A flexible scheduling framework for a radiology department

A multi-stage stochastic program on a rolling horizon integrating employee and resource scheduling based on demand

Graduation committee

Prof. Dr. R.J. Boucherie (UT)
Dr. Ing. A.B. Zander (UT)
Dr. M. Walter (UT)

Supervisors

Prof. Dr. R.J. Boucherie (UT)
J. Creemers (VMC)
Dr. I. Meuffels (Rhythm b.v.)
O. The (Rhythm b.v.)

Author

A.G.M. Vlooswijk

Abstract

Radiology departments in the Netherlands currently schedule their patient capacity and session schedule before the radiological technologist (RT) schedule is made. As per the collective agreement of labour of 2023, the RT schedule needs to be published three months in advance. This means that the session schedule must also be made three months in advance, based on limited information regarding the variable patient demand. This study develops a multi-stage stochastic model that describes the complete scheduling framework of a radiology department. The framework is adapted to allow for up to 60% of the session schedule to be determined one month in advance, enhancing scheduling flexibility based on patient demand and RT availability while considering fairness of the schedule. The model was approximated using stage aggregation on a rolling horizon, with independent operational offline decisions.

The performance of the month-based flexible model was compared to a myopic model, which fixates the schedule three months in advance. The results of a test case showed that such a myopic model outperforms the flexible models, unless the demand deviates from the mean for prolonged periods. The flexible framework was seen to introduce more fluctuations in the work backlog. Moreover, a case study regarding VieCuri Medical Centre showed that in a real-life example, a myopic method performs comparably to the two-stage stochastic method based on 50 scenarios. The heuristic using a regular mixed integer linear program and the flexible scheduling framework showed similar performance to the stochastic two-stage method.

Furthermore, VMC's current management maintains a more stable work backlog for most modalities than the proposed flexible model. However, our results show a proof of concept for a flexible method that can make operational adjustments earlier and offers the opportunity to adjust the session schedule while maintaining a decent work backlog and an equally or more stable RT schedule.

List of abbreviations

VMC VieCuri Medical Centre
CT computed tomography
MRI magnetic resonance imaging
Dexa dual X-ray absorptiometry
US ultrasound
FL fluoroscopy lab
FTE full time equivalent
GP general practitioner
TPM tactical planning meeting
OPM operational planning meeting
RH rolling horizon
RT radiological technologist
AHP analytic hierachy process
MILP mixed integer linear programs
SP sub problem
MP master problem
RMP relaxed master problem
KPI key performance indicators
SAA sample average approximation
AG stage aggregation
PI per-stage independent stage aggregation
CI confidence interval
TS two-stage stochastic program

Contents

Chapter 1

Introduction

In 2022, a shortage of 49,000 medical staff was reported in the Netherlands[1, 2]. Research conducted on behalf of the Dutch Ministry of Health, Welfare, and Sport projects that this shortage of medical staff will increase to 135.000 in 2031 [1]. Consequently, the optimization of workforce management in healthcare is more relevant than ever. At the heart of every hospital lies its radiology department, a supportive department specialised in different imaging techniques, modalities, vital for diagnostic and therapeutic purposes. Operating as a supportive department, it must exhibit flexibility to accommodate the patient demand generated by other departments and general practitioners (GPs). However, patient demand is variable and depends on a multitude of factors, which makes it hard to predict. Despite this unpredictability, a radiology department needs to effectively respond to the demand, otherwise either emergency care cannot be provided or the time until a patient can make an appointment will increase significantly. A delicate balance must be found between meeting the patient's needs while keeping the workload of employees manageable, and the costs within budget.

To effectively organize a radiology department, a structured framework derived from capacity management principles is used [3]. This framework optimizes the use of resources: radiology technologists (RTs), facilities, and equipment, to sustain quality of care and overall efficiency. Decision-making within this framework is divided into four levels: strategic, tactical, operational offline and operational online. Strategic decisions are long-term decisions related to resource allocation and investment planning to meet long-term goals. Such decisions include hiring RTs and investing in imaging modalities. Tactical decisions relate to session scheduling, radiological technologist (RT) rostering, and scheduling other activities such as education and maintenance. A session schedule outlines when imaging modalities are operated and thus opened for patients. Patient booking and last-minute changes to the RT schedule fall under operational decisions. Operational offline decisions are taken days to weeks in advance and online on the day itself. The strategic, tactical, and operational decisions guide all aspects of the department's operations: RT scheduling, session scheduling, and patient booking.

RTs are scheduled according to the session schedule. The collective labor agreement of hospitals mandates that RT schedules are published three months in advance [4]. Consequently, the session schedule needs to be available before. The session schedule is made using a prediction of the patient demand. The patient demand is determined by the work backlog and incoming imaging requests or orders for each modality. The work backlog exists of all orders that are planned, and can be planned and is influenced by the incoming orders and the executed orders. It is important to understand that the number of incoming orders is heavily influenced by the outpatient appointments of the referring departments. These appointments generate imaging orders and treatments, such as surgeries, requiring pre-operative and post-operative images. The number of outpatient appointments varies per week. For example, it depends on the availability of the specialists and holiday periods. Moreover, the correlation of the number of outpatient appointments and orders differs per physician and changes over time as treatment protocols are updated. In conclusion, accurately predicting patient demand remains a complex challenge.

The session schedule relies heavily on the accuracy of the work backlog prediction. If the prediction is too low, it results in patients having to wait longer to make an appointment, ultimately lowering healthcare standards and patient satisfaction. Conversely, overestimating demand leads to underutilization of modalities and staff, incurring unnecessary expenses for the hospital. Therefore, an accurate prediction is crucial. As more information is collected about patient demand over time, predictions tend to be more precise at the operational decision level compared to the tactical level. Nevertheless, the current scheduling framework limits the possibility of

adjustments to the session schedule and RT roster at the operational level. Instead, a flexible framework would allow these adjustments at the operational level. Hypothetically, a more flexible framework would enable a radiology department to better respond to varying demands, ultimately improving workforce management and patient care. The research question is formulated as follows:

What are the effects of introducing flexibility into the scheduling framework of a radiology department?

To answer this question, we need to answer the following subquestions:

1. How can the different decision levels of a radiology department and involved random variables be mathematically modelled?

2. How can flexibility be introduced into the scheduling framework of a radiology department?

3. What characteristics of a radiology department influence the effects of additional flexibility?

The first question is addressed through a general analysis, wherein we developed a multi-stage stochastic model that describes the decision stages of a radiology department. Given the practical challenges often encountered in healthcare research projects, a collaboration with VieCuri Medical Centre (VMC) was established at the start of the research. VMC is a tertiary hospital in Venlo in the south of the Netherlands. Additionally, we partnered with Rhythm b.v., a consultancy firm specializing in healthcare optimization.

Subsequently, to explore answers to the second and third subquestions, we adapted the proposed model to accommodate flexibility. We introduce a novel flexible scheduling framework that integrates session scheduling with staff scheduling, such that the resources can be adapted to the patient demand and RT availability. We approximated the model using a two-stage stochastic framework, aggregating stages and a case specific heuristic. To evaluate the efficacy of our approach, we conducted numerical experiments employing the integer L-shaped algorithm. The results of a test case showed that a myopic model that fixes the schedule three months in advance outperforms the flexible model unless the demand deviates from the mean for prolonged periods. Moreover, a case study regarding VieCuri Medical Centre showed that also in a real-life example, a myopic method performs comparably to the two-stage stochastic method. Furthermore, the model did not outperform VMC's current management but did show that the model is capable of introducing flexibility while maintaining a decent work backlog and an equally or more stable RT schedule.

This thesis is organised as follows: Chapter 2 provides a background on the workings of a radiology department, and Chapter 3 discusses related literature. Subsequently, Chapter 4 goes into more detail about the used methods and techniques. In Chapter 5, a mathematical formulation of the scheduling framework of a radiology department is presented, and the random variables and their properties are investigated. Moreover, the model is extended to allow flexibility. In Chapter 6, the results of the numerical experiments are discussed concerning both the test case and a case study regarding VMC. Finally, chapter 7 is used to draw conclusions, discuss the results, and suggest future research directions.

Chapter 2

Background

This section explores all facets and processes of a radiology department: the patient journey within a radiology department, the patient demand, and the scheduling framework. This background is necessary to develop a mathematical formulation of the scheduling framework in the next Chapters. The contents of this Section are based on the radiology department of VMC. While specific details may differ between different hospitals, the foundational principles of radiology departments are generally comparable.

2.1 Patient journey

A patient's journey begins with a visit to a physician, who may refer them to the radiology department. Patients are referred to a radiology department through one of three channels: first-line, second-line, and the emergency room. First-line referrals are initiated by a general practitioner (GP), while second-line referrals are generated by specialised physicians. Second-line patients can either be inpatients, currently hospitalised, or outpatients coming from home or another care facility. The second line encompasses many different departments, such as orthopedics, and cardiology.

A referral, or order, specifies the modality and describes the type of image to be made, including a predicted duration and a preferred execution date. A radiology department has different imaging modalities, such as computed tomography (CT), magnetic resonance imaging (MRI), ultrasound (US), Angiography, and X-ray. Once the correct protocol is assigned to an order, patient bookers – who are part of the administrative staff of a radiology department – may schedule an appointment. The time the patient needs to wait until the appointment, if the order is not urgent, is the access time. The order is booked on an available time slot. At the appointment time, the patient is called by the RT operating that modality during that time, according to the RT roster created by the RT-schedulers. Subsequently, the image is made by the RT, which may involve a radiologist depending on the type of scan. A radiologist assesses and reports the images afterward within a set time. Finally, the referring physician discusses the report with the patient.

2.2 Patient demand

The patient demand is determined by the current work backlog and the incoming orders per week. In more detail, the demand is the orders that should be executed that week according to the urgency level. The urgency of an order is determined by its desired execution date. Orders are classified into five urgency levels: emergency, urgent, expedited, elective, and long-term. Emergency orders need to be executed immediately. Urgent orders should be completed within the same week. Expedited orders aim for execution within two weeks. Elective orders are scheduled based on availability, and long-term orders are placed on a waitlist. The hospital usually aims for elective orders to be executed within a certain time frame: the target access time. This determines the height of the desirable work backlog. If the current work backlog is too high, then more elective orders should be scheduled to decrease the access time, if the work backlog is low, fewer elective orders should be scheduled to prevent empty sessions. Each week the work backlog is updated by adding the new orders of that week and subtracting all executed orders that week.

The number of incoming orders and the work backlog can be estimated to determine how much capacity is

needed for each modality. The orders originating from the first line and emergency department are generally quite stable throughout the year. However, the incoming orders from the second line are more variable. To illustrate: the incoming orders from the second line are correlated to which physicians are doing outpatient appointments and how many of these appointments the physicians do. In a holiday period, naturally, fewer physicians are working. Consequently, there are fewer outpatient appointments and thus fewer orders. Moreover, fewer surgeries will be performed, requiring fewer post-operative images and check-up images, further decreasing the demand in the weeks after. Therefore, the number of scheduled outpatient appointment sessions can be used to predict the number of incoming orders. More information regarding the outpatient appointment sessions: how many are scheduled, and how many orders they generate becomes known over time. The work backlog may be predicted by predicting the outflow using the planned sessions multiplied by the production (orders per session) and predicting the inflow of orders as described earlier.

2.3 Scheduling framework

Within the scheduling framework of a radiology department, three intertwined scheduling processes are distinguished: modality, employee (RT), and patient scheduling. To maintain a clear distinction between these three scheduling processes we refer to modality scheduling as session scheduling, RT scheduling as rostering, and patient scheduling as booking [3]. Per illustration, the scheduling timeline of VMC is summarised in Figure 2.1. In the next paragraphs, each scheduling process is explained in further detail.
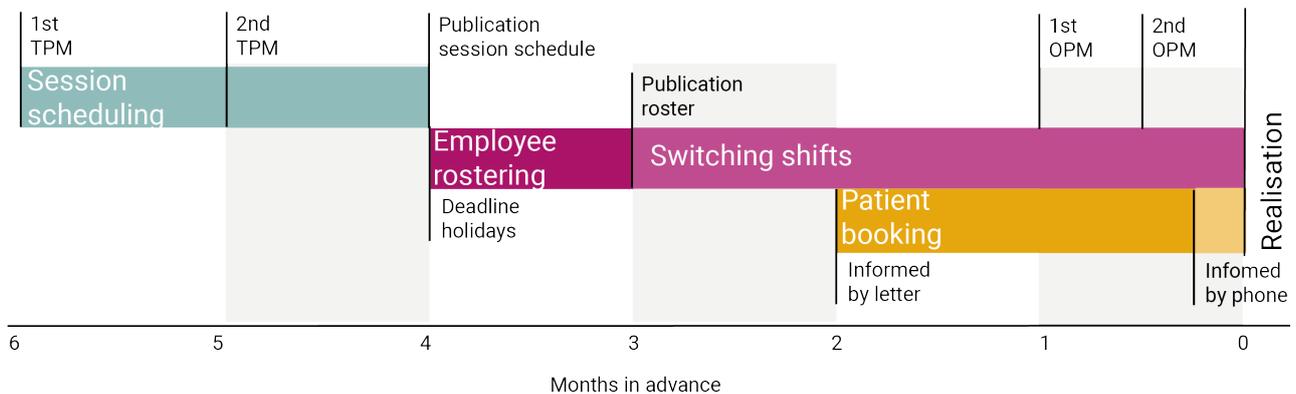

Figure 2.1: This figure visualizes the scheduling timeline of the radiology department of VMC.

2.3.1 Session scheduling

A session is a time period during which a modality is operated and thus available for patients. A session typically lasts half a day. The sessions for one period, typically a month, of each modality are determined several months in advance in a session schedule. A session schedule forms the basis for a time slot schedule, which is made after the session schedule has been determined. The time slot schedule further specifies which patients may be booked on a session. Adjustments to the modality and/or time slot plan can be made during two types of staff meetings: a tactical planning meeting (TPM) and an operational planning meeting (OPM).

During TPMs the session schedule can be adjusted based on the availability of RTs, the number of available apparatuses per modality, and the work backlog prediction. During OPMs the sessions per modality in the coming weeks are discussed. A session can be closed due to a shortage of personnel or due to low demand. In the latter, personnel is given the choice of taking a holiday or assigned to administrative activities. If there were any patients booked on the closed sessions, they need to be rebooked.

2.3.2 Employee rostering

The RT-schedulers roster the RTs and the interns within the time slot schedule of the modalities using the radiologist's roster. The radiologists are scheduled separately by one of the radiologists before the RT-schedulers start the rostering process. In some cases, the scheduled radiologist further specifies what scans can be done. The modality determines the needed qualifications and the number of RTs. Interns are always scheduled alongside RTs.

7

Every RT can make X-rays, but all other modalities require additional qualifications. Most RTs are qualified to operate one or two extra modalities. The modalities CT, MRI, US, and angiography require postgraduate professional education, while other qualifications can be obtained through internal courses. Additionally, in the case of MRI, CT, and US there are sub-qualifications, within the general qualification. For example, an RT might be qualified to make a CT of the abdomen, but not a CT of the heart. An RT needs to uphold their (sub)qualification by operating the modality a minimal amount of sessions each week or month. Consequently, the RT pool of less frequently needed qualifications should remain small. The diverse skill sets within the workforce influence the adaptability of the scheduling framework. Specifically, presumably a more diverse composition makes it easier to exchange capacity between modalities.

Besides modality work, several administrative activities need to be executed: adding a protocol to orders, helping with and updating the applications on devices, ordering supplies, and offering guidance to interns. Lastly, RTs are scheduled on indirect work activities, such as education and learning how to operate a new machine. The amount of indirect work is highly variable per month.

There exist three types of RT shifts: normal weekday shifts, on-call shifts (where the RT should be available), and irregular shifts in the evening and weekend. The on-call and irregular shifts are divided equally among the RTs first. Then the day shifts are rostered, taking into account the number of working hours per week, indirect work and the scheduled irregular shifts. This RT roster needs to be published at least three months in advance due to the collective labor agreement of hospitals [4]. The agreement only concerns the shifts and not activities. However, in the current scheduling framework, both are published simultaneously. The RTs are assigned to a specific activity, which specifies the specific starting time and the location of the modality: a modality can have multiple rooms and/or locations depending on the modality and hospital. Subsequently, the roster is checked for mistakes and fairness. The fairness of a roster is determined by factors that make a roster nicer for RTs, such as not working the same modality multiple days in a row, or being assigned more irregular shifts than others.

The finished roster, including activities, is published three months in advance. The RTs can request to switch shifts or activities with other RTs up to the day of the shift. The RT-schedulers check the rosters to see if these requests can be granted. The RT availability is determined by requested holidays, roster-free days, education, and illnesses. A roster-free day is a fixed weekday an RT, who works less than full-time, is preferably not scheduled. An RT may be asked to work on such day, but only if necessary. The hospital's policy typically requires scheduling extra staff to cover absences resulting from illness. This involves assigning additional RTs to a 'flex' position daily. Any additional unexpected staff shortages may be addressed operationally by reassigning RTs from administrative duties to a modality.

2.3.3 Patient booking

Patients can be booked on available time slots within a session. The number of time slots per session differs per modality and is determined by a standard appointment time. The appointments are made by the patient-bookers. The booking process varies per order, based on the modality of the order and its urgency level. A portion of time slots for each modality are set aside to accommodate emergency patients. Urgent, expedited, and elective patients are scheduled according to their preferred execution date. Long-term orders are placed on a waitlist and are incorporated into the work backlog once the preferred execution date falls within the booking horizon. The bookers can book appointments in published sessions only. Therefore, the booking horizon is usually approximately three months. When an appointment is made at least a week in advance the patient is notified by letter, if the appointment is made less than a week in advance the patient is contacted by telephone. A patient can rebook their appointment via telephone.

2.4 Conclusions and research direction

The described scheduling processes have distinct time horizons and happen in a fixed order: first the session schedule, second the RT roster and third, after publishing the RT roster and session schedule the patients can be booked. It is costly to cancel a session after it has been published because RTs cannot always be usefully repurposed and patients need to be rebooked. This limits the flexibility of the current scheduling framework.

The scheduling framework could be changed such that three months in advance the RTs are only assigned a shift and not a specific activity. This is allowed because the collective employment agreement mandates only the shifts to be published three months in advance [4]. This allows publishing a partly filled session schedule

three months in advance. Open scheduling blocks can be allocated later based on the current work backlog, enabling flexible modality assignment. Since RTs are not assigned to a specific activity, only shifts, they can be switched between modalities to facilitate different flexible sessions. The current patient booking horizon already offers flexibility to accommodate varying lead times. Therefore, the emphasis in this report will be on session scheduling and RT rostering. Moreover, this report focuses on the session schedule and leaves the time slot schedule out of scope, because the time slot schedule can be fit to the session schedule.

The performance of such a flexible scheduling framework can be measured by various key performance indicators (KPI). These include metrics related to the work backlog, session scheduling per modality, the changes in session schedules over time, the scheduled full time equivalent (FTE), and the fairness of the RT-rosters. The ultimate goal is to optimize scheduling to minimize patient access time while efficiently utilizing resources.

The problem at hand involves modeling a sequential decision-making process and introducing flexible scheduling strategies. Additionally, it is important to accurately model the stochastic demand and supply properties of the problem. These models are often high dimensional and complex, therefore efficient solution techniques should be researched as well.

Chapter 3

Literature review

In this section, we present relevant literature related to our research goals. We start by providing an overview of the literature about models used in healthcare scheduling. We then explore sequential decision-making under uncertainty in a more general context, report on related work and discuss various approximation methods. Finally, we present our contribution.

3.1 Operations research in healthcare

The scheduling framework of a hospital department consists of four decision levels: strategic, tactical, operational offline, and operational online. For an in-depth discussion of each of these decision levels in healthcare, we refer the reader to [3]. The problems related to each level have been extensively studied[5]. This report aims to capture the framework and thus these levels in a mathematical formulation, therefore a short overview of the mathematical optimization problem concerning each level is discussed first.

3.1.1 Optimization problems per decision level

The strategic decision relates to resource allocation, in the sense that the hospital should determine how many medical staff members and facilities are to be assigned to each department. The tactical decision problem involves resource allocation: session scheduling, time slot scheduling, and workforce scheduling. Workforce scheduling aims to schedule employees efficiently[6]. When the scheduling cycles or shift times are fixed in advance, we speak of shift scheduling. Scheduling of specifically RTs of a radiology department has not received much attention in literature. However, shift scheduling regarding physicians and nurses has been widely studied[7, 8, 5]. Often the scheduling problem is formulated as a mixed integer linear programs (MILP) that minimizes the cost of assigning staff to all shifts to meet the demand or man the activity schedule: covering constraints. The constraints of the MILP are determined by the unique characteristics of the hospital department. Often the workforce is heterogeneous, meaning that employees have different individual skill sets, contracts, and/or availability[9]. Moreover, shift schedules usually have to adhere to employer rules, such as government regulations.

A shift schedule does not necessarily assign a specific activity to the staff. This problem is captured in staff allocation problems. Staff allocation problems need not only to adhere to covering constraints, but should also take into account the fairness of the schedule, preferences of employees, and unwanted patterns[10]. Such aspects can be modelled as soft constraints, which may be violated, albeit at a cost. The relative importance of different soft constraints can be quantified with the analytic hierachy process (AHP)[11]. AHP quantifies the weights of soft constraints by leveraging the insights of individual experts; through pairwise comparisons, experts' experiences are harnessed to gauge the relative significance of various factors. Only when these softer constraints are modelled as well, then the model is suitable to be implemented in practice. Due to the added complexity and high variability among institutions of such constraints, these models are not very prevalent, resulting in many departments scheduling manually to this day[10].

The operational offline and online decisions relate to timetabling, and appointment scheduling[6, 12]. Timetabling refers to booking patient appointments. Appointment scheduling aims to use resources efficiently and to avoid queuing. Dynamic programming and Markov decision processes can be used to minimize waiting times[13]. Al-

ternatively, a MILP can be utilised to minimize the total time of a patient in the radiology department[14, 15].

3.1.2 Integrating multiple decision levels

The separate scheduling problems are well-researched. Moreover, some combinations of different components or levels have received plenty of attention. Van Lent et al. [16] used a Monte Carlo simulation to reduce the access time between outpatient appointments and CT-scan in a radiology department by changing the allocation of CT capacity to different patient groups. Vieira et al. [17] consider both workforce scheduling and timetabling to maximize the number of patients that are helped within a designated time in a radiotherapy department. They developed three MILPs that considered different components of the scheduling problem and were subjected to scenario s generated from historical data, which were solved in sequence based on hierarchy[17]. Chen et al. [11] integrated the medical staff allocation and the workforce scheduling of a radiology centre. They developed two sequentially solved Integer Linear Programs (ILPs), the first uses a worst-case scenario to determine the minimum required staff each month and the second generates the optimal staff schedule given the optimal solution of the first ILP as input[11].

However, Chen et al. [11] and Vieira et al. [17] do not integrate the different decision levels; they only solve them sequentially. Bentayeb et al. [9] do integrate two scheduling processes, namely they studied the combination of patient appointment scheduling and technologist scheduling, which allows flexible technologist coverage constraints depending on the patient. They developed a deterministic mixed-integer programming model of the two scheduling problems and solved them sequentially and integrated[9]. They found that the integrated model had better performance. We did not find any evidence of papers that integrate session scheduling with shift scheduling, and staff allocation.

3.1.3 Stochasticity and flexibility

The research that includes stochasticity often limits it to the demand, and the resources – staff and modalities – are assumed to be constant[6]. Moreover, the stochasticity regarding the demand is not utilised to adjust the resources to the demand, but rather to construct a robust planning[17, 11]. However, some papers investigate the effects of an flexible appointment scheduling approach[15, 18, 19, 13]. Abtahi et al. [15] formulate the waiting room of a radiology center as a flexible open shop scheduling problem and a stochastic programming model, to minimize the time patients spend in the radiology center. Moreover, Rohleder and Klassen [19] use a rolling horizon (RH) appointment scheduling approach with fluctuating demand to find a balance between minimizing access times and efficient patient scheduling[19]. Vermeulen et al. [18] present an adaptive approach to automatic optimization of the time slot schedule of a CT-scanner for a diverse patient group. They build a schedule with reservations for orders of different urgency levels and scan types. Consequently, as time slots are used or not, they reallocate resources over time to establish an efficient schedule. All authors report improved efficiency in comparison to static scheduling[15, 18, 19, 13].

Furthermore, Aydas et al. [20] and Huang et al. [21] investigate adjusting the staff allocation according to the patient demand on a short horizon. Aydas et al. [20] adjust the nurse schedule at the start of shifts to adjust to a dynamic patient demand. Similarly, Huang et al. [21] optimize the OR staff schedule upon arrival of surgical patients. These short-term adjustments contributed to more efficient schedules[20, 21]. The adjustments were made according to realizations of the patient demand. We did not find any proof of papers that adjust the staff allocation on a longer horizon using demand predictions in a healthcare setting. However, such sequential decision models are widely researched in other research areas.

3.2 Sequential decision making under uncertainty

Sequential decision models can use an estimate of the downstream impact of a decision to find the best decision[22]. These models are called look-ahead models. The value of the future can be approximated with a value function or a direct look ahead method. A value function can be hard to define in problems with complex iterations, such as scheduling problems. Therefore, a direct look-ahead policy is more suitable in this case. A direct look-ahead method tries to optimize over the entire horizon from time t until the end of the scheduling horizon T[22]. Two-stage stochastic programs and multi-stage stochastic programs are such direct look-ahead methods that also involve stochasticity. Two-stage and multi-stage stochastic programs are suitable for many different practical problems[22, 23, 24]. For example, Xie and Huang [25] utilize a multi-stage stochastic mixed integer model to support biofuel supply chain expansion under evolving uncertainties, and Chang et al. [26] use a two-stage stochastic program to plan hurricane relief logistic operations under uncertainty. Kadri et al. [27]

use a multi-stage stochastic program to find optimal locations for electric vehicle charging stations.

The stochasticity in two-stage and multi-stage stochastic is captured in scenarios. As outlined by Shapiro and Philpott [28] the scenarios are commonly identified using sample average approximation (SAA)[28, 29, 30]. SAA entails generating M independent random realizations ω from the given distribution via Monte Carlo simulation. Increasing the number of scenarios or the size of the scenario tree is essential to find a better solution than the deterministic case[22]. However, the size of the model increases significantly as the number of scenarios or the scenario tree grows. This increase poses a challenge in terms of computational time and computer memory requirements.

