time impact (TI) of each category during the observed hours. Hereafter, the total time impact was divided by the total capacity that had been present during these hours. For the category 'creating RVOs', the expected proportion of time spent on tasks of this category was for example 2.63%. A difference of 0.57 percentage points (pp) with the observed fraction of time spent on tasks of this category, which was 3.20%. In the next section the main takeaways from Table 8 are elaborated.

Figure 15: Pie chart of results work sampling study



5.1.3 Analysis work sampling study

From the results of the work sampling study, several things stand out. These findings are summarized in the bullet points below.

- **The proportions do not deviate much:** All predictions are within the maximum allowed deviation that was set before the incidents (see Table 6).
- **Relatively large proportion of time spent on other activities:** Beforehand, there was no estimate of time spent on other activities. Activities in this category comprise working on ancillary tasks, video calling with ICB to reflect on the shift, the shift evaluation with the OvD-I, direct evaluations of incident handling and direct communication with other departments at the OCCR (CHI, CaTo, VCNS, OvD-I, etc.).
- **Small proportion of time spent on transceiver communication:** It stands out that the time spent on communication via transceiver is even smaller than predicted. Therefore, one can conclude that this category does not have a substantial impact on the workload.
- **Large proportion of time is idle time:** It stands out that almost half of the time, dispatchers are waiting for work. This is a substantial amount of time and therefore there is a lot of potential there.

5.1.4 Adjustment of task times

According to (Groover, 2013), estimates of average task times based on a work sampling study are in general more accurate than when these estimates are based on expert knowledge. Therefore, the proportions obtained in the work sampling study are used to determine a new (improved) estimate for the average task times. This estimate is only changed for the categories 'incidents' and 'RVOs'. Since the data for the categories 'Transceiver' and 'Telephone' is fairly accurate and we do not have data for the number of calls that took place during the observed hours.

One can determine the average task time by computing the total time associated with one category and dividing it by the number of units that have been processed during the total time. The calculation is summarized in the following equation:

$$T_i = \frac{p_i(TT)}{Q_i} \quad (29)$$

In this equation, T_i is the average task time of task i, p_i is the proportion of observations associated with category i and Q_i is the number of units that have been processed during the total time. In this research, the number of units for the category 'RVOs' is for example the number of RVOs that are created by the MKS during the observed hours. The calculation of T_i is summarized in Table 9.

Table 9: Calculation of task times

Category	P_i	TT (minutes)	Q_i	T_i (minutes)
RVOs	0.0320	5640	73	2.472
Incidents	0.0907	5640	94	5.440

In Table 9, the calculation for the average task times of RVOs and incidents are shown. The proportions p_i were calculated by dividing the number of measurements in the category i during the work sampling study by the total number of measurements. The total time (TT) is calculated by looking at the total time that has been worked during the observation days. For 2 hours, 2 workers were present and during the other 30 hours, 3 workers were present. Therefore, the total time is equal to 2*2+30*3=94 hours or 90*60=5640 minutes. During these the observed hours, 73 RVOs and 94 SpoorWeb dossiers were created. Substituting all these numbers into Equation (29) shows that it took on average 2.47 minutes to create an RVO and 5.44 minutes to resolve an incident. This is the time spent on these categories when dispatchers were not on the phone. The category 'incidents' consists of two parts: the intake and follow-up. The intake is counted once, and the 'follow-up' part is included by setting one minute per logline. Knowing the number of incidents, loglines and time spent on the category incidents during the observed hours, we can calculate the average time per intake: The total time spent on incidents is equal to the total time (TT) multiplied with the proportion of time spent on incidents: 0.0907 * 5640 = 511.36 minutes. In the observation period, 353 lines have been logged for incidents by the MKS. Keeping the estimate of one minute per logline means that 511.36 – 353 = 158.36 minutes have been spent on intakes during the observations period (apart from the time spent on intakes while dispatchers were on the phone). Knowing that there were 94 incidents during the observation period means that on average 158.36 / 94 = 1.685 minutes of work have been spent per incident intake outside the telephone call.

In this study, no performance ratings were connected to observations. Since no two incidents or RVOs are the same, it is difficult to determine the performance of dispatchers when they process the tasks related to these categories. Furthermore, the researcher did not have enough expertise to reliably judge the performance of different dispatchers. Therefore, the 'normal' task times are set to be the same as the average task times.

In conclusion, all workload estimates were in an acceptable range when compared to the results from the work sampling study. The time spent per telephone or transceiver conversation remain based on the historical record, since these are deemed accurate based on the results from the work sampling study. The fixed time spent per incident and RVO have been updated compared to the values from Section 4.3.5. The new estimates have been calculated using the method from (Groover, 2013) and are: $FT_{incidents}$ = 1.685 minutes (1 minute and 41 seconds) and FT_{RVOs} = 2.472 minutes (2 minutes and 28 seconds).

5.1.5 Transformation from normal times to standard times

To set a standard time for tasks in a certain category, other factors that might impact the work pace must be incorporated. The normal times, which are in this study equal to the average task times, are multiplied with a factor A_{pfd}. This factor is the allowance for personal time, fatigue, and unavoidable delays (PFD).

The factor A_{pfd} in this research is set to 10%. Groover (2013) recommends using 15% for PFD. The reason that a lower value is chosen in this case is because we also incorporate 15% idle time per hour. This idle time is primarily meant to have a buffer to ensure that the workload can also be met when much of the work during an hour must be done at the same time. However, this idle time can also be used for personal time and therefore, a value lower than 15% is chosen for personal time, fatigue, and delay. The 15% time that is left open for idle time is based on the analysis of a multinational consulting firm: Voxco. This firm states that call centres should aim to maintain an occupancy rate of around 85% which translates to 15% idle time (Voxco, 2021).

The result of this for the experiments with the scheduling model is that the capacity that is required per hour is equal to the workload estimate multiplied by 1.10. Equation (30) illustrates how the normal task times are converted to standard times:

$$T_{stdi} = T_i(1 + A_{pfd}) \tag{30}$$

In this equation, T_{stdi} is equal to the standard time for tasks in category i, T_i is the normal time for tasks in category i, determined with Equation (29). After this multiplication, the required capacity per hour is determined by multiplying T_{stdi} with (1/0.85). This ensures that 15% of the hour remains free for idle time.

5.1.6 Role of other activities in the model

From the work sampling study, it became clear that dispatchers spent a substantial portion of their time on ‘other activities’ (see Figure 15). However, these other activities are not included in the workload estimation. This is due to the following reasons:

- **Lack of data:** The category ‘other activities’ includes among others: working on portfolio tasks, listening in on conversations of other colleagues, communicating with the OvD-I and team manager, checking e-mails, shift starts, shift evaluations and the shift transfers between consecutive shifts. For these activities, no data about the frequency and duration is available. The data gathered about these activities in the work sampling study is too limited to assign reliable task times to activities in each of these categories. Therefore, it is difficult to estimate the workload that stems from these work activities.
- **Non-urgent/movable tasks:** The tasks in the category ‘other activities’ are primarily non-urgent tasks. Most of the tasks can and will be executed at a quiet moments in case it is busy. Some of the tasks are simply omitted because they are either not that important or they resolve themselves. For example, listening in on another employee will normally only be done if there is an impactful incident and the dispatcher that listens in has no urgent tasks to do. Furthermore, there are tasks that must be executed once, but that are only carried out when it is quiet. These include for example answering e-mails.

In the regular experiments, the category 'other activities' is not included in the workload estimate. With the schedules created in these experiments, and their corresponding capacity levels, the capacity is on average more than twice as high as the average workload. Based on the results of the work sampling study, this should offer enough time to execute all the necessary tasks from the 'other activities' category.

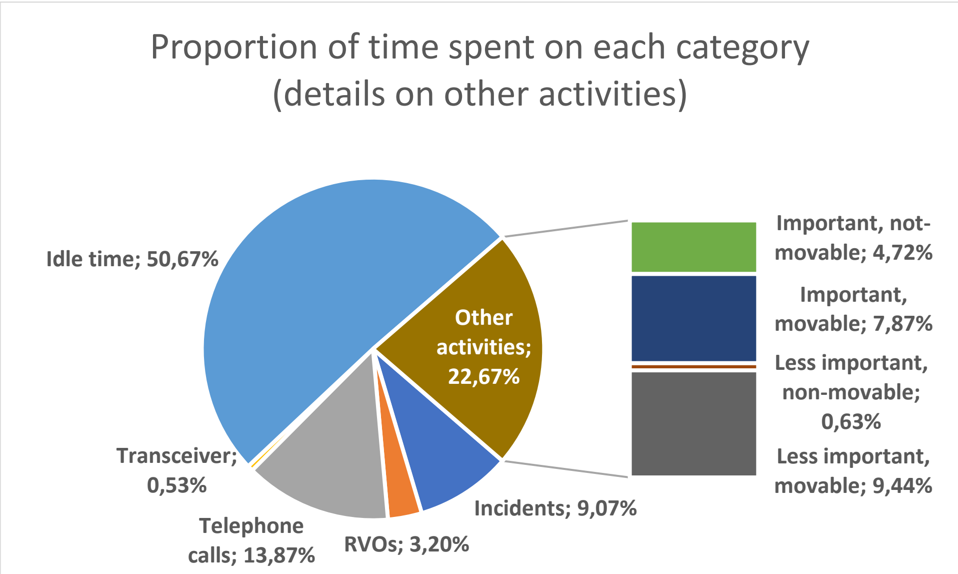
However, to determine the impact of incorporating valuable other activities, an additional set of experiments is run in the sensitivity analysis. The workload estimation that is used as input for these experiments is revised upwards by adapting the inflation factor A_{pfd} (see Section 5.1.5).

Lastly, an analysis has been performed into the other activities to find out whether there were activities that were important and non-movable, since these activities should be considered in the workload estimate. For this analysis, all 85 observations of 'other activities' were grouped into four categories:

- Important, non-movable
- Important, movable
- Less important, non-movable
- Less important, movable

Since there is on average enough overcapacity to fulfil these other activities, we only incorporate the important non-movable activities into the workload. These are activities like the shift evaluation, change of shifts and crucial discussions about incidents with colleagues. It turned out that on average 4.72% of the time is spent on important, non-movable activities. Therefore, this proportion of an hour is reserved for other activities in the workload estimate. The proportion of time spent on each category, including the details about the other activities is portrayed in Figure 16.

Figure 16: Proportion of time spend per category (including details about other activities)



5.2 Data preparation

In Section 5.1.4, the final required data for the workload estimation model was obtained. Inserting the adjusted task times into the model resulted in a new estimate that is used as input data for the experiments. However, the data had to be cleaned and split up in separate samples before the experiments can be run. In this section the steps that have been taken to clean and separate the data are described.

5.2.1 Outlier detection

Firstly, the outliers in the data were filtered out. Considering that a substantial amount of data is available and that each of the 168 hours of the week must appear with the same frequency, it was decided that in case of an outlier, a whole day of data was omitted. A period (one hour) is considered an outlier if the estimated capacity needed to handle the workload in this period is six workers or more. It is physically not possible at the MKS to have more than six people present at the same time and therefore, periods with a workload that high do not add value to the model. The data with all outliers included had 32,856 periods of one hour each. Six days were filtered out, which resulted that the data set after filtering had 32,856-6*24=32712 periods.

Secondly, a whole day of data is omitted when the workload estimate for the entire day (the sum of the hourly workload estimates for that day) was higher than Q3 + 1.5*IQR. This is a general approach for filtering outliers out. Normally, outliers at both sides of the median or filtered out, however Q1 – 1.5*IQR < 0 in this case and there are obviously no negative workload estimates. Therefore, only days with an exceptionally high workload are filtered out. This resulted in the aforementioned snow week in February ’21 being filtered out. In total 8 days were omitted, and therefore 32712-8*24=32,520 remained in the dataset.

5.2.2 Trend analysis

Hereafter, the trend in the data has been analysed and removed. The data had to be de-trended, so that older workload estimates could be used for the model as well. The monthly trend in the data after the outlier detection was plotted and based on this plot, a linear trend in the data was assumed. First a linear regression was executed on the 32,520 periods that were left after filtering out the outliers. Then, the workload estimates for the periods that were filtered out were replaced by interpolation using Holt-Winters multiplicative method without updating. The workload estimate (WE_p) for every period that was filtered out was estimated with the following formula:

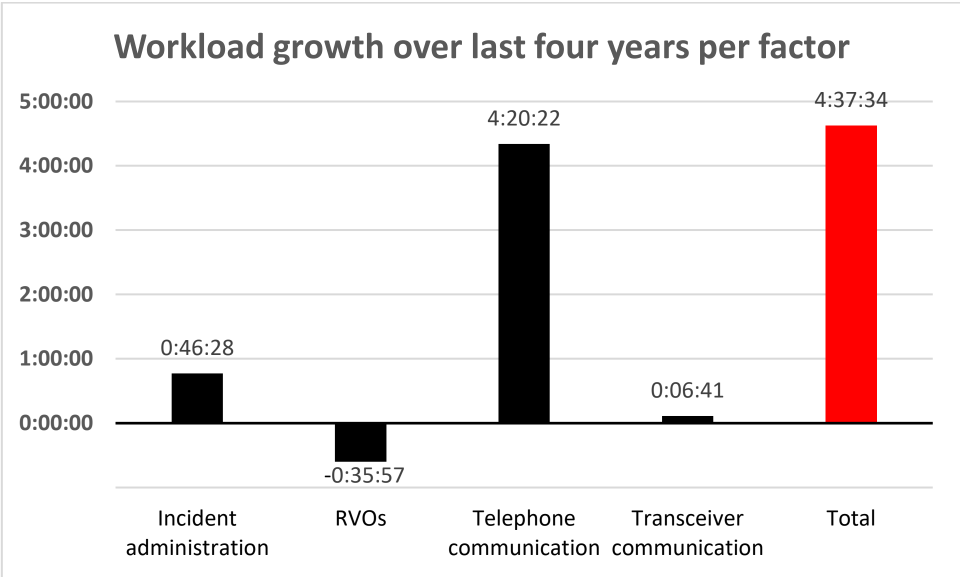
$$WE_p = (a + bt) * F_{p,hour} * F_{p,day} * F_{p,month} \tag{31}$$

Here, a is the level, determined by taking the average over all periods until period p. The factor b is the trend, determined by a linear regression based on the periods that were left after filtering out the outliers. The factor t is the number of periods that the trend has to be multiplied with. The factors $F_{p,hour}$, $F_{p, day}$ and $F_{p,month}$ are the seasonal indices that were determined using the method of (Silver, Pyke, & Thomas, 2016). Here F_p is calculated using the formula $F_p = D_p/\tilde{a}_p$ for every period p where a value of $\tilde{a}_p$ is available. The factor $\tilde{a}_p$ is the P period moving average centred at p. Here P is the number of seasonal factors for each seasonal indicator respectively. For example, for $F_{p,hour}$ the factor P=24, while for $F_{p,month}$ the factor P=12. After all values for F_p are determined, the different F_p’s have been averaged over similar periods in different seasons. For example, an estimate for F_p for July ’18, July ’19, July ’20 and July ’21 have been merged to one estimate of F_p for July in general. In the end, the estimates were normalized so that the sum of the estimates equals P for the hour seasonality (P=24), the day seasonality (P=7) and for the month seasonality (P=12). This normalization is achieved by applying the following formula for every seasonal factor:

$$F_{p,normalized} = \frac{F_p}{\sum_p F_p} \tag{32}$$

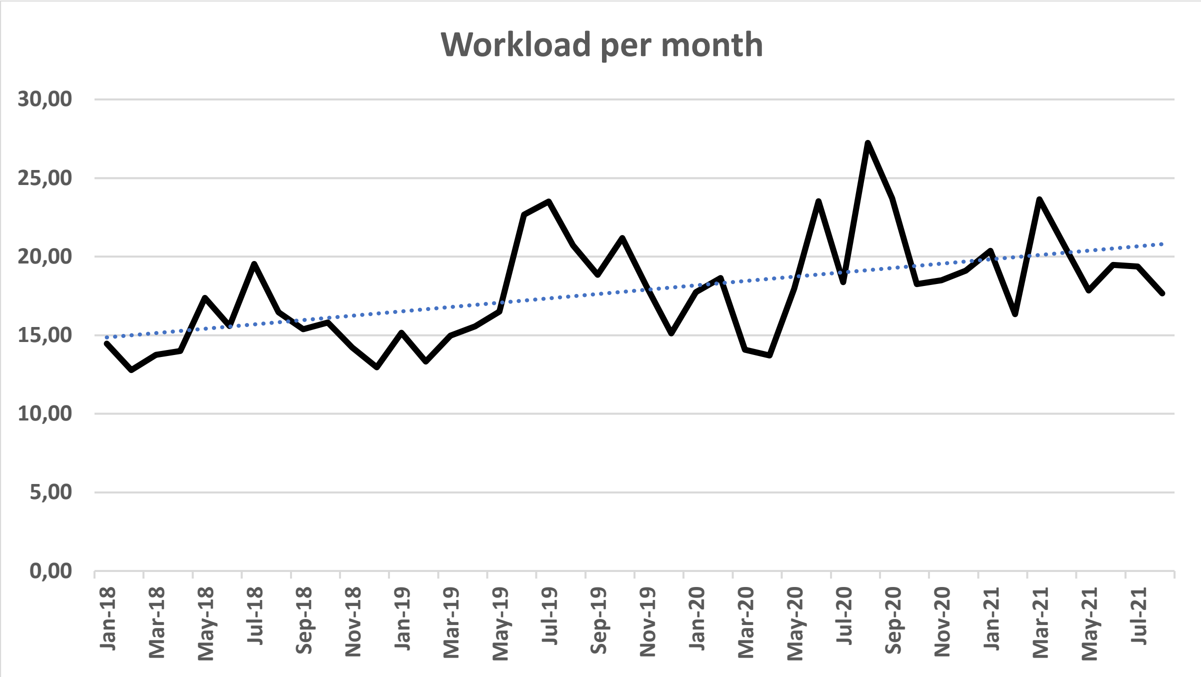
After the new workload estimates WE_p were calculated using Equation (31), a second linear regression was executed to obtain the final estimate for the linear trend in the data. The slope of the regression line per period was equal to $2.441*10^{-7}$. This means that the workload per period grew on average every period with $2.441*10^{-7}$ days. When we multiply this value with the total number of periods in the dataset, it becomes clear that on average $2.441*10^{-7}*32{,}856=8.020*10^{-3}$ days, which is approximately 11:34 minutes. This means that the workload per hour is around 11:34 minutes higher in November ’21 than in January ’18. In hours extra work per day, this is approximately 11:34 minutes * 24 = 4:37:34. To distinguish the different causes of this growth, separate regressions have been run on each workload factor. The results of these regressions are summarized in Figure 17.

Figure 17: Bar chart of the workload growth per workload factor (hours extra work per day)



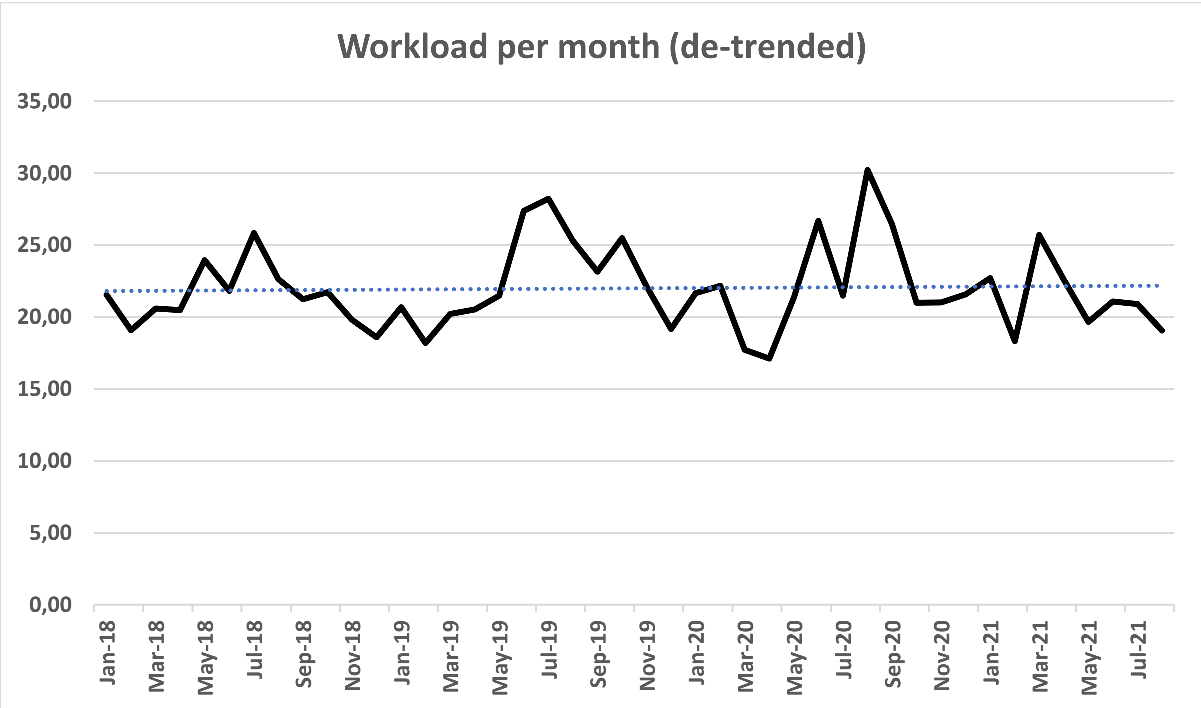
From this analysis it became clear that the upward trend in workload is primarily due to the rise in telephone communication over the years. The time spent on administrating incidents per month has grown with 46:28 minutes in total, the time spent on creating RVOs per month has declined with 35:57 minutes in total, the time spent on telephone communication per month has grown with 4:20:22 hours and the time spent on transceiver communication per month has grown with 6:41 minutes. These factors combined result in a total growth of 4:37:34 hours of work per day. There were in total 45 months in the dataset, with in total 1369 days. The average number of days per month was approximately 30.42, which means that the workload grew in total with 4:37:34 * 30.42 = 5.864 days between January ’18 and November ’21. The growth in days of work per month is shown in Figure 18. Here the trend is clearly visible, and one can also see that the average (based on the linear regression) has grown with 5.864 days (from 15 full days of work per month to more than 21 days).