The number of scenarios that need to be evaluated can be decreased while keeping the representation of the stochastic process as accurate as possible[22, 27]. This can be achieved by clustering the outcomes of SAA[31]. Each cluster contains a representative scenario that best reflects the scenarios in that cluster. A similar but simpler method is to decrease the granularity of the scenarios, meaning that the values are sorted into intervals and the scenarios are grouped that fall within one interval[22]. Alternatively, the number of necessary scenarios can be reduced by assuming some random variables to be deterministic[22]. In the current stage, the most up-to-date information may be used, but they are fixed in the look-ahead model.

The L-shaped method, an exact decomposition algorithm, is commonly used to solve two-stage stochastic models and different modifications have been proposed[32, 30, 33, 34, 35]. An interesting variation is proposed by Shi et al. [30] that applies cuts to only a subset of first-stage variables. The subset is formed by the first-stage variables that influence the value of the second-stage recourse function. This significantly reduces the number of nodes in the branching tree, and simplifies the computation of optimality cuts. The average running time was shown to be improved by roughly 10 times compared to the non-customised L-shaped method. Multi-stage stochastic programs can be solved using a nested decomposition method[25, 27]. Due to the exponential growth of the problem size as the horizon and scenario trees grow, instead metaheuristics and in particular genetic algorithms are often used[27]. Alternatively, a multi-stage stochastic program can be approximated by a two-stage program, lowering the numerical difficulties significantly[25].

Different methods to approximate multi-stage stochastic programs have been utilised. Chang et al. [26] approximate a multi-stage stochastic program with a two-stage stochastic program using a RH. They generate sample paths, solve a two-stage stochastic model to find the expected second-stage solution, implement the first-stage solution, update state variables, generate new sample paths, and repeat this until reaching the end of the time horizon[26]. This allows flexible planning over time. A RH is especially useful to solve multi-period problems where there's a need to make immediate decisions, and where the available data can be continuously updated over time[36]. Moreover, Xie and Huang [25] use stage aggregation to reduce a multi-period strategic expansion of a biofuel supply chain to a two-period program. Stage aggregation aggregates all but the first stage into one single second-stage decision; the first-stage decision is made, and then all future events are observed before making all remaining decisions at the same time. The nonanticipativity of information is lost when stages are aggregated because second-stage decisions can depend on future realizations of the data process[23]. This introduces inaccuracies, yet offers significant computational advantages.

3.3 Our contribution

This report contributes to the understanding of the multiple integrated decision levels at a radiology department by providing a mathematical model describing all decision layers and incorporating suitable random variables. We address the challenges associated with strategic resource allocation, tactical session scheduling and workforce scheduling, and operational scheduling, both offline and online.

In particular, we introduce a novel approach to flexible scheduling that dynamically responds to patient demand and staff availability. We incorporate resource allocation with staff scheduling, such that resources –the session schedule– can be adapted to the current patient demand. This innovative framework allows for the optimization of staff allocation considering both hard constraints, such as contractual constraints, and regulatory requirements, and soft constraints that influence the fairness of the schedule. This ensures the generated schedules are realistic and implementable in practice. Furthermore, we propose a suitable evaluation method using case-specific heuristics. The numerical experiments provide insights into the performance and adaptability of our flexible scheduling framework, shedding light on its potential benefits and challenges in real-world implementation.

Chapter 4

Methods and techniques

In this chapter, we discuss the methods and techniques used in this thesis in more detail. We introduce the basic properties of two-stage and multi-stage models. Moreover, we discuss the L-shaped method in more detail.

4.1 Two-stage stochastic programs

Two-stage stochastic programs are sometimes also called recourse models[24]. The first stage describes the decision x^0 that needs to be made now, at time t, and the second stage, or the recourse, describes the remaining decisions x^1 that need to be made at $t+1$. Often the second stage variable is written as y, however since we will introduce multi-stage linear programs later as well, we already adopt appropriate notation. The second stage involves a random variable ω with probability distribution P and support Ω. It is unknown at the first stage but known at the second stage. The realizations are also called scenarios. Commonly, a static information structure is assumed[23]:

Assumption 4.1.1 The probability distribution Ω of ω is known and independent of x^0.

However, in applications, the distribution is often unknown. Furthermore, even if it is known, often it still has to be approximated with a discrete distribution to find approximate solutions to a stochastic program. Therefore, the stability of stochastic programs with respect to perturbations of Ω is important. Under the condition that the approximating distributions converge weakly to the original distribution, the problem is functionally stable[37]. This justifies the use of discrete approximations of Ω[24].

The recourse decisions are made after observation of ω[24]. Therefore, the decision process is nonanticipative, meaning that decisions taken do not depend on future realizations of the data process[23]. The decisions in both stages are constrained, where $X^0 \in \mathbb{Z}^{\bar{p}^0} \times \mathbb{R}^{p^0 - \bar{p}^0}$ describes the feasible set of first stage actions and $X^1 \in \mathbb{Z}^{\bar{p}^1} \times \mathbb{R}^{p^1 - \bar{p}^1}$ describes the feasible set of recourse actions, the first $\bar{p}^0$ and $\bar{p}^1$ variables are integer variables. A two-stage stochastic program is defined as[24]:

$$\min_{x^0 \in X^0} c^0 x^0 \left(+\mathbb{E}_\omega \left[\min_{x^1 \in X^1} \left(c^1 x^1 : W x^1 \sim h(\omega) - T(\omega) x^0 \right) \right] : A x^0 = b \right). \tag{1}$$

Here, $Ax^0 = b$ represents m^0 deterministic equality constraints and $Wx^1 \sim h(\omega) - T(\omega)x^0$ represents m^1 random (in)equality constraints: $\sim$ indicates $=, \leq, \geq, <$ or $<$, and $T(\omega)$ is a $m^1 \times p^0$ matrix, $h(\omega)$ a $m^1 \times 1$ vector, and W a $m^1 \times p^1$ matrix[24]. The recourse structure is specified by (X^1, c^1, W). The recourse structure (X^1, c^1, W) is fixed if X^1, c^1 and W do not depend on the random vector ω.

Commonly, the two-stage problem is written more compactly using the recourse penalty cost function v and the expected minimum recourse cost Q defined as[24]:

$$v(z) = \min_{x^1 \in X^1} (c^1 x^1 : W x^1 \sim z), \quad z \in \mathbb{R}^m, \tag{2}$$

$$Q(x^0) = \mathbb{E}_\omega [v(h(\omega) - T(\omega)x^0)], \quad x^0 \in \mathbb{R}^n, \tag{3}$$

which enables us to write the two-stage problem as

$$\min_{x^0 \in X^0} \left(c^0 x^0 + Q(x^0) : Ax^0 = b \right). \tag{4}$$

The objective of this problem is nonlinear[24]. The properties of recourse models depend heavily on the properties of Q and the integrality conditions of the second stage. If the second stage is continuous, the objective can be shown to be convex. If the second stage contains integer variables, Q can be shown to be Lipschitz continuous in specific cases[37]. The concept of Lipschitz continuity is often useful in the proof of algorithms[24]. Other useful properties of recourse structures are[24]:

Definition 4.1.2 A mixed integer recourse problem has relative complete recourse if for all $\omega \in \Omega$ and al $x^0 \in X^0$ with $Ax^0 = b$ there exists a $x^1 \in X^1$, possibly depending on ω and x^0, such that $Wx^1 \sim h(\omega) - T(\omega)x^0$. Equivalently, if $v(h(\omega) - T(\omega)x^0) < +\infty$ for all $\omega \in \Omega, x^0 \in \{x^0 \in X^0 : Ax^0 = b\}$.

Relatively complete recourse is not unusual in stochastic programs[24]. In fact, often they satisfy the stronger property: complete recourse, which requires that for any feasible solution vector x^0 of the first state a feasible solution x^1 exists for the second stage[24]:

Definition 4.1.3 A mixed integer recourse problem has complete recourse if the recourse matrix W satisfies the condition: for all $z \in \mathbb{R}^{m^1}$ there exist $x^1 \in X^1$ such that $Wx^1 \sim z$. Equivalently, if $v(z) \leq +\infty$ for all $z \in \mathbb{R}^{m^1}$.

Definition 4.1.4 A fixed recourse structure (X^1, c^1, W) is called extremely inexpensive if there exists a $z \in \mathbb{R}^i$ such that $v(z) = -\infty$.

Extremely inexpensive recourse structures are not desirable, they may be avoided by setting $c^1 \geq 0$.

4.2 Multi-stage stochastic programs

A two-stage stochastic program may extended to a multi-stage stochastic program consisting of N stages[23]. A stage n is defined as a decision moment that impacts all subsequent stages $n+1 \ldots N$. Note, that a decision stage does not necessarily refer to a time period[38]. A multi-stage stochastic program is written as:

$$\min_{x^0 \in X^0} \left(c^0 x^0 + \mathbb{E}_{\omega^1} \left[\min_{x^1 \in X^1} \left(c^1 x^1 + \mathbb{E}_{\cdots} \left[\ldots + \mathbb{E}_{\omega^{N-1}} \left[\min_{x^{N-1} \in X^{N-1}} \left(c^{N-1} x^{N-1} + Q(x^{N-1}) \right] \right. \right. \right. \right. \right.$$
$$\left. \left. \left. : W^{N-1} x^{N-1} \sim h^{N-1}(\omega^{N-1}) - T^{N-1}(\omega^{N-1})x^{N-2} \right) \right] \ldots \right] \right)$$
$$\left. : Wx^1 \sim h^1(\omega^1) - T^1(\omega^1)x^0 \right] \right)$$
$$: Ax^0 = b \bigg), \tag{5}$$

where $X^n \in \mathbb{Z}^{\bar{p}^n} \times \mathbb{R}^{p^n - \bar{p}^n}$ describes the decision space with the first $\bar{p}^n$ integer variables. The variable ω is now a stochastic process $\omega = (\omega_0, \ldots, \omega_{N-1})$ with probability distribution P and support Ω[38]. The decision process is nonanticipative using Assumption 4.1.1. In a multistage setting, the definition of nonanticipativity should be extended to: x^n being independent of future realizations of ω, but may depend on past information and the probabilistic specification $(\Omega, \mathcal{F}, P)$ of the process ω, with $\mathcal{F}_{n-1} \subseteq \mathcal{F}$ the σ-field generated by the observations of $\omega^{n-1} = (\omega_0, \ldots, \omega_{n-1})$. The superscript n is added to the vectors and matrices h, W, T, since they may vary per decision stage. Their definitions remain the same as for two-stage stochastic programs.

A multi-stage stochastic program can be written more generically, without specifying the form of the constraints and simply including an objective function $\mathcal{O}$ that depends on the decision variable(s) and the stochastic variable(s)[23]:

$$\min_{x^0 \in X^0} \left(\mathcal{O}^0(x^0, \omega^0) + \mathbb{E}_{\omega^1} \left[\min_{x^1 \in X^1} \left(\mathcal{O}^1(x^1, \omega^1) + \mathbb{E}_{\cdots} \left[\ldots + \mathbb{E}_{\omega^{N-1}} \left[\min_{x^{N-1} \in X^{N-1}} \left(\mathcal{O}^{N-1}(x^{N-1}, \omega^{N-1}) \right) \right] \right] \right) \right] \right).$$
$$\tag{6}$$

with $x^n \in G^n$, where G^n is the subset of controls x^n that adhere to all constraints.

4.3   Solution methods

The simplest method to solve a two-stage or multi-stage stochastic program is to solve the deterministic equivalent of the problem[22]. First, a finite number of scenarios and their probabilities should be generated. The expected value can then be evaluated by taking the sum of the scenarios multiplied by their probability, transforming the problem into a linear program. A multi-stage program can be solved with a deterministic equivalent too by using a scenario tree. The scenario tree can be constructed by sampling scenarios starting at the root and continuing sampling scenarios for each branch until stage N. All nodes in the scenario tree should be linked using constraints in the model that connect the decisions and information about ω of connected nodes. The constraints can be designed such that nonanticipativity of information is kept[27]. The downside of the deterministic approach is that its size increases exponentially as the number of scenarios increases, causing computational challenges.

An alternative, more robust approach is the L-shaped method[24]. The L-shaped method takes advantage of the structure of a two-stage problem, similar to the Benders reformulation. The first stage decision is the master problem (MP) and the scenarios of the second stage are the sub problem (SP). The variable x^0 is the complicating factor, since if this was fixed, the first and second stages would be separable. The component in the objective function corresponding to the second stage variables: $Q(x^0)$ is replaced by a single variable ζ. The value of ζ is determined iteratively in a branch and cut process: branch on the relaxed master problem (RMP) until a feasible integer solution is found, then solve the SP to find a cut leveraging LP duality. The algorithm can be terminated when the desired accuracy level is achieved.

Laporte and Louveaux (1993)[39] extend this method for models with binary first-stage variables and mixed-integer second-stage variables. Although Q of such problems is non-convex, linear optimality cuts can still be constructed owing to the binary nature of the first-stage variables[39]. They show that for any problem for which a valid set of feasibility cuts and a valid set of optimality cuts exist, the Integer L-shaped method yields an optimal solution (when one exists) in a finite number of steps. In problems with (relative) complete recourse, only optimality cuts are necessary. The L-shaped method is implemented in a branch and cut process, as visualised in Figure 4.1.
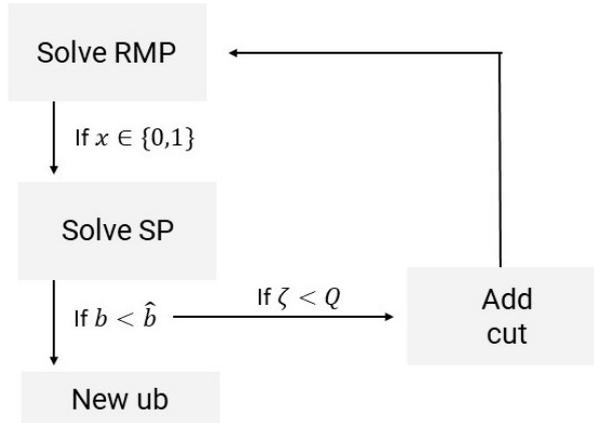

Figure 4.1: This figure depicts the integer L-shaped method. The RMP is solved until a feasible integer solution is found, then the SP is solved. If the solution of the current master problem and subproblem b is less than the current optimal solution $\hat{b}$, the upper bound is updated. If the objective value of the subproblem Q is also greater than ζ, then a cut is applied.

The optimality cut is defined as:

$$\zeta \geq \left(Q(x^{0,k}) - L\right) \left(\sum_{i \in V^k} x_i^0 - \sum_{i \notin V^k} x_i^0 - |V^k| + 1 \right) + L \tag{7}$$

with set $V^k = \{i : x_i^{0,k} = 1\}$, $x^{0,k} \in X^0 \subset \{0,1\}^{p^0}$ the solution of the kth iteration, $|V^k|$ the cardinality of V^k and L the lower bound of the second stage problem, this should be a finite value. The definition of $Q(x^{0,k}$

is in essence the same as in Equation (3): the sum of the objective values of the scenarios multiplied by their probability. This cut is sharp at $x^{0,k}$ and not larger than $Q(x^{0,k})$ for all feasible $x^{0,k}$[24].

Definition 4.3.1 A cut $Ax \leq b$ is called sharp if equality holds for at least one value of x.

If the first-stage integrality is discarded, nonlinear optimality cuts must be employed. These cuts are derived from the principles of duality theory in integer programming.

There are different ways to generate cuts[33]. The original integer L-shaped method is also known as the single-cut L-shaped method because it aggregates the information of all scenarios to generate a single cut. Alternatively, a multi-cut approach may be used. Then for each scenario, or sub-problem a cut is generated. This approach terminates in equal or fewer iterations than the original single-cut approach[40]. Alternatively, a hybrid version can be used: a cut is generated for each partition of subproblems[33]. The partitions can either be fixed during the optimization problem or vary each iteration. Another aggregation technique is cut consolidation. This may be applied together with any of the cut approaches mentioned in this paragraph. It reduces the size of the master problem by pruning historical cuts that have become inactive, but retain their aggregation to not lose all their information.

Chapter 5

Model

This chapter addresses sub-questions 1 and 2, which were introduced in Chapter 1. Sections 5.1-5.4 construct a general model step-by-step that describes a radiology department's current scheduling framework, as outlined in subquestion 1. Subsequently, in Section 5.5-5.6 the model is adjusted to accommodate flexible sessions, as outlined in subquestion 2.

In more detail, Section 5.1 discusses the general notation and definitions used in this Chapter. Section 5.2 introduces a simplified multi-stage decision model, under the assumption that decisions in each decision layer are taken at equal frequency. Section 5.3 relaxes this assumption, and it introduces two different approximation methods because the exact method is virtually intractable. In Section 5.4, objectives and constraints for the approximation methods are defined, culminating in the formulation of a general model for the existing scheduling framework. The constructed model is adjusted to accommodate flexible sessions in Section 5.5. Finally, Section 5.6 introduces precise definitions, constraints and objectives specific to the flexible model, contrasting with the more general constraints and objectives outlined in Sections 5.4 and 5.5. The models in Section 5.6 set the stage for the Numerical experiments in Chapter 6.

5.1 Notation and definitions

Let us first clarify notation and verbal terms. We use different decision levels and stages. A decision level is the highest level decision we are making at that moment, for example strategic. To make the strategic decision we have multiple decision stages that we take into account: tactical, operational offline, and operational online. Decision stages are decision moments considered within the main decision. The number of stages is counted by the number of decision moments. The time between the first and final decision is the horizon. The period a decision is made about is referred to as the planning period.

We define one decision vector x^n, which models the decision relevant for stage n. The meaning of x^n may differ per decision stage, namely in the strategic stage it might allocate general resources, while at the tactical stage, it can schedule sessions and RTs. The decision vector contains integer and continuous elements. Therefore, x^n of length p is contained in the set X^n: $X^n = \{x^{n,} \in \mathbb{Z}^{\bar{p}^n} \times \mathbb{R}^{p^n - \bar{p}^n}\}$, with $\bar{p}$ integer variables.

We introduce two random variables: supply S and demand D. We write lower case for their realizations and capital letters when they are random variables. The demand describes the patients in need of an appointment and the supply the radiology departments ability to provide the patients with an appointment. In more detail, the supply describes the availability of RTs. The availability of RTs is a random variable because the availability can change over time, due to for example education, pregnancy, sickness, and leave. The patient demand D is determined by the work backlog W and the newly incoming orders F, as explained in Section 2.2. The work backlog is predicted using the current work backlog (at the moment of the decision) w^c and the inflow of orders F^θ and the outflow of orders, the production P^θ per week θ until the period θ' we are predicting:

$$W^{\theta'} = w^c + \sum_{\theta=c}^{\theta'} F^\theta - P^\theta \tag{1}$$

The inflow and outflow of orders are both stochastic. As weeks pass, the inflow and outflow of these weeks become known; more information becomes available. Moreover, the predictions regarding the inflow of orders improve as more information becomes available regarding the number of outpatient sessions of the referring departments. In Figure 5.1 the inflow of information is visualised. All this information is nonanticipative. Therefore, the σ-field generated by (S, D) increases over time, known as filtration. The distribution of the supply and demand is inherently discrete because the supply describes RTs and the demand describes patients. The distributions of S and D are assumed to be independent of x, following the common Assumption 4.1.1. In practice, the decision x does influence how the resources can be used, however the availability S does not change.

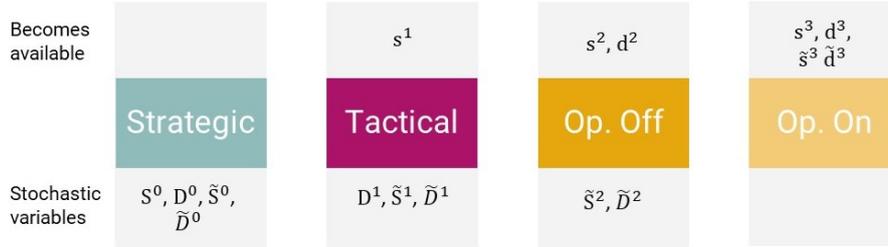| Becomes available		s^1	s^2, d^2	$s^3, d^3,$ $\tilde{s}^3 \tilde{d}^3$
	Strategic	Tactical	Op. Off	Op. On
Stochastic variables	$S^0, D^0, \tilde{S}^0,$ $\tilde{D}^0$	$D^1, \tilde{S}^1, \tilde{D}^1$	$\tilde{S}^2, \tilde{D}^2$	

Figure 5.1: *This figure depicts the information structure. For each level, the available variables and random variables are visualised.*

After the strategic decision, information regarding supply and demand will start arriving gradually. However, we can distinguish distinct moments in time where S and D become known. The supply becomes known just before making tactical decisions, as this timing coincides with RTs submitting their requests for vacation and education days. The demand is mostly known right before the operational offline decision because the rosters of the sessions of the referring departments become available. Moreover, the work backlog at the moment of the operational offline decision is quite representative of the work backlog during the planning period.

These distinguished moments in time are somewhat arbitrary, as they do not imply that the actual demand and supply will cease to change once they are available at time n. These remaining changes are captured in $\tilde{S}^n$ and $\tilde{D}^n$. These variables include all additional uncertainties for level n. Their expectations are set to zero because all expected changes are captured within S and D. Moreover, we do not consider $\tilde{S}^n$ and $\tilde{D}^n$ while their counterpart S or D are random variables. Namely, then the variability is captured in D or S. During the next decision $n + 1$ the known variables s^n and d^n are updated to the current knowledge s^{n+1} and d^{n+1}, the remaining variability is captured in $\tilde{S}^{n+1}$ and $\tilde{D}^{n+1}$. For example, s^1 is the supply as known at the tactical stage, and s^2 is the supply as known at the operational offline stage. During both these stages, the variability of the supply with respect to s^1 and s^2 is modelled using $\tilde{S}^1$ and $\tilde{S}^2$, respectively. Note that $\tilde{S}^1$ and $\tilde{S}^2$ describe the same last-minute variation in supply, just at different stages. Finally, at the final stage $n = 3$, operational online, all variables are known.

5.2 Multi-stage stochastic program

We model the different decision levels as a multi-stage stochastic program. In each stage n we minimize x^n over an objective function $\mathcal{O}^n$ using constraints, which we will define later in Section 5.4.1 and Section 5.4.2, respectively.

We make the following assumption to simplify the initial model formulation:

Assumption 5.2.1: The decisions on each level are taken with equal frequency and impact the same planning period.

Consequently, we have for, for example, the strategic decision level that we take the strategic decision, then the tactical decision, operational offline, and operational online decision all regarding the same period. Let us start with the formulation of the operational online decision. This is a single-stage deterministic decision, because all variables $s, d, \tilde{s}$ and $\tilde{d}$ are known. These variables can be used to make last-minute schedule changes. The operational online problem is formulated as:

$$\min_{x^3} \left(\mathcal{O}^3(x^3, s^3, \tilde{s}^3, d^3, \tilde{d}^3) \right). \tag{2}$$

The operational offline decision is a two-stage stochastic decision. The supply and demand are known and written as s^2 and d^2. Since both s and d are known in the first stage, they can be used in the second stage. Recall that the remaining variability is captured in the random variables $\tilde{S}$ and $\tilde{D}$. The operational offline problem is formulated as:

$$\min_{x^2} \left(\mathbb{E} \left[\mathcal{O}^2(x^2, s^2, \tilde{S}^2, d^2, \tilde{D}^2) + \mathbb{E} \left[\min_{x^3} \left(\mathcal{O}^3(x^3, s^2, \tilde{S}^3, d^2, \tilde{D}^3) \right) \right] \right] \right) \tag{3}$$

The tactical decision is a three-stage decision. The supply s is known and the demand D is a random variable, therefore $\tilde{D}^1$ is left out in the tactical stage within the tactical level. Since the supply is known, we use s^1 in all three stages. The tactical problem is formulated as:

$$\min_{x^1} \left(\mathbb{E} \left[\mathcal{O}^1(x^1, s^1, \tilde{S}^1, D^1) + \mathbb{E} \left[\min_{x^2} \left(\mathcal{O}^2(x^2, s^1, \tilde{S}^2, D^2, \tilde{D}^2) \right. \right. \right. \right.$$
$$\left. \left. \left. \left. + \mathbb{E} \left[\min_{x^3} \left(\mathcal{O}^3(x^3, s^1, \tilde{S}^3, D^3, \tilde{D}^3) \right) \right] \right) \right] \right] \right). \tag{4}$$

The strategic decision is a four-stage decision. The supply S and D are random variables in the strategic stage, therefore, $\tilde{S}^0, \tilde{D}^0$ are left out in the strategical stage within the strategical level. Similarly, $\tilde{D}^1$ is left out in the tactical stage. The strategic problem is formulated as:

$$\min_{x^0} \left(\mathbb{E} \left[\mathcal{O}^0(x^0, S^0, D^0) + \mathbb{E} \left[\min_{x^1} \left(\mathcal{O}^1(x^1, S^1, \tilde{S}^1, D^1) \right. \right. \right. \right.$$
$$\left. \left. \left. + \mathbb{E} \left[\min_{x^2} \left(\mathcal{O}^2(x^2, S^2, \tilde{S}^2, D^2, \tilde{D}^2) + \mathbb{E} \left[\min_{x^3} \left(\mathcal{O}^3(x^3, S^3, \tilde{S}^3, D^3, \tilde{D}^3) \right) \right] \right) \right] \right) \right] \right] \right) \tag{5}$$

5.3   Frequency

We relax assumption 5.2.1, because in reality not all four types of decisions are made as frequently and do not impact equal planning periods. Particularly, strategic decisions are made once a year for the entire next year. Tactical decisions are made once a month for a month at least three months in advance, and operational offline decisions are made regarding one week. Lastly, operational online decisions are made daily impacting just that day. The differences in frequency are visualised in Figure 5.2.

Figure 5.2: This figure visualizes the frequency at which the decisions of different levels are taken for a radiology department.