Figure 18: Workload estimate per month and trend (January '18 – November '21)



The data has been de-trended by subtracting the differences between the observed value and the linear regression and in the end add the level from November '21. In Figure 19, the de-trended workload per month is shown graphically. Notice that the slope of the trend-line has become 0.

Figure 19: Workload estimate per month de-trended (January '18 - November '21)



5.2.3 Seasonality

When we look at the seasonality of the workload, several things stand out:

- **Hour seasonality:** The seasonality impact on an hourly basis is the strongest. The average workload between 3:00 and 4:00 in the morning is around 30% of the average workload, while the average workload at the peak between 15:00 and 16:00 is around 1.4 times as big as the average workload. Furthermore, it stands out that the peak for the telephone communication is later in the day than the peak for the incidents and RVOs.
- **Day seasonality:** For the day seasonality, the difference between the weekdays and the weekend is very clear. The workload in the weekend is around 20% lower than during the week.
- **Month seasonality:** After omitting the data from February '21, it stands out that the summer period is busier than the winter. June, July, and August have by far the highest workload per month, with a workload that is on average 17% higher than the average. The months with the lowest workload are on average at the end of the year and during the spring

The y-axes of the graphs in Figure 21 and Figure 22 do not start at 0. Therefore, the intensity of the day and month seasonality might be overestimated when one bases their judgement only on the figures. In Table 10, the coefficient of variation of the three main workload factors and the workload estimates is provided. The coefficient of variation shows the relative dispersion of the datapoints and is therefore an indicator of the intensity of each seasonal factor. The results in the table show that the seasonal differences between the hours are clearly the largest. The month and day seasonality have an almost equal coefficient of variation.

Table 10: Coefficient of variation of seasonal factors per workload category

Season\Workload factor	Incidents	RVOs	Telephone	Workload
Month	0.060	0.081	0.090	0.085
Day	0.090	0.094	0.073	0.085
Hour	0.344	0.376	0.411	0.366

The seasonality on an hour, day and month basis is depicted in the figures below.

Figure 20: Hour seasonality of workload components (normalized)

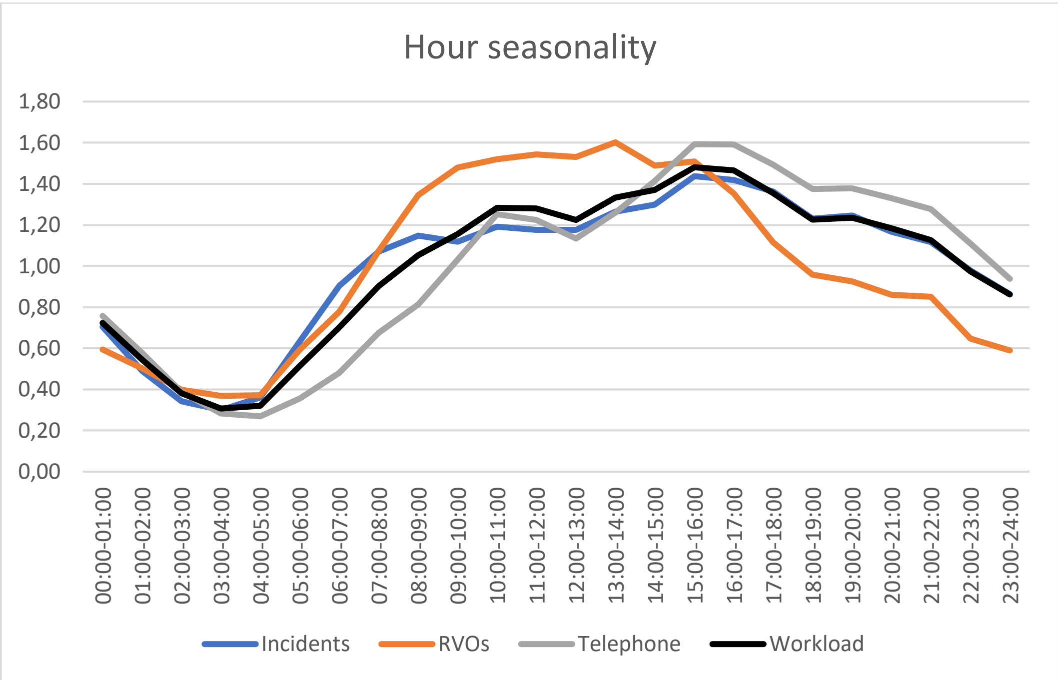


Figure 21: Day seasonality of workload components (normalized)

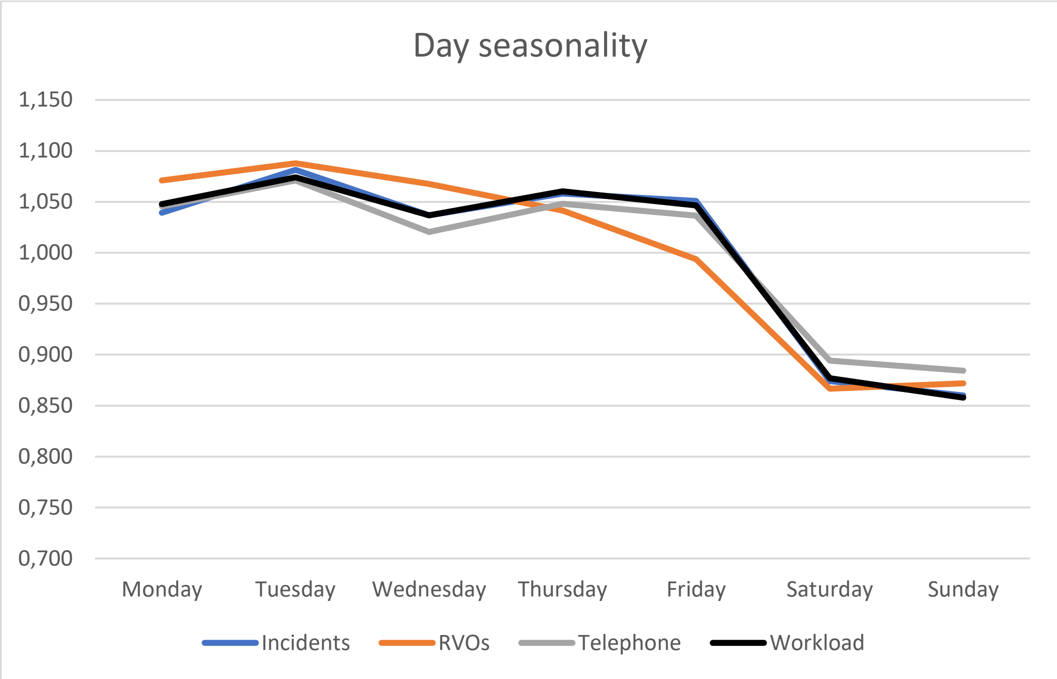
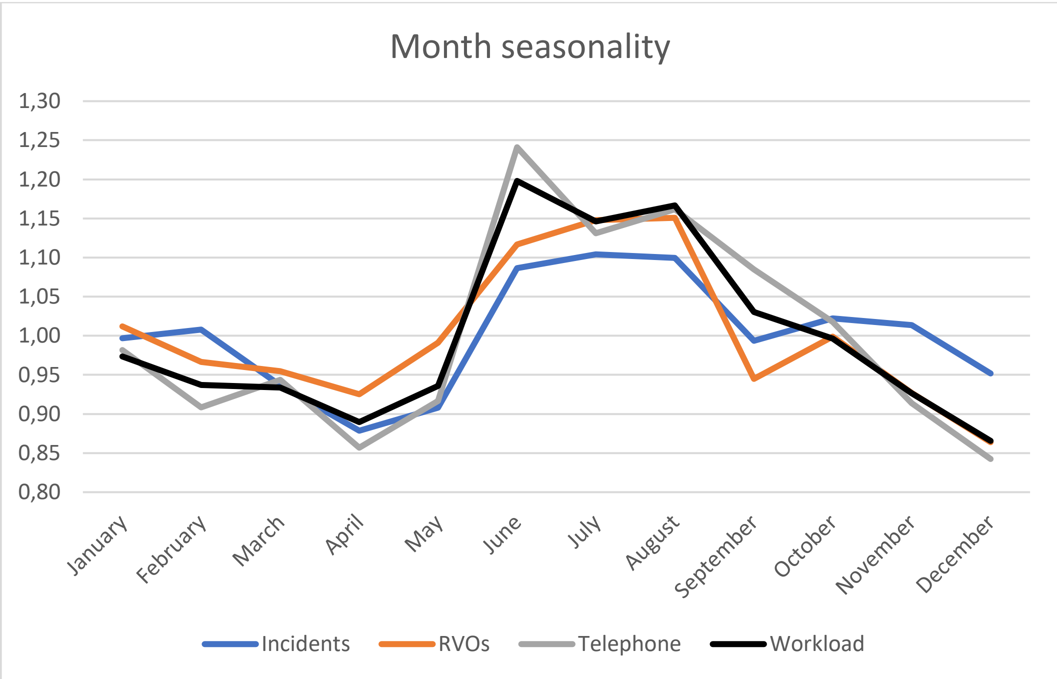


Figure 22: Month seasonality of workload components (normalized)



In the experiments, a weekly schedule is produced where every week has the same shifts. In these experiments, the model can already exploit the hour and day seasonality. However, it cannot exploit the seasonality differences between the different months. To create schedules that fit the seasonality well and that are also robust, it is decided to split the data in two seasons: summer and winter. For the summer period, the months June – September are included, which are the four months with the highest workload on average. For the winter period, the months October – May are included, which are the eight months with the lowest workload on average. To ensure that each season contains the same number of Mondays, Tuesdays, etc., 13 days of data were omitted from the winter season and 26 days of data from the summer season. In total, the winter season contains 868 days of data, and the summer season contains 462 days of data. The information about the two seasons is described in Table 11.

Table 11: Characteristics per season used in the experiments

	Summer	Winter
Months included	June – September (4 months)	October – May (8 months)
Average F_{month}	1.14 (14% higher than the average)	0.93 (7% lower than the average)
Periods in set (P)	11,088	20832
Days in set	462	868
Weeks in set (K)	66	124

5.3 Experiments

In this section, the experiments that were executed are explained. First, the setup of the experiments is explained. Hereafter, the results of the experiments are provided, analysed, and discussed.

5.3.1 Experimental setup

All experiments are run for the winter and summer sample. Furthermore, 2-fold cross validation is used, which means that both datasets are split up in two samples, where each sample functions once as a training sample and once as a test sample. Therefore, every experiment is executed with four different datasets. For all experiments, the capacity per hour needed is determined by dividing the workload estimate by

Five experiments are executed, where each experiment has different constraints regarding the start times. The following five experiments were executed with the four different datasets as input for each experiment:

- **Experiment 1:** In this experiment, the model has the most freedom in its options. There are no limitations on the start times (AST_i = 1 for all i) and there are no constraints that ensure that shifts on different days start at the same times (constraints (17) and (18) are not enforced).
- **Experiment 2:** In this experiment, the following limitations are in place: shifts are not allowed start or end during the night and the start times during the week are synchronized and during the weekend they are also synchronized (thus between the weekdays and weekends there may be differences). This means that constraint (18) is enforced and that AST_i = 1 for a limited number of i's. In Table 38, the allowed start times are shown.
- **Experiment 3:** In this experiment, the following limitations are in place: shifts are not allowed to start or end during the night and the start times are synchronized during the whole week. This means that constraint (17) is enforced and that AST_i = 1 for a limited number of i's. In Table 38, the allowed start times are shown.
- **Experiment 4:** In this experiment, the following limitations are in place: Shifts are allowed to start at the start times of the regular schedule and one hour earlier or later. Furthermore, four start times for a day shift are allowed, all allowed start times AST_i are shown in Table 38. Furthermore, the start times for every shift (morning, day, evening and night) must be the same every day, which is ensured by enforcing constraints (19) - (27).
- **Experiment 5:** In this experiment, the following limitations are in place: shifts are not allowed to start or end during the night and the start times are synchronized during the whole week.

Furthermore, a maximum of 21 start times is imposed. This means that constraint (17), (19), (20) and (21) are enforced and that $AST_i = 1$ for a limited number of i's. In Table 38, the allowed start times are shown.

- **Experiment 6:** In this experiment, the model is the most limited in its options. The open start times are so limited that the model needs all of them and therefore, it can only decide how much capacity is deployed at each start time. Dispatchers can only start at 7:00, 15:00 and 23:00. Therefore, the maximum number of start times is indirectly fixed at a total of 21 and the start times are indirectly linked since every day has the same start times. Therefore, none of the constraints of Section 4.5 must be enforced.

Concerning the parameters that are used in the scheduling model, the penalty given to one hour of overcapacity is 0.01, while the penalty given to one hour of undercapacity is 1. The service level is set at 98% for all experiments. The number of periods P and the number of weeks K of data per experiment are provided in Table 11.

Combining the 6 types of experiments with the 4 data samples, generates 24 different experiments, which are summarized in Table 12. Note that the left side of the table contains the input samples, and the right side contains the constraints for each experiment. Remember that the abbreviation AST stands for Allowed Start Times, TST for total start times and that (ind.) stands for indirect. Indirect in this context means that a certain constraint is not explicitly enforced, but it is attained due to the enforcement of another constraint.

Table 12: Set-up and input all experiments

Exp.	Season	Training sample	Test sample	Synchronization	AST_i	TST
1	Summer	S.1	S.2	None	All hours	Unlimited
2	Summer	S.1	S.2	Week & weekend	Scenario 1	Unlimited
3	Summer	S.1	S.2	Whole week	Scenario 1	Unlimited
4	Summer	S.1	S.2	Whole week	Scenario 2	28
5	Summer	S.1	S.2	Whole week	Scenario 1	21
6	Summer	S.1	S.2	Whole week (indirect)	7:00, 15:00 & 23:00	21 (ind.)
7	Summer	S.2	S.1	None	All hours	Unlimited
8	Summer	S.2	S.1	Week & weekend	Scenario 1	Unlimited
9	Summer	S.2	S.1	Whole week	Scenario 1	Unlimited
10	Summer	S.2	S.1	Whole week	Scenario 2	28
11	Summer	S.2	S.1	Whole week	Scenario 1	21
12	Summer	S.2	S.1	Whole week (indirect)	7:00, 15:00 & 23:00	21 (ind.)
13	Winter	W.1	W.2	None	All hours	Unlimited
14	Winter	W.1	W.2	Week & weekend	Scenario 1	Unlimited
15	Winter	W.1	W.2	Whole week	Scenario 1	Unlimited
16	Winter	W.1	W.2	Whole week	Scenario 2	28
17	Winter	W.1	W.2	Whole week	Scenario 1	21
18	Winter	W.1	W.2	Whole week (indirect)	7:00, 15:00 & 23:00	21 (ind.)
19	Winter	W.2	W.1	None	All hours	Unlimited
20	Winter	W.2	W.1	Week & weekend	Scenario 1	Unlimited
21	Winter	W.2	W.1	Whole week	Scenario 1	Unlimited
22	Winter	W.2	W.1	Whole week	Scenario 2	28
23	Winter	W.2	W.1	Whole week	Scenario 1	21
24	Winter	W.2	W.1	Whole week (indirect)	7:00, 15:00 & 23:00	21 (ind.)

All experiments were run on a Lenovo P51 ThinkPad, which has a 2.9GHz Quad-Core processor and 16GB RAM-memory. The experiments were set up in Excel using the OpenSolver from the University of Auckland (Mason, 2012). The solver engine that was used was that of Gurobi.

5.3.2 Results of the experiments

The solver was able to find an optimal solution for all experiments. In every experiment, the schedule is based on the data in the training set. The results are based on the performance of the schedule in the test set. For every experiment, the following information is provided:

- **Objective value:** This is the value of the objective function for this experiment. It is the sum of the sum of the shifts per week and the weighted overcapacity and undercapacity, calculated as in the objective function of the model in Section 4.4.
- **Service level:** The service level is the percentage of hours where the capacity present is greater than or equal to the capacity required to fulfil the demand in that hour. It is calculated by dividing the number of understaffed hours by the total number of hours and subtracting it from 1: SL = 1 – (understaffed hours / P).
- **Overcapacity per week:** This is the estimated time per week that are not required for meeting the workload that is a result of telephone and transceiver communication or incident administration expressed in hours.
- **Undercapacity per week:** When the required capacity in a specific period is lower than the capacity deployed by the model, the difference between the capacity present and the capacity required is called the undercapacity. The sum of the undercapacity over all period, divided by the number of weeks in the dataset is the undercapacity per week.
- **Number of separate start hours per week:** The practicality of the schedule is primarily defined by the number of separate start hours per week. The lowest number possible is 3, where every day has the same start times and separate shifts do not have any overlap.

In Table 13, The results of the 24 experiments are shown. The table is divided in four sections, where each section contains five experiments where the same set of input-data is used. For more information about the input data and constraints per experiment, see Section 5.3.1.

Table 13: Results of all 24 experiments

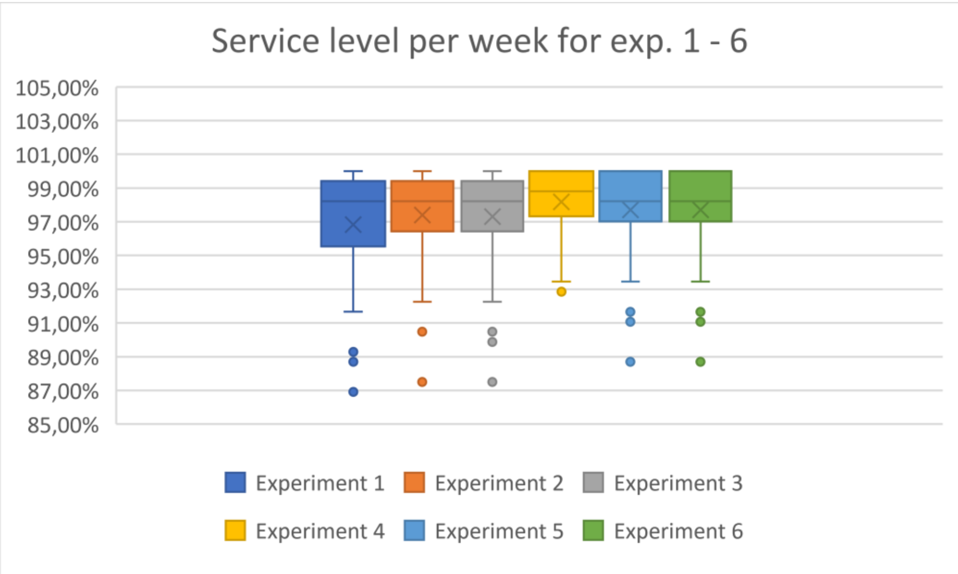
Exp.	Training Sample	Obj. value	Capacity / week	Service level	Overcapacity / week	Undercapacity / week	# Separate start hours
0	S.1	60,65	56	97,75%	267,21	1,98	3
1	S.1	60,58	55	96,83%	260,20	2,97	22
2	S.1	62,20	57	97,40%	275,67	2,44	7
3	S.1	62,24	57	97,31%	275,71	2,48	6
4	S.1	61,22	57	98,18%	274,70	1,47	4
5	S.1	61,90	57	97,71%	275,37	2,14	3
6	S.1	61,90	57	97,71%	275,37	2,14	3
0	S.2	60,72	56	97,55%	267,79	2,03	3
7	S.2	59,94	55	97,33%	260,09	2,34	16
8	S.2	60,89	56	97,42%	267,96	2,21	8
9	S.2	60,53	56	97,73%	267,60	1,85	6
10	S.2	62,73	58	97,60%	283,64	1,89	5
11	S.2	62,63	58	97,85%	283,54	1,79	3
12	S.2	62,71	58	97,73%	283,62	1,87	3
0	W.1	59,58	56	99,12%	287,92	0,70	3

13	W.1	53,00	49	97,69%	234,26	1,66	18
14	W.1	55,08	51	97,86%	250,19	1,58	7
15	W.1	55,08	51	97,86%	250,19	1,58	7
16	W.1	54,96	51	97,91%	250,07	1,46	4
17	W.1	54,95	51	98,02%	250,05	1,44	3
18	W.1	54,98	51	97,95%	250,09	1,48	3
0	W.2	59,49	56	99,16%	289,21	0,60	3
19	W.2	52,17	48	97,35%	225,13	1,92	17
20	W.2	55,26	51	97,93%	248,99	1,77	8
21	W.2	55,18	51	97,86%	248,90	1,69	7
22	W.2	54,86	51	98,12%	248,58	1,37	4
23	W.2	55,31	51	97,81%	249,03	1,82	3
24	W.2	56,19	52	97,91%	256,83	1,62	3

5.3.3 Service level per week

To get more insight into the robustness of the different schedules, the service level realizations for every separate week in the test data are analysed. This is done by means of box-whisker plots. When the spread in the service level realizations is higher, it means that the generated schedule is less robust. In Figure 23 - Figure 26, the realizations of the service level per week for every experiment are displayed in box-whisker plots.