Let us zoom in on the two-stage problem described in (3). This equation models the operational offline decision regarding one week θ, which contains 7 days t. Therefore, we need to adjust the objective of Equation (3) to

$$\min_{x^{2,\theta}} \left(\mathbb{E}\left[\mathcal{O}^2(x^{2,\theta}, s^{2,\theta}, \tilde{S}^{2,\theta}, d^{2,\theta}) + \mathbb{E}\left[\min_{x^{3,\theta}} \left(\mathcal{O}_W^3(x^{3,\theta}, s^{2,\theta}, \tilde{S}^{3,\theta}, d^{2,\theta}, \tilde{D}^{3,\theta}) \right) \right] \right] \right), \tag{6}$$

where we superscript all variables in the first stage with θ, indicating the planning period: week θ.

The objective $\mathcal{O}_W^3$ should take into account all 7 days considered in the operational offline decision. Ideally, we separate these decisions per day and in chronological order, such that outcomes of the first day can be considered in the decision for the second day and so forth. Consequently, $\mathcal{O}_W^3$ contains 7 separate operational online decisions, which are taken on different days t. If we reflect this in (3), we get:

$$\min_{x^{2,\theta}} \left(\mathbb{E}\left[\mathcal{O}^2(x^{2,\theta}, s^{2,\theta}, \tilde{S}^{2,\theta}, d^{2,\theta}, \tilde{D}^{2,\theta}) + \mathbb{E}\left[\min_{x_0^{3,\theta}} \left(\mathcal{O}^3(x_0^{3,\theta}, s_0^{2,\theta}, \tilde{S}_0^{3,\theta}, d_0^{2,\theta}, \tilde{D}_0^{3,\theta}) \right. \right. \right.$$
$$\left. \left. \left. + \mathbb{E}\left[\ldots + \mathbb{E}\left[\min_{x_6^{3,\theta}} \left(\mathcal{O}^3(x_6^{3,\theta}, s_6^{2,\theta}, \tilde{S}_6^{3,\theta}, d_6^{2,\theta}, \tilde{D}_6^{3,\theta}) \right) \right] \right] \right) \right] \right] \right), \tag{7}$$

where we superscript all variables in the first stage with θ, indicating the week the decision impacts and subscript the variables in the subsequent stages with $t = 0, 1, \ldots, 6$, indicating the day the decision impacts.

However, this problem is intractable for larger instances. Therefore, we propose an approximation: stage aggregation (AG). We assume that the operational online information, $s^{3,\theta}, \tilde{S}^{3,\theta}, d^{3,\theta}, \tilde{d}^{3,\theta}$ regarding all days in week θ is available at the beginning of the week. This allows us to aggregate all operational online decisions into a single decision stage taken at day 0. This results in the two-stage problem:

$$\min_{x^{2,\theta}} \left(\mathbb{E}\left[\mathcal{O}^2(x^{2,\theta}, s^{2,\theta}, \tilde{S}^{2,\theta}, d^{2,\theta}) + \mathbb{E}\left[\min_{x^{3,\theta}} \left(\sum_{t \in \theta} \mathcal{O}^3(x_t^{3,\theta}, s_t^{2,\theta}, \tilde{S}_t^{3,\theta}, d_t^{2,\theta}, \tilde{D}_t^{3,\theta}) \right) \right] \right] \right). \tag{8}$$

We can sum over the different days within the objective $\mathcal{O}_W^3$ because the operational online objective function $\mathcal{O}^3$ is the same for each day. The relations between days can be modeled by constraints. However, due to these constraints the nonanticipativity of information is lost, because now information regarding the last day is available for the decision regarding the first day. This incurs an error. The approximation method AG is a heuristic decision policy within the framework of multi-stage stochastic programming. Therefore, the well-established value of stochastic solutions of two-stage stochastic programs, may not directly translate to the multi-stage stochastic programming context when employing AG. Furthermore, it is evident that the objective function of AG will surpass that of (7). Nevertheless, our primary focus is to outline the general decision strategy, which we expect AG will exhibit. The performance of AG is analyzed using numerical experiments in Chapter 6.

Moreover, we can extend AG by assuming that the operational online decisions are independent in the operational online stage. This means that during the operational offline decision we decide on $x_t^{2,\theta}$: the resources available for day t. Subsequently, at the operational online decision these resources can only be redistributed over day t. Whereas before, only the aggregate decision $x^{2,\theta}$ was fixed and the resources could still be redistributed among the days in the week at the operational online level. The constraints in the operational offline stage that distribute the available resources among $x_t^{2,\theta}$, cause $x_t^{2,\theta}$ to be not mutually independent. Therefore, their assumed independence will incur an error. However, the independence assumption does make the information anticipative again, because only information regarding day t is used for the operational online decision regarding day t.

The minimum of the sum of functions of independent variables is equal to the sum of the minima of the functions, which is simply proven as:

Proof. Let x and y be independent variables and $f()$, $g()$ any functions. If we have $f(x) \geq a$ and $g(y) \geq b$, then we have that the sum of minima is: $\min_x(f(x)) + \min_y(g(y)) = a + b$. Then, if we define $h(x, y) = f(x) + g(y)$, then we have that $\min_{x,y} h(x, y) \geq a + b$, because x and y are independent. Thus proving that $\min_{x,y} f(x) + g(y) = \min_x f(x) + \min_y g(y)$; the minimum of the sum of functions with independent variables is equal to the sum of the minima. $\square$

Therefore, we can switch the order of the summation and minimum operator, subsequently using the additivity of expectation we can write Equation (8) as:

$$\min_{x^{2,\theta}} \left(\mathbb{E}\left[\mathcal{O}^2(x^{2,\theta}, s^{2,\theta}, \tilde{S}^{2,\theta}, d^{2,\theta}) + \sum_{t\in\theta} \mathbb{E}\left[\min_{x_t^{3,\theta}} \left(\mathcal{O}^3(x_t^{3,\theta}, s_t^{2,\theta}, \tilde{S}_t^{3,\theta}, d_t^{2,\theta}, \tilde{D}_t^{3,\theta}) \right) \right] \right] \right). \tag{9}$$

Let us call this the per-stage independent stage aggregation (PI) method. PI can reduce the computational burden of solving a two-stage stochastic model. This becomes clear when evaluating the stages sequentially, as done by a decomposition algorithm like the L-shaped method. Then, each component (day) of the sum in the operational online stage can be calculated independently. Consequently, the size of the SPs decreases while the number of SPs increases. If the original SP is sufficiently large, then solving multiple smaller SPs in parallel can be more computationally efficient than solving one large SP. However, if the SP is small the increased overhead of solving multiple SPs can diminish this effect.

By analogy of the previous arguments we can find expressions for the strategic and tactical decision for both AG and PI. However, AG is not suitable for decisions that encompass more than two stages because it is not able to separate decisions of the same level. This means that all decisions at the operational offline level are determined using one minimization problem within the tactical decision. This allows the redistribution of resources over the weeks within the month the tactical decision is about. However, it also means that all operational online decisions about all days of the month are taken in one minimization problem. This is impractical and does not capture the decision layers correctly.

In contrast, PI assumes that the different operational offline problems are independent and that the different operational online problems are independent in the tactical decision. This allows the following formulation that captures the different frequency and timing of the decisions of the different decisions levels more accurately:

$$\min_{x^{1,\Theta}} \left(\mathbb{E}\left[\mathcal{O}^1(x^{1,\Theta}, s^{1,\Theta}, \tilde{S}^{1,\Theta}, D^{1,\Theta}) + \sum_{\theta\in\Theta} \mathbb{E}\left[\min_{x^{2,\theta}} \left(\mathcal{O}^2(x^{2,\theta}, s^{1,\theta}, \tilde{S}^{2,\theta}, D_t^{2,\theta}, \tilde{D}^{2,\theta}) \right. \right. \right. \right.$$

$$\left. \left. \left. + \sum_{t\in\theta} \mathbb{E}\left[\min_{x_t^{3,\theta}} \left(\mathcal{O}^3(x_t^{3,\theta}, s_t^{1,\theta}, \tilde{S}_t^{3,\theta}, D_t^{3,\theta}, \tilde{D}_t^{3,\theta}) \right) \right] \right) \right] \right] \right), \tag{10}$$

$$\tag{11}$$

where $x^{1,\Theta}$ impacts the month Θ, $x^{2,\theta}$ the weeks in month Θ, and $x_t^{3,\theta}$ the days of week θ. Similarly, the strategic decision is defined as:

$$\min_{x^{0,\Xi}} \left(\mathbb{E}\left[\mathcal{O}^0(x^{0,\Xi}, S^{0,\Xi}, D^{0,\Xi}) + \sum_{\Theta^k\in\Xi} \min_{x^{1,\Theta^k}} \left(\mathbb{E}\left[\mathcal{O}^1(x^{1,\Theta^k}, S^{1,\Theta^k}, \tilde{S}^{1,\Theta^k}, D^{1,\Theta^k}) \right. \right. \right. \right.$$

$$\left. + \sum_{\theta\in\Theta^k} \mathbb{E}\left[\min_{x^{2,\Theta^k,\theta}} \left(\mathcal{O}^2(x^{2,\Theta^k,\theta}, S^{2,\Theta^k,\theta}, \tilde{S}_t^{2,\Theta^k,\theta}, D^{2,\Theta^k,\theta}, \tilde{D}^{2,\Theta^k,\theta}) \right. \right. \right.$$

$$\left. \left. \left. \left. + \sum_{t\in\theta} \mathbb{E}\left[\min_{x_t^{3,\Theta^k,\theta}} \left(\mathcal{O}^3(x_t^{3,\Theta^k,\theta}, S_t^{3,\Theta^k,\theta}, \tilde{S}_t^{3,\Theta^k,\theta}, D_t^{3,\Theta^k,\theta}, \tilde{D}_t^{3,\Theta^k,\theta}) \right) \right] \right) \right] \right) \right] \right] \right), \tag{12}$$

where $x^{0,\Xi}$ impacts year Ξ, x^{1,Θ^k} impacts months Θ^k in year Ξ, $x^{2,\theta}$ the weeks in month Θ^k, and $x_t^{3,\theta}$ the days of week θ.

5.4   General model

In the previous section, two approximations were formulated for the two-stage operational offline decision: AG and PI and for the multi-stage tactical and strategic decision just PI. In this section, we will further specify the objective and constraints for these approximations. The objective, constraints combined with Equations (8, 9), (10) and (12) are subsequently used to define a general formulation of the multi-stage model describing the radiology department's scheduling framework.

5.4.1 Objective

We define the general objective for all decision stages and levels as:

$$c^n * x^n + P^n, \tag{13}$$

where c^n the cost factor represents the costs of stage n incurred by x^n and P a continuous penalty variable. The penalty depends on x^n. We explicitly specify this additional penalty variable, because, as will be explained in subsequent Sections, the model is not only optimised over costs but also penalties generated by soft constraints. As c^n, P^n and the meaning of x^n may vary per stage, the objectives are also able to vary between stages. This is necessary because, for example, the strategic goals are different than the operational online goals. The coefficients c^n should be non-negative to ensure the problem is not extremely inexpensive, per definition 4.1.4.

5.4.2 Constraints

We define hard constraints that ensure the constructed schedule fulfills all minimal requirements to be a feasible schedule in practice. Additionally, we define soft constraints to promote the performance goals of the radiology department. A feasible solution at any stage should be feasible in another stage. Therefore, the model is constructed to have complete recourse. In practice, this is not too hard, in Section 5.6.3 we will show that the model constructed for the experiments indeed has complete recourse. For each stage, different constraints are needed. We define three types of constraints which together capture all needed constraints. These three different constraints also contain the definition(s) of P.

First, we need to constrain the problem according to hard constraints, such as the capacity and the budget. These constraints are of the form:

$$A^n x^n = b^n, \tag{14}$$

where A^n is the constraint matrix and b^n a vector of appropriate size.

Second, we need to include constraints that relate to the random variables, which are of the general form:

$$T^n(D^n, \tilde{D}^n, S^n, \tilde{S}^n)x^n \sim h^n(D^n, \tilde{D}^n, S^n, \tilde{S}^n), \tag{15}$$

where $T^n()$ is the constraint matrix and $h^n()$ a vector dependent on one or more of the random variables of appropriate size. The first stage cannot contain any random variables, therefore this constraint type is not needed in the first stage.

The third constraint links the decision variables x^n between the different stages. This constraint ensures continuity across stages, for example regarding resources, scheduled sessions, or budgetary commitments. We assume that only the decision in the previous stage should be taken into account. The constraints are of the general form:

$$E^n x^n \sim G^n x^{n+1} + k^n, \tag{16}$$

where E^n and G^n are matrices of appropriate size and k a vector of appropriate size. This constraint type is not necessary in the first stage as no previous decision has been taken yet.

Recall, the two general constraints that were introduced in Chapter 3 [24]:

$$A^n x^n = b^n, \tag{17}$$

$$W x^{n-1} n \sim h(\omega) - T(\omega)x^n, \tag{18}$$

where ω is a random variable.

Equation (18) can be used to model all three constraint types for all stages except the first stage. Namely, the matrices $T^n()$, $W^n()$ and vector h^n may be written such that part of the constraints do not depend on the random variables and x^{n-1}, and thus model the first constraint type. Furthermore, a part of the constraints may be independent of x^{n-1} and model the constraints of the second type. Lastly, by making a part of the matrices and vectors independent of the random variables, we can link the decision variables of subsequent stages and find constraints of type three. Equation (17) can be used to model the first stage constraints. The matrix A^n and vector b^n may depend on the expectation(s) of random variables, the realizations of random variables or the decisions taken in previous stages. Note, that the first stage differs depending on the decision level.

5.4.3 General multi-stage stochastic programs

Our fully defined, but generic multistage model can be constructed using the previously defined constraints, objective and aggregation methods. The operational online decision is a single stage decision and does not require any aggregation. The operational online decision that decides over one day t in week θ is defined as:

$$\min_{x_t^{3,\theta}} \left(c^3 x_t^{3,\theta} + P^3 : A^3(x_t^{2,\theta}, d_t^{3,\theta}, \tilde{d}_t^{3,\theta}, s_t^{3,\theta}, \tilde{s}_t^{3,\theta}) x_t^{3,\theta} = b^3(x_t^{2,\theta}, d_t^{3,\theta}, \tilde{d}_t^{3,\theta}, s_t^{3,\theta}, \tilde{s}_t^{3,\theta}) \right), \tag{19}$$

where $x_t^{2,\theta}$ is the previously determined resources in stage 2: operational offline. This decision does not contain any randomness.

The operational offline decision is a two-stage decision, which means we can apply both AG and PI. The operational offline decision decides over week θ that contains 7 days indexed by t and is defined for AG as:

$$\min_{x^{2,\theta}} \left(c^2 x^{2,\theta} + P^2 + \mathbb{E}\left[\min_{x_t^{3,\theta}, t \in \theta} \left(\sum_{t \in \theta} c^3 x_t^{3,\theta} + P^3 \right. \right. \right.$$
$$: W^3(d_t^{2,\theta}, \tilde{D}_t^{3,\theta}, s_t^{2,\theta}, \tilde{S}_t^{3,\theta}) x^{2,\theta} \sim h^3(d_t^{2,\theta}, \tilde{D}_t^{3,\theta}, s_t^{2,\theta}, \tilde{S}_t^{3,\theta}) - T^3(d_t^{2,\theta}, \tilde{D}_t^{3,\theta}, s_t^{2,\theta}, \tilde{S}_t^{3,\theta}) x_t^{3,\theta} \Big) \Big]$$
$$: A^2(x^{1,\theta}, d^{2,\theta}, s^{2,\theta}) x^{2,\theta} = b^2(x^{1,\theta}, d^{2,\theta}, s^{2,\theta}) \bigg), \tag{20}$$

and using PI as

$$\min_{x_t^{2,\theta}, t \in \theta} \left(\sum_{t \in \theta} c^2 x_t^{2,\theta} + P^2 + \sum_{t \in \theta} \mathbb{E}\left[\min_{x_t^{3,\theta}} \left(c^3 x_t^{3,\theta} + P_t^3 \right. \right. \right.$$
$$: W^3(d_t^{2,\theta}, \tilde{D}_t^{3,\theta}, s_t^{2,\theta}, \tilde{S}_t^{3,\theta}) x_t^{2,\theta} \sim h^3(d_t^{2,\theta}, \tilde{D}_t^{3,\theta}, s_t^{2,\theta}, \tilde{S}_t^{3,\theta}) - T^3(d_t^{2,\theta}, \tilde{D}_t^{3,\theta}, s_t^{2,\theta}, \tilde{S}_t^{3,\theta}) x_t^{3,\theta} \Big) \Big]$$
$$: A^2(x_t^{1,\theta}, d_t^{2,\theta}, s_t^{2,\theta}) x_t^{2,\theta} = b^2(x_t^{1,\theta}, d_t^{2,\theta}, s_t^{2,\theta}) \bigg), \tag{21}$$

where $x^{1,\theta}$ represents the previously determined resources in stage 1: tactical. Note, that A^2 and b^2 are not dependent on $\mathbb{E}[\tilde{S}_t^{3,\theta}]$ and $\mathbb{E}[\tilde{D}_t^{3,\theta}]$, because both these expectations were assumed zero and were therefore left out. We point out the difference between AG and PI of AG determining $x^{2,\theta}$ in the first stage and PI determining $x_t^{2,\theta}, \forall t \in \theta$.

The tactical decision is a three-stage decision that decides over month Θ. The tactical decision using PI is defined as:

$$\min_{x^{1,\theta}, \theta \in \Theta} \left(\sum_{\theta \in \Theta} c^1 x^{1,\theta} + P^1 + \sum_{\theta \in \Theta} \mathbb{E}\left[\min_{x_t^{2,\theta}} \left(\sum_{t \in \theta} c^2 x_t^{2,\theta} + P^{2,\theta} + \sum_{t \in \theta} \mathbb{E}\left[\min_{x_t^{3,\theta}} \left(c^3 x_t^{3,\theta} + P_t^{3,\theta} \right. \right. \right. \right. \right.$$
$$: W^3(D_t^{3,\theta}, \tilde{D}_t^{3,\theta}, s_t^{1,\theta}, \tilde{S}_t^{3,\theta}) x_t^{2,\theta} \sim h^3(D_t^{3,\theta}, \tilde{D}_t^{3,\theta}, s_t^{1,\theta}, \tilde{S}_t^{3,\theta}) - T^3(D_t^{3,\theta}, \tilde{D}_t^{3,\theta}, s_t^{1,\theta}, \tilde{S}_t^{2,\theta}) x_t^{3,\theta} \Big) \Big]$$
$$: W^2(D_t^{2,\theta}, \tilde{D}_t^{2,\theta}, s_t^{1,\theta}, \tilde{S}_t^{2,\theta}) x_t^{1,\theta} \sim h^2(D_t^{2,\theta}, \tilde{D}_t^{2,\theta}, s_t^{1,\theta}, \tilde{S}_t^{2,\theta}) - T^2(D_t^{2,\theta}, \tilde{D}_t^{2,\theta}, s_t^{1,\theta}, \tilde{S}_t^{2,\theta}) x^{2,\theta} \Big) \Big]$$
$$: A^1(x^{0,\theta}, \mathbb{E}[D^{1,\theta}], s^{1,\theta}) x^{1,\theta} = b^1(x^{0,\theta}, \mathbb{E}[D^{1,\theta}], s^{1,\theta}) \bigg), \tag{22}$$

where $x^{0,\theta}$ is the resources decided on previously in stage 0: strategical.

The strategic decision is a four-stage decision that decides over year Ξ is defined as:

$$
\min_{x^{0,\Theta^k},\Theta^k\in\Xi}\left(\sum_{\Theta^k\in\Xi}c^0x^{0,\Theta^k}+P^0+\sum_{\Theta^k\in\Xi}\mathbb{E}\min_{x^{1,\Theta^k,\theta}}\left(\sum_{\theta\in\Theta^k}c^1x^{1,\Theta^k,\theta}+P^{1,\Theta^k}+\sum_{\theta\in\Theta^k}\mathbb{E}\Bigg[\tag{23}\right.\right.
$$

$$
\min_{x_t^{2,\Theta^k,\theta}}\left(\sum_{t\in\theta}c^2x_t^{2,\Theta^k,\theta}+P^{2,\Theta^k,\theta}+\sum_{t\in\theta}\mathbb{E}\left[\min_{x_t^{3,\Theta^k,\theta}}\left(c^3x_t^{3,\Theta^k,\theta}+P_t^{3,\Theta^k,\theta}\right.\right.\right.
$$

$$
: W^3(D_t^{3,\Theta^k,\theta},\tilde{D}_t^{3,\Theta^k,\theta},S_t^{3,\Theta^k,\theta},\tilde{S}_t^{3,\Theta^k,\theta})x_t^{2,\Theta^k,\theta}\sim h^3(D_t^{3,\Theta^k,\theta},\tilde{D}_t^{3,\Theta^k,\theta},S_t^{3,\Theta^k,\theta},\tilde{S}_t^{3,\Theta^k,\theta})
$$

$$
\left.\left.-T^3(D_t^{3,\Theta^k,\theta},\tilde{D}_t^{3,\Theta^k,\theta},S_t^{3,\theta},\tilde{S}_t^{3,\Theta^k,\theta})x_t^{3,\Theta^k,\theta}\right)\right]
$$

$$
: W^2(D_t^{2,\Theta^k,\theta},\tilde{D}_t^{2,\Theta^k,\theta},S_t^{2,\Theta^k,\theta},\tilde{S}_t^{2,\Theta^k,\theta})x_t^{1,\theta}\sim h^2(D_t^{2,\theta},\tilde{D}_t^{2,\theta},S_t^{2,\theta},\tilde{S}_t^{2,\Theta^k,\theta})
$$

$$
\left.\left.-T^2(D_t^{2,\Theta^k,\theta},\tilde{D}_t^{2,\Theta^k,\theta},S_t^{2,\theta},\tilde{S}_t^{2,\Theta^k,\theta})x_t^{2,\Theta^k,\theta}\right)\right]
$$

$$
: W^1(D^{1,\Theta^k,\theta},\tilde{D}^{1,\Theta^k,\theta},S^{1,\Theta^k,\theta},\tilde{S}^{1,\Theta^k,\theta})x^{0,\Theta^k,\theta}\sim h^1(D^{1,\Theta^k,\theta},\tilde{D}^{1,\Theta^k,\theta},S^{1,\Theta^k,\theta},\tilde{S}^{1,\Theta^k,\theta})
$$

$$
\left.\left.-T^1(D_t^{1,\Theta^k,\theta},\tilde{D}_t^{1,\Theta^k,\theta},S_t^{1,\Theta^k,\theta},\tilde{S}^{1,\Theta^k,\theta})x^{1,\Theta^k,\theta}\right)\right]
$$

$$
\left.: A^0(\mathbb{E}[D^{0,\Theta^k}],\mathbb{E}[S^{0,\Theta^k}])x^{0,\Theta^k}=b^0(\mathbb{E}[S^{0,\Theta^k}],\mathbb{E}[D^{0,\Theta^k}])\right).
$$

5.5 General model including flexible sessions

The scheduling framework is now adjusted to allow flexibility in the schedule to adapt to a varying demand and RT availability. The strategic and operational online decisions are removed from the scope because they are not as relevant: strategic decisions are made only once a year and operational online decisions are highly variable and based on small unpredictable changes $\tilde{S}$ and $\tilde{D}$. This leaves us to focus on tactical decisions and operational offline decisions. Since it is a two-stage problem, both AG and PI are suitable approximation methods.

Recall, that the current scheduling framework regarding the tactical decision and operational offline decision, consists of the tactical decision: making a session schedule and then the RT roster and the operational offline decision: adjusting the RT schedule such that it remains feasible and canceling sessions in case of low demand. Flexibility can be introduced into this model by changing the scheduling framework and timeline. We still distinguish a tactical decision taken three months in advance and an operational offline decision, but change their meaning, which is explained in the next subsections.

5.5.1 Tactical decision

We integrate the construction of the session schedule with the RT-schedule and make the session schedule flexible. Flexible sessions are essentially open scheduling blocks that are temporarily assigned to a modality but may be changed in the operational offline decision. Thus, the tactical decision determines, based on the demand and supply, a session schedule with fixed and flexible sessions and an RT schedule in accordance with this session schedule. In the tactical decision RTs are scheduled on activities to ensure feasibility, but only the shift times are fixed. Fixing the shift times in the tactical decision three months in advance is necessary to oblige to the collective labor agreement. The RTs may be switched between activities in the operational offline decision.

Different tactical decisions are related. We model this relation hierarchically. A hierarchical relation means that to make a tactical decision three months in advance, we take the outcome, the scheduled sessions, of the months before. In more detail, the work backlog prediction for the tactical decision regarding month 4 is based on the operational offline decision concerning month 1 and the tactical decisions concerning month 2 and month 3.