Figure 23: Box whisker plots of weekly SL realizations exp. 1 – 6 (Summer 1)



ProRail

Figure 24: Box whisker plots of weekly SL realizations exp. 7 – 12 (Summer 2)

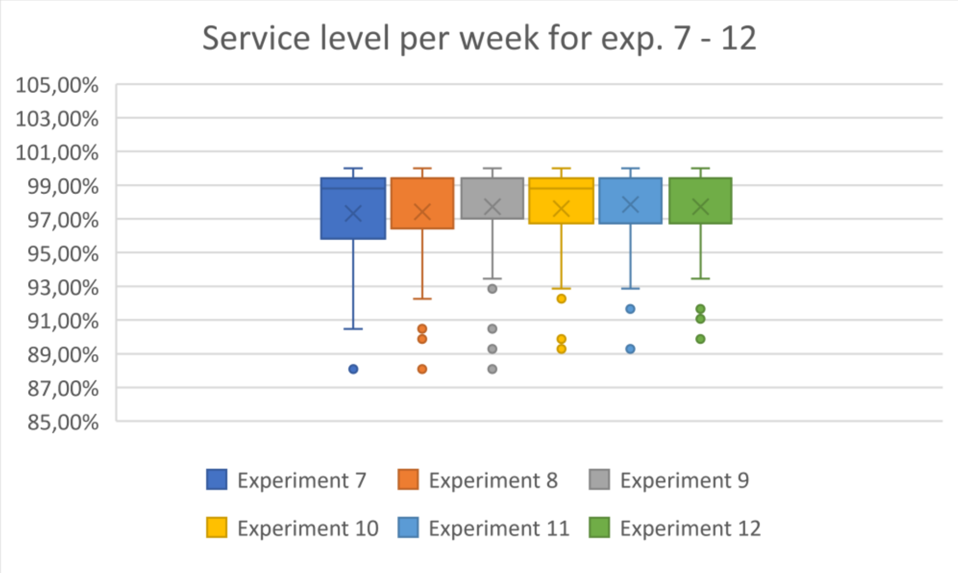


Figure 25: Box whisker plots of weekly SL realizations exp. 13 – 18 (Winter 1)

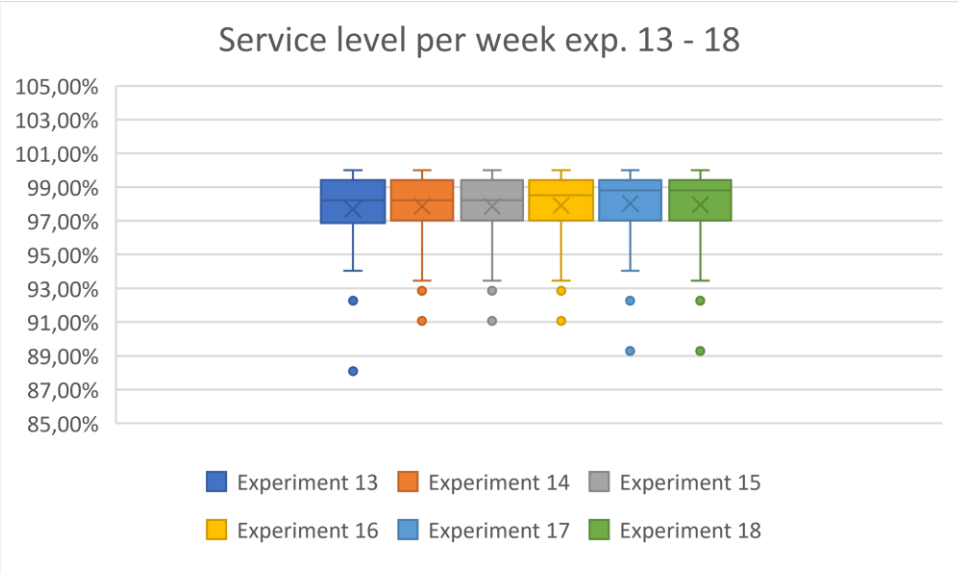
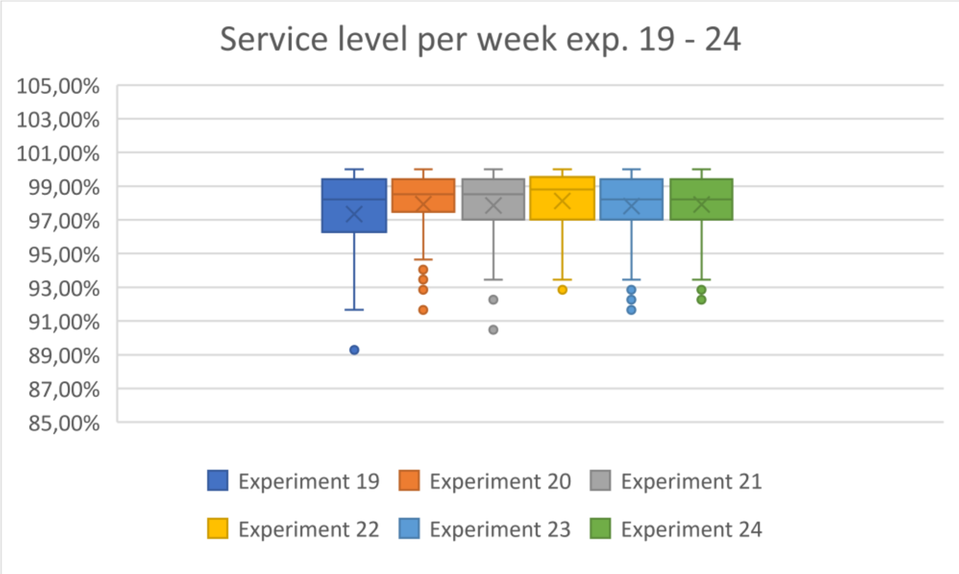


Figure 26: Box whisker plots of weekly SL realizations exp. 19 - 24 (Winter 2)



From the four box-whisker plots above, one can conclude that the spread in Service Level realizations per experiments does not differ greatly. The experiments in which the model is allotted the most freedom (exp. 1, 7, 13 and 19), the spread is usually the largest. Furthermore, these experiments produce the lowest outliers. For the summer experiments, the schedule that is produced in experiment 4 has a low spread and only one outlier, which is less than all other schedules that were generated in the summer experiments. Therefore, this schedule seems the most robust solution from the summer schedules.

For the winter schedules, it stands out that the schedule generated in experiment 13 has a low spread, even while this schedule has 18 different start times, from which you might suspect some overfitting issues. Furthermore, experiment 20 also has a low spread in the service level realizations. However, there are four outliers for this schedule, which is more than for any other winter experiment. Lastly, experiment 22 only has one outlier (shared lowest), while having a relatively low spread and high average service level.

5.3.4 Analysis of results

In this section, the most important results are discussed.

- **Anticipated results:** In general, the results in Table 13, appear to be logical: The experiments in which the model has been allotted the most freedom (1, 7, 13 and 19) have the lowest (best) best objective value. That is mostly due to the low number of dispatchers that were needed to meet all requirements and the low amount of overcapacity. Furthermore, we see that the model deploys the most capacity when it has the least freedom (in experiment 6, 12, 18 and 24). In most cases it also has the most overcapacity per week. Additionally, the schedules created by the model in the experiments with less freedom generally have a higher service level than the schedules created in the experiments where the model is allotted more freedom. This is probably because the model needs more capacity to reach the service level due to the tighter constraints. Because more capacity has been used, the model scores better in terms of service level on the test set.
- **Low number of shifts in the winter compared to regular schedule:** When we compare the number of shifts assigned by the scheduling model compared to the general schedule, there is a clear difference in the number of shifts per week. While 56 shifts are currently deployed

every week (excluding the relief shifts), the model only assigns on average 51 shifts in the experiments with the winter samples. The number of shifts assigned to the summer is slightly higher than in the current situation: 56,75 instead of 56.

- **More shifts lead to higher service levels and less undercapacity:** The schedules with the highest service levels and the lowest undercapacity always have the (shared) highest number of shifts per week. What further stands out is that the schedule with the fixed start times (at 7:00, 15:00 and 23:00, experiment 6, 12, 18 and 24) is invariably outperformed by other schedules when the number of shifts is equal. For the S.1 sample, schedule 6 has an equal performance as schedule 5. With the S.2 sample, schedule 12 is outperformed by schedule 11 on every metric. With the W.1 sample, schedule 18 is outperformed by schedule 17 on every metric and for the W.2 sample, schedule 22's service level and undercapacity are very close to that of schedule 24, with 1 shifts less per week. This shows that the start times of the regular schedule might not be optimal.
- **Overfitting:** In all experiments, the service level constraint was binding, meaning that this constraint was just narrowly met in each experiment. As one can see in in Table 13, the minimum service level of 98% is often not met on the test set. Therefore, we can conclude that overfitting occurs at least to a certain extent. Overfitting means that the model focuses on specific datapoints or patterns in the training dataset that do not appear (to the same extent) in the entire dataset. Therefore, the model performs better on the training set than on the test dataset.
- **Experiment 4 & 22 showing good results:** Experiments 4 and 22 both resulted in a schedule that had a service level higher than 98% on the test dataset. Furthermore, both experiments scored well in terms of robustness of their results, which is explained in Section 5.3.3. The schedule generated in experiment 4 meets the service level of 98% on the test dataset, while the existing schedule does not. The schedule generated in experiment 22 meets the 98% service level with 5 shifts less per week than the current schedule.

Concerning the distribution over the day and over the week, a few things stand out. The model assigns the least capacity to the weekend. This holds for the schedules made for the summer months as well as those for the winter months. However, the difference between the weekdays and the weekend are larger in the summer than in the winter. These results are clearly visible in Figure 27 and Figure 28, which look at the average capacity assigned per schedule to each day over all experiments for each season. Concerning the hour seasonality, it stands out that the model assigns on average an equal number of workers to shifts in the night during the summer as during the winter. The difference in capacity between the seasons is thus made during the daytime. The capacity patterns in Figure 28 and Figure 29 clearly follow a similar pattern as the hour seasonality pattern of Figure 20. However, this concerns the average over a multitude of schedules. Unique schedules can obviously not follow the pattern so meticulously since we are bound to shifts of eight hours.

Figure 27:Average capacity assigned per day by the scheduling model (summer)

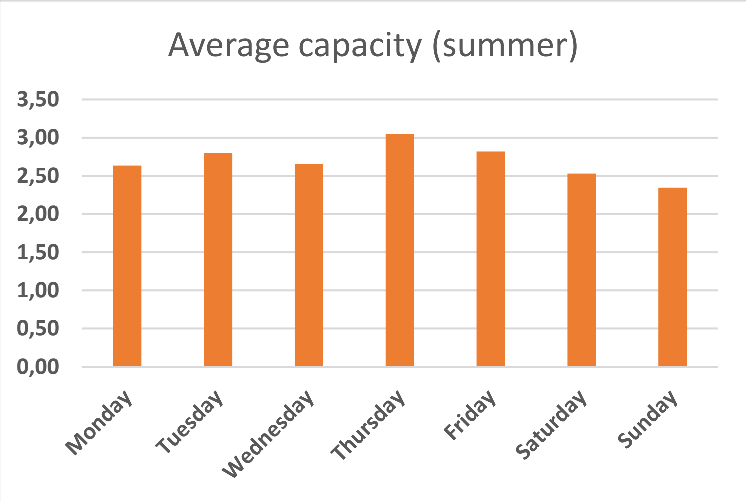


Figure 28:Average capacity assigned per day by the scheduling model (winter)

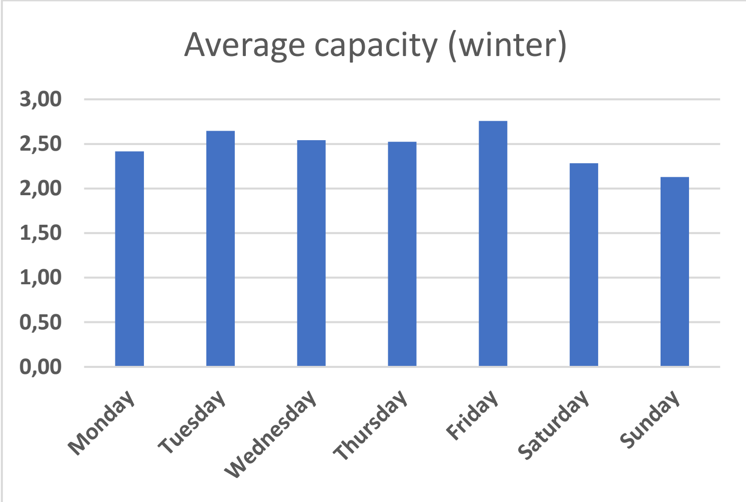


Figure 29: Average capacity assigned per hour by the scheduling model (summer)

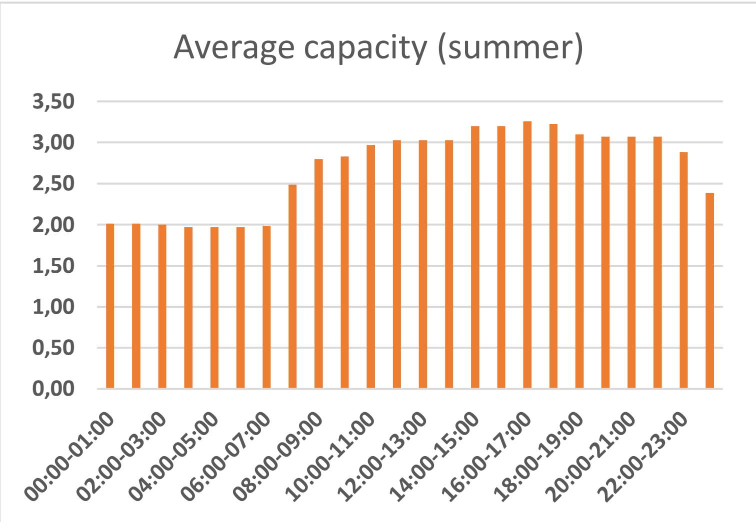
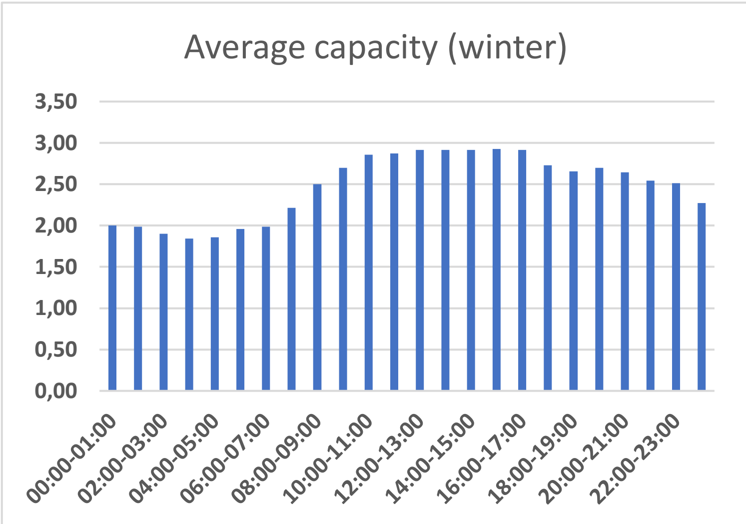


Figure 30: Average capacity assigned per hour by the scheduling model (winter)



5.3.5 Modelling decisions

When it comes to individual schedules, the exact start times of all shifts generated are stated in Table 39. Since it might be difficult to understand why specific modelling decisions are taken, several decisions are elaborated in this section.

Regarding the experiments with the first winter sample as training set, the schedule that had the highest service level was the schedule generated in experiment 17. The capacity distribution and workload distribution over the week are shown in Figure 34. In this case, the model keeps the base capacity always at two workers and increases it to three workers on every weekday. It stands out that the capacity is always higher than the average workload and that the capacity pattern does not always follow the average workload. In that sense, the average workload can be misleading, because the outliers are important.

For example, the estimated workload on Monday between 16:00 and 00:00, was 8.81 hours, while the average estimated workload on Wednesday between 16:00 and 00:00 was 8.72 hours, yet only two

ProRail

workers were assigned to the shift on Monday, while three workers were assigned to the shift on Wednesday. This seems implausible. However, when we look at the number of instances where a capacity of two workers is not enough per shift according to the workload estimate, the decision does make sense. Since the training data contained 25 instances where a capacity of three workers was required to meet the workload for the Monday evening shift, while there were 35 instances where a capacity of three workers was required to meet the workload for the Wednesday evening shift. This tilted the balance in favour of the Wednesday evening shift and therefore the model assigned the capacity here. Even while the average workload on Wednesday evening was lower than on the Monday evening.

In Table 14 all instances where a capacity of at least three workers was needed for the W.1 sample is shown. The instances for the Tuesday morning and Saturday evening shifts are highlighted.

Table 14: Instances where a capacity of three workers is needed for sample W.1

Hour\Day	Mond.	Tuesd.	Wedn.	Thurs.	Friday	Saturd.	Sund.	Total:
00:00-01:00	0	0	1	1	0	0	0	2
01:00-02:00	1	0	0	0	0	0	0	0
02:00-03:00	0	0	0	1	0	0	0	0
03:00-04:00	0	0	1	0	0	1	0	1
04:00-05:00	0	0	0	0	0	0	0	1
05:00-06:00	2	1	0	0	0	0	0	0
06:00-07:00	1	1	2	1	0	0	0	1
07:00-08:00	1	2	1	2	1	2	2	3
08:00-09:00	4	6	5	2	3	3	1	12
09:00-10:00	5	8	2	7	2	2	2	14
10:00-11:00	6	5	4	8	7	3	4	14
11:00-12:00	9	6	9	8	8	4	3	18
12:00-13:00	9	6	6	4	5	4	2	6
13:00-14:00	9	5	10	9	6	3	3	17
14:00-15:00	12	9	11	4	9	6	2	19
15:00-16:00	6	11	8	13	10	8	6	20
16:00-17:00	8	9	10	13	15	6	4	16
17:00-18:00	4	8	7	6	8	4	1	20
18:00-19:00	4	7	2	6	4	1	1	18
19:00-20:00	2	5	6	7	5	5	2	13
20:00-21:00	1	7	6	2	7	5	3	10
21:00-22:00	5	3	3	1	7	5	1	12
22:00-23:00	1	1	0	0	4	2	0	8
23:00-24:00	0	0	1	0	1	0	0	7
Total:	32	42	31	46	44	27	10	

Figure 31: Average workload estimate and capacity deployment regular schedule

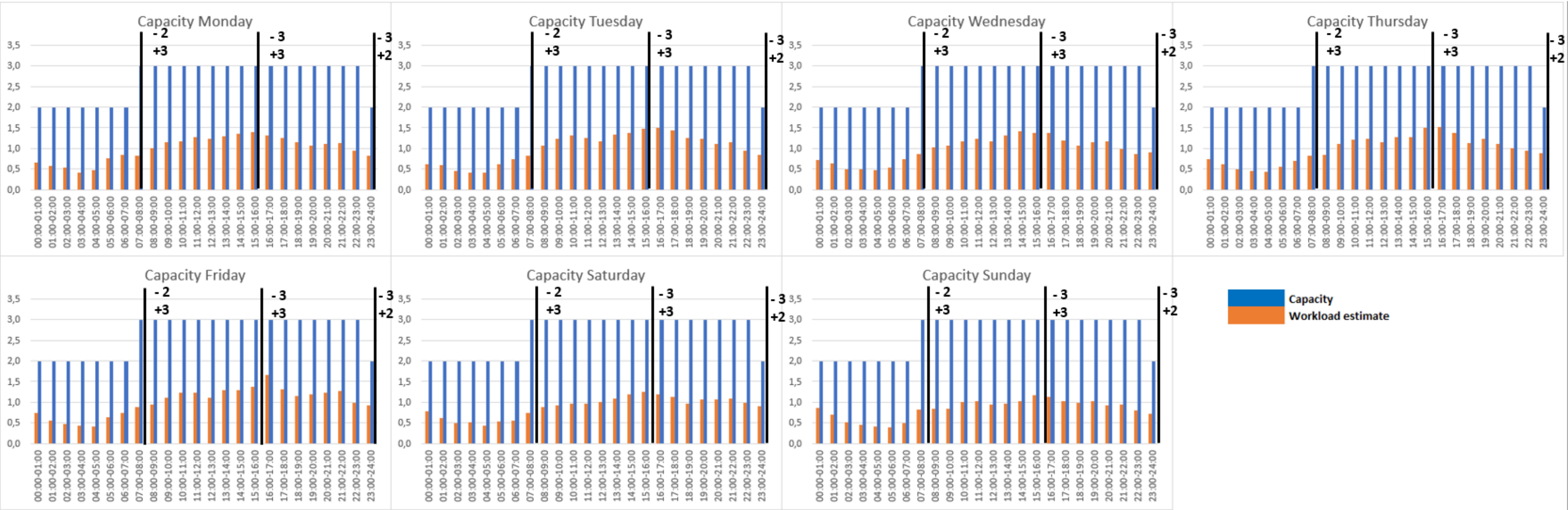
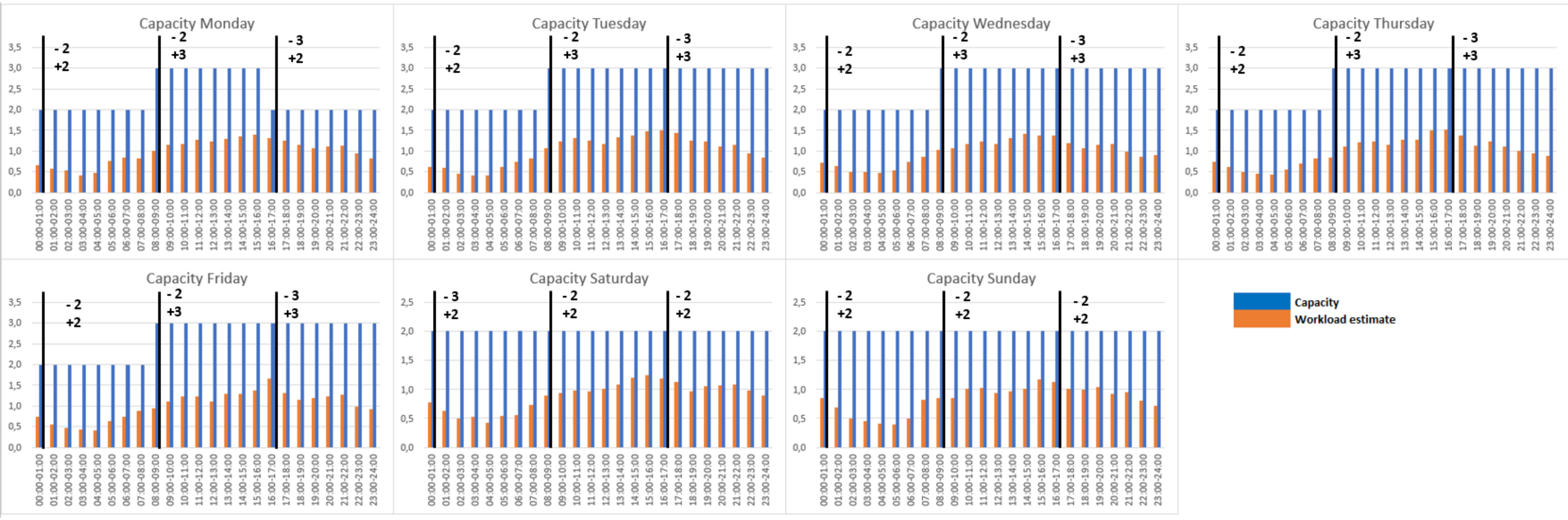


Figure 32: Average workload estimate and capacity deployment result of experiment 17 (winter 1)



5.4 Recommended schedules

From the experiments in Section 5.3, two schedules stood out in terms of performance on the test dataset: the schedules generated in experiments 4 and 22. These schedules meet the required service level, are robust and have a convenient number of different start times. Therefore, these schedules are recommended to ProRail. In the following sections, the change in performance that is the result from implementing these schedules is shown as well as an analysis that shows the reduction in required FTEs that follows from implementing the schedules. The allocation of capacity in the schedules is depicted in Figure 35 and Figure 36.