The tactical decision can be formulated by adjusting (10) by removing operational online stage. For AG we define:

$$
\min_{x^{1,\Theta}}\left(\mathbb{E}\left[\mathcal{O}^1(x^{1,\Theta},s^{1,\Theta},\tilde{S}^{1,\Theta},D^{1,\Theta})+\mathbb{E}\left[\min_{x^{2,\theta},\theta\in\Theta}\left(\sum_{\theta\in\Theta}\mathcal{O}^2(x^{2,\theta},s^{1,\theta},\tilde{S}^{2,\theta},D^{2,\theta},\tilde{D}^{2,\theta})\right)\right]\right]\right), \tag{24}
$$

which can be extended to:

$$\min_{x^{1,\Theta}} \left(c^1 x^{1,\Theta} + P^1 + \mathbb{E}\left[\min_{x^{2,\theta},\theta\in\Theta} \left(\sum_{\theta\in\Theta} c^2 x^{2,\theta} + P^2 + \right.\right.\right.$$

$$: W^2(D^{2,\theta},\tilde{D}^{2,\theta},s^{1,\theta},\tilde{S}^{2,\theta})x^{1,\Theta} \sim h^2(D^{2,\theta},\tilde{D}^{2,\theta},s^{1,\theta},\tilde{S}^{2,\theta}) - T^2(D^{2,\theta},\tilde{D}^{2,\theta},s^{1,\theta},\tilde{S}^{2,\theta})x^{2,\theta} \left.\left.\bigg)\right]\right.$$

$$: A^1(x^{0,\Theta},\mathbb{E}[D^{1,\Theta}],s^{1,\Theta})x^{1,\Theta} = b^1(x^{0,\Theta},\mathbb{E}[D^{1,\Theta}],s^{1,\Theta}) \bigg) \tag{25}$$

and for PI:

$$\min_{x^{1,\Theta}} \left(\mathbb{E}\left[\mathcal{O}^1(x^{1,\Theta},s^{1,\Theta},\tilde{S}^{1,\Theta},D^{1,\Theta}) + \sum_{\theta\in\Theta} \mathbb{E}\left[\min_{x^{2,\theta}} \left(\mathcal{O}^2(x^{2,\theta},s^{1,\theta},\tilde{S}^{2,\theta},D^{2,\theta},\tilde{D}^{2,\theta}) \right)\right]\right]\right), \tag{26}$$

which can similarly be extended to:

$$\min_{x^{1,\theta},\theta\in\Theta} \left(\sum_{\theta\in\Theta} c^1 x^{1,\theta} + P^1 + \sum_{\theta\in\Theta} \left(\mathbb{E}\left[\min_{x^{2,\theta}} c^2 x^{2,\theta} + P^{2,\theta} + \right.\right.\right.$$

$$: W^2(D^{2,\theta},\tilde{D}^{2,\theta},s^{1,\theta},\tilde{S}^{2,\theta})x^{1,\theta} \sim h^2(D^{2,\theta},\tilde{D}^{2,\theta},s^{1,\theta},\tilde{S}^{2,\theta}) - T^2(D^{2,\theta},\tilde{D}^{2,\theta},s^{1,\theta},\tilde{S}^{2,\theta})x^{2,\theta} \left.\left.\bigg)\right]\right.$$

$$: A^1(x^{0,\theta},\mathbb{E}[D^{1,\theta}],s^{1,\theta})x^{1,\theta} = b^1(x^{0,\theta},\mathbb{E}[D^{1,\theta}],s^{1,\theta}) \bigg). \tag{27}$$

5.5.2 Operational offline decision

The operational offline decision needs to be made at least 4 weeks in advance, due to practical constraints. Therefore, the interval between the tactical and operational offline decisions spans approximately two months. During this time, more information about the supply and demand becomes known. While this could prompt the introduction of more decision points to leverage this new information, practical considerations suggest otherwise. Since all obligatory sessions for emergency and long-term booking are already accounted for in the three-month tactical decision, there is no immediate need to allocate more sessions until 4 weeks in advance. Thus, deferring operational offline decisions until this point is preferred; postponing these decisions is always better than taking additional decisions earlier.
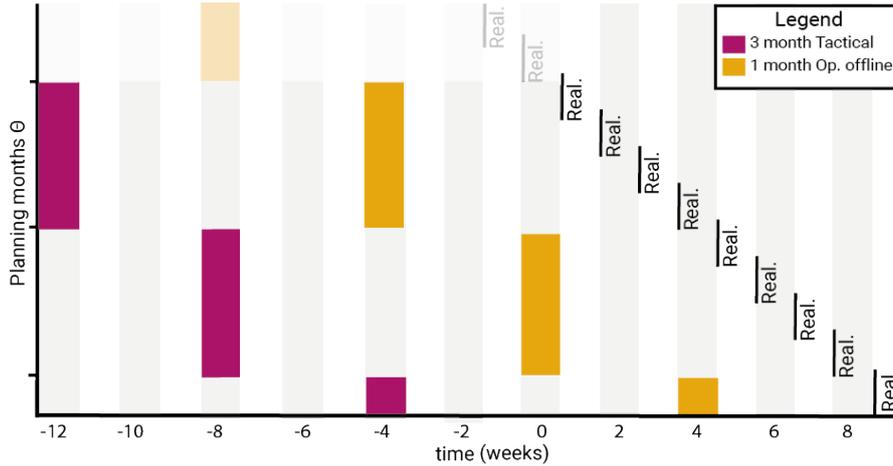

Figure 5.3: The pink blocks represent the tactical decisions made for one month (four weeks) three months in advance. The yellow-toned blocks represent the operational offline decisions taken for one month one month in advance.

The operational offline decision or decisions need to be made four weeks in advance. The decisions can be made on a month-basis. This new timeline is visualised in Figure 5.3. The operational offline decision can be modeled using an operational offline MP and operational offline SPs. The operational offline decision still involves random variables. Therefore, it is useful to model operational offline SPs, because then still different outcomes of the random variables can be considered.

The operational offline decision using AG is defined as:

$$\min_{x^{2,\Theta}} \left(\mathbb{E} \left[\mathcal{O}^2(x^{2,\Theta}, s^{2,\Theta}, \tilde{S}^{2,\Theta}, d^{2,\Theta}, \tilde{D}^{2,\Theta}) + \mathbb{E} \left[\min_{x^{3,\theta}, \theta \in \Theta} \left(\sum_{\theta \in \Theta} \mathcal{O}^3(x^{3,\theta}, s^{2,\theta}, \tilde{S}^{3,\theta}, d^{2,\theta}, \tilde{D}^{3,\theta}) \right) \right] \right] \right), \qquad (28)$$

which can be extended to:

$$\min_{x^{2,\Theta}} \left(c^2 x^{2,\Theta} + P^2 + \mathbb{E} \left[\min_{x^{3,\theta}, \theta \in \Theta} \left(\sum_{\theta \in \Theta} c^3 x^{3,\theta} + P^3 + \right. \right. \right.$$
$$: W^3(d^{2,\theta}, \tilde{D}^{3,\theta}, s^{2,\theta}, \tilde{S}^{3,\theta}) x^{2,\Theta} \sim h^3(d^{2,\theta}, \tilde{D}^{3,\theta}, s^{2,\theta}, \tilde{S}^{3,\theta}) - T^3(d^{2,\theta}, \tilde{D}^{3,\theta}, s^{2,\theta}, \tilde{S}^{3,\theta}) x^{3,\theta} \bigg) \bigg]$$
$$: A^2(x^{1,\Theta}, d^{2,\theta}, s^{2,\theta}) x^{2,\Theta} = b^2(x^{1,\Theta}, d^{2,\theta}, s^{2,\theta}) \bigg) \qquad (29)$$

and for PI:

$$\min_{x^{2,\Theta}} \left(\mathbb{E} \left[\mathcal{O}^2(x^{2,\Theta}, s^{2,\Theta}, \tilde{S}^{2,\Theta}, d^{2,\Theta}, \tilde{D}^{2,\Theta}) + \sum_{\theta \in \Theta} \mathbb{E} \left[\min_{x^{3,\theta}} \left(\mathcal{O}^3(x^{3,\theta}, s^{2,\theta}, \tilde{S}^{3,\theta}, d^{2,\theta}, \tilde{D}^{3,\theta}) \right) \right] \right] \right), \qquad (30)$$

which can similarly be extended to:

$$\min_{x^{2,\theta}, \theta \in \Theta} \left(\sum_{\theta \in \Theta} c^2 x^{2,\theta} + P^2 + \sum_{\theta \in \Theta} \mathbb{E} \left[\min_{x^{3,\theta}} \left(c^3 x^{3,\theta} + P^{3,\theta} + \right. \right. \right.$$
$$: W^3(d^{2,\theta}, \tilde{D}^{3,\theta}, s^{2,\theta}, \tilde{S}^{3,\theta}) x^{2,\theta} \sim h^3(d^{2,\theta}, \tilde{D}^{3,\theta}, s^{2,\theta}, \tilde{S}^{3,\theta}) - T^3(d^{2,\theta}, \tilde{D}^{3,\theta}, s^{2,\theta}, \tilde{S}^{3,\theta}) x^{3,\theta} \bigg) \bigg]$$
$$: A^2(x^{1,\theta}, d^{2,\theta}, s^{2,\theta}) x^{2,\theta} = b^2(x^{1,\theta}, d^{2,\theta}, s^{2,\theta}) \bigg). \qquad (31)$$

Note that $n = 3$ is used for the operational offline SP. However, this SP is not of the operational online level as was previously indicated with $n = 3$, because last-minute decision-making is not considered.

However, only regarding the process on a month-basis is not necessarily optimal, because the operational offline decisions regarding all but the first week of the month are made earlier than necessary and use relatively old data. Instead, the operational offline decision can be taken weekly. This means that every week an operational offline decision is made regarding the week which is four weeks away, using the most up-to-date data regarding the work backlog and the sessions. This second possible timeline is visualised in Figure 5.4.
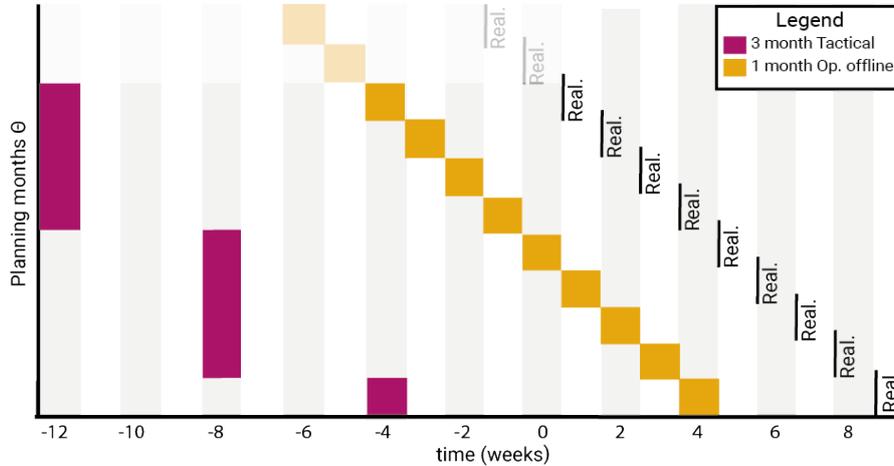

Figure 5.4: The pink blocks represent the tactical decisions made for one month (four weeks) three months in advance. The yellow-toned blocks represent the operational offline decisions taken for one week one month in advance.

The operational decision on a week-basis is the same for AG and PI, because only one week is considered. We find:

$$\min_{x^{2,\theta}} \left(\mathbb{E}\left[\mathcal{O}^2(x^{2,\theta}, s^{2,\theta}, \tilde{S}^{2,\theta}, d^{2,\theta}) + \mathbb{E}\left[\min_{x^{3,\theta}} \left(\mathcal{O}^3(x^{3,\theta}, s^{2,\theta}, \tilde{S}^{3,\theta}, d^{2,\theta}, \tilde{D}^{3,\theta}) \right) \right] \right] \right), \tag{32}$$

which can similarly be extended to:

$$\min_{x^{2,\theta}} \left(c^2 x^{2,\theta} + P^2 + \mathbb{E}\left[\min_{x^{3,\theta}} \left(c^3 x^{3,\theta} + P^3 + \right.\right.$$

$$: W^3(d^{2,\theta}, \tilde{D}^{3,\theta}, s^{2,\theta}, \tilde{S}^{3,\theta}) x^{2,\theta} \sim h^3(d^{2,\theta}, \tilde{D}^{3,\theta}, s^{2,\theta}, \tilde{S}^{3,\theta}) - T^3(d^{2,\theta}, \tilde{D}^{3,\theta}, s^{2,\theta}, \tilde{S}^{3,\theta}) x^{3,\theta} \bigg) \bigg]$$

$$: A^2(x^{1,\theta}, d^{2,\theta}, \mathbb{E}[D^{2,\theta}], s^{2,\theta}) x^{2,\theta} = b^2(x^{1,\theta}, d^{2,\theta}, \mathbb{E}[\tilde{D}^{2,\theta}], s^{2,\theta}) \bigg). \tag{33}$$

The best performing timeline should be verified by numerical experiments.

5.6 Specified model including flexible sessions

This section refines our previously formulated models setting the stage for the numerical experiments. We have identified two aggregation methods—-stage aggregation (AG) and per-stage independent aggregation (PI)——along with two timelines, each containing a tactical two-stage decision and an operational offline two-stage decision. This yields a total of eight two-stage stochastic programs, which share similar components and objectives.

The objective of each of these stochastic programs should reflect the performance of the radiology department in terms of the goals of the department: minimizing the costs, ensuring a low and stable access time for patients, minimizing deviations from ideal scheduling scenarios to maintain model feasibility, and optimizing the fairness of the RT schedule. Consequently, the objectives contain 4 distinct elements: the session costs and three types of penalties: demand-related penalties, feasibility penalties, and fairness penalties. Before we define the objective in more detail, we introduce all necessary elements, including sets, subsets of variables and decision variables in Subsection 5.6.1. Followed by the definition and description of the constraints that model the problem and calculate the different types of penalties in Subsection 5.6.2. Finally, we construct the models and their objectives using these components in Subsection 5.6.3.

5.6.1 Definitions

We define multiple sets to index our decision and penalty variables. Furthermore, we define several subsets, notated as $\mathcal{A}_b^B$ with $\mathcal{A}$ and $\mathcal{B}$ both sets of distinct variables. Then $\mathcal{A}_b^B$ is a subset of $\mathcal{A}$, existing of all elements of $\mathcal{A}$ that are related to b in the set $\mathcal{B}$. For example, $\mathcal{I}_l^L$ contains all modalities $i \in \mathcal{I}$ that are available at location $l \in \mathcal{L}$. All sets and subsets are listed in Table 5.1.

Set	Description
$\mathcal{I}$	Contains all modalities i that should be scheduled.
$\mathcal{J}$	Contains all RTs j that may be scheduled.
$\mathcal{T}$	Contains all scheduling blocks t in a period. A day has a morning scheduling block and an afternoon scheduling block.
$\mathcal{L}$	Contains all modality locations l.
Θ	Contains all weeks θ belonging to scheduling period Θ.
θ	Contains all scheduling blocks t belonging to scheduling week θ.
$\mathcal{I}^{\delta}$	Contains all modalities i that are scheduled based on demand.
$\mathcal{I}^{\eta}$	Contains all modalities i that require postgraduate professional education (CT, MRI, US, and angiography).
$\mathcal{I}^{\kappa}$	Contains all modalities i that do not require postgraduate professional education.
$\mathcal{I}_l^L$	Contains all modalities i on location l.
$\mathcal{I}_j^J$	Contains all modalities i that RT j is qualified to operate.
$\mathcal{J}_i^I$	Contains all RTs j qualified to operate modality i
$\mathcal{J}_l^L$	Contains all RTs j that work on location l
$\mathcal{J}_t^R$	Contains all RTs j that are roster-free at t, roster-free means that the RT is available but preferably not scheduled.
$\mathcal{J}_t^{\Theta,n} (\mathcal{J}_t^{\theta,n})$	Contains all RTs j that may be scheduled during t in period $\Theta(\theta)$ in stage n.
$\Theta_j^{J,n} (\theta_j^{J,n})$	Contains all scheduling blocks t in period $\Theta(\theta)$ that RT j may be scheduled in stage n.
$\Theta^o (\theta^o)$	Contains all morning scheduling blocks t in period $\Theta(\theta)$.
$\Theta_{i,l}^C (\theta_{i,l}^C)$	Contains all scheduling blocks t that modality i should be closed at location l in period $\Theta(\theta)$.
$\mathcal{M}$	Contains all scenarios m.

Table 5.1: This table describes the definitions of necessary sets and subsets.

Using these sets and subsets we define our decision variables as given in Table 5.2 and the stochastic inputs as given in Table 5.3.

Variable	Description
$x_{i,t,l}^n \in \{0,1\} \, \forall i \in \mathcal{I}, \forall t \in \theta, \forall l \in \mathcal{L}_i^I,$ $(\forall \theta \in \Theta)$	Binary decision variable that indicates if there is a regular session scheduled for modality i during scheduling block t at location l in stage n.
$z_{i,t,l}^n \in \{0,1\} \, \forall i \in \mathcal{I}, \forall t \in \theta, \forall \theta \in \Theta, \forall l \in \mathcal{L}_i^I,)$ $(\forall \theta \in \Theta)$	Binary decision variable that indicates if there is a flexible session scheduled for modality i during scheduling block t at location l in stage n.
$y_{i,j,t,l}^n \in \{0,1\} \, \forall i \in \mathcal{I}, \forall j \in \mathcal{J}_i^I, \forall t \in \theta_j^{J,n},$ $, \forall l \in \mathcal{L}_i^I \cap \mathcal{L}_j^J, (\forall \theta \in \Theta)$	Binary decision variable that indicates if RT j is scheduled to work modality i during scheduling block t at location l in stage n.
$\Delta_i^{+n}, \Delta_i^{-n}, \delta_{i,\theta}^{+n}, \delta_{i,\theta}^{-n}, \pi_{i,\theta}^n \in \mathbb{R}_{\geq 0} \, \forall i \in \mathcal{I}^{\delta},$ $(\forall \theta \in \Theta)$	Continuous penalty variables regarding the demand in stage n.
$\kappa_{i,j,\theta}^n \in \mathbb{R}_{\geq 0} \, \forall j \in \mathcal{J}, \forall i \in \mathcal{I}_j^J \cup \mathcal{I}^{\kappa}, (\forall \theta \in \Theta),$ $\gamma_{i,j,t,\theta}^n \in \mathbb{R}_{\geq 0} \, \forall t \in \theta^o, \forall j \in \mathcal{J}_t^{\theta} \cap \mathcal{J}_{t+1}^{\theta},$ $\forall i \in \mathcal{I}_j^J, (\forall \theta \in \Theta), \rho^n \in \mathbb{R}_{\geq 0}, \psi_j^n \in \mathbb{R}_{\geq 0} \, \forall j \in \mathcal{J},$ $\eta_{i,j,\theta}^n \in \mathbb{R}_{\geq 0} \, \forall j \in \mathcal{J}, \forall i \in \mathcal{I}_j^J \cup \mathcal{I}^{\eta}, (\forall \theta \in \Theta)$	Continuous penalty variables regarding fairness in stage n.
$\alpha_j^{+n}, \alpha_j^{-n} \in \mathbb{R}_{\geq 0} \, \forall j \in \mathcal{J}, \beta_{i,t,l}^n \in \mathbb{R}_{\geq 0} \, \forall i \in \mathcal{I}_l^L,$ $\forall l \in \mathcal{L}_i^I, \forall t \in \theta, (\forall \theta \in \Theta), \xi_{j,t}^{+n}, \xi_{j,t}^{-n} \in \mathbb{R}_{\geq 0}$ $\forall j \in \mathcal{J}, \forall t \in \theta_j^J, (\forall \theta \in \Theta)$	Continuous penalty variables regarding feasibility in stage n.

Table 5.2: This table describes the definitions of necessary variables. If the total period is one week, $(\forall \theta \in \Theta)$ is left out.

Inputs	Description
$\mathbb{E}[F_{i,\theta}^n]$	The expected orders in sessions for modality i in week θ in the master problem in stage n.
$f_{i,\theta}^{n,m}$	The orders in sessions for modality i in week θ in scenario m in stage n.
$\mathbb{E}[W_i^n]$	The expected work backlog for modality i at the start of the planning period in the master problem in stage n.
$w_i^{n,m}$	The work backlog for modality i at the start of the planning period in scenario m in stage n.
ϕ_m	The probability of scenario m.
$\mathcal{J}_t^{\Theta,n}$ $(\mathcal{J}_t^{\theta,n})$	The up-to-date information of the RT availability $\mathcal{J}_t^{\Theta,n}$ $(\mathcal{J}_t^{\theta,n})$ in stage n.
$\Theta_j^{J,n}$ $(\theta_j^{J,n})$	The up-to-date information of the RT availability $\Theta_j^{J,n}$ $(\theta_j^{J,n})$ in stage n.

Table 5.3: This table describes the definitions of the (stochastic) inputs.

5.6.2 Constraints

First, in Table 5.4 we introduce the additional inputs and functions needed for the constraints. Second, we introduce the necessary constraints to model the problems, using either AG or PI, correctly. The constraints are split into general constraints, constraints for the tactical MP and SP, and constraints for the operational offline MP and SP. The constraints that are equivalent across problems are marked with the same letter and the additions of n for the stage, and T for tactical and O for operational. Subsequently, a description is provided for each of these constraints. Third, the demand constraints and fairness constraints are introduced, each accompanied by a motivation and description.

Functions and constants

Entity	Description
$\underline{x}(i,l)$	Returns the minimum number of sessions for modality i that is required per week.
$\bar{x}(i,l)$	Returns the weekly capacity of modality i.
$b^\Theta(j)$	Returns the number of sessions an RT j is available for in period Θ. This is the contracted FTE minus indirect work and a reservation for irregular shifts.
$b^\theta(j)$	Returns the number of sessions an RT j is available for per week θ. This is the contracted FTE minus indirect work and a reservation for irregular shifts.
$V_{i,t,l}$	The set of the scheduling blocks t during which it is required to have a session of modality i at location l. The size of the set is $|V_{i,l}|$.
$e(i,t,l)$	Returns the number of RTs needed to staff a session of modality i at scheduling block t at location l
$m(i)$	Returns the minimum number of sessions that should be scheduled for a modality i per month.
$u^n(i)$	Returns the minimum number of sessions that should be fixed in stage n for a modality i per week.
$\bar{q}(i)$	Returns the average production (orders/session) of modality i.
$q^p(i,l)$	Returns the average production (orders/session) of modality i at location l.
w_i*	The desired work backlog for modality i.
$k(i)$	The minimum number of sessions an RT should desirably be scheduled for each modality they are qualified for.
g^I, g^{III}, g^{IV}	The number of sessions you want to do in addition to the incoming orders in case the work backlog is in this interval: I, III, IV.
a	The number of sessions the model can deviate from the number of sessions determined by the incoming orders and goal before applying the additional penalty π.

Table 5.4: This table describes the definitions of necessary functions and constants.

General

$$\sum_{t\in\theta} x_{i,t,l}^n \geq \underline{x}(i,l) \qquad \forall l \in \mathcal{L}, \forall i \in \mathcal{I}_l^L, \quad \text{(a)},$$
$$(\forall \theta \in \Theta), \forall n$$

$$\sum_{l\in\mathcal{L}} \sum_{i\in\mathcal{I}_l^L \cap \mathcal{I}_j^J} y_{i,j,t,l}^n \leq 1 \qquad \forall j \in \mathcal{J}, \forall t \in \theta_j^{J,n}, \quad \text{(b)},$$
$$(\forall \theta \in \Theta), \forall n$$

$$\sum_{t\in V_{i,t,l}} x_{i,t,l}^n = |V_{i,l}| \qquad \forall l \in \mathcal{L}, \forall i \in \mathcal{I}_l^L : \quad \text{(c)},$$
$$V_{i,l} \neq \emptyset, (\forall \theta \in \Theta), \forall n$$

$$\sum_{l\in\mathcal{L}} \sum_{i\in\mathcal{I}^L \cap \mathcal{I}_j^J} y_{i,j,t,l}^n - y_{i,j,t+1,l}^n = 0 \qquad \forall t \in \theta^o, \forall j \in \mathcal{J}_t^\theta \cap \mathcal{J}_{t+1}^\theta, \quad \text{(d)},$$
$$(\forall \theta \in \Theta), \forall n$$

$$\sum_{i\in\mathcal{I}_{l'}^L \cap \mathcal{I}_j^J} \left(y_{i,j,t,l'}^n + \right. \leq \sum_{i\in\mathcal{I}_{l'}^L \cap \mathcal{I}_j^J} \left(y_{i,j,t+1,l'}^n + \right. \qquad \forall l' \in \mathcal{L}, \forall t \in \theta^o, \quad \text{(e)},$$
$$\left. \sum_{l\in\mathcal{L}\setminus l'} y_{i,j,t+1,l}^n \right) \quad \left. \sum_{l\in\mathcal{L}} y_{i,j,t+1,l}^n \right) \qquad \forall j \in \mathcal{J}_t^R \cup \mathcal{J}_{t+1}^R,$$
$$(\forall \theta^o \in \Theta^o), \forall n$$

Remark: if the constraints are used on a weekly basis $(\forall \theta \in \Theta)$ is left out.

Tactical MP

$$\sum_{j\in\mathcal{J}_t^\Theta} y_{i,j,t}^1 - e(i,t,l)(x_{i,t,l}^1 + z_{i,t,l}^1) = 0 \qquad \forall l \in \mathcal{L}, \forall i \in \mathcal{I}_l^L, \forall t \in \Theta \qquad \text{(f.1.T)},$$

$$x_{i,t,l}^1 + z_{i,t,l}^1 \leq 1 \qquad \forall l \in \mathcal{L}, \forall i \in \mathcal{I}_l^L, \forall t \in \Theta \qquad \text{(g.1.T)},$$

$$\sum_{t\in\theta} x_{i,t,l}^1 + z_{i,t,l}^1 \leq \bar{x}(i,l) \qquad \forall l \in \mathcal{L}, \forall i \in \mathcal{I}_l^L, \forall \theta \in \Theta \qquad \text{(h.1.T)},$$

$$\sum_{t\in\Theta} \sum_{l\in\mathcal{L}_i^I} x_{i,t,l}^1 + z_{i,t,l}^1 \geq m(i) \qquad \forall i \in \mathcal{I} \qquad \text{(i.1.T)},$$

$$\sum_{t\in\theta} \sum_{l\in\mathcal{L}_i^I} x_{i,t,l}^1 \geq u(i)\mathbb{E}[F_{i,\theta}^1] \qquad \forall i \in \mathcal{I}^\delta, \forall \theta \in \Theta \qquad \text{(j.1.T)},$$

$$\alpha_j^{+1} - \alpha_j^{-1} = b^\Theta(j) - \qquad \forall j \in \mathcal{J} \qquad \text{(k.1.T)},$$
$$\sum_{l\in\mathcal{L}} \sum_{i\in\mathcal{I}_l^L \cap \mathcal{I}_j^J} \sum_{t\in\Theta_j^{J,1}} y_{i,j,t,l}^1$$

$$\sum_{l\in\mathcal{L}} \sum_{i\in\mathcal{I}_l^L \cap \mathcal{I}_j^J} \sum_{t\in\theta_j^{J,1}} y_{i,j,t,l}^1 \leq b^\theta(j) + 1 \qquad \forall j \in \mathcal{J}, \forall \theta \in \Theta \qquad \text{(l.1.T)},$$

$$\sum_{i\in\mathcal{I}} \sum_{l\in\mathcal{L}_i^I} \sum_{t\in\Theta_{i,l}^C} x_{i,t,l}^1 + z_{i,t,l}^1 = 0 \qquad \text{(m.1.T)}.$$

Tactical SP

$$\sum_{j\in\mathcal{J}_t^\theta} y_{i,j,t}^2 - e(i,t,l)x_{i,t,l}^2 = 0 \qquad \forall l \in \mathcal{L}, \forall i \in \mathcal{I}_l^L, \forall t \in \theta, \qquad \text{(f.2.T)},$$
$$(\forall \theta \in \Theta)$$

$$\sum_{t\in\theta} x_{i,t,l}^2 \leq \bar{x}(i,l) \qquad \forall l \in \mathcal{L}, \forall i \in \mathcal{I}_l^L, \qquad \text{(h.2.T)},$$
$$(\forall \theta \in \Theta)$$

$$\sum_{i\in\mathcal{I}} \sum_{l\in\mathcal{L}_i^I} \sum_{t\in\theta_{i,l}^C} x_{i,t,l}^2 = 0 \qquad (\forall \theta \in \Theta) \qquad \text{(m.2.T)},$$

$$\sum_{l\in\mathcal{L}} \sum_{i\in\mathcal{I}_l^L \cap \mathcal{I}_j^J} y_{i,j,t,l}^1 - y_{i,j,t,l}^2 = 0 \qquad \forall j \in \mathcal{J}, \forall t \in \theta_j^J, \qquad \text{(n.2.T)},$$
$$(\forall \theta \in \Theta)$$

$$\beta_{i,t,l}^2 = x_{i,t,l}^2 - 2x_{i,t,l}^1 + 1 \qquad \forall l \in \mathcal{L}, \forall i \in \mathcal{I}_l^L, \forall t \in \theta, \qquad \text{(o.2.T)}.$$
$$(\forall \theta \in \Theta)$$

Remark: if the constraints are used on a weekly basis $(\forall \theta \in \Theta)$ is left out.

Operational MP

$$\sum_{j\in\mathcal{J}_t^\theta} y_{i,j,t}^2 - e(i,t,l)x_{i,t,l}^2 = 0 \qquad \forall l\in\mathcal{L}, \forall i\in\mathcal{I}_l^L, \forall t\in\Theta \qquad \text{(f.2.O)},$$

$$\sum_{t\in\theta} x_{i,t,l}^2 \leq \bar{x}(i,l) \qquad \forall l\in\mathcal{L}, \forall i\in\mathcal{I}_l^L, \forall\theta\in\Theta \qquad \text{(h.2.O)},$$

$$\sum_{t\in\Theta}\sum_{l\in\mathcal{L}_i^I} x_{i,t,l}^2 \geq m(i) \qquad \forall i\in\mathcal{I} \qquad \text{(i.2.O)},$$

$$\alpha_j^{+2} - \alpha_j^{-2} = b^\Theta(j) - \qquad \forall j\in\mathcal{J} \qquad \text{(k.2.O)}$$
$$\sum_{l\in\mathcal{L}}\sum_{i\in\mathcal{I}_l^L\cap\mathcal{I}_j^J}\sum_{t\in\Theta_j^{J,1}} y_{i,j,t}^2$$

$$\sum_{l\in\mathcal{L}}\sum_{i\in\mathcal{I}_l^L\cap\mathcal{I}_j^J}\sum_{t\in\theta_j^{J,1}} y_{i,j,t}^2 \leq b^\theta(j)+1 \qquad \forall j\in\mathcal{J}, \forall\theta\in\Theta \qquad \text{(l.2.O)},$$

$$\sum_{i\in\mathcal{I}}\sum_{l\in\mathcal{L}_i^I}\sum_{t\in\Theta_{i,t,l}^C} x_{i,t,l}^2 = 0 \qquad \text{(m.2.O)},$$

$$\xi_{j,t}^{+2} - \xi_{j,t}^{-2} = \sum_{l\in\mathcal{L}}\sum_{i\in\mathcal{I}_l^L\cap\mathcal{I}_j^J} y_{i,j,t}^1 - y_{i,j,t}^2 \qquad \forall j\in\mathcal{J}, \forall t\in\Theta_j^J \qquad \text{(n.2.O)},$$

$$\beta_{i,t,l}^2 = x_{i,t,l}^2 - 2x_{i,t,l}^1 + 1 \qquad \forall l\in\mathcal{L}, \forall i\in\mathcal{I}_l^L, \forall t\in\Theta) \qquad \text{(o.2.O)}.$$

Operational SP

$$\sum_{j\in\mathcal{J}_t^\theta} y_{i,j,t}^3 - e(i,t,l)x_{i,t,l}^3 = 0 \qquad \forall l\in\mathcal{L}, \forall i\in\mathcal{I}_l^L, \forall t\in\theta, \qquad \text{(f.3.O)},$$
$$(\forall\theta\in\Theta)$$

$$\sum_{t\in\theta} x_{i,t,l}^3 \leq \bar{x}(i,l) \qquad \forall l\in\mathcal{L}, \forall i\in\mathcal{I}_l^L, \qquad \text{(h.3.O)},$$
$$(\forall\theta\in\Theta)$$

$$\sum_{i\in\mathcal{I}}\sum_{l\in\mathcal{L}_i^I}\sum_{t\in\theta_{i,t,l}^C} x_{i,t,l}^3 = 0 \qquad (\forall\theta\in\Theta) \qquad \text{(m.3.O)},$$

$$\sum_{l\in\mathcal{L}}\sum_{i\in\mathcal{I}_l^L\cap\mathcal{I}_j^J} y_{i,j,t,l}^2 - y_{i,j,t,l}^3 = 0 \qquad \forall j\in\mathcal{J}, \forall t\in\theta_j^J, \qquad \text{(n.3.O)},$$
$$(\forall\theta\in\Theta)$$

$$\beta_{i,t,l}^{3,m} = x_{i,t,l}^3 - 2x_{i,t,l}^2 + 1 \qquad \forall l\in\mathcal{L}, \forall i\in\mathcal{I}_l^L, \forall t\in\theta, \qquad \text{(o.3.O)}.$$
$$(\forall\theta\in\Theta)$$

Remark: if the constraints are used on a weekly basis $(\forall\theta\in\Theta)$ is left out.

Description

(a) Ensures that at least $\underline{x}(i,l)$ sessions are scheduled for modality i at location l per week.

(b) Ensures that every RT j is scheduled for at most one modality i and one location l during every scheduling block t.

(c) Ensures that modalities that have obligatory sessions $t\in V_{i,l}$ are opened during these scheduling blocks. The set $V_{i,l}$ contains $|V_{i,l}|$ scheduling blocks.

(d) Ensures that every RT j, who is available and does not have the morning or afternoon roster-free, is scheduled for the morning and afternoon and works both sessions at the same location.

(e) Ensures that every RT j that is available in the morning and afternoon and has the morning or/and afternoon roster-free, is scheduled at the same location if they are scheduled in the morning and afternoon.

(f) Ensures that exactly the number of needed RTs $e(i,t,l)$ are scheduled for a session of modality i during t at location l.

- (f.1.T): Enforced for regular sessions x and flexible sessions z.

- (f.2.O), (f.2.T), (f.3.O): Enforced only for regular sessions x .

(g) Ensures that either a regular session x or a flexible session z is scheduled.

(h) Ensures that equal or fewer regular (and flexible) sessions than the maximum capacity $\bar{x}(i,t)$ are scheduled for each modality every week θ.

- (h.1.T): Enforced for regular sessions x and flexible sessions z.

- (h.2.O), (h.2.T), (h.3.O): Enforced only for regular sessions x .

(i) Ensures that for every modality i at least a monthly minimum $m(i)$ of sessions is scheduled.

- (i.1.T): Enforced for regular sessions x and flexible sessions z.
- (i.2.O): Enforced only for regular sessions x.

(j) Ensures that for every demand-based scheduled modality i at least $u(i)$ of the expected orders in sessions are dedicated in the tactical decision.

(k) Computes a penalty if an RT j is scheduled more than their monthly availability $b^\Theta(j)$: α_j^{+n} or if an RT is scheduled less than their availability: α_j^{-n}.

(l) Ensures that an RT is not scheduled more than one scheduling block extra per week than their availability $b^\theta(j)$ describes.

(m) Ensures that modality i at location l is closed during the scheduling blocks in $\Theta_{i,l}^C$.

- (m.1.T): Enforced for regular sessions x and flexible sessions z.
- (m.2.O), (m.2.T), (m.3.O): Enforced only for regular sessions x .

(n) Ensures that if an RT j is scheduled in the first stage during scheduling block t, they are still scheduled during the second stage.

- (k.2.O): Enforced as a soft constraint, computing the penalty ξ^+ or ξ^- for a change in the operational RT shift schedule versus the tactical RT shift schedule.
- (k.2.T), (k.3.O): Enforced as a hard constraint by setting the right-hand side of the equality to zero. Therefore, allowing no changes between the first and second stage decision.

(o) Computes the penalty $\beta_{i,t,l}^n$ for closing a fixed session $x_{i,t,l}^1$ in the second stage.

Demand constraints

The demand constraints need to ensure that the sessions are scheduled according to the demand. Recall that the demand is determined by the work backlog and the incoming orders. We decided to formulate the demand constraints by determining a 'goal' number of sessions based on the orders and the current work backlog. A penalty is applied to the objective if these goals are not met. Note that the work backlog or deviation from the desired work backlog is not part of the objective function. This method, utilizing a goal, is a heuristic method to have the number of sessions adapt to the demand and is not an optimal method. However, we propose that when appropriate goals and intervals are used, they should motivate the model to schedule a suitable number of sessions.

Particularly, we determined four intervals of the work backlog and a corresponding number of goal sessions. The first interval I, is when the work backlog is lower than the desirable work backlog. In that case, fewer sessions than the incoming orders should be scheduled to increase the work backlog, resulting in a negative goal: g^I. The second interval II, is when the work backlog is within the desirable range, then the sessions should cover the incoming orders, resulting in a zero goal. The third interval III, applies when the predicted work backlog is slightly higher than the desirable interval. In that case, more sessions should be scheduled to decrease the work backlog, resulting in a small positive goal g^{III}. The fourth interval IV, applies when the predicted work backlog is further outside of the desirable range. This results in a bigger additional goal g^{IV}.

The monthly goals g are scaled to weekly goals by dividing it by the number of weeks of the month. The monthly constraints ensure that the monthly goal is met, allowing distribution of the demand over a month. The weekly constraints ensure that the orders and goal is met as well as possible every week, ensuring the sessions are spread over different weeks in a month.

Moreover, we introduce a steering parameter C^s and tolerance constraints to support the model in situations where it can not meet the goal sessions exactly. Firstly, the steering parameter $C^s > 1$ reduces the penalty applied if fewer than the goal sessions are scheduled for a modality with a work backlog that is too low. Moreover, C^s reduces the penalty applied if more than the goal sessions are scheduled for a modality with a work backlog that is too high. Secondly, the tolerance constraints apply an additional penalty if the number of scheduled sessions of a week deviates more from the goal sessions than the tolerance a. These additional constraints are not optimised, they are heuristics, motivated by practical reasons and observations.

Below, the demand constraints are specified and an additional description of the constraints is provided.

$$C^s\Delta_i^{+n} - \Delta_i^{-n} = \bar{q}(i)\left(\sum_{\theta\in\Theta}\mathbb{E}[F_{i,\theta}^n] + g^I\right) \qquad \forall i \in \mathcal{I}^\delta : \mathbb{E}[W_i^n] \tag{D.I.A}$$
$$\qquad\qquad -\sum_{l\in\mathcal{L}_i^I} q(i,l)\left(\sum_{t\in\Theta} x_{i,t,l}^n(+z_{i,t,l}^n)\right) \qquad < w_i^* - C_i^w$$

$$\Delta_i^{+n} - \Delta_i^{-n} = \bar{q}(i)\left(\sum_{\theta\in\Theta}\mathbb{E}[F_{i,\theta}^n]\right) \qquad \forall i \in \mathcal{I}^\delta : w_i^* - C_i^w \le \mathbb{E}[W_i^n] \tag{D.II.A}$$
$$\qquad\qquad -\sum_{l\in\mathcal{L}_i^I} q(i,l)\left(\sum_{t\in\Theta} x_{i,t,l}^n(+z_{i,t,l}^n)\right) \qquad < w_i^* + C_i^w$$

$$\Delta_i^{+n} - C^s\Delta_i^{-n} = \bar{q}(i)\left(\sum_{\theta\in\Theta}\mathbb{E}[F_{i,\theta}^n] + g^{III}\right) \qquad \forall i \in \mathcal{I}^\delta : w_i^* + C_i^w \le \mathbb{E}[W_i^n] \tag{D.III.A}$$
$$\qquad\qquad -\sum_{l\in\mathcal{L}_i^I} q(i,l)\left(\sum_{t\in\Theta} x_{i,t,l}^n(+z_{i,t,l}^n)\right) \qquad < w_i^* + 2C_i^w$$

$$\Delta_i^{+n} - C^s\Delta_i^{-n} = \bar{q}(i)\left(\sum_{\theta\in\Theta}\mathbb{E}[F_{i,\theta}^n] + g^{IV}\right) \qquad \forall i \in \mathcal{I}^\delta : \mathbb{E}[W_i^n] \ge w_i^* + 2C_i^w \tag{D.IV.A}$$
$$\qquad\qquad -\sum_{l\in\mathcal{L}_i^I} q(i,l)\left(\sum_{t\in\Theta} x_{i,t,l}^n(+z_{i,t,l}^n)\right)$$

$$C^s\delta_{i,\theta}^{+n} - \delta_{i,\theta}^{-n} = \bar{q}(i)\left(\mathbb{E}[F_{i,\theta}^n] + \frac{g^I}{|\Theta|}\right) \qquad \forall i \in \mathcal{I}^\delta : \mathbb{E}[W_i^n] < w_i^* - C_i^w, \tag{D.I.B}$$
$$\qquad\qquad -\sum_{l\in\mathcal{L}_i^I} q(i,l)\left(\sum_{t\in\theta} x_{i,t,l}^n(+z_{i,t,l}^n)\right) \qquad (\forall\theta \in \Theta)$$

$$\delta_{i,\theta}^{+n} - \delta_{i,\theta}^{-n} = \bar{q}(i)\mathbb{E}[F_{i,\theta}^n] \qquad \forall i \in \mathcal{I}^\delta : w_i^* - C_i^w \le \mathbb{E}[W_i^n] \tag{D.II.B}$$
$$\qquad\qquad -\sum_{l\in\mathcal{L}_i^I} q(i,l)\left(\sum_{t\in\theta} x_{i,t,l}^n + z_{i,t,l}^n\right) \qquad < w_i^* + C_i^w, (\forall\theta \in \Theta)$$

$$\delta_{i,\theta}^{+n} - C^s\delta_{i,\theta}^{-n} = \bar{q}(i)\left(\mathbb{E}[F_{i,\theta}^n] + \frac{g^{III}}{|\Theta|}\right) \qquad \forall i \in \mathcal{I}^\delta : w_i^* + C_i^w \le \mathbb{E}[W_i^n] \tag{D.III.B}$$
$$\qquad\qquad -\sum_{l\in\mathcal{L}_i^I} q(i,l)\left(\sum_{t\in\theta} x_{i,t,l}^n(+z_{i,t,l}^n)\right) \qquad < w_i^* + 2C_i^w, (\forall\theta \in \Theta)$$

$$\delta_{i,\theta}^{+n} - C^s\delta_{i,\theta}^{-n} = \bar{q}(i)\left(\mathbb{E}[F_{i,\theta}^n] + \frac{g^{IV}}{|\Theta|}\right) \qquad \forall i \in \mathcal{I}^\delta : \mathbb{E}[W_i^n] \ge w_i^* + 2C_i^w, \tag{D.IV.B}$$
$$\qquad\qquad -\sum_{l\in\mathcal{L}_i^I} q(i,l)\left(\sum_{t\in\theta} x_{i,t,l}^n(+z_{i,t,l}^n)\right) \qquad (\forall\theta \in \Theta)$$

$$\delta_{i,\theta}^{-n} + a \le \pi_{i,\theta}^n \qquad \forall i \in \mathcal{I}^\delta : \mathbb{E}[W_i^n] < w_i^* - C_i^w, \tag{D.I.C}$$
$$\qquad (\forall\theta \in \Theta)$$

$$\delta_{i,\theta}^{+n} - \delta_{i,\theta}^{-n} + a \le \pi_{i,\theta}^n \qquad \forall i \in \mathcal{I}^\delta : w_i^* - C_i^w \le \mathbb{E}[W_i^n] \tag{D.II.C}$$
$$\qquad < w_i^* + C_i^w, (\forall\theta \in \Theta)$$

$$\delta_{i,\theta}^{+n} - \delta_{i,\theta}^{-n} + a \le \pi_{i,\theta}^n \qquad \forall i \in \mathcal{I}^\delta : w_i^* + C_i^w \le \mathbb{E}[W_i^n] \tag{D.III.C}$$
$$\qquad < w_i^* + 2C_i^w, (\forall\theta \in \Theta)$$

$$\delta_{i,\theta}^{+n} + a \le \pi_{i,\theta}^n \qquad \forall i \in \mathcal{I}^\delta : \mathbb{E}[W_i^n] \ge w_i^* + 2C_i^w, \tag{D.IV.C}$$
$$\qquad (\forall\theta \in \Theta)$$

If the constraints are used for a SP or the operational offline MP the $(+z)$ component is left out and $\mathbb{E}[F_{i,\theta}^n]$ is replaced for $f_{i,\theta}^{n,m}$ and $\mathbb{E}[W_i^n]$ for $w_i^{n,m}$ in scenario m. Additionally, if the constraints are used on a weekly basis $(\forall\theta \in \Theta)$ is left out.

Description

(D.I) Computes a demand penalty for modality i if more or fewer sessions are scheduled than the expected inflow of orders minus the goal related to interval I. This constraint is only active when the current work backlog is less than $w_i^* - C_i^w$. The steering parameter C^s reduces the penalty if fewer sessions are scheduled.

- D.I.A) Enforced for period Θ, computed as Δ_i^{+n} or Δ_i^{-n}, with goal g^I.

- D.I.B) Enforced for period θ, computed as $\delta_{i,\theta}^{+n}$ or $\delta_{i,\theta}^{-n}$, with goal $\frac{g^I}{|\Theta|}$.

- D.I.C) Computes an additional penalty $\pi_{i,\theta}^n$ for week θ if $\delta_{i,\theta}^{+n}$ is higher than the tolerance a.

(D.II) Computes a demand penalty if more or fewer sessions are scheduled than the expected inflow of orders. This constraint is only active when the current work backlog is contained in the interval $[w_i^* - C_i^w, w_i^* + C_i^w]$.

- D.II.A) Enforced for period Θ, computed as Δ_i^{+n} or Δ_i^{-n}.

- D.II.B) Enforced for period θ, computed as $\delta_{i,\theta}^{+n}$ or $\delta_{i,\theta}^{-n}$.

- D.II.C) Computes an additional penalty $\pi_{i,\theta}^n$ for week θ if $\delta_{i,\theta}^{+n}$ is higher than the tolerance a.

(D.III) Computes a demand penalty if more or fewer sessions are scheduled than the expected inflow of orders plus the goal related to interval III, to decrease the work backlog to the desired work backlog. This constraint is only active when the current work backlog is contained in the interval $[w_i^* + C_i^w, w_i^* + 2C_i^w]$. The steering parameter C^s reduces the penalty if more sessions are scheduled.

- D.III.A) Enforced for period Θ, computed as Δ_i^{+n} or Δ_i^{-n}, with goal g^{III}.

- D.III.B) Enforced for period θ, computed as $\delta_{i,\theta}^{+n}$ or $\delta_{i,\theta}^{-n}$, with goal $\frac{g^{III}}{|\Theta|}$.

- D.III.C) Computes an additional penalty $\pi_{i,\theta}^n$ for week θ if $\delta_{i,\theta}^{+n}$ is higher than the tolerance a.

(D.IV) Computes a demand penalty if more or fewer sessions are scheduled than the expected inflow of orders plus the goal related to interval IV, to decrease the work backlog to the desired work backlog. This constraint is only active when the current work backlog is greater than $w_i^* + 2C_i^w$. The steering parameter C^s reduces the penalty if more sessions are scheduled.

- D.IV.A) Enforced for period Θ, computed as Δ_i^{+n} or Δ_i^{-n}, with goal g^{IV}.

- D.IV.B) Enforced for period θ, computed as $\delta_{i,\theta}^{+n}$ or $\delta_{i,\theta}^{-n}$, with goal $\frac{g^{IV}}{|\Theta|}$.

- D.IV.C) Computes an additional penalty $\pi_{i,\theta}^n$ for week θ if $\delta_{i,\theta}^{+n}$ is higher than the tolerance a.

Fairness constraints

The fairness constraints are constructed to stimulate the model to construct a fair schedule. The fairness constraints are implemented as soft constraints and as a result, do not eliminate any feasible schedules. However, adding this functionality allows us to create realistic schedules, which makes the model more suitable and more comparable to practice. Below, the fairness constraints are specified and an additional description of the constraints is provided.

$$
\begin{aligned}
\rho^n &= \sum_{t \in \Theta} \sum_{j \in \mathcal{J}_t^R \cup \mathcal{J}_t^{\Theta,n}} \sum_{i \in \mathcal{I}_j^J} \sum_{l \in \mathcal{L}_i^I \cup \mathcal{L}_j^J} y_{i,j,t,l}^1 && \text{(F.I)}\\
\eta_{i,j,\theta}^n &\geq k(i) - \sum_{t \in \theta} \sum_{l \in \mathcal{L}_i^I} y_{i,j,t,l}^n && \forall j \in \mathcal{J}, \forall i \in \mathcal{I}_j^J \cup \mathcal{I}^\eta, (\forall \theta \in \Theta) && \text{(F.II)}\\
\kappa_{i,j,\theta}^n &\geq k(i) - \sum_{t \in \theta} \sum_{l \in \mathcal{L}_i^I} y_{i,j,t,l}^n && \forall j \in \mathcal{J}, \forall i \in \mathcal{I}_j^J \cup \mathcal{I}^\kappa, (\forall \theta \in \Theta) && \text{(F.III)}\\
\psi_j^n &\geq 1 - \sum_{t \in \theta} \sum_{i \in \mathcal{I}_|^{\mathcal{J}}} y_{i,j,t,l'} && \forall j \in \mathcal{J}, (\forall \theta \in \Theta) && \text{(F.IV)}\\
\gamma_{i,j,t,\theta}^n &= \sum_{l \in \mathcal{L}_i^I} y_{i,j,t,l}^n - y_{i,j,t+1,l}^n && \forall t \in \theta^o, \forall j \in \mathcal{J}_t^\theta \cap \mathcal{J}_{t+1}^\theta, \forall i \in \mathcal{I}_j^J, && \text{(F.V)}\\
& && (\forall \theta \in \Theta)
\end{aligned}
$$

Remark: if the constraints are used on a weekly basis $(\forall \theta \in \Theta)$ is left out.

Description

(F.I) Computes the penalty ρ if RT j is scheduled during their roster-free scheduling block. This fairness constraint is only included in the MPs because it is only effective when the shifts are being determined.

(F.II) Computes the penalty η if RT j is scheduled less than $k(i)$ times per week on modality i, which requires postgraduate professional education (CT, MRI, US, and angiography).

(F.III) Computes the penalty κ if RT j is scheduled less than $k(i)$ times per week on modality i, which does not require postgraduate professional education.

(F.IV) Computes the penalty ψ if RT j is not scheduled at location l' during period Θ.

(F.V) Computes the penalty γ if an RT is not scheduled for the same modality during the entire day.

5.6.3 Specified two-stage stochastic programs including flexible sessions

Now, that all components of the models have been introduced, the models are constructed. An overview of all the models is given in Table 5.5. To reduce the problem size, the fairness constraints are only used in decisions in which the RT schedule is final: the operational MP and the SPs. Except for ρ, which is considered in the tactical and operational MP because it concerns shift scheduling instead of activity scheduling.

	Level	Problem	Vars	General	Stage specific	Demand	Fairness
AG	Tactical	MP	x, y, z	(a)-(e)	(f.1.T)-(m.1.T)	(D.I-IV.1), (D.I-IV.2)	(F.I)
		SP	x, y	(a)-(e)	(f.2.T)-(o.2.T)	(D.I-IV.1), (D.I-IV.2), (D.I-IV.3)	(F.II)-(F.V)
AG	Operational	MP	x, y	(a)-(e)	(f.1.O)-(o.1.O)	(D.I-IV.1), (D.I-IV.2)	(F.I)-(F.V)
		SP	x, y	(a)-(e)	(f.2.O)-(o.2.O)	(D.I-IV.1), (D.I-IV.2), (D.I-IV.3)	(F.II)-(F.V)
PI	Tactical	MP	x, y, z	(a)-(e)	(f.1.T)-(m.1.T)	(D.I-IV.1), (D.I-IV.2)	(F.I)
		SP	x, y	(a)-(e)	(f.2.T)-(o.2.T)	(D.I-IV.2), (D.I-IV.3)	(F.II)-(F.V)
AG	Operational	MP	x, y	(a)-(e)	(f.1.O)-(o.1.O)	(D.I-IV.1), (D.I-IV.2)	(F.I)-(F.V)
		SP	x, y	(a)-(e)	(f.2.O)-(o.2.O)	(D.I-IV.2), (D.I-IV.3)	(F.II)-(F.V)

Table 5.5: *This table provides an overview of the variables and constraints per two-stage stochastic program including flexible sessions.*

We can show that all the resulting models have complete recourse. Let us look at the tactical two-stage model using AG. Let x^1, y^1, z^1 be a feasible solution to the MP. Then, x^1, y^1, z^1 adhere to the constraints (f.1.T)-(m.1.T). We can then, construct a solution to the SP x^2, y^2 by taking $x^2 = x^1, y^2 = y^1$ and setting $x^2_{i,t,l} = 1$ for any $z^1_{i,t,l} = 1$. This solution x^2, y^2 automatically adheres to constraints (f.2.T)-(m.2.T) due to its construction. Moreover, this solution also adheres to constraints (n.2.T) and (o.2.T), as these are designed to keep continuity between the first and second stages. Furthermore, any fairness and/or demand constraints that are added to the constraint set of the SP are soft. Therefore, they do not influence the polyhedron of feasible solutions of the SP. Conclusively, we can always construct a feasible solution to the SP using the feasible solution of the tactical problem; the tactical model using AG has complete recourse. The difference between AG and PI lies in demand constraints, and therefore these arguments hold also for PI. Following similar reasoning, it can easily be shown that the operational models also have complete recourse.