5.4.1 Comparison in performance

In the scheduling model, the goal was to minimize the overcapacity as well as the undercapacity and the number of shifts deployed per week, while meeting the service level of 98%. The schedules of experiments 4 and 22 meet these requirements with less shifts on average. However, this reduction in shifts does decrease the overall performance of the schedule in terms of service level and average undercapacity. In Table 15 and Table 16, the performance of the current schedule and the performance of the proposed schedules is summarized.

Table 15: Performance current schedule over the year

Current Schedule	Shifts per week	Service Level	Avg. Overcapacity	Avg. Undercapacity
Winter	56	99.14%	288.64	0.64
Summer	56	97.65%	267.50	2.01
Whole year	**56**	**98.64%**	**281.59**	**1.10**

Table 16: Performance of proposed schedules over the year

Proposed Schedules	Shifts per week	Service Level	Avg. Overcapacity	Avg. Undercapacity
Winter	51	98.12%	248.58	1.37
Summer	57	98.18%	274.70	1.47
Whole year	**53**	**98.14%**	**257.29**	**1.40**

From this comparison, one can conclude that the service level decreases and the average undercapacity increases. However, the goal for the service level of 98% is still met, and the average undercapacity per week is still low. The average undercapacity increases with 0.3 hours per week, which is 18 additional minutes per week where the capacity is not sufficient. Furthermore, the average overcapacity drops from 281.59 – 257.29 = 24.3 hours. This is more than the combination of the three shifts that are not deployed in the proposed schedule, which shows that the match between the capacity and workload has also improved.

5.4.2 FTE reduction

During the larger portion of the year (the eight months in the winter sample), the model allocates 5 shifts less per week in the proposed schedule than in the current schedule. In the summer, it allocates one more shift in the proposed schedule than in the current schedule to meet the required service level. Since the decrease in the winter is larger than the increase in the summer and that the winter period in this research is defined as longer than the summer period, the average number of shifts allocated per week reduces.

The winter period contains 8 months, while the summer period contains 4 months. One FTE at the MKS works 36 hours per week, which is equal to 4.5 shifts per week. The gross salary for one dispatcher currently ranges between €2658 and €3725. The labour costs for ProRail for employees at the MKS are approximately 40% higher than the gross salary. Therefore, the labour costs for ProRail range between €3721 and €5215.

Taking all these parameters into account, one can calculate that the number of FTEs required per year drops with 0.667 FTE, which translates to a cost reduction that ranges between €29,770 and €41,720 per year. The parameters are summarized in Table 17. The calculation for the reduction in FTE is summarized in Table 18.

Table 17: Parameters for cost calculation

Parameter	Value
Length winter (months):	8
Length summer (months):	4
Contract in hours per week:	36
Shifts 1 FTE:	4,5
Salary 1 FTE (LB, month):	€2.658,00
Salary 1 FTE (UB, month):	€3.725,00
Labour cost multiplier:	1,4

Table 18: Calculations for reduction in FTE and costs

	Diff. Shifts per week	FTE/week	FTE/year	Savings / year (LB):	Savings / year (UB):
Winter	5	1.111	0.750	€33,077.33	€46,355.56
Summer	-1	-0.222	-0.070	-€3,307.73	-€4,635.56
Total:			**0.667**	**€29,769.60**	**€41,720.00**

5.4.3 Schedule explanation

The figures on the next two pages show how much capacity is deployed during every hour (blue bars), how much capacity was on average required (orange bars) and when shifts start and end (black bars). Next to each bar, the number of shifts that end at that time and the number of shift that start are noted.

From the figures, one can derive that both during the winter and during the summer shifts start at 7:00, 10:00, 15:00 and 23:00. During the night, there are always two dispatchers present, while during the day, the capacity varies between 2 and 4 dispatchers. In the winter, the model never deploys more than 3 dispatchers at the same time. In the summer, 4 dispatchers are deployed at the same time on Tuesday and Friday. In the summer, more capacity is deployed on Tuesday and Friday and less on Saturday when compared to the current schedule. In the winter, less capacity is deployed on Monday, Wednesday, Thursday, Saturday, and Sunday.

Figure 33: Capacity allocation for the recommended summer schedule

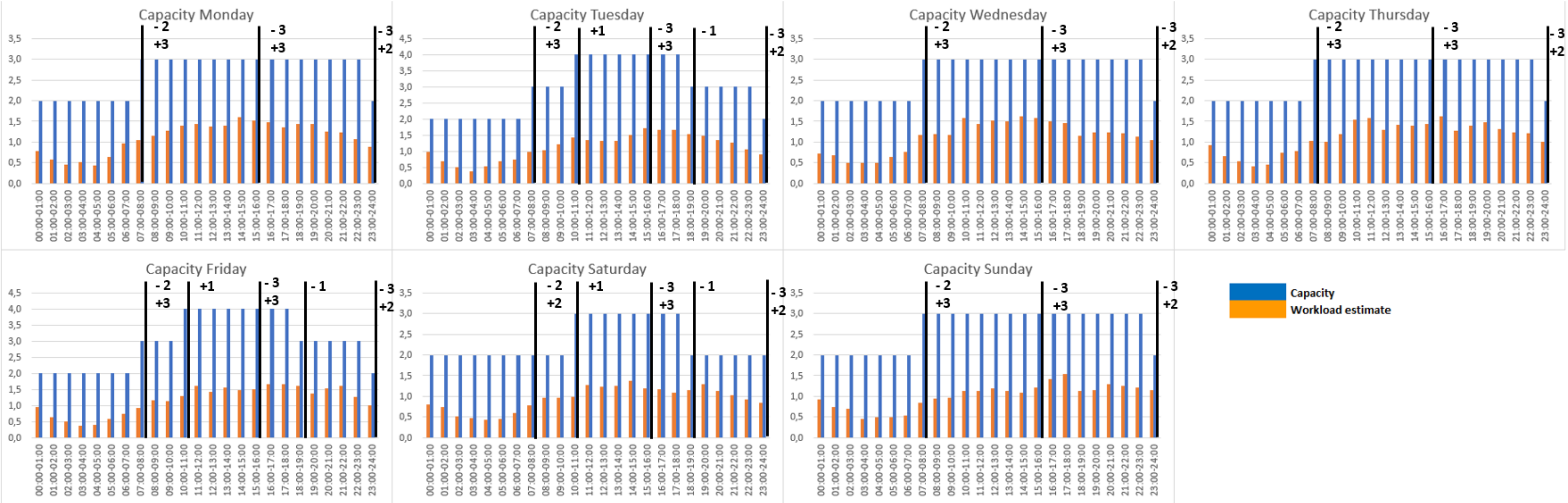
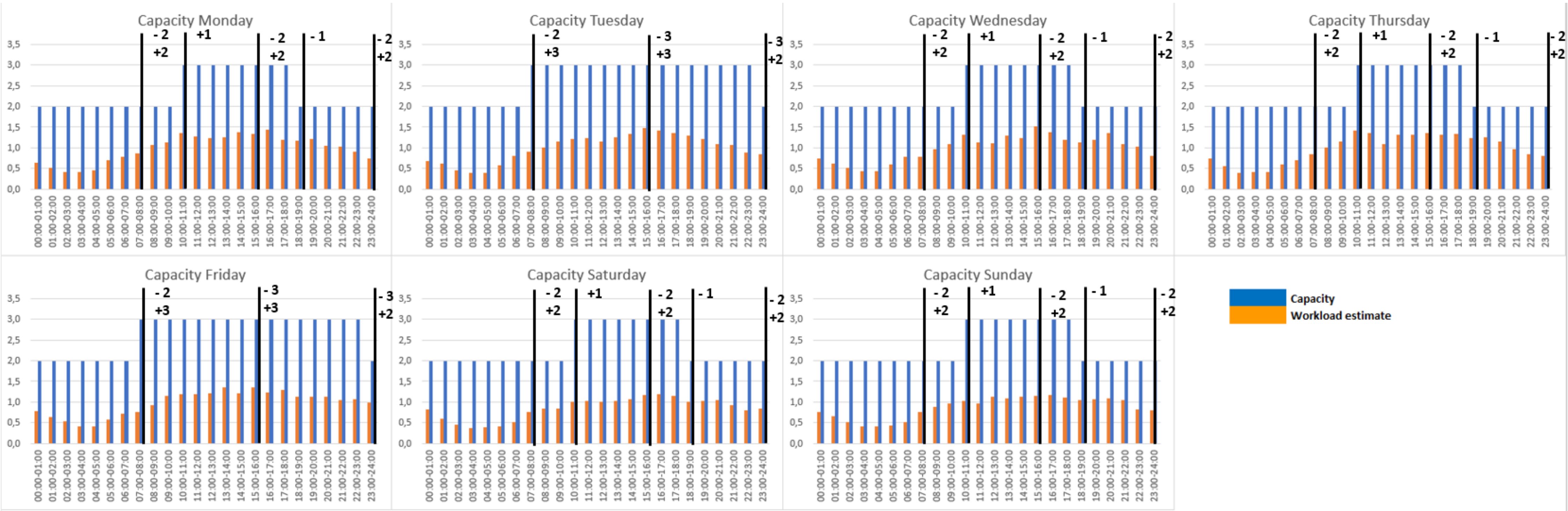


Figure 34: Capacity allocation for the recommended winter schedule



5.5 Experiments for practical decision making

In Section 5.3, the scheduling model is used to create optimal schedules based on a set of input parameters, constraints, and an objective. The schedules that are created in this section provide insights about how one should deploy capacity. However, it does not provide clear-cut answers for practical day-to-day decisions. Therefore, this section contains a series of experiments, meant to provide insights that can be used when the decision maker faces specific capacity problems. This section is split up in three parts: determining optimal start times, dayshifts, and information about which shifts should be omitted first.

Since the examined schedules are not created by a model, there is no split in training and testing samples needed. All experiments are conducted for both the integral summer sample (S.1 + S.2) and the integral winter sample (W.1 + W.2). For all experiments, the same parameter setup is used as in the experiments of Section 5.3.1.

5.5.1 Optimal Start times

For the experiments concerning the optimal start times, a limited set of experiments is run. It is convenient to have a schedule where shifts start 8 hours after one another. Since it is not favourable to have shifts starting and ending during the night, the sets of convenient start times can essentially be limited to three sets:

- **Set 1:** Morning shift starts at 6:00, afternoon shift starts at 14:00, night shift starts at 22:00.
- **Set 2:** Morning shift starts at 7:00, afternoon shift starts at 15:00, night shift starts at 23:00.
- **Set 3:** Morning shift starts at 8:00, afternoon shift starts at 16:00, night shift starts at 00:00.

In the 6 experiments for this section, the capacity levels of the regular schedule are used (3 in the morning, 3 in the evening and 2 in the night). For the first free experiments, integral summer sample is used and for the last three experiments, the integral winter simple is used as input. The results of the experiments are depicted in Table 19.

Table 19: Results experiments for practical decisions (optimal Start Times)

Experiment	Input sample	Set of shifts	Service level	Overcapacity / week	Undercapacity / week
1	Summer	1	97.45%	267.65	2.16
2	Summer	2	97.65%	267.50	2.01
3	Summer	3	97.67%	267.48	1.99
4	Winter	1	99.09%	288.67	0.68
5	Winter	2	99.14%	288.64	0.64
6	Winter	3	99.12%	288.66	0.67

For the experiments with the summer sample as input, the third set of start times has the highest performance on all three KPIs. However, the differences between the second and third set of shifts is minimal. Therefore, it cannot be said with certainty that this ranking will hold in the future.

For the experiments with the winter sample as input, the second set of start times slightly outperforms the other two sets for all three KPIs. Since the winter dataset contains more data than the summer dataset, these results are more reliable. Finally, since it is convenient to have the same start times throughout the whole year and since the difference in performance between the second and third set of shifts is very small, it is recommended to keep using the second set of start times for the whole year.

5.5.2 Optimal day shifts

In this section, the experiments that are used to determine optimal day shifts are explained. It is important to keep in mind that day shifts differ from relief shifts. A relief shift is used to provide dispatchers the opportunity to have a break while the capacity is kept at the same level. When none of the regular dispatchers is having a break, the dispatcher in the relief shift can work on ancillary tasks, such as portfolio tasks.

A day shift is a regular shift that has overlap with both the morning and the evening shift. A day shift is often used as a replacement for two regular shifts (1 morning and 1 evening shift). To illustrate this with an example: It can occur that one of the dispatchers calls in sick on Friday for his shift on Saturday morning. The planner cannot find a replacement at this short notice. To better match the capacity to the workload, the planner will try to let one of the dispatchers in the evening shift start earlier, so that the busiest hours of the day are covered with a capacity of three workers. The question raised by the management of the MKS is: what is the best time to let the day shift start? This example is presented graphically in Figure 35.

Figure 35: Illustration of the use of a day shift when a drop in capacity cannot be covered

To answer the question raised by the management, 8 experiments are conducted. The integral summer sample serves as input for the first four experiments, while the integral winter sample serves as input for the last four experiments. For every experiment, the capacity for all hours is set to 2 workers, except for the hours that are covered by the day shift, where the capacity is set to 3 workers. The following four start times for the day shift are examined: 9:00, 10:00, 11:00 and 12:00.

Table 20:Results experiments for practical decisions (optimal day shifts)

Experiment	Input sample	Start of day shift	Service level	Overcapacity / week	Undercapacity / week
1	Summer	9:00	94.65%	214.65	5.07
2	Summer	10:00	94.94%	214.12	4.63
3	Summer	11:00	94.92%	214.01	4.52
4	Summer	12:00	94.86%	214.17	4.68
5	Winter	9:00	97.46%	233.94	1.95
6	Winter	10:00	97.48%	233.84	1.85
7	Winter	11:00	97.45%	233.86	1.87
8	Winter	12:00	97.46%	233.91	1.92

From the experiments it becomes clear that starting a day shift at 10:00 results in the highest service levels. However, when looking at all three KPIs, starting a dayshift at 11:00 also provides good results.

In the four experiments where the model had to choose for one of the day shifts (experiments 4, 10, 16 & 22), all four different start times for the day shift are chosen once. Therefore, these experiments do not offer clarity either. Over all 24 experiments from Section 5.3, a day shift starting at 10:00 is scheduled the most often from the four possible day shifts; in 8 of the 24 experiments capacity is deployed at this time. Furthermore, it has the highest service level for both samples and the lowest overcapacity and undercapacity in the winter sample. Therefore, starting the relieve shift at 10:00 seems the most promising.

To illustrate why moving the capacity from the morning or evening shift (in case of a drop in capacity at one of these shifts) to a day shift, is preferable compared to keeping everything the same, the day shift at 10:00 is compared to keeping capacity at the morning or evening shift. For this analysis, the regular start times are used, since these appeared the most preferable from the analysis in Section 5.5.1.

Table 21: Comparison between moving capacity to a day shift and keeping it the same

Experiment	Input sample	Start of shift with capacity of 3	Service level	Overcapacity / week	Undercapacity / week
1	Summer	7:00	93.76%	215.57	6.07
2	Summer	10:00	94.94%	214.12	4.63
3	Summer	15:00	94.58%	214.88	5.38
4	Winter	7:00	96.67%	234.59	2.60
5	Winter	10:00	97.48%	233.84	1.85
6	Winter	15:00	96.88%	234.44	2.45

The results from Table 21 should be read in the following way: there is a situation that the capacity at the morning or evening shift drops with one worker and therefore there will only be one shift of 8 hours where there is a capacity of 3 workers. The third column in the table shows the decision about when the 8-hour period of a capacity of 3 workers starts. The fourth to the sixth column show the KPI values that are the result of this decision. The KPI values are an average over the total sample and are not tied to a specific day.

From the results in Table 21, it is clear that deploying a day shift instead of either a morning or an evening shift is preferable. The service level is significantly higher, while both the overcapacity and undercapacity are lower. This means that there is a better match between the workload and capacity for the day shift than for a single morning or evening shift. Furthermore, one can conclude from this table that evening shifts in the current schedule are on average busier than morning shifts, since deploying extra capacity at an evening shift provides better results than on a morning shift.

5.5.3 Impact of omitting shifts

In section 5.5.2, a situation is outlined where the planner had to make a last-minute decision about a change in the capacity deployment. Yet, it occurs often that the available capacity drops for a longer time due to a variety of reasons. Examples of these situations are when an employee has left the company or has retired and his or her replacement has not completed the training to be a dispatcher yet. Or when employees are at home due to long-term sickness etc. For these situations, the planner has a longer time and more options to adjust to the capacity loss.

In this section, we investigate the impact of omitting shifts. The results of the experiments in this section can be used as a guideline for the decision maker, for various situations where there is less capacity available than the standard.

To determine what the impact is of omitting different shifts, 28 different experiments were executed: 14 with the summer sample as input and 14 with the winter sample. In the first 7 experiments for each sample, the combination of one morning and one evening shift is replaced by one day shift. For the second group of 7 experiments for each sample, the combination of one morning and one evening shift is removed without replacement. For each experiment, the performance difference on the three main KPIs compared to the base schedule is given. The experiments are ranked from least to most impact. In the third and fourth column, scenarios indicate the change that is made. Here, scenario 1 is the regular situation in the base schedule, scenario 2 is the situation where a base shift replaces one morning and evening shift and scenario 3 is when a morning and evening shift are removed with no replacement. The results of the experiments are stated in Table 22.

Table 22: Results experiments for practical decisions (impact of omitting shifts)

Exp.	Day	Input sample	From	To	Δ Service level	Δ Over-capacity	Δ Under-capacity	Ranking (group)
1	Monday	Summer	Scen. 1	Scen. 2	-0.49pp	-7.60	0.40	4
2	Tuesday	Summer	Scen. 1	Scen. 2	-0.57pp	-7.52	0.48	5
3	Wednesday	Summer	Scen. 1	Scen. 2	-0.28pp	-7.72	0.28	3
4	Thursday	Summer	Scen. 1	Scen. 2	-0.43pp	-7.47	0.53	6
5	Friday	Summer	Scen. 1	Scen. 2	-0.46pp	-7.45	0.55	7
6	Saturday	Summer	Scen. 1	Scen. 2	-0.23pp	-7.84	0.16	1
7	Sunday	Summer	Scen. 1	Scen. 2	-0.25pp	-7.79	0.21	2
8	Monday	Summer	Scen. 1	Scen. 3	-1.15pp	-14.99	1.01	3
9	Tuesday	Summer	Scen. 1	Scen. 3	-1.22pp	-14.68	1.32	5
10	Wednesday	Summer	Scen. 1	Scen. 3	-1.05pp	-14.97	1.03	4
11	Thursday	Summer	Scen. 1	Scen. 3	-1.21pp	-14.59	1.41	6
12	Friday	Summer	Scen. 1	Scen. 3	-1.13pp	-14.46	1.54	7
13	Saturday	Summer	Scen. 1	Scen. 3	-0.63pp	-15.39	0.61	2
14	Sunday	Summer	Scen. 1	Scen. 3	-0.58pp	-15.47	0.53	1
15	Monday	Winter	Scen. 1	Scen. 2	-0.23pp	-7.85	0.16	3
16	Tuesday	Winter	Scen. 1	Scen. 2	-0.30pp	-7.76	0.25	7
17	Wednesday	Winter	Scen. 1	Scen. 2	-0.25pp	-7.82	0.19	4
18	Thursday	Winter	Scen. 1	Scen. 2	-0.26pp	-7.88	0.19	5
19	Friday	Winter	Scen. 1	Scen. 2	-0.29pp	-7.80	0.20	6
20	Saturday	Winter	Scen. 1	Scen. 2	-0.18pp	-7.82	0.13	2
21	Sunday	Winter	Scen. 1	Scen. 2	-0.16pp	-7.86	0.08	1
22	Monday	Winter	Scen. 1	Scen. 3	-0.69pp	-15.44	0.58	3
23	Tuesday	Winter	Scen. 1	Scen. 3	-0.79pp	-15.30	0.71	6
24	Wednesday	Winter	Scen. 1	Scen. 3	-0.77pp	-15.47	0.55	5
25	Thursday	Winter	Scen. 1	Scen. 3	-0.70pp	-15.49	0.66	4
26	Friday	Winter	Scen. 1	Scen. 3	-0.83pp	-15.39	0.63	7
27	Saturday	Winter	Scen. 1	Scen. 3	-0.46pp	-15.54	0.35	2
28	Sunday	Winter	Scen. 1	Scen. 3	-0.41pp	-15.61	0.28	1

The results stated in Table 22 must be read in the following way: when the planner replaces one evening and one morning shift (scenario 1) with a day shift (scenario 2) on a Wednesday in the summer (experiment 3), then the service level drops with 0.10 percentage points, the overcapacity decreases with 7.94 hours and the undercapacity increases with 0.058 hours. From the seven experiments in the summer where a day shift is used to replace a morning and evening shift, this change on the Wednesday had the third least impact.

As expected, removing shifts from the weekend has the least impact on the expected undercapacity. For the summer, when one shift should be removed, replacing the morning and evening shift on

Saturday with a day shift decreases the performance of the schedule the least. For the winter, the same action on Sunday decreases the performance of the schedule the least.

Several things stand out from Table 22: First, there are differences between the seasons. The difference between the workload during the week and the weekend is larger in the summer than in the winter. This was also visible in Figure 27 to Figure 30 and it can also be derived from Table 22, by looking at the additional undercapacity that is the result from removing a shift.

In Figure 36 up to and including Figure 39, the impact of replacing two shifts with one dayshift or removing two shifts without replacement on the average amount of undercapacity per week is shown. From these figures, one can conclude that removing a shift in the summer generally causes more undercapacity than in the winter.

Figure 36: Impact of replacing two shifts by a day shift per day (summer)

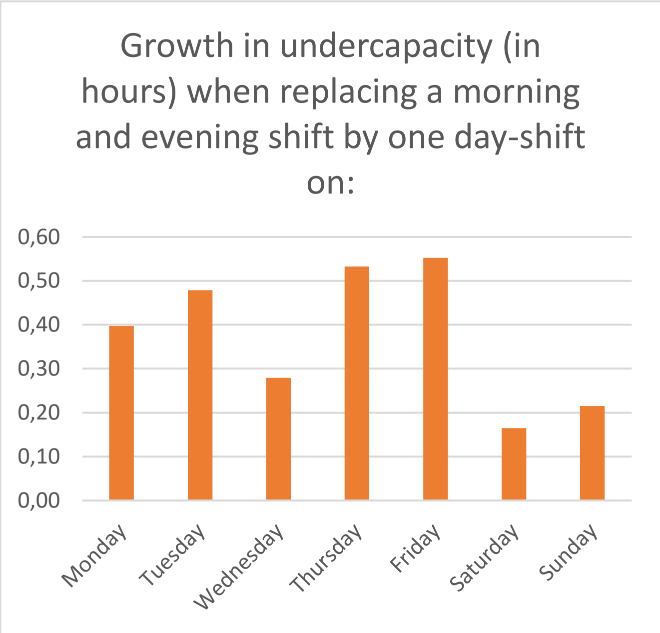


Figure 37: Impact of replacing two shifts by a day shift per day (winter)

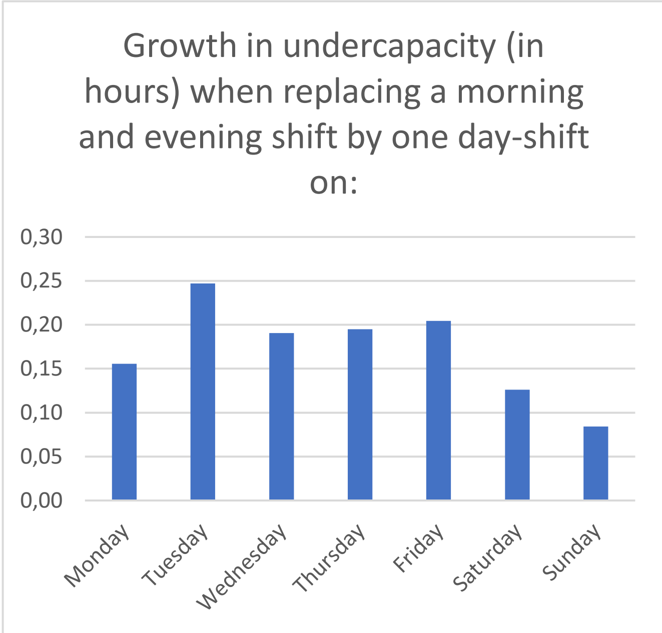


Figure 38: Impact of removing two shifts (per day, summer)

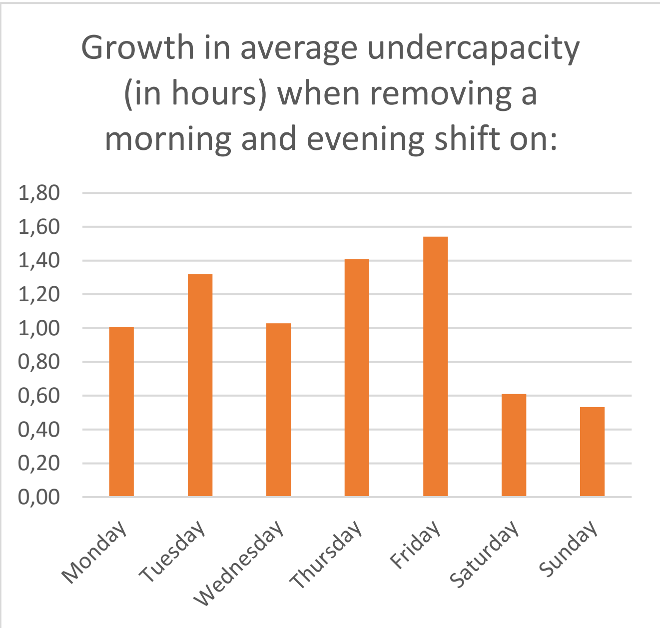
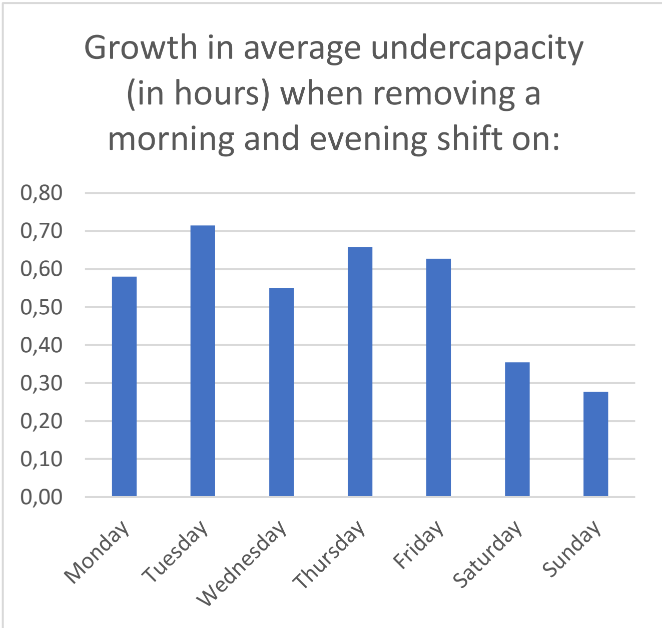


Figure 39: Impact of removing two shifts (per day, winter)



5.5.4 Break opportunities

This section ends with an analysis into the break opportunities. Currently, dispatchers take a half-an-hour break during their shifts if a dispatcher in the relief shift can take over his or her tasks. If there is no relief shift present, then the dispatchers do not take an official break. Both the management of the MKS and the dispatchers prefer that breaks can be taken. If the dispatchers do not take an official break, they are compensated for the extra time they work. In this analysis, the effects of breaks on the service level are analysed for every day of the week. In the end an overview is given for which days a break can be taken without violating the service level constraint and on which days this is not the case. The possibilities for breaks with the current schedule in the summer and in the winter are analysed as well as the possibilities for breaks in the proposed summer and winter schedules.

In Section 5.5.2, it was determined that the optimal day shifts should start at 10:00 in the morning. Furthermore, in the proposed schedules, the day shifts start at 10:00 as well. Therefore, breaks should not take place before 10:00 and after 18:00 (the end of the day shift). This prevents the capacity to drop to one person at times, which would lead to poor performance. From the workload analysis, it appeared that the workload drops slightly between 11:00 and 13:00, which offers an opportunity for breaks in the morning shift to take place in this time window. In the evening, the average workload decreases every hour and breaks in the evening should therefore take place as late as possible. In case there is a dayshift, this would mean between 16:30 and 18:00. Considering these break times, the performance of the current schedule (during both seasons) and that of the proposed summer and winter schedules is analysed. In the experiments, capacity is deducted during the breaks to simulate the effect of workers not being present at the MKS. There is no replacement (no relief shift) because these experiments are created to check at which days a relief shift is required. In Table 23, the results of the experiments are depicted.

Table 23: Results of experiments for evaluating break opportunities

Day	Service level (summer, curr.)	Service level (winter, curr.)	Service level (summer, proposal)	Service level (winter, proposal)
Monday	96,78%	98,02%	96,78%	96,43%
Tuesday	96,21%	97,95%	98,11%	97,95%
Wednesday	95,83%	97,92%	95,83%	96,16%
Thursday	95,01%	98,03%	95,01%	96,26%
Friday	94,32%	97,95%	97,60%	97,95%
Saturday	97,54%	99,00%	95,96%	97,70%
Sunday	97,92%	99,30%	97,92%	98,20%

From Table 23, it stands out that with the current schedule, it is not possible to take breaks during the summer while meeting the service level of 98% without a relieve shift that can take over. During the winter however, it is possible to take breaks without a relieve shift on Monday, Thursday, Saturday, and Sunday. Therefore, we conclude that a relieve shift on one of these days during the winter is superfluous.

With the proposed summer schedule, it is possible to take breaks without a relief shift on Tuesday, but not on any other day. The fact that this is possible on Tuesday is the result of the extra shift that is allocated to Tuesday in the proposal. With the proposed winter schedule, it is possible to take breaks without a relief shift on Sundays, but not on any other day.

In the experiments, all dispatchers that work during the day were expected to take a break. In reality, one might opt for only letting dispatchers during either the morning or evening shift go on a break. On days where the service level from the experiments is close to 98%, one can be sure that the service level will be met when only dispatchers of one shift take a break. This is for example the case for Tuesdays and Fridays in the proposed winter schedule.

If dispatchers would take breaks instead of continuing to work, would save ProRail costs in terms of compensation. Furthermore, if a dispatcher in a relief shift would not be needed to take over tasks while dispatchers take breaks, then he or she could spend this time on other tasks. Using the parameters from the previous cost savings calculation (see Table 18), one can calculate a lower and upper bound for the potential savings. Someone who has a gross salary of €2,658 / month, for a 36 hour contract earns €18.74 / hour. The labour costs for ProRail / hour are €18.74 * 1.4 = €26.24. The costs of a half-an-hour break are then €26.24 / 2 = €13.12. The 8 months in the winter sample are 2/3rd of the year. If 6 dispatchers would not have to be compensated for their breaks every week, then 6 * 52 * (2/3) * €13.12 = €2,728.59 in savings. This is the lower bound calculation for the winter. For the summer, the factor 2/3rd is replaced with 1/3rd and for the upper bound calculation, a gross salary of €3,725 is used. Repeating the calculation for all situations where breaks can be taken without a relief shift, leads to the results depicted in Table 24 and Table 25.

Table 24: Cost savings resulting from break opportunities (lower bound)

Day	Current schedule (summer)	Current schedule (winter)	Proposed summer schedule	Proposed winter schedule
Monday	€ -	€ 2,728.59	€ -	€ -
Tuesday	€ -	€ -	€ 1,364.29	€ -
Wednesday	€ -	€ -	€ -	€ -
Thursday	€ -	€ 2,728.59	€ -	€ -
Friday	€ -	€ -	€ -	€ -
Saturday	€ -	€ 2,728.59	€ -	€ -
Sunday	€ -	€ 2,728.59	€ -	€ 2,728.59
Total	**€ -**	**€ 10.914,35**	**€ 1,364.29**	**€ 2,728.59**

Table 25: Cost savings resulting from break opportunities (upper bound)

Day	Current schedule (summer)	Current schedule (winter)	Proposed summer schedule	Proposed winter schedule
Monday	€ -	€ 3,823.92	€ -	€ -
Tuesday	€ -	€ -	€ 1,911.96	€ -
Wednesday	€ -	€ -	€ -	€ -
Thursday	€ -	€ 3,823.92	€ -	€ -
Friday	€ -	€ -	€ -	€ -
Saturday	€ -	€ 3,823.92	€ -	€ -
Sunday	€ -	€ 3,823.92	€ -	€ 3,823.92
Total	**€ -**	**€ 15.295,70**	**€ 1,911.96**	**€ 3,823.92**

From the results in the tables above, it becomes clear that with the current schedule, a lot of costs can be saved during the winter (between €10.914 and €15.296). With the proposed schedules, both during the summer and during the winter, costs can be saved, corresponding with a total that lies between € 4,092.88 and € 5,735.89. Keep in mind that for the proposed schedules this total is in addition to the previous savings mentioned in Section 5.4.2.

5.6 Sensitivity Analysis

In the experiments of Section 5.3, six different experiments were conducted for every input sample. The experiments differed from each other in terms of when shifts were allowed to start, how many different start times were allowed and whether the shifts on different days had to start at the same time. The impact of different parameters, such as the service level, weights for under- and overcapacity and the inflation factor (allowance for personal time, fatigue, and delay: A_{pfd}). That these parameters were not investigated in the regular experiments was mostly due to the run times. The run times of the experiments were relatively high, with on average 1.5 hours / experiment with one of the summer samples as input and 2.5 hours / experiment with one of the winter samples as input.

In the following sections, the results of a limited number of experiments are explained to analyse the impact that different parameters have. The parameters that are analysed are the service level, the inflation factor, and the weights for over- and undercapacity. Lastly, the impact of processing RVOs during the night is analysed.

In every set of experiments where one of the parameters is analysed, the experiment where the parameter has the standard value is highlighted by bold numbers. This experiment is called the base experiment and it has the following parameter values: the service level is 98%, the inflation factor for personal time, fatigue, and delay is 15%, the weight for overcapacity is 0.01 and the weight for undercapacity is 1.

For the sensitivity analysis, a specific 'sensitivity-sample' has been created. This sample contains many periods (17,544). The periods are selected in such a way that every year, month, and day is represented equally to prevent a bias. The settings of experiment 4 (see section 5.3.1) are used, since the model was able to create useful schedules while also providing the model with adequate freedom.

5.6.1 Impact of the service level

To evaluate the impact that the service level has on the model, four experiments were conducted. In the four experiments, the service level changes from 95% - 99%. For each experiment, the schedule that results from the experiment is provided in Appendix G. The main characteristics as well as the performance of the schedule are provided in Table 26.

Table 26: Experiments sensitivity analysis (Service level)

Exp.	Minimum Service Level	Shifts scheduled / week	Realized service level	Overcapacity / Week	Undercapacity / Week
1	99%	54	99.03%	294.86	0.75
Base	**98%**	**50**	**98.15%**	**255.64**	**1.43**
2	97%	47	97.08%	232.54	2.47
3	96%	43	96.04%	201.44	3.39

From the results in Table 26, it is clear that the set service level has a considerable impact on the decisions made by the schedule. Therefore, it is particularly important that the service level is set at the desired level. One way to make the service level less impactful would be by changing the weights given to the amount of under- and overcapacity per week. If the weight for the undercapacity is increased sufficiently, then a higher service level than the minimum service level will be attained anyway. Increasing the weight for the overcapacity would not make much difference in this situation, because the model already attempts to minimize the number of shifts and with that the overcapacity in the current situation.

5.6.2 Impact of the inflation factor

In section 5.1.4, the task times for incidents and RVOs were determined using the results of the work sampling study. It is normal to inflate these task times with a certain factor to incorporate personal time, fatigue, and delays. In section 5.1.5, the choice for an allowance factor of 15% is elaborated. However, it is difficult to determine the most accurate value for this allowance factor for the MKS. Therefore, this impact of changes in this factor are evaluated by a sensitivity analysis. To evaluate the impact that the inflation factor has on the model, five experiments were conducted, in which the inflation factor ranged between 5% and 25%. The schedules that resulted from these experiments are provided in Appendix G. The results of these five experiments are provided in Table 27.

Table 27: Experiments sensitivity analysis (Inflation factor)

Exp.	Inflation factor	Shifts scheduled / week	Realized service level	Overcapacity / Week	Undercapacity / Week
4	25%	54	98.05%	268.48	1.62
5	20%	53	98.11%	270.72	1.60
Base	**15%**	**50**	**98.15%**	**255.64**	**1.43**
6	10%	49	98.04%	255.69	1.46
7	5%	45	98.03%	230.83	1.40

The results in Table 27 show that the inflation factor also has a significant impact on the results. The model needs multiple shifts more to meet the service level with a higher inflation factor. The number of shifts in the optimal schedule also decreases quickly when the inflation factor is lower. These results are further discussed in the limitations.

5.6.3 Impact of weights for under- and overcapacity

In the objective function that the model uses to determine the optimal schedule, different weights are attached to the average under- and overcapacity. Since it is a much bigger issue when there is not enough capacity, then when there is some more idle time, undercapacity is weighed much more heavily than overcapacity. In the standard experimental settings, undercapacity is weighed 100 times as heavy as overcapacity. This has ensured that the penalty's given for under- and overcapacity in the objective function are in the same order of magnitude.

To evaluate the effect of different weights, the following experiments were executed: In the second experiment, the weight for overcapacity is omitted completely. In the second experiment, the weight for the overcapacity is set equal to the weight for undercapacity, both are set to 1. In the third experiment, the weight for undercapacity is set to 0.01, while the weight for undercapacity is again 1.

Table 28: Experiments sensitivity analysis (Weights for under- and overcapacity)

Exp.	Weight overcapacity	Weight undercapacity	Shifts scheduled / week	Realized service level	Overcapacity / Week	Undercapacity / Week
Base	**0.01**	**1**	**50**	**98.15%**	**255.64**	**1.43**
1	0	1	50	98.15%	255.64	1.43
2	1	1	50	98.15%	255.64	1.43
3	1	0	50	98.02%	255.61	1.56

From the results in Table 28, one can conclude that the impact of the selected weights on the under- and overcapacity do not have much impact on the created schedules. Only when the weight for the undercapacity is left out, then the schedule changes slightly, decreasing the overcapacity per week while keeping the number shifts per week the same. These experiments show that, with the base settings, the model tries to minimize the number of shift that is needed to meet the service level constraint. The side effect of this approach is that both the over- and undercapacity are (almost) minimized. This is also expected, since minimizing the number of periods where there is undercapacity also minimizes the total undercapacity when the undercapacity is evenly distributed over the periods with undercapacity. Furthermore, minimizing the number of shifts reduces the overcapacity per week, since more shifts generally means more overcapacity.

5.6.4 Impact of processing RVOs during the night

One of the suggestions of the management of the MKS was that the peak workload could be flattened by removing tasks to later times. The most obvious category of tasks that could be postponed is the category of 'low priority RVOs'. For every RVO, its priority is registered in the SAP system and the database. The priority values range from 1 to 9, where 1 means the highest priority and 9 the lowest. All RVOs with a priority of 5 or lower are considered low priority. These low priority RVOs could generally be processed later if necessary. Therefore, the impact of moving the low priority RVOs to quieter hours is evaluated.

The model keeps the capacity during the nights generally at 2 workers. While this is necessary to meet the occasional peak in the workload, it is much higher than the average workload. Especially between 02:00 and 05:00, the workload per dispatcher is much lower than anywhere during the day. An additional set dataset for the sensitivity analysis is created that contains a new estimate for the workload per hour, where the workload that stems from low priority RVOs during the day is moved to the period from 02:00 to 05:00 in the night. The workload from the low priority RVOs is only moved when the workload during the hour in which the RVOs were created was at a level that a capacity of more than two workers would have been necessary. To explain this with an example, if the workload during an hour requires a capacity of three workers, then several low priority RVOs is moved to ensure that the capacity required for that hour drops to two workers. If only one or two workers were needed according to the workload estimate, then no RVO is moved at all. The workload that stems from the moved RVOs is moved to the night at the following day.

Two experiments were executed to see the effects of processing RVOs during the night. The first experiment is created to show how many shifts are needed per week to maintain the same quality level (98% service level) when the RVOs are moved. For this experiment, the settings from the base experiment are used (see the introduction of section 5.6). The only difference is the input dataset, as the values of WE_p have changed. The second experiment evaluates how the schedule created by the base experiment performs when the low priority RVOs are moved to the night. The results of the two experiments are depicted in Table 29.

Table 29: Experiments sensitivity analysis (processing RVOs during the night)

Exp.	Input sample	Schedule	Shifts scheduled / week	Realized service level	Overcapacity / Week	Undercapacity / Week
Base	**Sensitivity sample**	**Schedule from base model**	**50**	**98,15%**	**255,64**	**1,43**
8	Sensitivity sample with moved RVOs	Created by scheduling model	49	98,22%	253,25	1,45

9	Sensitivity sample with moved RVOs	Schedule from base model	50	98,37%	255,52	1,31

From the results in Table 29, one can conclude that processing RVOs during the night has a limited effect. The service level can be attained with one shift less. If the low priority RVOs are processed during the night with the schedule from the base model, then the service level is 98.37% - 98.15% = 0.22pp higher than when the RVOs are processed directly. Furthermore, the undercapacity decreases with 1.43 – 1.31 = 0.12 hours per week. However, contractors would often have to wait longer for instructions and therefore it is questionable if implementing this policy is valuable

5.7 Conclusions

In this section, the conclusions from the work sampling study, the workload estimation model, the experiments, and the sensitivity analysis are provided. In Chapter 0, these conclusions are linked to the other chapters.

The results of the work sampling study showed that in the current schedule slightly more than 25% of the time is contributed to the four core tasks of the MKS: communication by telephone and transceiver, administration of incidents and processing RVOs. Furthermore, slightly less than 25% is spent on other activities and nearly half of the time is idle time. Using the results from the work sampling study, it was determined that on average 1.685 minutes were spent on the intake of an incident (excluding the follow-up and the work done while the dispatcher was still on the phone). Additionally, 2.472 minutes were spent on processing one RVO. The task times were inflated to normal times with a factor of 10%, to allow for personal time, fatigue, and delay. Furthermore, 15% of the time per hour is reserved for idle time as a buffer for situations where much of the work must be done at the same time. Besides, from an analysis into the 'other activities', it turned out that 4.72% of the time is spent on important, non-movable other activities. Therefore, an additional 4.72% of the time is reserved for these activities.

After the parameters obtained in the work sampling study were inserted in the workload estimation model, a more thorough analysis of the workload could be conducted. From this analysis, it appeared that there was a positive trend in the workload, which means that the average workload has grown over the last four year. In addition, a seasonality analysis showed that there are strong seasonal influences during the day, the week and between the summer and the winter.