In the remainder of this chapter, all models are discussed in more detail and their objectives are defined. The objectives consist of two parts. The first part defines the cost of the MP: the session costs, feasibility penalties, demand penalties and in the case of the operational MP fairness penalties. The second part defines the costs of the subproblem for each scenario multiplied by its probability: the session costs, feasibility penalties, demand penalties and fairness penalties.

The coefficients of the objectives are the same for each model. The session costs x^n are weighed using the coefficient C^x and the demand penalties Δ^n and δ^n by C^Δ, and π^n by C^π. Moreover, the feasibility penalties $\alpha^{+n}, \alpha^{-n}, \beta^n$, and ξ^n are weighed by $C^{\alpha^+}, C^{\alpha^-}, C^\beta$, and C^ξ, respectively. Lastly, the fairness penalties $\rho^n, \eta^n, \kappa^n, \psi^n$, and γ^n are weighed by $C^\rho, C^\eta, C^\kappa, C^\psi$, and C^γ, respectively. Note, that we chose the coefficients to be independent of stage n.

Aggregated model

The objective of the tactical problem is:

$$
\min \sum_{x, \Delta, \alpha, \rho} \left(C_x x^1 + C_\Delta \left(\Delta^{+1} + \Delta^{-1} \right) + C_{\alpha^+} \alpha^{+1} + C_{\alpha^-} \alpha^{-1} + C_\rho \rho^1 \right) \tag{34}
$$
$$
+ \sum_{m=1}^{M} \phi_m \cdot \left(\sum_{x, \Delta, \delta, \pi, \beta, \gamma, \eta, \kappa, \psi} \left(C_x x^{2,m} + C_\Delta \left(\Delta^{+2,m} + \Delta^{-2,m} + \delta^{+2,m} + \delta^{-2,m} \right) + C_\pi \pi^{2,m} + C_\beta \beta^{2,m} \right. \right.
$$
$$
\left. \left. + C_\gamma \gamma^{2,m} + C_\eta \eta^{2,m} + C_\kappa \kappa^{2,m} + C_\psi \psi^{2,m} \right) \right),
$$

where we write all variables without subscripts to alleviate notation. Note that the SP contains monthly demand penalties. The tactical problem is defined using the general constraints, the tactical master constraints for the MP, and the general constraints and the tactical SP constraints for the SP. Additionally, the demand constraints (D.I-IV.A) and (D.I-IV.B) are added to the MP and (D.I-IV.A), (D.I-IV.B), and (D.I-IV.C) without z to the SP. The fairness constraint (F.I) is added to the MP and (F.II)-(F.V) are added to the SP.

The objective of the operational problem is:

$$
\min_{x^2,\Delta^2,\alpha^2,\beta^2,\xi^2,\rho^2,\gamma^2,\eta^2,\kappa^2,\psi^2} \left(C_x x^2 + C_\Delta \left(\Delta^{+2} + \Delta^{-2}\right) + C_{\alpha^+}\alpha_j^{+2} + C_{\alpha^-}\alpha_j^{-2} + C_\beta \beta^2 + C_\xi \left(\xi^{+2} + \xi^{-2}\right) \right)
$$
(35)
$$
+ C_\rho \rho^2 + C_\gamma \gamma^2 + C_\eta \eta^2 + C_\kappa \kappa^2 + C_\psi \psi^2 \Big)
$$
$$
+ \sum_{m=1}^{M} \phi_m \cdot \left(\sum_{x^3,\Delta^3,\delta^3,\pi^3,\beta^3,\gamma^3,\eta^3,\kappa^3,\psi^3} \left(C_x x^{3,m} + C_\Delta \left(\Delta^{+3,m} + \Delta^{-3,m} + \delta^{+3,m} + \delta^{-3,m}\right) + C_\pi \pi^{3,m} \right. \right.
$$
$$
\left. \left. + C_\beta \beta^{3,m} + C_\gamma \gamma^{3,m} + C_\eta \eta^{3,m} + C_\kappa \kappa^{3,m} + C_\psi \psi^{3,m} \right) \right).
$$

Note that the feasibility penalties are included in the MP and SP, and the additional feasibility penalty ξ as calculated by constraints (n.2.O) is included. The operational problem is defined using the general constraints, the operational master constraints for the MP and the general constraints and the operational SP constraints for the SP. Additionally, the demand constraints (D.I-IV.A) and (D.I-IV.B) without z are added to the MP and (D.I-IV.A), (D.I-IV.B), and (D.I-IV.C) without z to the SP. The fairness constraints (F.I)-(F.V) are added to the MP and (F.II)-(F.V) to the SP.

Per-stage independent aggregation model

The objective of the tactical problem is:

$$
\min_{x^1,\Delta^1,\alpha^1,\rho^1} \left(C_x x^1 + C_\Delta \left(\Delta^{+1} + \Delta^{-1}\right) + C_{\alpha^+}\alpha^{+1} + C_{\alpha^-}\alpha^{-1} + C_\rho \rho^1 \right)
$$
(36)
$$
+ \sum_{m=1}^{M} \phi_m \cdot \left(\sum_{x^2,\delta^2,\pi^2,\beta^2,\gamma^2,\eta^2,\kappa^2,\psi^2} \left(C_x x^{2,m} + C_\Delta \left(\delta^{+2,m} + \delta^{-2,m}\right) + C_\pi \pi^{2,m} + C_\beta \beta^{2,m} + C_\gamma \gamma^{2,m} \right. \right.
$$
$$
\left. \left. + C_\eta \eta^{2,m} + C_\kappa \kappa^{2,m} + C_\psi \psi^{2,m} \right) \right).
$$

Note, that the monthly demand penalties are not included in the SP. The tactical problem is defined using the general constraints, the tactical master constraints for the MP and the general constraints and the tactical SP constraints for the SP. Additionally, the demand constraints (D.I-IV.A) and (D.I-IV.B) are added to the MP and (D.I-IV.B), and (D.I-IV.C) without z to the SP. The fairness constraint (F.I) is added to the MP and (F.II)-(F.V) to the SP.

The objective of the operational problem is:

$$
\min_{x^2,\Delta^2,\alpha^2,\beta^2,\xi^2,\rho^2,\gamma^2,\eta^2,\kappa^2,\psi^2} \left(C_x x^2 + C_\Delta \left(\Delta^{+2} + \Delta^{-2}\right) + C_{\alpha^+}\alpha_j^{+2} + C_{\alpha^-}\alpha_j^{-2} + C_\beta \beta^2 + C_\xi \left(\xi^{+2} + \xi^{-2}\right) \right)
$$
(37)
$$
+ C_\rho \rho^2 + C_\gamma \gamma^2 + C_\eta \eta^2 + C_\kappa \kappa^2 + C_\psi \psi^2 \Big)
$$
$$
+ \sum_{m=1}^{M} \phi_m \cdot \left(\sum_{x^3,\delta^3,\pi^3,\beta^2,\gamma^3,\eta^3,\kappa^3,\psi^3} \left(C_x x^{3,m} + C_\Delta \left(\delta^{+3,m} + \delta^{-3,m}\right) + C_\pi \pi^{3,m} + C_\beta \beta^{3,m} + C_\gamma \gamma^{3,m} + C_\eta \eta^{3,m} \right. \right.
$$
$$
\left. \left. + C_\kappa \kappa^{3,m} + C_\psi \psi^{3,m} \right) \right).
$$

Note that the feasibility penalties are included in the MP and SP, and the additional feasibility penalty ξ as calculated by constraints (n.2.O) is included. The operational problem is defined using the general constraints, the operational master constraints for the MP and the general constraints and the operational SP constraints for the SP. Additionally, the demand constraints (D.I-IV.A) and (D.I-IV.B) without z are added to the MP and (D.I-IV.A), (D.I-IV.B), and (D.I-IV.C) without z to the SP. The fairness constraint (F.I)-(F.V) are added to the MP and (F.II)-(F.V) to the SP.

Chapter 6

Numerical experiments

We conducted extensive numerical experiments to evaluate the performance of the proposed flexible scheduling framework for a monthly and weekly timeline, using AG and PI and various scenario methods. We used SAA to generate scenarios and used the integer L-shaped method to find solutions. The performance of the models is analyzed using a test case and a case study regarding VMC. The test case was designed to mimic the most important characteristics of a radiology department. All tests are coded in Python and deployed with Gurobi 10.1 on a server with 16 cores and 64 GB of memory [41]. The multiprocessing library was used to allow parallel computations where necessary. The code used in this project is available on Github[1]. While the datasets concerning the VieCuri case are confidential, the datasets used for the test case are provided for evaluation purposes.

6.1 Scenario generation

We use scenarios to capture the uncertainty in our models. The tactical decision involves the variables: $\tilde{S}$, D, and the operational decision $\tilde{S}$ and $\tilde{D}$.

6.1.1 Demand

As explained in Section 2.2 the demand is determined by the current work backlog and the incoming orders. An overview of the inflow of orders is depicted in 6.1. Scenarios regarding the inflow of orders and work backlog can be generated using the distributions of each source.
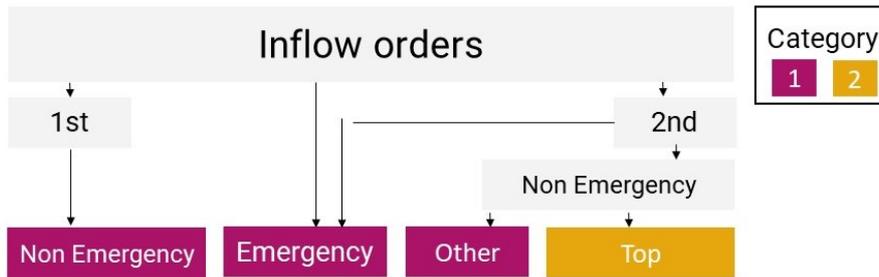
Figure 6.1: This figure visualizes the different sources of inflow of orders of a radiology department.

The four different sources can be divided into two main categories. The first consists of all emergency orders, the first line, and not frequently referring second line departments. This inflow is more or less constant over the year or hard to predict. The second category is the elective orders generated by the top referring hospital departments $\mathcal{R}$. Their referrals are strongly related to the number of outpatient consult sessions they do. Each appointment during a session is in essence a Bernoulli trial, where a success is a referral for a modality i. The scheduled number of sessions, which is almost identical to the realised sessions, of each referring department

---

[1]https://github.com/Ansvlooswijk/thesispublic

is unknown until the operational offline stage. The plan that describes the number of expected sessions for a referral department is available at the tactical decision.

Therefore, we need three distributions to generate scenarios: the distribution of the first category orders F_i^1, the distribution $F_{r,i}^2$ of the number of orders for modality i per outpatient session of each top referring department r, and the distribution of the deviation of the realised number of outpatient sessions from the plan F_r^3 for each referring department. This last distribution F_r^3 enables us to sample the number of outpatient sessions by summing the planned number of outpatient sessions $A_r^{2,\theta}$ in week θ with a sample from F^3. We define the target access time to be two weeks, therefore, we suppose that elective orders should be executed two weeks after they are received. As a result, the relevant elective orders generated by the top referring departments in week θ can be predicted using the number of outpatient sessions of two weeks earlier. The number of orders f of the week θ for scenarios in the tactical decisions for modality i is sampled using:

$$f_i^{2,\theta} = F_i^1(k=1) + \sum_{r \in \mathcal{R}} \left(F_{r,i}^2(k = A_r^{2,\theta-2} + F^{3,r}(k=1)) \right), \tag{1}$$

where k is the number of samples from the respective distribution.

The weekly orders for the operational offline decisions are generated similarly, only now the scheduled outpatient sessions $A_r^{3,\theta}$ of the referral departments are used to determine how many samples are drawn from $F_{r,i}^2$, providing the operational offline decision with more accurate information. The number of orders f of week θ for scenarios in the operational offline decision for modality i is sampled using:

$$f_i^{3,\theta} = F_i^1(k=1) + \sum_{r \in \mathcal{R}} \left(F_{r,i}^2(k = A_r^{3,\theta-2}) \right), \tag{2}$$

where k is the number of samples from the respective distribution.

The future work backlog for modality i is sampled by summing the current work backlog w_i^c with the incoming orders in the weeks within the horizon and subtracting the outflow of orders within the horizon. The horizon is the time between the time the decision is made $\theta = c$ and the start of the period that is being scheduled. The session schedule and the average production (orders per session) are used to predict the outflow p_i per week. The session schedule can be determined using the fixed sessions and currently assigned flexible sessions. The inflow f_i^θ is predicted by sampling orders for each week for each source. The elective orders from the top referring departments are not shifted by two weeks as before, because the work backlog monitors all received orders, not when they should be executed. The scheduled outpatient sessions of the first four weeks of the horizon are known. The planned number of outpatient sessions and samples from F^3 are used to sample the incoming orders for the remaining weeks of the horizon. In line with this logic, the work backlog at the start of a planning period θ' is found using:

$$w_i^{\theta'} = w_i^c + \sum_{\theta=c}^{\theta'} f_i^\theta - p_i^\theta \tag{3}$$

6.1.2 Supply
We are modelling the tactical decision and the operational offline decision. In these stages, s is known as explained in earlier sections. Therefore, the RTs can be rostered according to the most recent s. We use a rolling horizon, meaning that s is updated in the operational offline decision. If the updated availability leads to an infeasible operational offline MP the violated constraint can be relaxed. Namely, in practice the model is always feasible because the RT-schedulers can find many different practical solutions, such as asking someone who is not working to take an extra shift, making an exception in the number of needed RTs to work a modality, partly closing a certain session et cetera. We decided not to model all these heuristic rules, because these are not relevant in the tactical and operational offline stage and increase the dimensionality of the problem significantly. We implemented the feasibility-relax functionality offered by Gurobi [41]. This was initialised to minimise the number of constraints by which the solution violates the original model. Subsequently, these violated constraints were saved to be inspected later.

The remaining deviations from s were captured in $\tilde{S}$, with $\mathbb{E}[\tilde{S}] = 0$. We do not model any scenarios regarding $\tilde{S}$ because most changes captured in $\tilde{S}$ happen in the operational online stage and therefore are outside the scope of our model.

6.2 Solution method

We use the L-shaped method to solve the full problem. Since our problem has complete recourse, we only have to find optimality cuts, as defined in Equation (7). We can find a lower bound L of the second stage by summing the costs of the sessions of the demand-based scheduled modalities needed to meet the demand and the work backlog goal. If this goal exceeds the maximum capacity, we also sum the minimal incurred penalty for not meeting the demand and goal.

Our first-stage decision variables x, y, and z are binary, which would allow us to apply the integer L-shaped method. The variable z does not appear in the second stage, however, it does influence the optimal solution of the second stage. The set of continuous penalty variables P is defined by x, y, z. We propose that we can still use the integer L-shaped method if we do not apply cuts to the penalty variables, but only to x, z, and y. The continuous penalty variables do not affect the second stage. Shi et al. [30] showed, that the cuts do not need to be applied to variables that do not influence the second stage for the algorithm to converge. They showed that the integer L-shaped method converges faster if cuts are applied to strictly the subset of integer variables that influence the second stage objective [30]. Consequently, the integer L-shaped problem is suitable to solve our model.

The problem structure, evaluating multiple scenarios is visualised in Figure 6.2. We take advantage of this structure by evaluating each scenario in parallel using the multiprocessing library, reducing the computation time. Moreover, when using the PI aggregation method, the SPs can be evaluated per week instead of per scenario. This significantly decreases the size of the SPs that need to be solved.
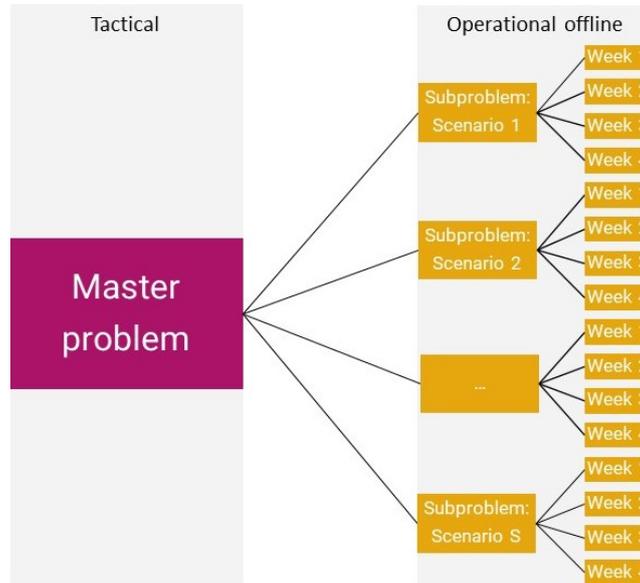

Figure 6.2: This figure visualizes the problem structure. The pink block represents the tactical decision: the MP. The yellow blocks represent the operational offline decisions regarding each scenario: the SPs. The SPs can be split up into weeks when using the PI method.

Given the above description, we assemble the components into Algorithm 1 [30, 24]. The current best objective value of the master is referred to as $\hat{b}$, the objective value excluding the second stage component is referred to as $\mathcal{O}$.

Algorithm 1 Integer L-shaped method

Input	: Master problem, sub problem, scenarios $\mathcal{M}$
Algorithm parameters:	C, Aggregation method: PI or AG
Initialize	: $\bar{B} = \infty$, $\zeta = -\infty$, Compute L, Start branch and bound on MP ignoring ζ to create a single pendant node without any optimality cut. Append this node as the root node to the branch-and-bound tree.

Select a pendant node in the branching tree.
if <u>none exists</u> **then**
     Stop
else
     Solve the current MP, find $\hat{x}, \hat{y}, \hat{z}, \hat{\zeta}$
     if <u>the current problem has no feasible solution</u> **then**
         Fathom the current node.
     else
         if <u>$\hat{x}, \hat{y}, \hat{z} \in \{0,1\}$</u> **then**
             if <u>PI</u> **then**
                 Solve weekly SPs in parallel, find $Q^{m,\theta}$
                 $Q = \sum_{m \in \mathcal{M}} \sum_{\theta \in \Theta} Q^{m,\theta}$
             else if <u>AG</u> **then**
                 Solve SPs in parallel, find Q^m
                 $Q = \sum_{m \in \mathcal{M}} Q^m$
             $b = \mathcal{O}(\hat{x}, \hat{y}, \hat{z}) + Q$
             if <u>$b < \hat{b}$</u> **then**
                 Update $\hat{b} = b$ and record current $\hat{x}, \hat{y}, \hat{z}$ as the best solution so far
                 if <u>$\hat{\zeta} < Q$</u> **then**
                     Add cut to all pendant nodes in the tree:

$$\zeta \geq (Q - L) \left(\sum_{i \in V_{\hat{x}}^k} x_i^0 + \sum_{i \in V_{\hat{z}}^k} z_i^0 + \sum_{i \in V_{\hat{y}}^k} y_i^0 - \sum_{i \notin V_{\hat{x}}^k} x_i^0 - \sum_{i \notin V_{\hat{z}}^k} z_i^0 - \sum_{i \notin V_{\hat{y}}^k} y_i^0 \right.$$
$$\left. - |V_x^k| - |V_z^k| - |V_y^k| + 1 \right) + L$$

6.3 Experimental setup

The models are tested using different timelines, scenario generation methods and aggregation methods. The models are evaluated over a RH including multiple tactical and operational decisions. The work backlog is initialised using historical or sampled data. During the first two months of the horizon, only tactical decisions are taken. The work backlog is updated during these months with historical or sampled sessions, production and orders. Starting from the third month of the horizon an operational offline is taken. The sessions scheduled by the model regarding this month and historical or sampled data regarding the production and orders can then be used to update the work backlog after this month has taken place.

The performance can be measured using the KPIs established in Section 2.4: the number of modalities with a work backlog within the desired interval, the mean deviation of the work backlog from the desired work backlog, the maximum deviation of the work backlog from the desired work backlog, the scheduled FTE, the roster fairness score, the scheduled sessions and the difference in scheduled sessions between the tactical and operational decision. We use Algorithm 2 to evaluate the timeline:

Algorithm 2 Evaluate model over RH

Input	: Master problem, SP, distributions to sample scenarios, and for the horizon: months to evaluate, initial work backlog w^i, order data per week, production data per week, availability RTs per week, sessions referring departments per week, sampled/historical sessions first three months

Algorithm parameters: C, Aggregation method, timeline, scenario generation method

Initialize : $w^0 = w^i$, K using sampled/historical sessions

for <u>month h in horizon</u> **do**
   Update the Θ^T to be the month 3 months in advance and Θ^O to be one month in advance
   if $h < 3$ **then**
      for <u>for θ in Θ</u> **do**
         Update w^θ using sampled/historical sessions, and order and production data
   else
      Operational offline decision:
      if <u>week-based</u> **then**
         for <u>θ in Θ^O</u> **do**
            Sample M scenarios m using w^θ, and currently scheduled sessions K
            Solve the operational 2-stage model for week θ using Algorithm 1 given $x^{1,\theta}, y^{1,\theta}$
            Update sessions $K[\theta]$ using $x^{2,\theta}$
            Calculate KPIs and w^θ using order and production data

      if <u>month-based</u> **then**
         Sample M scenarios m using w^θ, and currently scheduled sessions K
         Solve the operational 2-stage model for month Θ^O using Algorithm 1 given $x^{1,\Theta^O}, y^{1,\Theta^O}$
         for <u>for θ in Θ^O</u> **do**
            Update $K[\theta]$ using $x^{2,\theta}$
            Calculate KPIs and w^θ using order and production data
   Tactical decision:
      Sample M scenarios m using w^θ, and currently scheduled sessions K
      Solve the tactical 2-stage model for month Θ^T using Algorithm 1, save $x^{1,\Theta^T}, y^{1,\Theta^T}$
      Save sessions per modality per week in $K[\theta]$

6.4 Experiments on test set

The test set entailed a radiology department split over two locations. Location 1 has two CT and MRI scanners and an X-ray, or Bucky room and location 2 only has a Bucky room. One RT can also be scheduled on the administrative task: protocol. The department has 5 employees who can all work Bucky and have one or two additional skills: two RTs are qualified for the CT, two RTs are qualified for the MRI, and one RT is qualified for both. The employees work full-time and are always available. We did not assign any roster-free days to keep the test set simple.

The department is only open two days a week. Each day has two scheduling blocks: morning and afternoon. A scheduling period/month is four weeks. The Bucky must always be operative at both locations, and at least one CT and MRI scanner must be operative in the morning. The Bucky at location 2, the CT and MRI are operated by one RT, and the Bucky at location 1 is operated by two RTs. At least one protocol session should be scheduled monthly and a maximum of two weekly. Therefore, three employees are always operating the Bucky, leaving two employees to work the CT, MRI, or protocol. The maximum capacity of the CT and MRI is a total of 8 sessions per week and 31 per month. We determined, arbitrarily, that the desired work backlog for both CT and MRI is 12 sessions, with C^w set initially at 4. The values for $u(i)$ and $k(i)$ where chosen for MRI and CT as $u(i) = 0.4$ and $k(i) = 1$.

Entity	Distribution
CT orders resulting from one session of referring departments	$\sim \mathcal{N}(2.5, \frac{1}{2})$
CT emergency orders	$\sim \mathcal{N}(20, 2)$
MRI orders resulting from one session of referring departments	$\sim \mathcal{N}(7, \frac{1}{2})$
MRI emergency orders	$\sim \mathcal{N}(5, 2)$
Realized production	$\sim \mathcal{N}(15, \frac{1}{2})$
Deviation rostered from planned sessions referring departments	$\sim \mathcal{N}(0, 1)$

Table 6.1: This table includes the distributions used to generate datasets for the test case.

The test case demand was generated based on distributions outlined in Table 6.1. We chose to use normal distributions because the distributions of referrals and orders of VMC resembled normal distributions the most. Any samples with negative values regarding the number of orders were discarded. Each referring department had ten scheduled outpatient sessions weekly, which was reduced to six sessions during holiday weeks $(1, 7, 18 - 19, 28 - 34, 40, 52)$. The distributions were designed to ensure that the mean orders, calculated as $(10(2.5 + 7) + 5 + 20)/15 = \frac{120}{15} = 8$, could be accommodated by the total capacity of 8 sessions for MRI and CT combined. We chose to construct the demand and supply to match instead of having an excessive supply because in practice radiology departments rarely have excessive supply.

A sample of the orders is depicted in Figure 6.3. The order pattern resembles an actual radiology department compared to the VMC order pattern discussed later in Section 6.5.
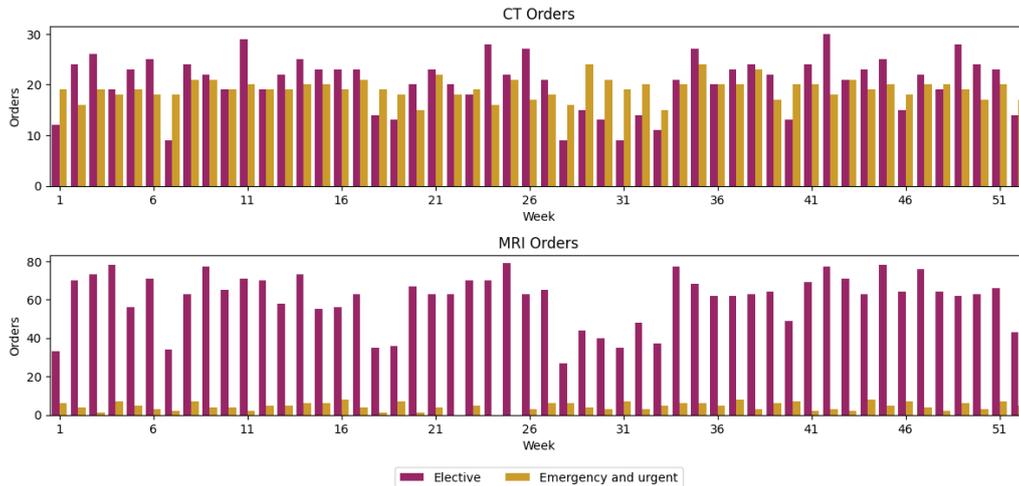
Figure 6.3: This figure depicts the orders generated for CT and MRI over one year using the distributions as defined in Table 6.1.

6.4.1 Model verification and validation

In this section, we describe how the model was tuned, analysed, and validated. First, we briefly discuss the tuning process and present the results of the sensitivity analysis of the tuned model. These results help to gain insight into how variations in demand affect the model's decisions. Second, we discuss a two-stage example of the model to highlight that the model is implemented correctly.