For the experiments, historical estimates for the workload were used for the input samples. To create suitable samples, the data first had to be prepared. The outliers were filtered out by omitting days with an extremely high workload based on the interquartile range rule. The data has been de-trended based on a linear trend. Lastly, the data has been split into a summer and winter season, so that the differences between these seasons could be exploited.

In Section 5.3, the setup and results of 24 experiments are explained. In these experiments, the scheduling model creates an optimal schedule in which it minimizes the number of shifts while meeting the service level and penalizing both under- and overcapacity. In the results, one can see that the model needs less shifts when it is provided with more freedom. However, when these schedules are examined on the test sample, they score higher on the expected undercapacity and have a lower service level than schedules with more shifts. This shows that the model is to a certain extent affected by overfitting (see Section 5.3.3). Furthermore, it was remarkable that significantly less shifts were scheduled by the model than the standard number of shifts the current schedule has. Moreover, the model assigns around 6 shifts more with the summer sample as input than with the winter sample, while the current schedule is the same in the summer and the winter. Therefore, one can conclude that ProRail currently does not exploit the workload differences between the seasons.

In Section 5.4, a series of experiments were executed that can support the decision-making process. An analysis into the optimal set of start times with 8 hours in-between concluded that the start times of 7:00, 15:00 and 23:00 were the most preferable. An analysis into the optimal start time for a dayshift as replacement for the combination of a morning and evening shift showed that starting this dayshift at 10:00 in the morning was the most preferable. Based on these start times for the regular shifts and day shifts, the impact of removing two shifts or replacing two shifts with a day shift has been analysed. The analysis was based on the current schedule with 56 shifts and showed which shifts should be omitted first in case of a capacity disturbance. The results of this analysis are shown in Table 22.

In Section 5.6, a sensitivity analysis is used to determine the impact of parameter settings on the results. From this analysis, it became clear that the inflation factor and the service level have a big impact on the schedule and its performance. A lower service level means that multiple shifts less are needed per week and the same holds for a lower inflation factor. The sensitivity analysis also showed that the weights for the over- and undercapacity do not have a big impact on the schedule and its performance. Lastly, the effect of processing low priority RVOs during the night was investigated. From this analysis it appeared that removing RVOs had an impact, but not a major one. One shift less would be needed per week to meet the service level, or the service level would increase with 0.22 percentage points in case the schedule from the base experiment for the sensitivity analysis would be used.

6 Conclusion & Discussion

In this chapter, the results of Chapter 5 are explained and interpreted. The following research question is centra in this chapter:

How should the results of the performance be interpreted and what consequences would the implementation of the models have?

In Section 6.1, the different findings in this study are connected to answer the central research question. In Section 6.2, the validity and limitations of this research are discussed, even as the suggestions for further research. Finally, in Section 6.3, the recommendations for ProRail are provided.

6.1 Conclusions

This research aimed at providing ProRail insights about how the capacity at the MKS could be deployed efficiently. The research question that was answered in this research was the following:

"How can ProRail deploy the capacity at the MKS in the most efficient and effective way and thus reduce both undercapacity and overcapacity while meeting the service level"

In this research, a scheduling model was created that minimized the required capacity per week while meeting the service level. This scheduling model is based on the elastic set-covering model, where both undercapacity and overcapacity are allowed. The model uses estimates of the workload that are based on historical data as input. To create reliable estimates for the workload, a work sampling study has been conducted at the MKS.

From the results of the work sampling study, it appeared that in the current situation, approximately half of the available time can be categorized as idle time. Furthermore, the study showed that only a quarter of the available time was spent on the core tasks.

The results of the work sampling study were used to improve the accuracy of the workload estimation model. From the general workload analysis, it became clear that there is a positive trend in the workload per month, meaning that the average workload grows each month. Furthermore, there are significant seasonal effects in the workload. The analysis in Section 5.2.3 showed that the seasonal effect is the strongest on an hourly level. However, there were also significant seasonal effects from the day of the week and the month of the year. On an hourly level, the workload is the lowest during the night, it rises quickly in the morning and reaches a peak in the afternoon, to then decrease gradually decrease toward the night. On a daily level, the seasonality analysis showed that the workload in the weekend is significantly lower than during the week. On a monthly level, the workload during the summer (from June – September) is significantly higher than during the rest of the year (October – May).

Currently, these seasonal effects are only exploited moderately. The hour seasonality is exploited by having a lower capacity during the night. The day seasonality is exploited even less, since the schedule is the same every day. The only difference between the weekdays and the weekend is that no relief shift is scheduled during the weekend. The month seasonality is currently not exploited.

The experiments of Section 5.3 showed that the 98% service level could be met with on average less shifts than the 56 of the current schedule. If the schedule must meet the same requirements for practicality as the current schedule (at most four separate start times that do not vary between different days), then 57 shifts are needed during the summer and 51 shifts are needed for the winter to meet the service level. Keep in mind that the data is detrended to represent the workload in 2022. The increasing trend in the workload is incorporated in the workload estimates and therefore, 56 shifts might have been sufficient for the summer of 2021, while it is not expected to be sufficient for the summer of 2022.

Implementing the schedules recommended in Section 5.4, improves the service level during the summer and reduces the overcapacity during the winter. The differences in performance between the seasons are reduced and the number of FTE required to meet this service level is decreased with 0.667 FTE. With the current labour costs, this would save ProRail between €29,769.60 and €41,720.00 per year.

In conclusion: the management of the MKS can deploy the capacity at the MKS more efficiently by exploiting the seasonality effects during the week and during the year. For the upcoming summer, it is estimated that one more shift is needed to meet the required service level compared to the current schedule, while during the winter, the service level can be met with 5 shifts less than in the current situation. The practical implications of these results discussed are discussed in Section 6.3: Recommendations. However, an asterisk that must be placed by these results is that the scheduling model was sensitive for changes in the inflation factor and the service level.

6.2 Discussion

In this section, the results and conclusions are interpreted. The validity of this research and its limitations are discussed. This section ends with several suggestions for further research.

6.2.1 Validity

This research aimed at matching the capacity to the workload at the MKS of ProRail. Determining an accurate estimate of the workload was one of the major challenges in this research. An estimate based only on historical data and expert estimates was deemed not sufficiently reliable. Therefore, the work sampling study was conducted to verify the accuracy of the estimates. As described in Section 5.1.3, the estimates were close to the measured proportions. The estimates for four out of five categories were within one percentage point of the measured proportion. In the end, the task times were based on the work sampling study and the expert estimates may be seen as validation.

Regarding the external validity of the work sampling study, the following are relevant: 11 out of the 22 workers at the MKS have been observed for some period in the work sampling study. This is 50% of the total population and therefore one can conclude that the population validity in this dimension is guaranteed. However, the work sampling study was conducted during four different days in January and February '22. This is a relatively small number of sampling days, which can be a threat for the external validity of the results. However, the number of SpoorWeb dossiers and RVOs that were created during the observed hours is known. In total 94 SpoorWeb dossiers and 73 RVOs were created during the observed period, which is 94 hours of worktime. Here, worktime means the number of workers that are present during an hour multiplied by the number of hours worked. This equates to exactly 1 SpoorWeb dossier per hour and 0.78 RVOs per hour. Over the last four years, on average 0.85 SpoorWeb dossiers 0.63 RVOs were created per hour of worktime. This shows that the observed hours were only slightly busier than normally.

Furthermore, the results of the work sampling study might have been influenced by the Hawthorne-effect. This is the tendency of participants to behave differently because they know they are being investigated. However, as researcher I have been physically present at the MKS at many other instances, and I did not notice any difference in behaviour during the work sampling study compared to other instances when I was there.

Furthermore, to guarantee the internal validation of the experiments, 2-fold cross validation was used. This means that a schedule was created based on a set of training data and the performance was evaluated on a set of test data. The training and testing sets were built up in such a way that the seasonal factors were equally prevalent in both sets, but the sets themselves were mutually exclusive.

6.2.2 Interpreting the results

Regarding the results of this study, there were some expected and some unexpected results. First, the work sampling study showed that only a relatively small proportion of the available worktime is spent on the core tasks (approximately 26.5%). This was very close to the estimate of 27.3%, which was based on a combination of historical data and expert estimates. Furthermore, the seasonal effects throughout the day and throughout the week were expected. The dispatchers as well as the management of the MKS was aware that the workload during the weekend was generally lower than during the week and that the workload was higher during the day than during the night, the early morning, and the late evening.

The difference in average workload and peaks in the workload between the summer and the winter was bigger than expected. This might have been the result from the fact that the last few winters have been mild, while the summers were very hot and very hot. Besides, the week of snow weather in February ’21 has been filtered out, which would otherwise have decreased some of the difference in workload between the two seasons.

Moreover, the schedules that resulted from the experiments with the scheduling model showed that the required number of shifts per week was significantly lower than the number of shifts that are currently deployed each week. It might be the case that the hourly workload estimate is not the most accurate predictor of the actual workload. This theory is further discussed in the limitations (Section 6.2.3). The shifts that were often removed by the scheduling model were anticipated upfront. As indicated in Figure 27 tot Figure 30, the model assigns less capacity to the night than to the day, less capacity to the weekend than to the weekdays and less capacity to the winter than to the summer, which is completely in line with the results from the seasonality analysis in Section 5.2.3.

Regarding the model used in this research: the elastic set-covering model was an understandable choice, since this model allows both over- and undercapacity. This matched very well with the volatile workload, since a lot of capacity would be required when the model would not allow undercapacity. However, the requirement to meet a high service level resulted in a situation where the average capacity was always well above the average workload, resulting in a situation where the penalties for over- and undercapacity became almost irrelevant. Furthermore, when the dataset used for the model was small, the model suffered from overfitting. This is visible in the results in Table 13 as well: the experiments with the smaller input samples (experiments 1 – 12) always had a service level on the test set that was lower than 98% (except for experiment 4), the experiments with the larger input sample (experiment 13 – 24) suffered less from overfitting and for these experiments, the service level was in some cases even higher on the test set than on the training set. The experiments also took a long time to run, which is due to the larger number of variables, parameters, and constraints. A model that was better suited to deal with a substantial difference between the average capacity and the average demand (workload) might have been more appropriate.

6.2.3 Limitations

There are several limitations to consider when interpreting the results of this study. The first concerns the category of ‘other activities’. There was no data available to accurately estimate the workload that stems from these activities. Based on descriptions of the other activities during the work sampling study, the researcher determined that almost all these activities could be either removed or omitted in a busy situation. However, this might be an oversimplification, since it is based on only four days of observations. Therefore, the workload that stems from other activities and its impact on the total workload should be determined in later research. Removing these other activities from busy periods to quieter periods (e.g., the night) might reduce the peaks in the workload.

The workload estimates were created to determine how much capacity was needed for every hour in a sample. While the workload estimate gives a relatively accurate estimate for the total workload in an hour, the timing of tasks within the hour is not considered. There are certainly cases where much of the work arrives at the same time (e.g., within the same 15 minutes of an hour). It might be that the workload that arrives within these 15 minutes is not enough for the model to assign a capacity of more than two workers to this hour, since the model only examines the workload per hour. Even while the workload could never be handled with two workers during this 15-minute period. Therefore, one might argue that the workload estimation is an underestimation of the actual capacity that is needed. However, reducing the time buckets further comes with disadvantages as well. There might have been situations in the past when dispatchers have created several relatively simple RVOs and SpoorWeb dossiers in a short period of time, that might cause the workload estimate to be very high for that period if small time buckets are used. The scheduling model could for example conclude that a capacity of more than 6 people would be needed to meet the workload, even while this is not realistic either.

Furthermore, the reader should bear in mind that the model turned out to be rather sensitive to changes in two parameters: the inflation factor and the service level. The service level is a parameter/KPI that is set by the management. A trade-off between the service level and the performance of the schedule is logical and not an issue. However, the inflation factor that is used to inflate the workload to compensate for personal time, fatigue and delay also had a big impact on the characteristics of the schedule. While the choice for the value of 10% is substantiated in this research, it is not certain if the value is accurate for the MKS. This uncertainty in a parameter with a considerable impact on the results is therefore a limitation of this research.

6.2.4 Suggestions for further research

The first suggestion for further research was already mentioned in the limitations section. There is a lack of data about the other activities that dispatchers perform. Mapping what these activities are, with which frequency they occur, how long they take and whether they can be moved would strengthen the workload estimate of this research. Based on the labelled observations from the work sampling study, it is expected that most of the tasks in this category can be moved and therefore, tasks in this category could be used to flatten the peaks in the workload.

In this research, a few experiments have been conducted to evaluate the impact of processing low priority RVOs during the night. While this work category was the most obvious to move to a later, quieter moment, there might be other tasks that can be moved or outsourced when there is not enough capacity at the MKS. It has for example been suggested that train dispatchers, who are the most frequent interlocutor of the MKS, could fill in the SpoorWeb form for frequently occurring incidents themselves. This would save the workload of a phone call and registering the incident every time that this would occur. Therefore, this is an interesting subject for further research.

From the workload analysis, it was obvious that the workload during the summer was significantly higher than during the winter. Dispatchers at the MKS explained that the high number of lateral distortions of the track might cause this difference. It is currently not clear what the main causes of the difference are, except that more incidents happened, more RVOs were created, and more telephone calls occurred. It might for example also be related to the covid-19 crisis, since society was more open during the summer than during the winter during the pandemic and therefore more people have travelled by train during the summer than during the winter. It is important to know whether this is the main cause, because if that is the case, then this seasonal effect will decrease or even vanish know that the pandemic has probably come to an end. Therefore, investigating the cause of the difference is suggested for further research.

6.3 Recommendations

The results of the scheduling model in Section 5.3, show that it is possible to meet the service level during the summer with 57 shifts and during the winter with 51 shifts. This is the case when the schedule must meet the strict requirements of having at most 4 different start times per day that do not deviate between different days of the week. It is difficult to determine the most preferable schedule for the summer or winter based on the results of the 24 experiments in Table 13. There are no schedules that dominate all other schedules. Furthermore, there is the issue that the model suffers from overfitting to a certain extent and that the workload estimate might underestimate the actual capacity needed in some cases (as explained in Section 6.2.3).

However, two schedules that prompted from the experiments had a relatively robust service level and scored the required service level of 98% on the test set, while also meeting the requirement of having at most four different start times. Therefore, these two schedules are recommended to ProRail in Section 5.4. In these schedules, dispatchers start at either 7:00, 10:00, 15:00 or 23:00, which match the optimal start times for regular shifts and day shifts that were determined in Section 5.5. These schedules are recommended, since the requested service level can be attained with these schedules, with less shifts on average than the current schedule. The exact schedules are portrayed on page 72 and page 73, and are shown again in smaller versions underneath in Figure 42 and Figure 43.

Figure 40: Copy of the recommended summer schedule

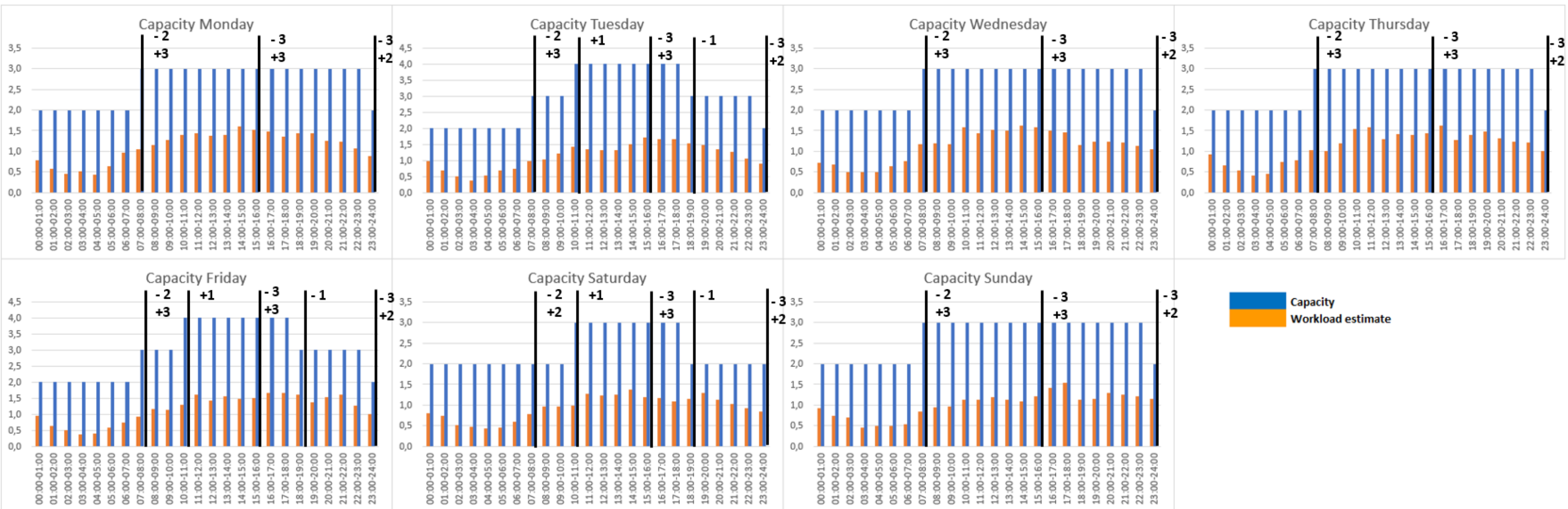
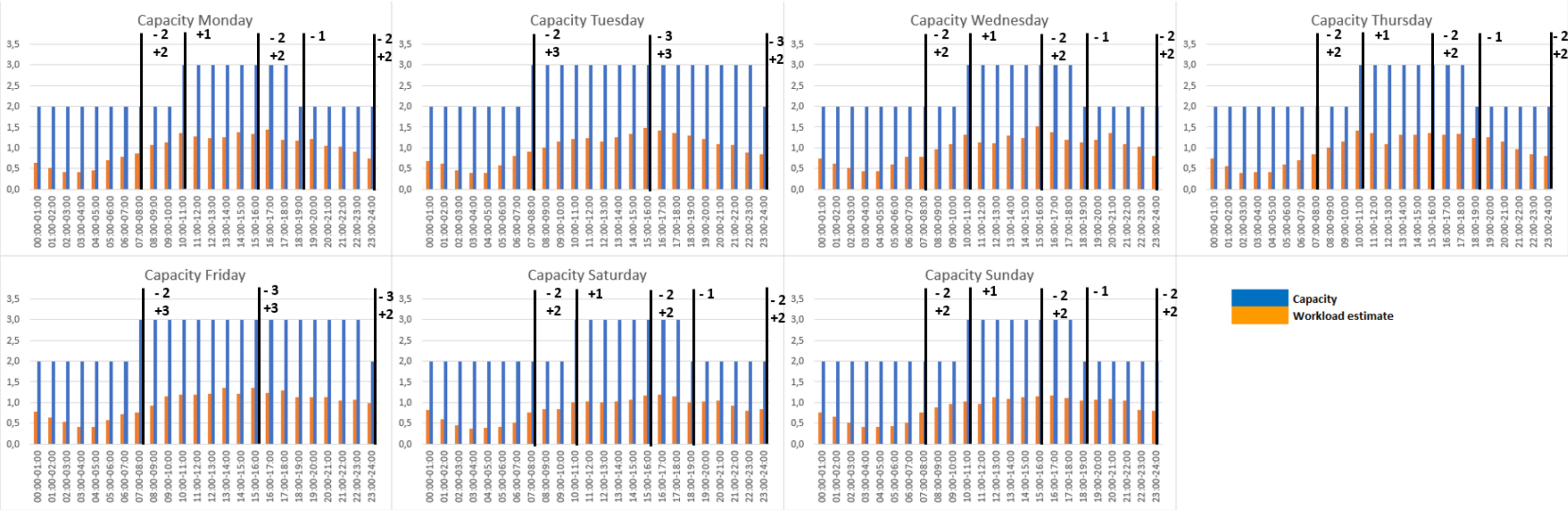


Figure 41: Copy of the recommended winter schedule



7 References

Aksin, Z., Armony, M., & Mehrotra, V. (2009). The Modern Call Center: A Multi-Disciplinary Perspective on Operations Management Research. *Production and Operations Management, 16*(6), 665–688. doi:10.1111/j.1937-5956.2007.tb00288.x

Atlason, J., Epelman, M., & Henderson, S. (2008). Optimizing Call Center Staffing Using Simulation and Analytic Center Cutting-Plane Methods. *Management Science, 54*(2), 295-309. doi:10.1287/mnsc.1070.0774

Avramidis, A., Chan, W., & L'Ecuyer, P. (2009). Staffing multi-skill call centers via search methods and a performance approximation. *IIE Transactions, 41*(6), 483-497. doi:10.1080/07408170802322986

Bailyn, L., Collins, R., & Song, Y. (2007). Self-scheduling for hospital nurses: an attempt and its difficulties. *Journal of Nursing Management, 15*(1), 72-77. doi:10.1111/j.1365-2934.2006.00633.x

Cao, Y., & Shen, Z. (2019). Quantile forecasting and data-driven inventory management under nonstationary demand. *Operations Research Letters, 47*(6), 465-472. doi:10.1016/j.orl.2019.08.008

Cezik, M., & L'Ecuyer, P. (2008). Staffing Multiskill Call Centers via Linear Programming and Simulation. *Management Science, 54*(2), 310-323. doi:10.1287/mnsc.1070.0824

De Bruijn, D., Zwaagstra, A., & Van de Weide, R. (2005). *Werkbelastingmethodieken gewogen.* Utrecht. Retrieved October 28, 2021

Ernst, A., Jiang, H., Krishnamoorthy, M., & Sier, D. (2004). Staff scheduling and rostering: A review of applications, methods and models. *European journal of operational research, 153*(1), 3-27. doi:10.1016/S0377-2217(03)00095-X

Eveborn, P., & Rönnqvist, M. (2004). Scheduler - A System for Staff Planning. *Annals of Operations Research, 128*(1-4), 21-45. doi:10.1023/b:anor.0000019097.93634.07

Gans, N., Shen, H., Zhou, Y., Korolev, N., McCord, A., & Ristock, H. (2015). Parametric Forecasting and Stochastic Programming Models for Call-Center Workforce Scheduling. *Manufacturing & Service Operations Management, 17*(4), 571-588. doi:10.1287/msom.2015.0546

Groover, M. P. (2013). *Work Systems: The Methods, Measurement & Management of Work.* Pearson Education Limited.