Tuning and sensitivity analysis

The goal of sensitivity analysis was to identify the coefficients that promote desirable behaviour to enhance overall outcomes. It is important to balance the coefficients well to achieve the right balance of meeting the demand while simultaneously minimising the fairness and feasibility penalties. We initialised the session costs and fairness coefficients in the same order of magnitude: $C_x = 1$ and the fairness coefficients between 0 and 1, multiplied by the number of sessions (4), so between 0-4. The values of the individual fairness constraints were determined according to the outcome of an AHP analysis conducted for the case study, which is discussed in Section 6.5. The feasibility penalties were set relatively high at $C = 40$ because they should not be violated unless the model is otherwise infeasible. Subsequently, we adapted the demand penalty coefficients through an iterative process of trial and error until desirable behaviour was observed. Moreover, we tested different values

for g. If g was set too high, the model was overcompensating, while if set too low, the model did not correct the work backlog enough. We established: $g^I = -6, g^{III} = 3, g^{IV} = 6$.

The sensitivity analysis on a sub problem using PI and AG, after tuning, is depicted in Figure 6.4. This figure visualises in A) and C) what decisions the model makes in different situations. Particularly, the model must schedule at least 2 additional sessions and, at most, 4 additional sessions for CT and MRI combined. The figures in A) and C) show the demand in sessions on the x-axis and the additionally scheduled sessions on the y-axis. Moreover, B) and D) visualise the fairness penalty caused by the chosen schedule, calculated as

$$\sum_{\rho,\gamma,\eta,\kappa,\psi} \left(C_\rho \rho^2 + C_\gamma \gamma^2 + C_\eta \eta^2 + C_\kappa \kappa^2 + C_\psi \psi^2 \right). \tag{4}$$

Recall that the fairness penalty indicates how well the fairness constraints are obliged and gives an indication of how realistic this schedule is. Namely, if the fairness penalty is too high, such a schedule would not be adopted in practice and therefore it is important to take the fairness of the schedule into account.

We formulated four cases using the work backlog intervals as introduced in Section 5.6.3. The first case entailed a work backlog for CT in interval I and for MRI in II. The second case entailed an equal work backlog for both modalities in interval II. In the third case, CT had a work backlog in interval III and MRI in II. Lastly, in the fourth case, the work backlog for CT was in interval IV and MRI in II. PI and AG were subsequently tested for two order patterns: one with an equally increasing number of weekly orders and one with a contrasting number of weekly orders for CT and MRI. The combination of cases and order patterns provides an extensive overview of situations that may occur.

The results of AG and PI were the same across all cases and order patterns. In the equal orders experiment in 6.4A) we expect that as the demand increases, more additional sessions will be scheduled. The distribution of the additional sessions depends on the case. In the first case fewer sessions should be scheduled for CT, which has a lower work backlog than desired. In the second case, equal sessions for CT and MRI should be scheduled, and in the third and fourth cases, more sessions should be scheduled for CT, which has a higher work backlog. We see that in Figure 6.4.A) indeed, this behaviour is exhibited. Therefore, we see that the model makes sensible and expected decisions.

In Figure 6.4.C), we see that the number of scheduled sessions for CT and MRI intersect at different points. In the first case, 4 additional sessions are scheduled for MRI until 6 ordered sessions; at 10 ordered sessions, 4 additional sessions are scheduled for CT. This is in line with the goal of ordered sessions minus 4 sessions with a work backlog in interval I. The additional sessions are divided equally in the middle of the range in the second case. Moreover, the scheduled sessions in the third and fourth cases intersect on the left side of the figure, in line with CT having a higher work backlog and thus a higher goal.

In Figure 6.4.B) and D) we see that when the additional sessions are not divided equally between CT and MRI, this incurs a fairness penalty. Namely, in this situation, constraints (F.II) and (F.III) cannot be fulfilled. If the sessions are divided more equally over CT and MRI, we see that the fairness penalty is zero, indicating a fair schedule. This shows that the model is correctly tuned because the fairness constraints are obliged if possible, but answering to the demand is favoured.

In a subsequent analysis that evaluated the model's performance within a two-stage framework, the coefficient C^s was adjusted to promote an improved allocation of sessions between modalities: if more than the available additional sessions were needed, they were divided more equally between MRI of CT. Moreover, the incorporation of an additional penalty, specified in constraint (D.I-IV.C) with $a = 2$, further improved the division of additional sessions over modalities. The coefficient C^π was determined accordingly through trial and error.
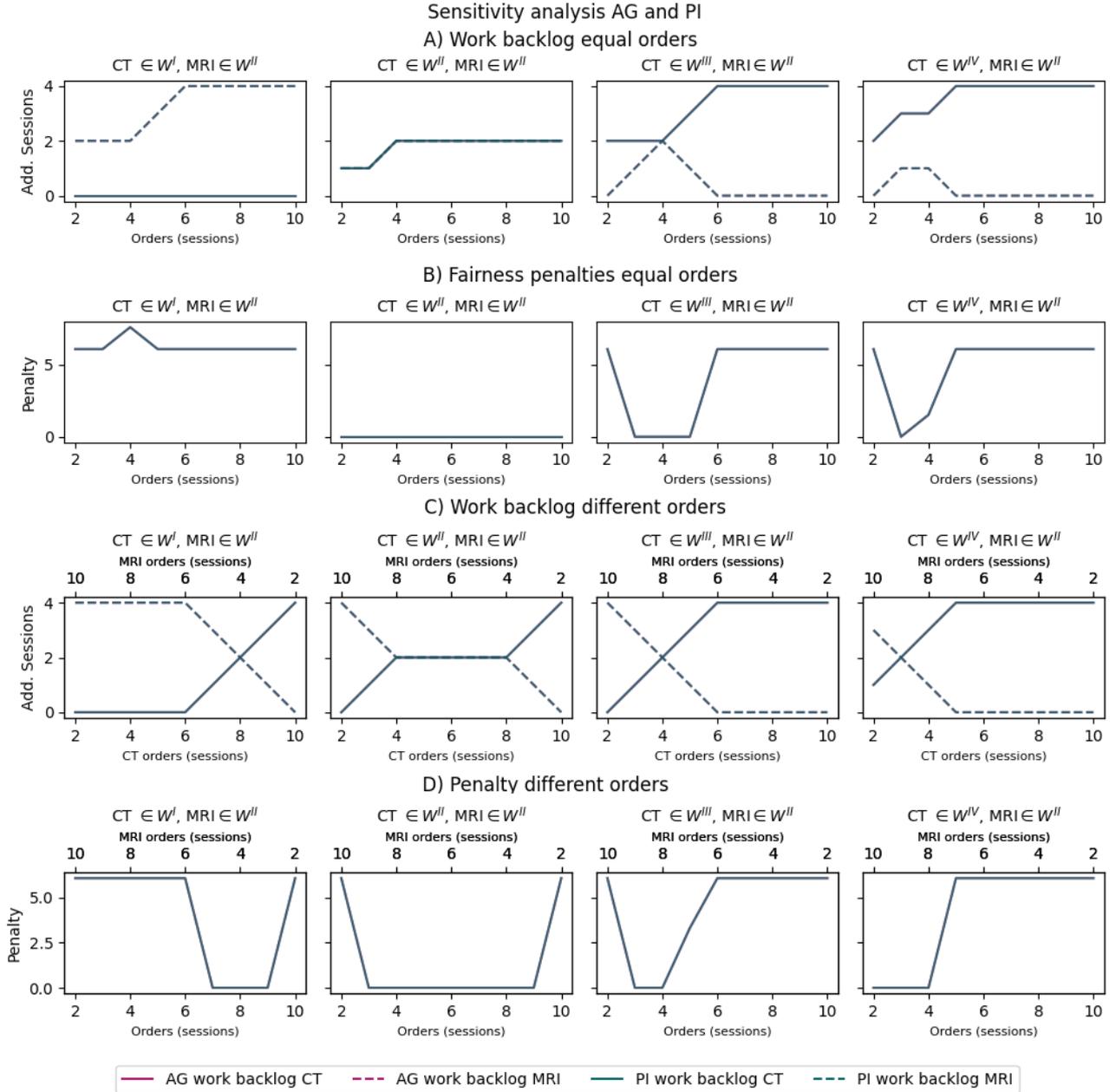
Figure 6.4: This figure shows the results of the sensitivity analysis of AG and PI. It depicts the additionally scheduled sessions per week for MRI and CT and the fairness penalties in four different work backlog cases and two different order patterns: an equally increasing number of weekly orders and a contrasting number of weekly orders.

Two-stage example

In Figure 6.5, the schedule resulting from a tactical decision based on three scenarios is illustrated. The AG and PI method gave the same results, with some arbitrary switches in RTs. The expected work backlog, expected orders, and scenarios for the tactical decision are given in Table 6.2. We initialised the work backlog for CT in interval IV and MRI in interval II. The scenarios were generated as outlined in Section 6.1 using the distribution as given in Table 6.1, where the σ values of all distributions, except the production, were doubled to generate more distinguished scenarios to illustrate the workings of the model better. Each horizontal bar in Figure 6.5 represents the schedule of a modality at a certain location over time. The sessions are indicated by a coloured block, where each unique colour represents an RT working that modality during that scheduling block. Multiple colours in one scheduling block indicate that multiple RTs are scheduled on that modality. On the right side of the figure, the number of fixed and regular sessions is printed. The flexible sessions are hatched.
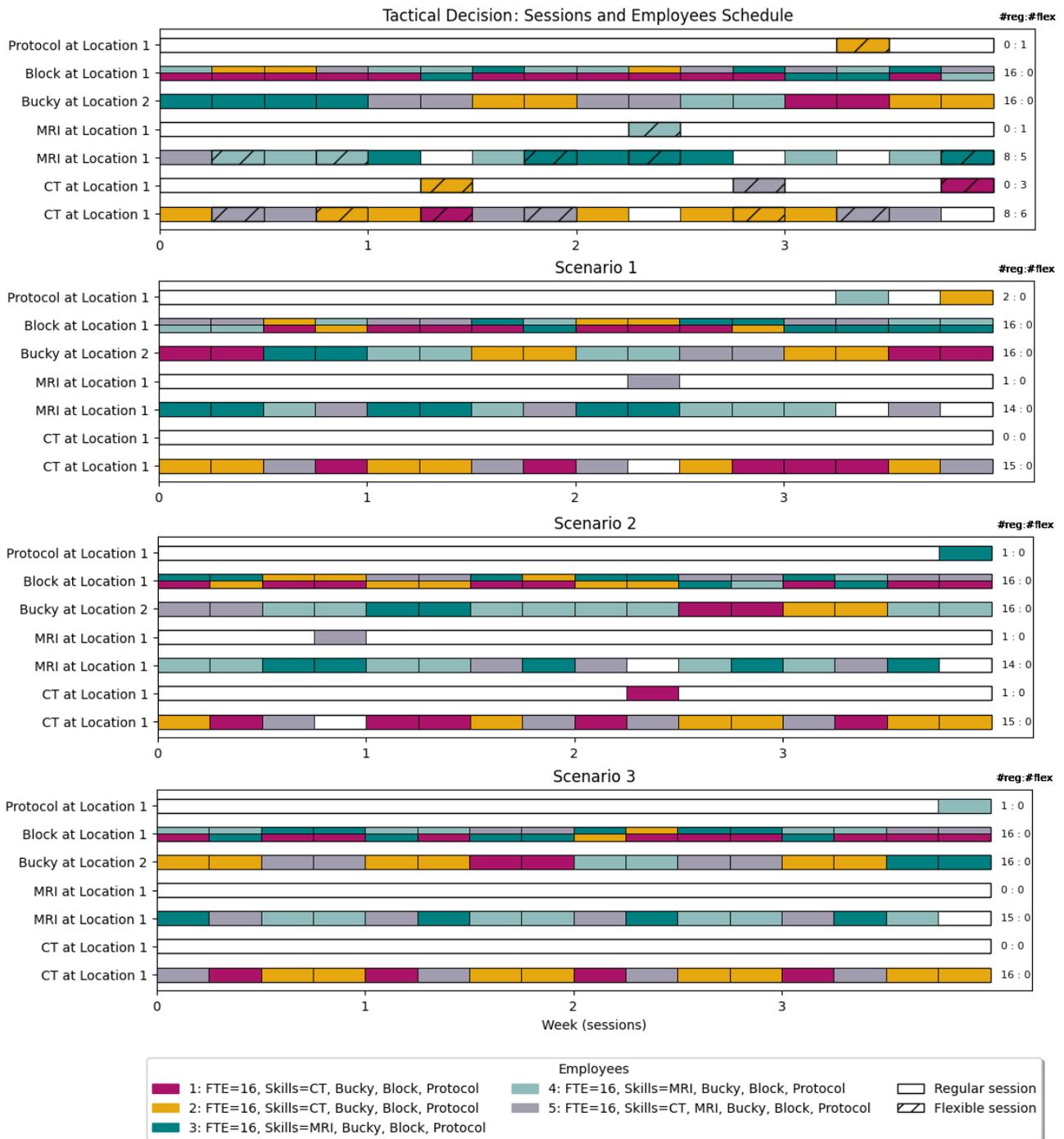
Figure 6.5: This figure visualizes the resulting schedule of a tactical decision based on three scenarios.

Modality		MP (mean)	Scenario 1	Scenario 2	Scenario 3
CT	Work backlog	22,	25	23	24
	Orders	$(3, 3, 3, 2)$	$(3, 3, 2, 2)$	$(2, 3, 4, 2)$	$(3, 3, 3, 2)$
MRI	Work backlog	15	13	11	14
	Orders	$(5, 5, 5, 3)$	$(5, 5, 5, 2)$	$(6, 5, 5, 3)$	$(6, 4, 5, 3)$

Table 6.2: This table includes the means and scenarios used for the tactical decision in the two-stage example. The orders are given per week, for week 1-4 as $(1, 2, 3, 4)$.

First, we observe in Figure 6.5 that the correct number of RTs is scheduled to operate the modalities. The assignment of specific RTs with the same qualifications is arbitrary in the test case because the RTs are always available. Moreover, the obligatory sessions for Bucky, CT, and MRI and the monthly minimum of protocol are scheduled. The fairness constraints were applied to the tactical subproblems. We see that the number of employees scheduled on Bucky at location 2 is maximal, minimising ψ. Furthermore, in many cases an employee is scheduled on one modality during the entire day, minimising γ. All employees are scheduled on Bucky and depending on their qualifications on CT, MRI, or both, at least one session per week in all scenarios. Therefore, κ and η are also minimised. We did not include roster-free days in employees' calendars. Therefore, ρ cannot be measured.

The tactical decision resulted in a total of 17 CT sessions, compared to 14 MRI sessions. All non-obligatory sessions for both MRI and CT, as well as for the protocol, are flexible, allowing for adjustments as needed. In the scenarios, the number of scheduled MRI sessions per week is equal to the number of orders or 1-2 sessions fewer. In contrast, the number of scheduled CT sessions is 1-2 sessions more than the number of ordered sessions, matching the weekly goal, which was $\frac{g^{IV}}{|\Theta|} = \frac{6}{4}$. Furthermore, the mandatory protocol session is consistently scheduled during the last week, which typically has fewer orders, as seen in Table 6.2.

The operational decision taken for this same month is visualised in Figure 6.6 for PI and AG separately because the results were slightly different. As all sessions are fixed in the operational offline decision, the schedules of the scenarios are the same and, therefore, omitted. The expected work backlog and orders and scenarios for the operational decision are given in Table 6.3.
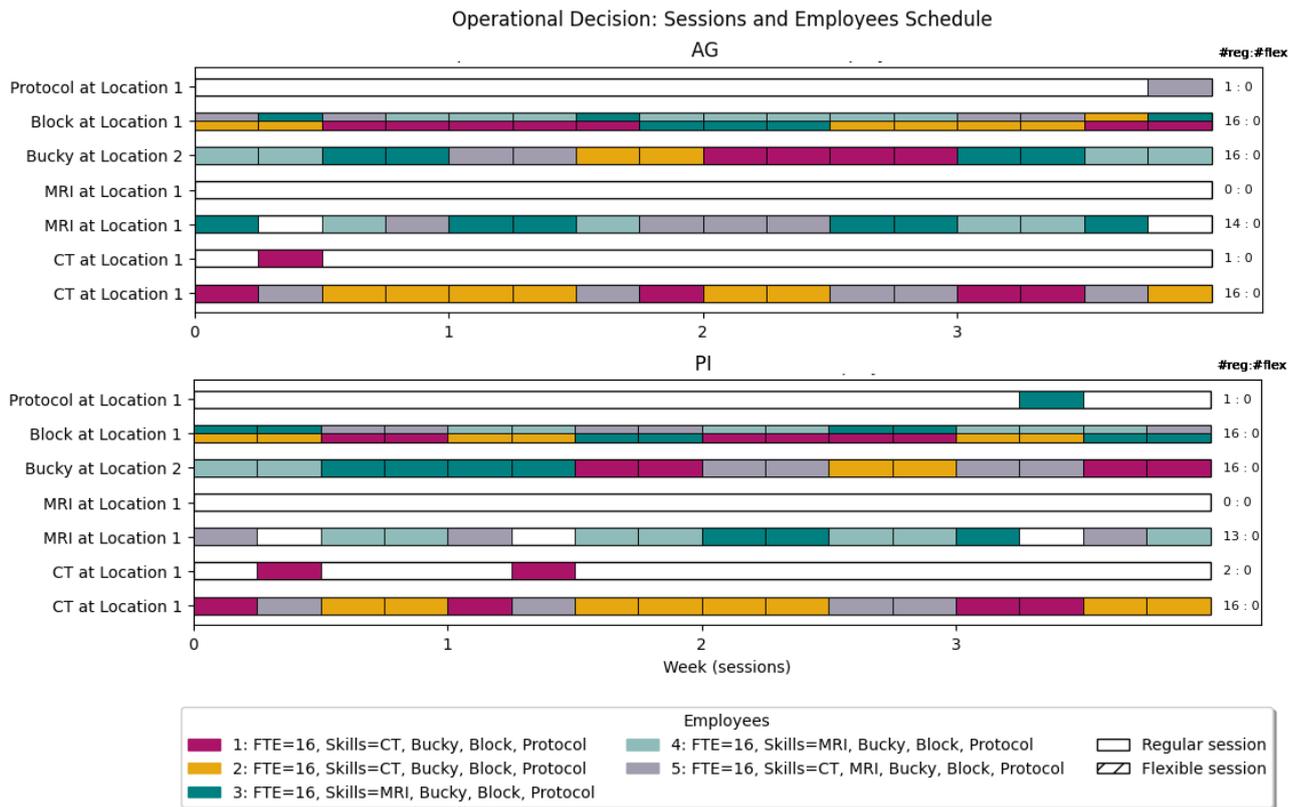
Figure 6.6: This figure depicts the resulting schedules of an operational offline decision based on three scenarios using AG and PI.

Modality		MP (mean)	Scenario 1	Scenario 2	Scenario 3
CT	Work backlog	22,	22	21	20
	Orders	$(3,3,3,2)$	$(4,3,3,2)$	$(3,3,3,2)$	$(3,3,3,3)$
MRI	Work backlog	11	11	12	11
	Orders	$(5,5,5,3)$	$(5,5,5,3)$	$(4,6,5,3)$	$(5,4,5,3)$

Table 6.3: This table includes the means and scenarios used for the operational decision in the two-stage example. The orders are given per week, for week 1-4 as $(1,2,3,4)$.

In the first week, PI and AG schedule 3 MRI sessions, while the orders ranged from 4 to 5 sessions. In the second week, PI schedules 3 MRI sessions and AG schedules 4 MRI schedules, with 4-6 MRI orders. Therefore, the PI schedule exceeds the tolerance of 2 sessions for MRI in one scenario, as described in constraint (D.II.C). In week 4, PI and AG schedule 3 MRI sessions, which matches the number of ordered sessions. Consequently, the obligatory protocol session is scheduled for this week. The number of CT orders ranges from 11-12 sessions, which is increased with $g^{IV} = 6$ to find the ideal number of sessions as described in constraint (D.IV.A): 17-18. PI schedules 18 CT sessions, and AG schedules 17 CT sessions. Similarly, the ideal number of MRI sessions is to match the number of orders, which is 17-18; PI schedules 13 MRI sessions, and AG schedules 14 CT sessions. Consequently, the work backlog for CT will decrease by 6-7 sessions, while the work backlog for MRI will increase by 3-5 sessions. Therefore, both PI and AG schedule an appropriate amount of sessions for each modality.

6.4.2 Model comparison

We tested the proposed models in different configurations in terms of timeline, aggregation method, and scenario method. We tested two timelines, month-based and week-based, and two aggregation methods, AG and PI. Moreover, in total, five different scenario generation methods were tested: sampling 20 and 50 scenarios and using a single scenario based on the mean, the first quartile (q1), and the third quartile (q3). The scenarios were generated as outlined in Section 6.1. The performance of the different methods was evaluated using Algorithm 2 over a horizon of 15 months for five different datasets generated according to the distributions as given in Table 6.1. A period of 15 months includes 12 tactical and operational decisions. We expect the model to take effect within this time. Five episodes are too little to draw definitive conclusions on the performance or to judge the standard sample deviation. However, it does indicate how well the methods perform. The limited number of episodes was chosen due to the large number (20) of different configurations that were being tested.

These tests were executed with an equal initial work backlog in interval II for case 1 and the work backlog for CT in interval IV and MRI in interval II for case 2. The two different cases allow us to evaluate how the model corrects a work backlog outside the desired interval and its ability to maintain a work backlog within the desirable interval. The results are given in Table 6.4 for the month-based timeline and in Table 6.5 for the week-based timeline.

Case	Scenario method	$w \in w^*$		$\Delta \bar{w}$		$\Delta \max\{w\}$	
		AG	PI	AG	PI	AG	PI
Case 1	q1	(1.06,0.06)	(1.03,0.15)	(0.94,0.16)	(0.92,0.15)	(1.27,0.25)	(1.25,0.21)
	mean	(**1.78**,0.13)	(1.68,0.17)	(0.52,0.16)	(0.57,0.18)	(0.71,0.20)	(0.74,0.21)
	q3	(1.70,0.13)	(1.75,0.13)	(**0.36**,0.11)	(**0.39**,0.14)	(**0.53**,0.18)	(**0.56**,0.20)
	$S = 20$	(**1.78**,0.14)	(**1.82**,0.13)	(0.53,0.17)	(0.51,0.14)	(0.72,0.25)	(0.70,0.20)
	$S = 50$	(1.73,0.27)	(1.77,0.13)	(0.51,0.16)	(0.55,0.17)	(0.76,0.22)	(0.79,0.23)
Case 2	q1	(0.66,0.14)	(0.69,0.14)	(0.73,0.15)	(0.74,0.15)	(1.13,0.26)	(1.12,0.21)
	mean	(1.37,0.27)	(1.20,0.43)	(**0.32**,0.16)	(0.36,0.17)	(**0.45**,0.23)	(0.50,0.23)
	q3	(**1.70**,0.07)	(**1.67**,0.10)	(0.37,0.15)	(0.35,0.16)	(0.49,0.20)	(0.47,0.21)
	$S = 20$	(1.40,0.29)	(1.45,0.26)	(0.35,0.20)	(**0.30**,0.14)	(0.51,0.33)	(**0.44**,0.24)
	$S = 50$	(1.37,0.14)	(1.35,0.13)	(0.33,0.19)	(0.32,0.17)	(0.49,0.29)	(0.47,0.25)

Table 6.4: This table offers an overview of test results comparing different scenario methods, cases, and aggregation methods for a month-based timeline. Column $w \in w^*$ contains the number of modalities out of two with a work backlog in the desired range, column $\Delta \bar{w}$ contains the mean deviation of the work backlog from the desired work backlog in weeks, and column $\Delta \max\{w\}$ contains the maximum deviation of the work backlog from the desired work backlog in weeks. These KPIs are calculated across all weeks and five different episodes. They are presented in the format: $(\bar{x}, s)$, with $\bar{x}$ the sample mean and s the sample standard deviation. The highest value in the first column and the lowest value in the second and third columns are printed in bold font.

In Table 6.4, we see the standard deviation of $w \in w^*$, $\Delta \bar{w}$ and $\Delta \max\{w\}$ lies around 0.20 on average across all methods and cases. The differences between the mean values are mostly smaller than 0.20. Therefore, we can not draw definitive conclusions. However, we can conclude that q1 performs significantly worse than the other scenario methods.

In the first case, the mean and $M = 20$ generate $w \in w^* = 1.78$, indicating that in 90% of the weeks the work backlog for MRI and CT were in the desired range. Furthermore, we see that in general the work backlog's mean $\Delta \bar{w}$ and maximum deviation $\Delta \max\{w\}$ is below 1, except for q1. This indicates that the work backlog of both CT and MRI does not deviate more than 1 week from w^* on average. The method q3 scores the lowest values for $\Delta \bar{w}$ and $\Delta \max\{w\}$ for both PI and AG. The values of the mean scenario, $M = 20$ and $M = 50$ are slightly higher. In the second case, q3 has the highest $w \in w^*$, but the mean scenario and $M = 20$ have lower $\Delta \bar{w}$ and $\Delta \max\{w\}$ for AG and PI, respectively. Combining the results from both cases, for AG q3 and the mean scenario seem to perform the best, while for PI q3 and $M = 20$ seem to perform the best.

The discrepancies in values between PI and AG are minor, remaining mostly below 0.05. Therefore, we can not draw conclusions about which method is better.