Harper, S., & Mousa, F. (2013, January 28). *Time and Motion Studies.* doi:10.1093/OBO/9780199846740-0027

Heerkens, H., & Van Winden, A. (2012). *Geen Probleem: Een aanpak voor alle bedrijfskundige vragen en mysteries.* Nieuwegein: Van Winden Communicatie.

Human Reliability. (2021). *Human Reliability*. Retrieved October 27, 2021, from Human Reliability: https://www.humanreliability.com/human-cognitive-workload-assessment-tool/

Inspectie Verkeer en Waterstaat. (2005). *Onderzoek naar FNV-zorgpunten.* Utrecht. Retrieved October 28, 2021, from https://docplayer.nl/10548313-Onderzoek-naar-fnv-zorgpunten.html

Khatoon, S., Ibraheem, Singh, A., & Priti. (2014). Analysis and Comparison of Various Methods Available for Load Forecasting: An Overview. *Innovative Applications of Computational Intelligence on Power, Energy and Controls with their impact on Humanity (CIPECH)*, 243-247. doi:10.1109/cipech.2014.7019112

Kirkpatrick, S., Gelatt, C., & Vecchi, M. (1983). Optimization by Simulated Annealing. *Science, 220*(4598), 671-680. doi:10.1126/science.220.4598.671

Kraai, T. (2021). *Incident end time prediction during the incident recovery process.* University of Twente. Retrieved 10 27, 2021, from https://essay.utwente.nl/85779/1/Kraai_MA_BMS.pdf

Liao, S., Koole, G., van Delft, C., & Jouini, O. (2011). Staffing a call center with uncertain non-stationary arrival rate and flexibility. *OR Spectrum, 34*(3), 691-721. doi:10.1007/s00291-011-0257-0

Liao, S., van Delft, C., & Vial, J. (2013). Distributionally robust workforce scheduling in call centres with uncertain arrival rates. *Optimization Methods and Software, 28*(3), 501-522. doi:10.1080/10556788.2012.694166

Lokad. (2021). *Forecasting Technology FAQ - Inventory Optimization Software*. Retrieved november 17, 2021, from Lokad.com: https://www.lokad.com/forecasting-technology-faq#How_accurate_are_your_forecasts_1

Mason, A. J. (2012). OpenSolver - An Open Source Add-in to Solve Linear and Integer Programmes in Excel. *Operations Reserach Proceedings 2011*, 401-406. doi:10.1007/978-3-642-29210-1_64

Mehrotra, V., & Fama, J. (2003). Call center simulation modeling: methods, challenges, and opportunities. *Proceedings of the 2003 Winter Simulation Conference: Driving Innovation*, 135-143. doi:10.1109/wsc.2003.1261416

PetroWiki. (2016, May 31). *Probabilistic verses deterministic in production forecasting*. Retrieved October 28, 2021, from Petrowiki: https://petrowiki.spe.org/Probabilistic_verses_deterministic_in_production_forecasting

ProRail. (2019). *Jaarverslag 2019.* Retrieved October 27, 2021, from ProRail: https://www.prorail.nl/over-ons/organisatie/jaarverslagen

ProRail. (2020). *Jaarverslag ProRail.* Retrieved October 27, 2021, from Jaarverslag ProRail: https://www.jaarverslagprorail.nl/verslag/het-jaar-2020

ProRail. (2021). *ProRail - Over Ons*. Retrieved October 27, 2021, from ProRail: https://www.prorail.nl/overons/organisatie

Silver, E. A., Pyke, D. F., & Thomas, D. J. (2016). *Inventory and Production Management in Supply Chains.* Boca Raton: CRC Press. doi:10.1201/9781315374406

van Buuren, M., Kommer, G., van der Mei, R., & Bhulai, S. (2017). EMS call center models with and without function differentiation: A comparison. *Operations Research for Health Care, 12*, 16-28. doi:10.1016/j.orhc.2016.12.001

Veldhoven, M., Jonge, J., Broersen, S., Kompier, M., & Meijman, T. (2002). Specific relationships between psychosocial job conditions and job-related stress: A three-level analytic approach. *Work & Stress*, 207-228. doi:10.1080/02678370210166399

Verweij, B., Ahmed, S., Kleywegt, A. J., Nemhauser, G., & Shapiro, A. (2003). The sample average approximation method applied to stochastic routing problems: A computational study. *Computational Optimization and Applications*(24), 289-333. doi:10.1023/a:1021814225969

Voxco. (2021). *Idle Time in Contact Centers | Why it is important?* Retrieved March 2022, from voxco.com: https://www.voxco.com/idle-time/

Appendices

Appendix A

In the table below, every TIS-code between TIS 1.1 and TIS 5.4 is explained. This is a translation of the official table that is used by ProRail.

Table 30: TIS-matrix (ProRail, 2020)

Scenario	Situation	Severity	Consequences
Disruption train service			
TIS 1.1	Disruption train service (Due to, for example, a derailment without casualties, defect material, a power failure, bad weather, suspension of train service on the orders of emergency services, stranded train at a platform or a switch failure)	Very limited	-Despite the disruption, the train service can be carried out almost according to plan. Structural delay of 5 minutes or more, for 30 minutes or longer. -1 or more trains can be cancelled. In the event of a failure of complete train series, upscaling to TIS 1.2 will follow -Possibly an evacuation of a stranded train must be carried out
TIS 1.2	Disruption train service (Due to, for example, derailment without casualties, defect material, a power failure, infrastructure failure, bad weather, urgent unplanned work or suspension of train service on the orders of emergency services)	Limited	-Due to the disruption, the train service can no longer be carried out according to plan. Structural delay of 5 minutes or more for 30 minutes or longer. One or more train series are suspended. -Possibly an evacuation of a stranded train must be carried out
TIS 1.3	Total blockage. Train service no longer feasible (Due to, for example, derailment without casualties, defect material, a power failure, infrastructure failure, bad weather, urgent unplanned work or suspension of train service on the orders of emergency services)	Severe	-Due to the disruption, the train service can no longer be delivered. Total blockage. -Possibly an evacuation of a stranded train must be carried out
TIS 1.4	Total blockage. Train service no longer possible in at least one post or busy junction (Due to, for example, a failure of the central infrastructure control system or suspension of train service on the orders of emergency services	Very severe	-Due to the disruption, the train service can no longer be delivered within the entire service area of a station. -Possibly an evacuation of a stranded train must be carried out
Fire			
TIS 2.1	- Automatic fire alarm (at station) in tunnel, without stationary train - Smoke development and/or fire symptoms at or under a train on an open track or yard - Fire in station(building) that does not spread - Fire near or in the track with possible influence on train traffic (e.g. a sleeper on fire or a roadside fire)	Very limited	-Depending on the approach of the fire brigade, a short-term disruption of the train service. -Possibly rescuers in the track, possibly causing decommissioning of the track.
TIS 2.2	- Fire (at station) in tunnel, without stationary train - Smoke development and/or fire phenomena near and/or under train at station or in tunnel	Limited	-Possibly rescuers in the track, possibly causing decommissioning of the track. -Possibly an evacuation of a stranded train must be carried out

	- Fire in train does not spread (e.g. fire in trash can or smouldering fire in train) - Fire in station(building) spreading		
TIS 2.3	Fire in train, spreading on a free track or yard	Severe	-As a result of fire, no or limited train traffic is possible on one or more tracks -Rescuers in the track, possibly causing decommissioning of the track. -Procedure 'electrocution-safe workplace' can be started by IB or emergency services. Complete line shutdown (CLU) is started automatically. -Possibly an evacuation of a stranded train must be carried out
TIS 2.4	- (Automatic) fire (notification) in tunnel, with stationary train - Fire in train breaking out at station or in tunnel - Train stoppage in tunnel, no audio contact possible with train staff	Very severe	-As a result of fire, no or limited train traffic is possible on one or more tracks -Rescuers in the track, possibly causing decommissioning of the track. -Procedure 'electrocution-safe workplace' can be started by IB or emergency services. Complete line shutdown (CLU) is started automatically. -Possibly an evacuation of a stranded train must be carried out
Collision, or derailment with casualties			
TIS 3.1	Collision of train with: - person or large livestock - (moped) cyclist - infra element or object - small road vehicle, such as a car, motorcycle, delivery van or tractor (verifiably without casualties)	Very limited	-Possibility of some casualties and emotionally affected persons. -Rescuers in the track, possibly causing decommissioning of the track. -Crossings in malfunction, possibly slightly damaged infrastructure and equipment. -Possibly stranded trains with evacuation scenarios and impact on the timetable.
TIS 3.2	-Collision between shunting vehicles - Hard coupling (train with train or shunting vehicle - Collision of a train or shunting vehicle with a small road vehicle (with casualties or casualties unknown) or large road vehicle, such as a bus or truck (without casualties)	Limited	-Possibility of some casualties and emotionally affected persons. -Rescuers in the track, possibly causing decommissioning of the track. -Possibly stranded trains with evacuation scenarios and impact on the timetable. -Some media attention -Possibly slightly damaged infrastructure and equipment (but trains can leave on their own)
TIS 3.3	Derailment with victims in a train, or collision of a train with: - train or shunting part - large road vehicle (with casualties or casualties unknown), but carriages are not deformed, tilted or stacked and the overhead line group has not failed	Severe	-Possibly multiple victims and emotionally affected persons. -Rescuers in the track, possibly causing decommissioning of the track. -Severely damaged infrastructure and equipment -Possibly stranded trains with evacuation scenarios and high impact on the timetable. -A lot of media attention
TIS 3.4	Derailment with victims in a train, or collision of a train with: - train or shunting part - large road vehicle (with casualties or casualties unknown), causing carriages to be deformed, tilted or stacked or the overhead line group has failed	Very severe	-Possibly many victims and emotionally affected persons. -Rescuers in the track, possibly causing decommissioning of the track. -Severely damaged infrastructure and equipment -Possibly stranded trains with evacuation scenarios and high impact on the timetable. -A lot of media attention -Procedure 'electrocution-safe workplace' can be started by IB or emergency services. Complete line shutdown (CLU) is started automatically.
Hazardous substances			

TIS 4.1	-Small outflow of hazardous substance -Outflow of unknown substance -GEVI code starts with 7 -Collision/derailment of freight train with danger of outflow of hazardous substances	Very limited	-Display: drip, hiss, stink, slight leakage valve. -Impact: source area, the immediate vicinity of the incident. For example, the area immediately around the incident vehicle. -Possibly environmental damage -Rescuers in the track, possibly causing decommissioning of the track.
TIS 4.2	Fire in freight train involving hazardous substances	Limited	-Impact: source area, the immediate vicinity of the incident, but a life-threatening situation can also arise outside the source area. -Procedure 'electrocution-safe workplace' can be started by IB or emergency services. Complete line shutdown (CLU) is started automatically. -Rescuers in the track, possibly causing decommissioning of the track. -Possibly environmental damage
TIS 4.3	Large outflow of hazardous material of which GEVI code starts with 3, 4, 5, 6, 8 or 9	Severe	-Possibly multiple casualties in the source and effect area. -Impact: source area, the immediate vicinity of the incident and there is a danger to public health outside the source area. -Procedure 'electrocution-safe workplace' can be started by IB or emergency services. Complete line shutdown (CLU) is started automatically. -Rescuers in the track, possibly causing decommissioning of the track.
TIS 4.4	Large outflow of hazardous substance of which GEVI code starts with 2	Very severe	-Possibly multiple casualties in the source and effect area. -Impact: source area, the immediate vicinity of the incident and there is a danger to public health outside the source area due to the rapid spread of the gas. -Procedure 'electrocution-safe workplace' can be started by IB or emergency services. Complete line shutdown (CLU) is started automatically. -Rescuers in the track, possibly causing decommissioning of the track.
Suspicious object/behaviour or bomb threat			
TIS 5.1	-Bomb threat -Suspicious behaviour -Suspicious object along a free track -Unexploded conventional explosives are found on a railway terrain	Very limited	-Evaluated and considered serious by the police. -Possibility of impact on train service.
TIS 5.2	Bomb discovery or suspicious object in train on the open track	Limited	-Evaluated and considered serious by the police. -Possibility of impact on train service.
TIS 5.3	Bomb discovery or suspicious object: serious threat - in train at station - in tunnel or station	Severe	-Evaluated and considered serious by the police. -Chance of long-term decommissioning due to investigation of explosives search command.
TIS 5.4	Bomb Explosion: - in train, station or tunnel	Very severe	-Chance on many casualties. -A lot of damage to infrastructure, e.g. tunnel out of service for a long time. -Procedure 'electrocution-safe workplace' can be started by IB or emergency services. Complete line shutdown (CLU) is started automatically. -Rescuers in the track, possibly causing decommissioning of the track.

Appendix B

The table below shows which percentage of the incidents gets assigned every TIS-code.

Table 31: Percentage of incidents per TIS-code

TIS	Percentage:
1	72,40%
1,1	8,25%
1,2	11,06%
1,3	4,73%
1,4	0,08%
2,1	0,34%
2,2	0,14%
2,3	0,00%
2,4	0,02%
3,1	2,42%
3,2	0,10%
3,3	0,01%
3,4	0,01%
4,1	0,27%
4,2	0,00%
4,3	0,00%
5,1	0,05%
5,2	0,00%
5,3	0,09%
5,4	0,00%

Appendix C

The questions below are a translation of the real questions that were asked to dispatchers at the MKS. The first section contains a mix of open and MC questions. The second section contains questions where dispatchers were asked to indicate how much work incidents with a specific incident-label are on a scale of 1 (Not much work) to 5 (A lot of work).

1. **On average, how often do you have to deal with a workload that you experience as unpleasant?**
   a. Multiple times per shift
   b. Once per shift
   c. Once per week
   d. Less than once per week
2. **Which of the following situations do you find the most unpleasant?**
   a. A **very** high workload for ten minutes (not enough people to always answer the phone and process everything properly in Spoorweb and SAP)
   b. A high workload for an hour (enough people to answer the phone within 10 seconds, but not enough time to process everything calmly and keep an eye on forecasts)
   c. A longer period (e.g. several hours) with a very low workload (virtually no significant incidents or other activities)
3. **Do you have enough time to work on portfolio tasks?**
   a. More than enough
   b. Just enough
   c. Not enough
   d. By far not enough
4. **Which of the following answers describes your situation best: I work on portfolio tasks …**
   a. during my own time
   b. during a day shift
   c. during a relief shift
   d. during a (quiet) regular shift
5. **Does the composition of the team in a shift have a significant impact on the perceived workload?**
   a. Strong impact
   b. Little impact
   c. No impact
   d. Don't know / no opinion
6. **Does the perceived workload influence the pace of the actions that you execute? (e.g., you try to process a telephone call quicker when the telephone rings and no one can take it at that moment)**
   a. Strong impact
   b. Little impact
   c. No impact
   d. Don't know / no opinion
7. **Does the perceived workload influence the quality of your work? (e.g., you are less accurate in registering information in Spoorweb or SAP when the workload is high)**
   a. Strong impact
   b. Little impact
   c. No impact
   d. Don't know / no opinion

8. **Which situations / incidents that occur regularly cause a high workload?** *(Open question)*

9. **What could be a solution to lower the perceived workload in those situations?** *(Open question)*

10. **How long does it take you to do the intake of an incident?** *(Open question)*

11. **How long does it take you to process an RVO?** *(Open question)*

12. **How long does it take you to fill in a form related to the calls from PuVo, DVP and Storing Publiek?** *(Open question)*

13. **Which registered variable is probably the best indicator for determining the time you spent on an incident?**
    a. The number of loglines (logged by the MKS)
    b. The TIS-code of the incident
    c. The incident-label
    d. The total time the dossier was active

Second section of the survey:

For each of the following incident-labels: indicate how much work stems from incidents of that type in your opinion. One stands for a small amount of work and five stands for a lot of work.

14. **Defect material**

	1	2	3	4	5	
Small amount of work	○	○	○	○	○	A lot of work

15. **Hindrance caused by people on or close to the rails**

	1	2	3	4	5	
Small amount of work	○	○	○	○	○	A lot of work

16. **Section malfunction**

	1	2	3	4	5	
Small amount of work	○	○	○	○	○	A lot of work

17. **Switch malfunction**

	1	2	3	4	5	
Small amount of work	○	○	○	○	○	A lot of work

18. **Crossing malfunction**

	1	2	3	4	5	
Small amount of work	◯	◯	◯	◯	◯	A lot of work

19. Hindrance caused by emergency services

	1	2	3	4	5	
Small amount of work	◯	◯	◯	◯	◯	A lot of work

20. Hindrance caused by an object, vehicle, or animals on or close to the rails

	1	2	3	4	5	
Small amount of work	◯	◯	◯	◯	◯	A lot of work

21. State of the rails

	1	2	3	4	5	
Small amount of work	◯	◯	◯	◯	◯	A lot of work

22. Hindrance caused by behaviour of (a) traveller(s) or personnel

	1	2	3	4	5	
Small amount of work	◯	◯	◯	◯	◯	A lot of work

23. Slippery tracks

	1	2	3	4	5	
Small amount of work	◯	◯	◯	◯	◯	A lot of work

24. Other infrastructure-related problems

	1	2	3	4	5	
Small amount of work	◯	◯	◯	◯	◯	A lot of work

25. Hindrance caused by a logistical problem or error

	1	2	3	4	5	
Small amount of work	◯	◯	◯	◯	◯	A lot of work

26. (Potential) IT failure

	1	2	3	4	5	
Small amount of work	○	○	○	○	○	A lot of work

27. Infrastructure problem related to the environment/surroundings

	1	2	3	4	5	
Small amount of work	○	○	○	○	○	A lot of work

28. Almost a collision

	1	2	3	4	5	
Small amount of work	○	○	○	○	○	A lot of work

29. Hindrance caused by (a) traveller(s) or personnel due to their health

	1	2	3	4	5	
Small amount of work	○	○	○	○	○	A lot of work

30. Collision with an (infrastructure) object

	1	2	3	4	5	
Small amount of work	○	○	○	○	○	A lot of work

31. Signal failure

	1	2	3	4	5	
Small amount of work	○	○	○	○	○	A lot of work

32. Collision with a person

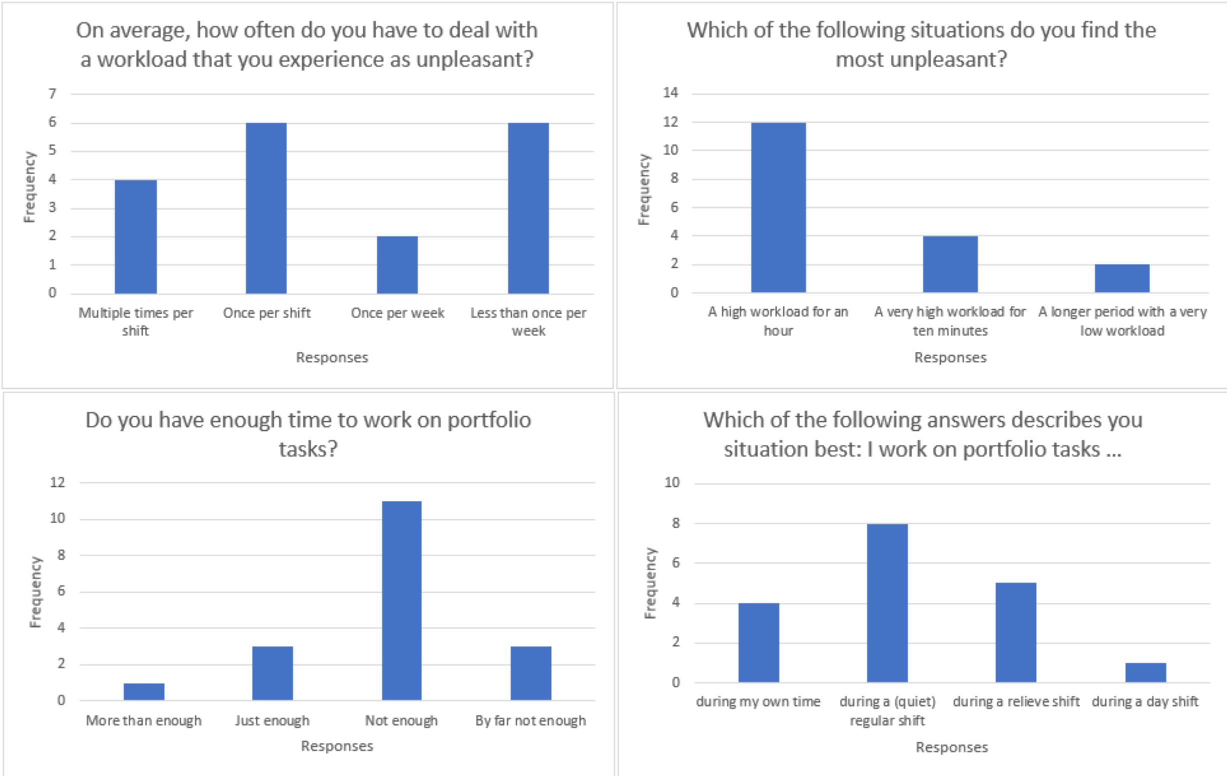
	1	2	3	4	5	
Small amount of work	○	○	○	○	○	A lot of work

Appendix D

This appendix contains the (categorized) answers to the questionnaire used in this research. All questions are translated. For the multiple-choice questions, the answers are translated as well. For the open questions, the answers are categorized, and the categorization is translated.

Questions 1, 2, 3 and 4:

Figure 42: Bar Charts of Q1-4 of questionnaire MKS



Questions 5, 6, 7 and 13:

Figure 43: Bar charts of Q5-7 & 13 of questionnaire MKS

Questions 8 and 9:

Table 32: Categorized answers to Q8 of questionnaire MKS

Question 8:	
Which situations / incidents that occur regularly cause a high workload?	
Elements that appear in multiple answers:	**Frequency:**
Multiple (large) incidents happening at the same time	10
The addition of AKI-dossiers	2
A lack of capacity	2
People are working on portfolio tasks during a regular shift	5
Collisions	2
Incidents with suicidal people	2
A lack of relief shifts	2

Table 33: Categorized answers to Q9 of questionnaire MKS

Question 9:	
What could be a solution to lower the perceived workload in those situations?	
Elements that appear in multiple answers:	**Frequency:**
More time for portfolio tasks	3
Execute portfolio tasks only during night- and relief shifts	2
Improve SpoorWeb (multiple solutions were indicated)	2
Add capacity	6
Change the intake process (e.g., have a colleague listening in on conversations)	2

Questions 10, 11 and 12

Table 34: Categorized answers to Q10 of questionnaire MKS

Question 10:			
How long does it take you to do the intake of an incident?			
Answers:	**LB**	**UB**	**Average**
1	0	1	0.5
2	1	2	1.5
3	1	2	1.5
4	1	3	2
5	2	3	2.5
6	1	4	2.5
7	1	5	3
8	1	5	3
9	3	3	3
10	1	5	3
11	3	4	3.5
12	3	4	3.5
13	4	4	4
14	5	5	5
		Average:	2.79

Table 35: Categorized answers to Q11 of questionnaire MKS

Question 11:			
How long does it take you to process an RVO?			
Answers:	**LB**	**UB**	**Average**
1	1	1	1
2	1.5	1.5	1.5
3	1	2	1.5
4	1	2	1.5
5	1	2	1.5
6	0.33	3	1.67
7	2	2	2
8	2	2	2
9	2	3	2.5
10	3	3	3
11	2	4	3
12	3	3	3
13	2	5	3.5
14	3	4	3.5
15	4	4	4
		Average	2.93

Table 36: Categorized answers to Q12 of questionnaire MKS

Question 12:			
How long does it take you to fill in a form related to the calls from PuVo, DVP and Storing Publiek?			
Answers:	**LB**	**UB**	**Average**
1	1	1	1
2	1	1	1
3	1	1	1
4	1	1	1
5	1	2	1.5
6	1	3	2
7	2	2	2
8	2	2	2
9	2	3	2.5
10	3	3	3
11	3	3	3
12	3	3	3
13	3	3.5	3.25
14	3	5	4
15	5	5	5
16	5	5	5
17	5	5	5
18	5	10	7.5

Questions 14 up to and including 32

Table 37: Average workload per incident-label from questionnaire MKS

Incident-label	Average:
Collision with a person	4.44
Hindrance caused by people on or close to the rails	3.89
Hindrance caused by an object, vehicle, or animals on or close to the rails	3.72
Collision with an (infrastructure) object	3.61
Hindrance caused by emergency services	3.56
Hindrance caused by (a) traveller(s) or personnel due to their health	3.06
Hindrance caused by behaviour of (a) traveller(s) or personnel	3.00
Crossing malfunction	2.94
Almost a collision	2.89
Defect material	2.72
State of the rails	2.72
Section malfunction	2.67
Other infrastructure-related problems	2.61
Signal failure	2.61
(potential) IT failure	2.56
Switch malfunction	2.39
Infrastructure problem related to the environment/surroundings	2.28
Hindrance caused by a logistical problem or error	2.22
Slippery tracks	1.11

Appendix E

This appendix contains the start times that are allowed under various scenarios.

Scenario 1: No shifts allowed to start during the end of the night.

Scenario 2: Three sets of three start times that start around the regular start times and four possible start times for the day shift

Table 38: Allowed start times scenario 1

Scenario: **Hour:**	**Scenario 1:** **Shift allowed to start?**	**Scenario 2:** **Shift allowed to start?**
00:00-01:00	Yes	Yes
01:00-02:00	No	No
02:00-03:00	No	No
03:00-04:00	No	No
04:00-05:00	No	No
05:00-06:00	No	No
06:00-07:00	Yes	Yes
07:00-08:00	Yes	Yes
08:00-09:00	Yes	Yes
09:00-10:00	Yes	Yes
10:00-11:00	Yes	Yes
11:00-12:00	Yes	Yes
12:00-13:00	Yes	Yes
13:00-14:00	Yes	No
14:00-15:00	Yes	Yes
15:00-16:00	Yes	Yes
16:00-17:00	Yes	Yes
17:00-18:00	No	No
18:00-19:00	No	No
19:00-20:00	No	No
20:00-21:00	No	No
21:00-22:00	No	No
22:00-23:00	Yes	Yes
23:00-24:00	Yes	Yes

Appendix F

Table 39: Schedules generated in all 24 experiments

Experiment Start Time	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
1	0	2	1	0	0	0	0	0	0	0	2	0	0	1	1	2	2	0	0	1	1	0	2	0
2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	2	1	1	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0
8	1	1	2	3	3	3	1	2	2	3	0	3	0	1	1	0	0	3	0	0	0	2	0	3
9	1	2	1	0	0	0	0	0	0	0	3	0	0	1	1	2	3	0	0	1	1	0	3	0
10	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	1	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	1	0	0	0	0	2	2	2	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0
16	2	1	2	3	3	3	1	1	1	3	0	3	0	1	1	0	0	2	0	0	0	2	0	3
17	0	1	1	0	0	0	0	0	0	0	3	0	0	1	1	2	2	0	0	1	1	0	2	0
18	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	1	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0
24	1	1	1	2	2	2	1	1	1	2	0	2	0	1	1	0	0	2	0	0	0	2	0	2
25	0	1	1	0	0	0	0	0	0	0	2	0	0	1	1	2	2	0	0	1	1	0	2	0
26	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0
31	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0
32	1	1	1	3	3	3	1	2	2	2	0	3	0	1	1	0	0	3	0	0	0	3	0	3
33	1	2	2	0	0	0	0	0	0	0	3	0	1	1	1	3	3	0	0	1	1	0	3	0
34	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
35	1	0	0	1	0	0	1	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0
36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0
38	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0
39	1	1	0	0	0	0	1	1	2	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0
40	1	1	2	3	3	3	1	2	2	3	0	4	1	2	2	0	0	3	0	0	0	3	0	3
41	1	2	2	0	0	0	1	0	0	0	4	0	0	1	1	3	3	0	0	1	1	0	3	0
42	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
43	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0
44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
45	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

46	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
47	1	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0
48	0	1	1	2	2	2	0	1	1	2	0	2	1	1	1	0	0	2	0	0	0	2	0	2
49	1	1	1	0	0	0	1	0	0	0	2	0	0	1	1	2	2	0	0	1	1	0	2	0
50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
51	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
52	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
53	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
54	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
55	1	0	0	0	0	0	1	1	1	0	0	0	1	0	0	0	0	0	1	1	1	0	0	0
56	0	1	1	3	3	3	1	2	2	3	0	4	0	1	1	0	0	3	0	0	0	2	0	2
57	1	2	2	0	0	0	1	0	0	0	3	0	0	1	1	2	3	0	0	2	1	0	3	0
58	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
59	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1	0	0	1	0	0
60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
61	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
62	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
63	1	0	0	0	0	0	2	2	2	0	0	0	1	0	0	0	0	0	1	1	2	0	0	0
64	0	1	1	3	3	3	0	1	1	3	0	3	0	1	1	0	0	3	0	0	0	2	0	3
65	1	2	2	0	0	0	1	0	0	0	3	0	0	1	1	2	3	0	0	2	1	0	3	0
66	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
67	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
68	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
69	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
70	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
71	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0
72	1	1	1	2	2	2	1	1	1	2	0	2	0	1	1	0	0	2	0	0	0	2	0	2
73	0	1	1	0	0	0	0	0	0	0	2	0	0	1	1	2	2	0	0	1	1	0	2	0
74	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
75	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
76	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
77	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
78	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
79	0	0	0	0	0	0	1	1	1	0	0	0	1	0	0	0	0	0	1	1	1	0	0	0
80	1	1	1	3	4	4	1	2	2	3	0	3	0	1	1	0	0	3	0	0	0	2	0	4
81	2	3	3	0	0	0	0	0	0	0	3	0	0	1	1	3	3	0	0	2	2	0	3	0
82	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0
83	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0
84	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
85	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
86	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
87	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0
88	1	1	1	3	4	4	1	2	2	3	0	3	1	1	1	0	0	3	0	0	0	2	0	3
89	2	3	3	0	0	0	0	0	0	0	3	0	0	1	1	3	3	0	0	2	2	0	3	0
90	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
91	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
92	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0
93	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
94	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
95	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

96	0	0	1	2	2	2	0	1	1	2	0	2	1	1	1	0	0	2	0	0	0	2	0	2
97	1	2	1	0	0	0	0	0	0	0	2	0	0	1	1	2	2	0	0	2	2	0	2	0
98	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
99	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
101	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
102	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0
103	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
104	0	0	1	3	3	3	1	2	2	4	0	4	1	1	1	0	0	3	0	0	0	3	0	3
105	1	3	2	0	0	0	0	0	0	0	4	0	0	1	1	2	3	0	0	3	3	0	3	0
106	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0
107	0	0	0	1	0	0	1	0	0	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0
108	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
109	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
110	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0
111	0	0	0	0	0	0	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
112	1	1	2	3	3	3	2	2	2	4	0	4	1	1	1	0	0	3	0	0	0	3	0	3
113	0	2	1	0	0	0	0	0	0	0	4	0	0	2	2	3	3	0	0	3	3	0	2	0
114	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
115	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0
116	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
117	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
118	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
119	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
120	0	0	1	2	2	2	1	1	1	2	0	2	0	1	1	0	0	2	0	0	0	2	0	2
121	1	2	1	0	0	0	0	0	0	0	2	0	0	1	1	2	2	0	0	2	2	0	2	0
122	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
123	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
124	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
125	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
126	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
127	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0
128	0	0	1	2	3	3	1	1	1	2	0	3	0	2	1	0	0	2	0	1	0	2	0	2
129	1	3	2	0	0	0	0	0	0	0	3	0	0	1	2	2	2	0	0	1	1	0	2	0
130	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
131	0	0	0	1	0	0	0	1	0	1	0	0	1	0	0	1	0	0	0	0	0	1	0	0
132	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
133	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0
134	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
135	0	0	0	0	0	0	1	1	2	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0
136	0	0	1	2	3	3	0	0	1	2	0	2	0	2	2	0	0	2	0	0	0	2	0	2
137	1	3	2	0	0	0	0	1	0	0	2	0	0	1	1	2	2	0	0	1	1	0	2	0
138	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
139	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
140	0	0	0	0	0	0	1	0	0	0	0	0	2	0	0	0	0	0	1	0	0	0	0	0
141	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
142	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
143	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0
144	0	0	1	2	2	2	0	1	1	2	0	2	0	1	1	0	0	2	0	0	0	2	0	2
145	0	2	1	0	0	0	0	0	0	0	2	0	0	1	1	2	2	0	0	1	1	0	2	0

146	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
147	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
148	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
149	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
150	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
151	0	0	0	0	0	0	2	1	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0
152	1	0	1	3	2	2	0	1	1	2	0	2	1	1	1	0	0	2	1	1	0	2	0	2
153	0	2	1	0	0	0	0	0	0	0	3	0	0	1	1	2	2	0	0	2	1	0	3	0
154	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
155	0	0	0	0	0	0	1	1	0	2	0	0	1	0	0	0	0	0	0	0	0	1	0	0
156	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
157	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
158	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
159	0	0	0	0	0	0	2	1	1	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0
160	1	0	1	3	3	3	0	0	2	0	0	3	1	1	1	0	0	2	0	0	0	2	0	2
161	0	2	1	0	0	0	1	2	0	3	3	0	0	1	1	2	2	0	0	1	1	0	2	0
162	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
163	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
164	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
165	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
166	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
167	0	0	0	0	0	0	2	1	1	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0
168	1	0	1	2	2	2	0	1	1	2	0	2	1	1	1	0	0	2	0	0	0	2	0	2

Appendix G

Table 40: Schedules generated in sensitivity analysis experiments

Experiment Start Time	0	1	2	3	4	5	6	7	8	9	10	11	12
1	0	2	0	0	2	0	0	0	2	2	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0
8	3	0	2	2	0	3	2	2	0	0	3	3	3
9	0	3	0	0	3	0	0	0	3	3	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0
16	2	0	3	3	0	3	3	2	0	0	2	2	3
17	0	3	0	0	2	0	0	0	2	2	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0
24	2	0	2	2	0	2	2	2	0	0	2	2	2
25	0	2	0	0	2	0	0	0	2	2	0	0	0
26	0	0	0	0	0	0	0	0	0	0	0	0	0
27	0	0	0	0	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0	0	0	0	0
29	0	0	0	0	0	0	0	0	0	0	0	0	0
30	0	0	0	0	0	0	0	0	0	0	0	0	0
31	0	0	0	0	0	0	0	0	0	0	0	0	0
32	2	0	3	3	0	3	3	2	0	0	2	2	2
33	0	3	0	0	2	0	0	0	3	3	0	0	0
34	0	0	0	0	0	0	0	0	0	0	0	0	0
35	0	0	0	0	0	0	0	0	0	0	0	0	0
36	0	0	0	0	0	0	0	0	0	0	0	0	0
37	0	0	0	0	0	0	0	0	0	0	0	0	0
38	0	0	0	0	0	0	0	0	0	0	0	0	0
39	0	0	0	0	0	0	0	0	0	0	0	0	0
40	3	0	3	2	0	3	2	2	0	0	3	3	3
41	0	3	0	0	3	0	0	0	3	3	0	0	0
42	0	0	0	0	0	0	0	0	0	0	0	0	0
43	0	0	0	0	0	0	0	0	0	0	0	0	0

44	0	0	0	0	0	0	0	0	0	0	0	0	0
45	0	0	0	0	0	0	0	0	0	0	0	0	0
46	0	0	0	0	0	0	0	0	0	0	0	0	0
47	0	0	0	0	0	0	0	0	0	0	0	0	0
48	2	0	2	2	0	2	2	2	0	0	2	2	2
49	0	2	0	0	2	0	0	0	2	2	0	0	0
50	0	0	0	0	0	0	0	0	0	0	0	0	0
51	0	0	0	0	0	0	0	0	0	0	0	0	0
52	0	0	0	0	0	0	0	0	0	0	0	0	0
53	0	0	0	0	0	0	0	0	0	0	0	0	0
54	0	0	0	0	0	0	0	0	0	0	0	0	0
55	0	0	0	0	0	0	0	0	0	0	0	0	0
56	3	0	3	2	0	3	2	2	0	0	3	3	3
57	0	3	0	0	2	0	0	0	3	3	0	0	0
58	0	0	0	0	0	0	0	0	0	0	0	0	0
59	0	0	0	0	0	0	0	0	0	0	0	0	0
60	0	0	0	0	0	0	0	0	0	0	0	0	0
61	0	0	0	0	0	0	0	0	0	0	0	0	0
62	0	0	0	0	0	0	0	0	0	0	0	0	0
63	0	0	0	0	0	0	0	0	0	0	0	0	0
64	3	0	3	3	0	3	2	2	0	0	3	3	2
65	0	3	0	0	2	0	0	0	2	2	0	0	0
66	0	0	0	0	0	0	0	0	0	0	0	0	0
67	0	0	0	0	0	0	0	0	0	0	0	0	0
68	0	0	0	0	0	0	0	0	0	0	0	0	0
69	0	0	0	0	0	0	0	0	0	0	0	0	0
70	0	0	0	0	0	0	0	0	0	0	0	0	0
71	0	0	0	0	0	0	0	0	0	0	0	0	0
72	2	0	2	2	0	2	2	2	0	0	2	2	2
73	0	2	0	0	2	0	0	0	2	2	0	0	0
74	0	0	0	0	0	0	0	0	0	0	0	0	0
75	0	0	0	0	0	0	0	0	0	0	0	0	0
76	0	0	0	0	0	0	0	0	0	0	0	0	0
77	0	0	0	0	0	0	0	0	0	0	0	0	0
78	0	0	0	0	0	0	0	0	0	0	0	0	0
79	0	0	0	0	0	0	0	0	0	0	0	0	0
80	3	0	3	3	0	3	2	2	0	0	3	3	3
81	0	3	0	0	3	0	0	0	3	3	0	0	0
82	0	0	0	0	0	0	0	0	0	0	0	0	0
83	0	0	0	0	0	0	0	0	0	0	0	0	0
84	0	0	0	0	0	0	0	0	0	0	0	0	0
85	0	0	0	0	0	0	0	0	0	0	0	0	0
86	0	0	0	0	0	0	0	0	0	0	0	0	0
87	0	0	0	0	0	0	0	0	0	0	0	0	0
88	3	0	3	3	0	3	3	2	0	0	3	3	2
89	0	3	0	0	2	0	0	0	2	2	0	0	0
90	0	0	0	0	0	0	0	0	0	0	0	0	0
91	0	0	0	0	0	0	0	0	0	0	0	0	0
92	0	0	0	0	0	0	0	0	0	0	0	0	0
93	0	0	0	0	0	0	0	0	0	0	0	0	0

94	0	0	0	0	0	0	0	0	0	0	0	0	0
95	0	0	0	0	0	0	0	0	0	0	0	0	0
96	2	0	2	2	0	2	2	2	0	0	2	2	2
97	0	2	0	0	1	0	0	0	2	2	0	0	0
98	0	0	0	0	0	0	0	0	0	0	0	0	0
99	0	0	0	0	0	0	0	0	0	0	0	0	0
100	0	0	0	0	0	0	0	0	0	0	0	0	0
101	0	0	0	0	0	0	0	0	0	0	0	0	0
102	0	0	0	0	0	0	0	0	0	0	0	0	0
103	0	0	0	0	0	0	0	0	0	0	0	0	0
104	3	0	3	3	0	3	3	3	0	0	3	3	3
105	0	3	0	0	3	0	0	0	3	3	0	0	0
106	0	0	0	0	0	0	0	0	0	0	0	0	0
107	0	0	0	0	0	0	0	0	0	0	0	0	0
108	0	0	0	0	0	0	0	0	0	0	0	0	0
109	0	0	0	0	0	0	0	0	0	0	0	0	0
110	0	0	0	0	0	0	0	0	0	0	0	0	0
111	0	0	0	0	0	0	0	0	0	0	0	0	0
112	3	0	3	3	0	3	3	2	0	0	3	3	3
113	0	3	0	0	3	0	0	0	3	3	0	0	0
114	0	0	0	0	0	0	0	0	0	0	0	0	0
115	0	0	0	0	0	0	0	0	0	0	0	0	0
116	0	0	0	0	0	0	0	0	0	0	0	0	0
117	0	0	0	0	0	0	0	0	0	0	0	0	0
118	0	0	0	0	0	0	0	0	0	0	0	0	0
119	0	0	0	0	0	0	0	0	0	0	0	0	0
120	2	0	2	2	0	2	2	2	0	0	2	2	2
121	0	2	0	0	2	0	0	0	2	2	0	0	0
122	0	0	0	0	0	0	0	0	0	0	0	0	0
123	0	0	0	0	0	0	0	0	0	0	0	0	0
124	0	0	0	0	0	0	0	0	0	0	0	0	0
125	0	0	0	0	0	0	0	0	0	0	0	0	0
126	0	0	0	0	0	0	0	0	0	0	0	0	0
127	0	0	0	0	0	0	0	0	0	0	0	0	0
128	2	0	2	2	0	3	2	2	0	0	2	2	3
129	0	3	0	0	2	0	0	0	2	2	0	0	0
130	0	0	0	0	0	0	0	0	0	0	0	0	0
131	0	0	0	0	0	0	0	0	0	0	0	0	0
132	0	0	0	0	0	0	0	0	0	0	0	0	0
133	0	0	0	0	0	0	0	0	0	0	0	0	0
134	0	0	0	0	0	0	0	0	0	0	0	0	0
135	0	0	0	0	0	0	0	0	0	0	0	0	0
136	2	0	3	2	0	3	2	2	0	0	2	2	2
137	0	2	0	0	2	0	0	0	2	2	0	0	0
138	0	0	0	0	0	0	0	0	0	0	0	0	0
139	0	0	0	0	0	0	0	0	0	0	0	0	0
140	0	0	0	0	0	0	0	0	0	0	0	0	0
141	0	0	0	0	0	0	0	0	0	0	0	0	0
142	0	0	0	0	0	0	0	0	0	0	0	0	0
143	0	0	0	0	0	0	0	0	0	0	0	0	0

144	2	0	2	2	0	2	2	2	0	0	2	2	2
145	0	2	0	0	1	0	0	0	2	2	0	0	0
146	0	0	0	0	0	0	0	0	0	0	0	0	0
147	0	0	0	0	0	0	0	0	0	0	0	0	0
148	0	0	0	0	0	0	0	0	0	0	0	0	0
149	0	0	0	0	0	0	0	0	0	0	0	0	0
150	0	0	0	0	0	0	0	0	0	0	0	0	0
151	0	0	0	0	0	0	0	0	0	0	0	0	0
152	2	0	3	2	0	3	2	2	0	0	2	2	2
153	0	3	0	0	2	0	0	0	2	2	0	0	0
154	0	0	0	0	0	0	0	0	0	0	0	0	0
155	0	0	0	0	0	0	0	0	0	0	0	0	0
156	0	0	0	0	0	0	0	0	0	0	0	0	0
157	0	0	0	0	0	0	0	0	0	0	0	0	0
158	0	0	0	0	0	0	0	0	0	0	0	0	0
159	0	0	0	0	0	0	0	0	0	0	0	0	0
160	2	0	2	2	0	2	2	2	0	0	2	2	2
161	0	2	0	0	2	0	0	0	2	2	0	0	0
162	0	0	0	0	0	0	0	0	0	0	0	0	0
163	0	0	0	0	0	0	0	0	0	0	0	0	0
164	0	0	0	0	0	0	0	0	0	0	0	0	0
165	0	0	0	0	0	0	0	0	0	0	0	0	0
166	0	0	0	0	0	0	0	0	0	0	0	0	0
167	0	0	0	0	0	0	0	0	0	0	0	0	0
168	2	0	2	2	0	2	2	2	0	0	2	2	2