Case	Scenario method	$w \in w^*$		$\Delta \bar{w}$		$\Delta \max\{w\}$	
		AG	PI	AG	PI	AG	PI
Case 1	q1	(0.72,0.19)	(0.92,0.25)	(0.97,0.11)	(0.95,0.12)	(1.18,0.19)	(1.16,0.19)
	mean	(1.37,0.20)	(**1.73**,0.23)	(0.71,0.13)	(0.64,0.13)	(0.98,0.18)	(0.90,0.17)
	q3	(**1.75**,0.25)	(1.66,0.20)	(**0.54**,0.14)	(**0.52**,0.14)	(**0.70**,0.18)	(**0.70**,0.19)
	$S = 20$	(1.48,0.27)	(1.53,0.30)	(0.66,0.16)	(0.66,0.17)	(0.90,0.19)	(0.88,0.20)
	$S = 50$	(1.31,0.22)	(1.41,0.24)	(0.61,0.13)	(0.59,0.12)	(0.81,0.14)	(0.79,0.16)
Case 2	q1	(0.37,0.07)	(0.39,0.08)	(0.84,0.10)	(0.74,0.16)	(1.02,0.17)	(0.96,0.17)
	mean	(0.88,0.20)	(1.00,0.22)	(0.48,0.16)	(**0.32**,0.16)	(0.61,0.19)	(**0.44**,0.22)
	q3	(**1.38**,0.25)	(**1.45**,0.26)	(**0.37**,0.12)	(0.40,0.12)	(**0.52**,0.17)	(0.54,0.17)
	$S = 20$	(0.97,0.39)	(1.00,0.49)	(0.48,0.20)	(0.45,0.19)	(0.70,0.30)	(0.66,0.29)
	$S = 50$	(1.13,0.18)	(1.19,0.16)	(0.51,0.18)	(0.47,0.17)	(0.72,0.20)	(0.67,0.18)

Table 6.5: This table offers an overview of test results comparing different scenario methods, cases, and aggregation methods for a week-based timeline. Column $w \in w^*$ contains the number of modalities out of two with a work backlog in the desired range, column $\Delta \bar{w}$ contains the mean deviation of the work backlog from the desired work backlog in weeks, and column $\Delta \max\{w\}$ contains the maximum deviation of the work backlog from the desired work backlog in weeks. These KPIs are calculated across all weeks and five different episodes. They are presented in the format: $(\bar{x}, s)$, with $\bar{x}$ the sample mean and s the sample standard deviation. The highest value in the first column and the lowest value in the second and third columns are printed in bold font.

In Table 6.5, we see similar trends as were identified in Table 6.4. Namely, q1 performs significantly worse than

the other scenario methods. Moreover, in the first case for AG, q3 has the highest values for $w \in w^*$, and lowest values for $\Delta \bar{w}$ and $\Delta \max$. After q3, $M = 20$ scores highest $w \in w^*$ for AG . PI has the highest $w \in w^*$ for the mean scenario in the first case and q3 scores lowest values for $\Delta \bar{w}$ and $\Delta \max$. In the second case, q3 has the highest $w \in w^*$ for PI and AG. The $w \in w^*$ of $M = 50$ is the closest second. Using AG, the scenario method q3 generates the lowest values of $\Delta \bar{w}$ and $\Delta \max\{w\}$. The mean, $M = 20$ and $M = 50$ deviate only 0.03 for $\Delta \bar{w}$, while the mean has $\Delta \max\{w\}$ 0.09 lower. Using PI, the mean scenario generates a lower $\Delta \bar{w}$ and $\Delta \max\{w\}$. These differences are too small to be significant. Nonetheless, we draw the initial conclusion that q3 and $M = 20$ seem to perform the best for AG, while q3 and the mean scenario perform the best for PI.

Lastly, the highest values in the first column and the lowest values in the second and third columns across scenario methods are generated by PI. However, the differences with AG are too small to draw final conclusions.

When comparing Table 6.4 with Table 6.5, we see that $w \in w^*$ is consistently higher and the values of $\Delta \bar{w}$ and $\Delta \max\{w\}$ are consistently lower for the month-based timeline in comparison to the week timeline for all scenario methods and both PI and AG. The best scoring scenario methods are further investigated for the month-based timeline in Figure 6.7 and for the week-based timeline in Figure 6.9.

Month-based timeline

In Figure 6.7, we notice that the mean work backlog for CT is, in general, more stable than the work backlog for MRI. The work backlog for CT is initiated in interval IV in case 2 and is seen to gradually decrease to inside the desirable interval for all tested scenarios and aggregation methods. The mean of q3 decreases faster and stabilises at a lower work backlog, explaining the higher values of $w \in w^*$ in Table 6.4.
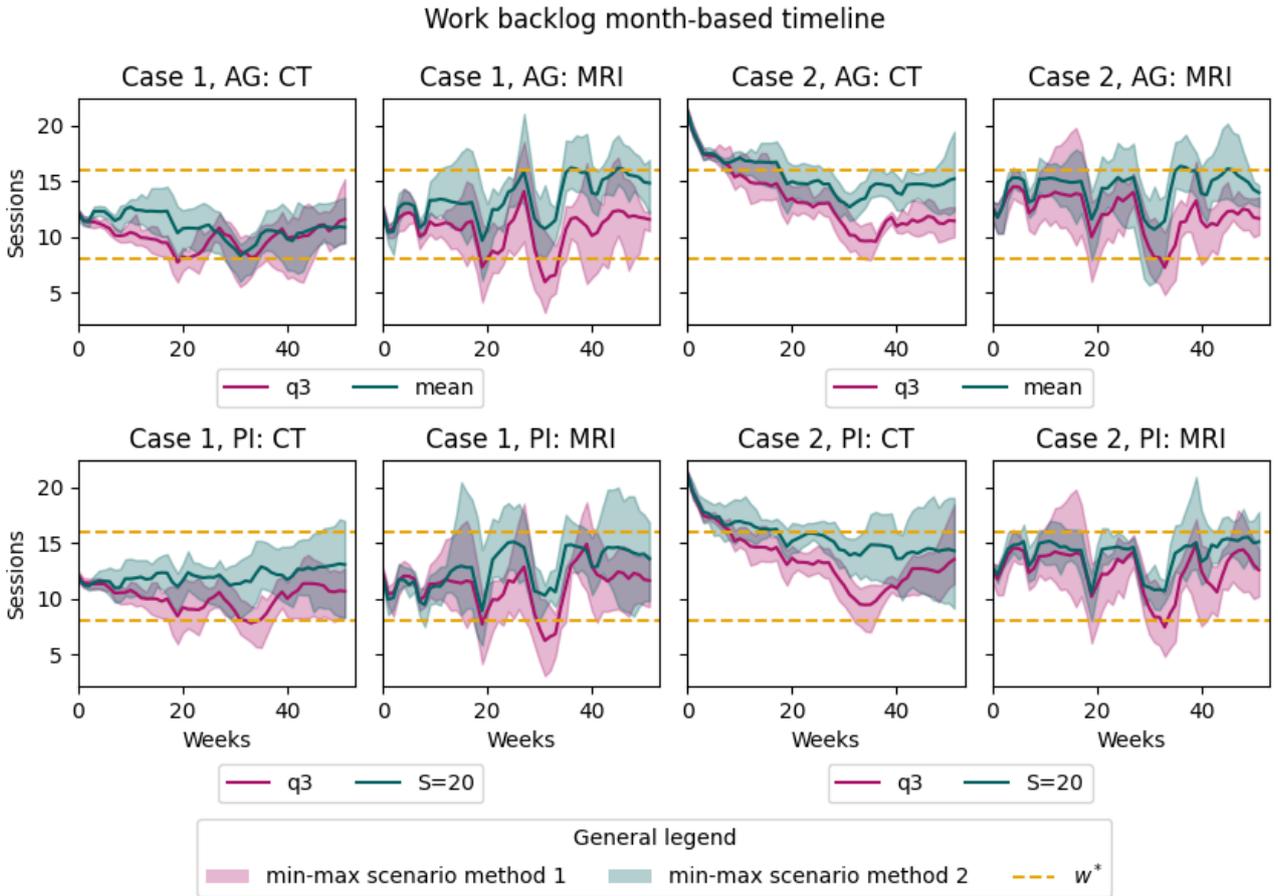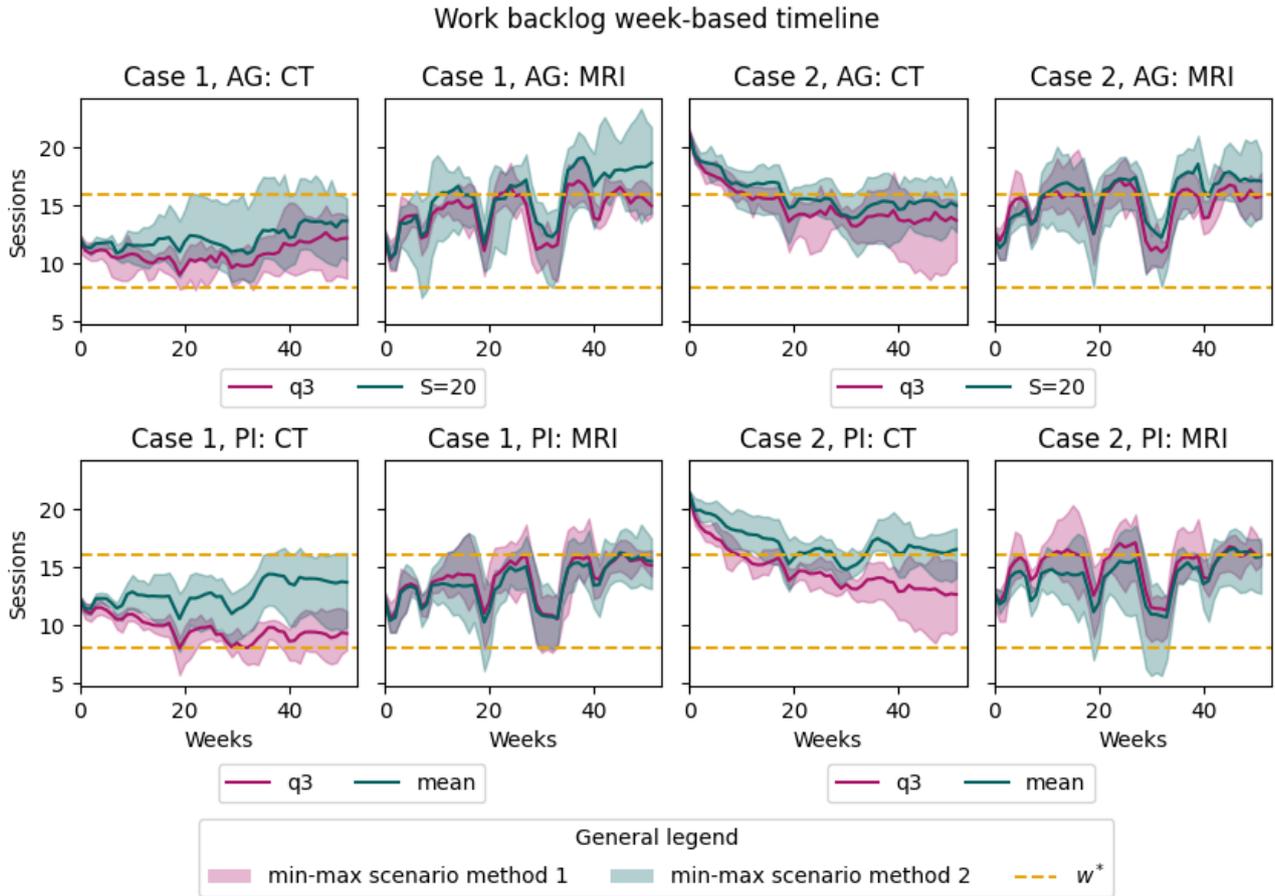

Figure 6.7: *This figure depicts the work backlog over time for CT and MRI in cases 1 and 2 as generated by PI and AG using a month-based timeline. The mean work backlog over 5 episodes is plotted with a line, and the spread between the minimum and maximum work backlog over 5 episodes is shaded. Yellow dashed lines indicate the desirable work backlog range.*

Furthermore, the mean of q3 is generally lower than the other visualised scenario methods in all cases for both AG and PI for CT and MRI. This observation aligns with the understanding that q3 consistently tends to overestimate demand, scheduling more sessions as a result. The mean of q3 dips around weeks 35-40, the summer holidays. The work backlog of the MRI dips below the desirable range, after which it increases to the desirable range again. In contrast, the other depicted methods have less prominent dips during the summer holidays. The spread between the min and max values across the different episodes is more narrow for the single scenario methods q3 and the mean than $M = 20$. The differences between AG and PI do not seem to show a pattern. Note that the difference in spread between min and max values of the methods plotted in green are likely due to the difference in scenario method.

The scheduled FTE and fairness scores are plotted in Figure 6.8. We see in the upper figure that predominantly 5 FTE are scheduled. However, during the holiday weeks, this is reduced to 4.5-4.75. This incurs a feasibility penalty of 80 and explains the opposite peaks in the fairness penalty depicted in the lower figure. The fairness penalty fluctuates around zero to 5 in normal weeks. The scheduled FTE differs per case but is similar for AG and PI. However, AG has an additional peak around week 8 in case 2, and PI has an additional peak around weeks 12-15 in case 1. These differences translate to additional peaks in the fairness plots. The fairness penalty is predominantly less for q3 than the other scenario methods.
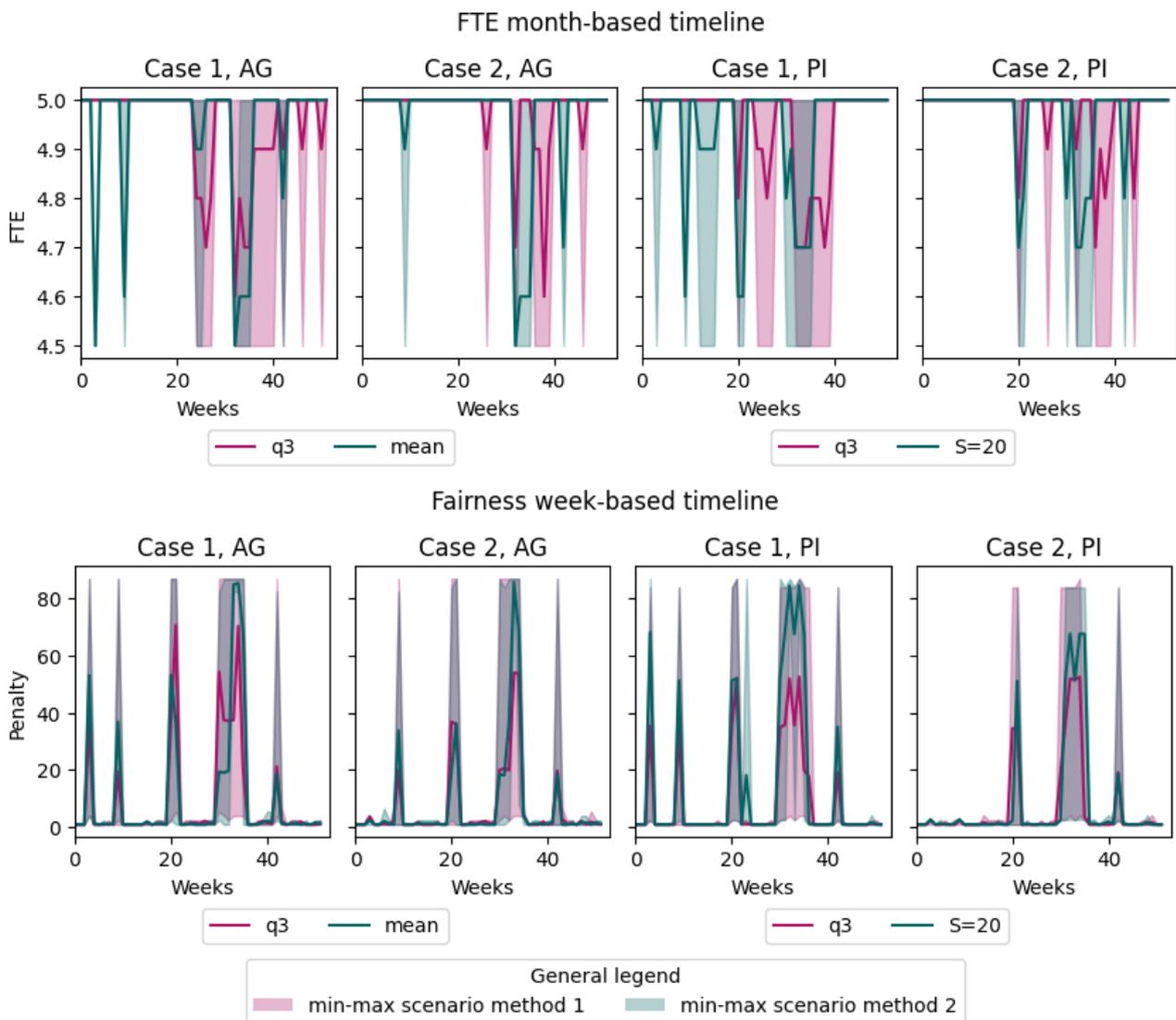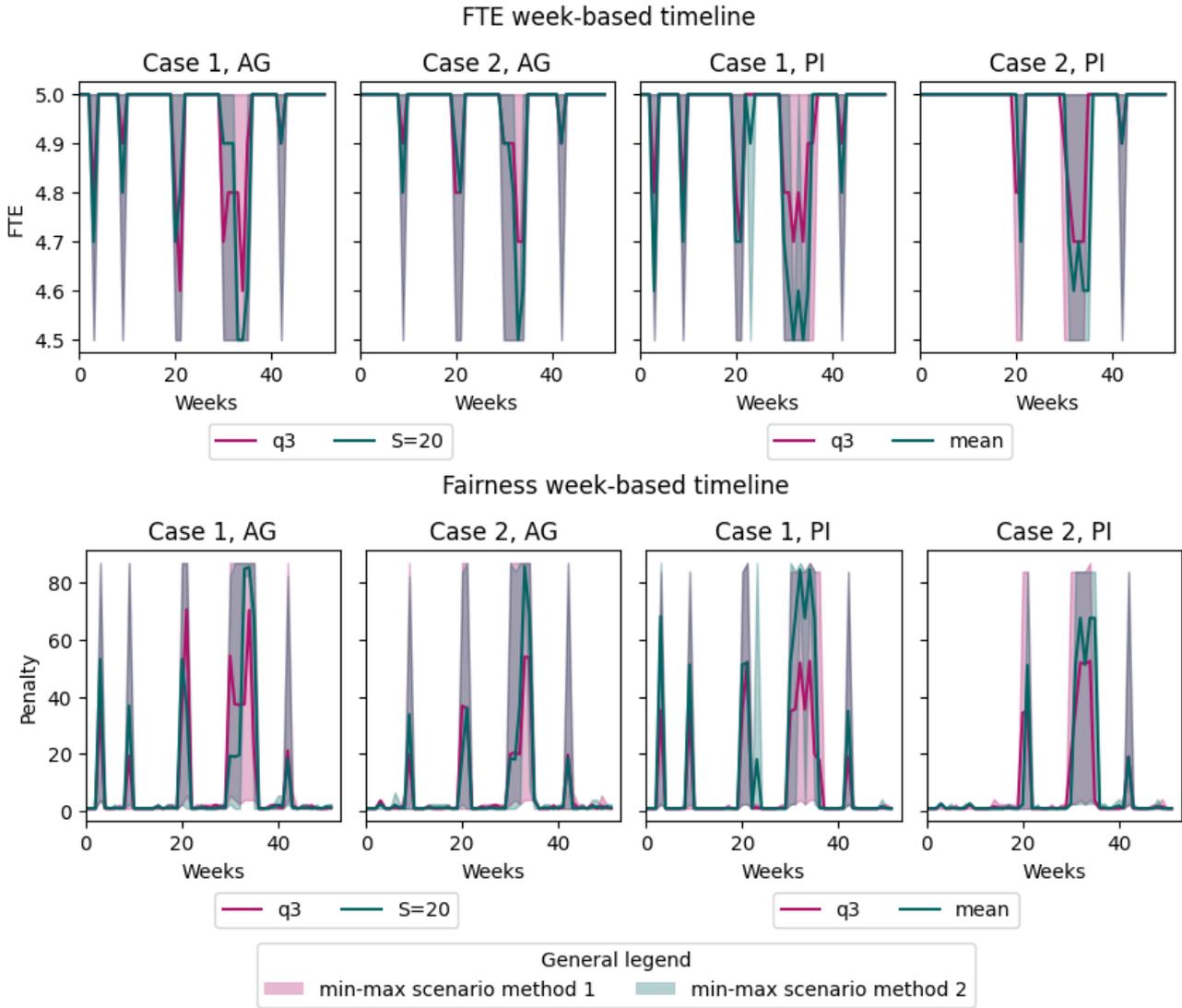

Figure 6.8: *This figure depicts the scheduled FTE and fairness quality for the month-based method over the tested horizon. The mean and the spread between the minimum and maximum work backlog over 5 episodes for two scenario generation methods are plotted.*

Week-based timeline

In Figure 6.9, we notice that the general trends of the work backlog are the same as seen in Figure 6.7. The mean of the work backlog for CT is relatively stable in case 1, where q3 generates a work backlog that is consistently lower than the work backlog generated by the other plotted scenario methods. In case 2, the mean of q3 of the CT work backlog decreases towards the center of the desirable interval, whereas the work backlog of the other scenario methods remains at the top or slightly above the desirable interval.

We still see strong fluctuations in the work backlog for MRI. Thus, shortening the timeline to week-based does not generate a more stable work backlog. Moreover, the fluctuations are stronger compared to Figure 6.7, causing the work backlog to be more frequently outside the desirable range. The spread between the minimum and maximum values across the episodes is wider for the $M = 20$ method, as was seen before in the month-based method. The differences between PI and AG are again not clear.

Figure 6.9: This figure depicts the work backlog over time for CT and MRI in cases 1 and 2 as generated by PI and AG using a week-based timeline. The mean work backlog over 5 episodes is plotted with a line, and the spread between the minimum and maximum work backlog over 5 episodes is shaded. Yellow dashed lines indicate the desirable work backlog range.

In Figure 6.10, the scheduled FTE and fairness penalty are plotted over the tested horizon of the week-based method. PI and AG show very similar results, scheduling less FTE in the holiday weeks. Furthermore, we see that q3 schedules more FTE than the other methods, incurring a lower feasibility penalty and also generating a lower work backlog as was seen in Figure 6.9.

Figure 6.10: This figure depicts the scheduled FTE and fairness quality for the week-based method over the tested horizon. The mean and the spread between the minimum and maximum work backlog over 5 episodes for two scenario generation methods are plotted.

The differences between the week- and month-based timelines are not very pronounced. However, the month-based method achieves a slightly more stable work backlog. Especially, the work backlog of MRI is more within the desirable interval. We conclude that a month-based method shows the best performance. Furthermore, the scenario method q3 showcases the strongest performance for both timelines. However, using $M = 20$ and the mean scenario works decently as well.

Execution time

The execution time of 20 different tactical decisions is evaluated in Figure 6.11. The computation time increases as the number of scenarios increases, but they are all relatively short. The computation time of PI is generally longer than that of AG. The biggest difference is measured when using 50 scenarios: the computation time of PI is almost twice as long as AG. We expect that this is due to the SPs being sufficiently small that they are solved instantly also for AG. Then, the increased computation overhead caused by initiating and solving the SPs separately causes the computation time of PI to be longer.
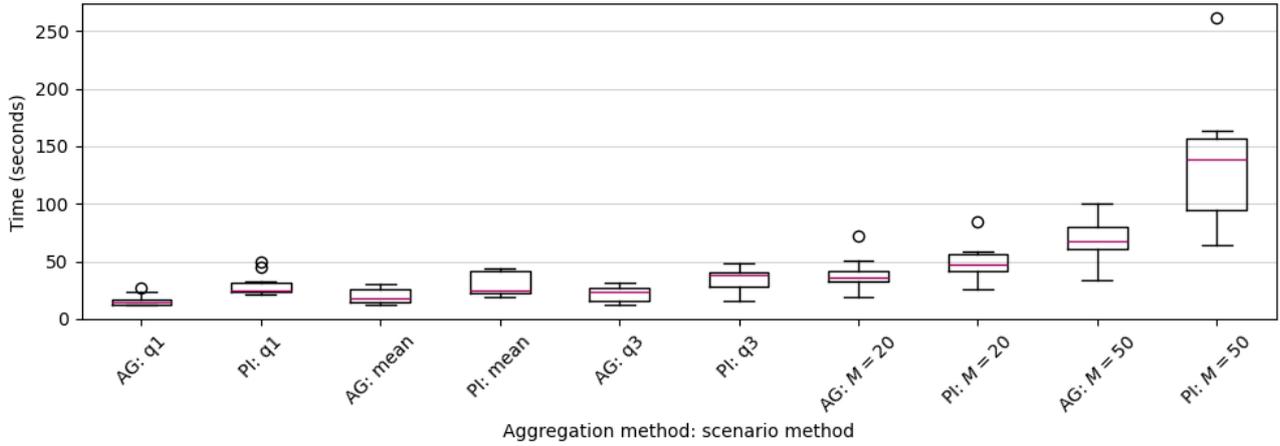
Figure 6.11: This figure depicts the computation time of one tactical decision for AG and PI for different scenario methods.

The performance of PI and AG were very similar overall. This is supposedly due to the characteristics of the tested case. Therefore, we conclude that this test case fails to showcase the possible differences in performance between AG and PI.

6.4.3 Model performance

We further explore the functionalities of the developed model on a month-based timeline. We test only PI because the performance of AG and PI was similar and because this method is better suited for large cases, as will become clear in the Case Study. In the results of the previous section, we saw that the model is capable of decreasing and/or maintaining a work backlog within the desirable range of $[8, 16]$. To further investigate the model's ability to maintain a desirable interval and possibly decrease the fluctuations in the work backlog of MRI, we narrowed this interval to $[10, 14]$ by setting $C^w = 2$. The results are plotted in Figure 6.12. We employ the q3 method due to its previously shown superior performance. Moreover, we also test $M = 20$ to investigate potential variations in performance in different experiments.

In addition, in Figure 6.12 the PI two-stage (TS) model is compared to two myopic models. The first myopic model uses a single-stage tactical decision that fixes the schedule 3 months in advance. Therefore, the first myopic model can be used to evaluate the added value of using a flexible framework. The second myopic model uses the proposed flexible framework but does not incorporate stochasticity. In more detail, it takes the tactical decision, including flexible sessions based on a single stage, a regular mixed integer program. Subsequently, it takes the operational decision one month in advance, again based on a single stage. Consequently, the second myopic model can be used to evaluate the added value of using a two-stage stochastic program.

We see in Figure 6.12, that the 95% confidence intervals are mostly the same width for all tested methods and the methods generate a similar work backlog over time. In comparison to Figure 6.7, we see two small differences: the work backlog for CT stabilises at a lower level for $M = 20$ than before. Moreover, the fluctuations of the work backlog for MRI are different but not better or worse. We can conclude that using a smaller desirable interval does not decrease the fluctuations in the work backlog for MRI. We suggest that, since the coefficients and goals were not adjusted, this could possibly be improved by tuning the model for this new interval of w^*.
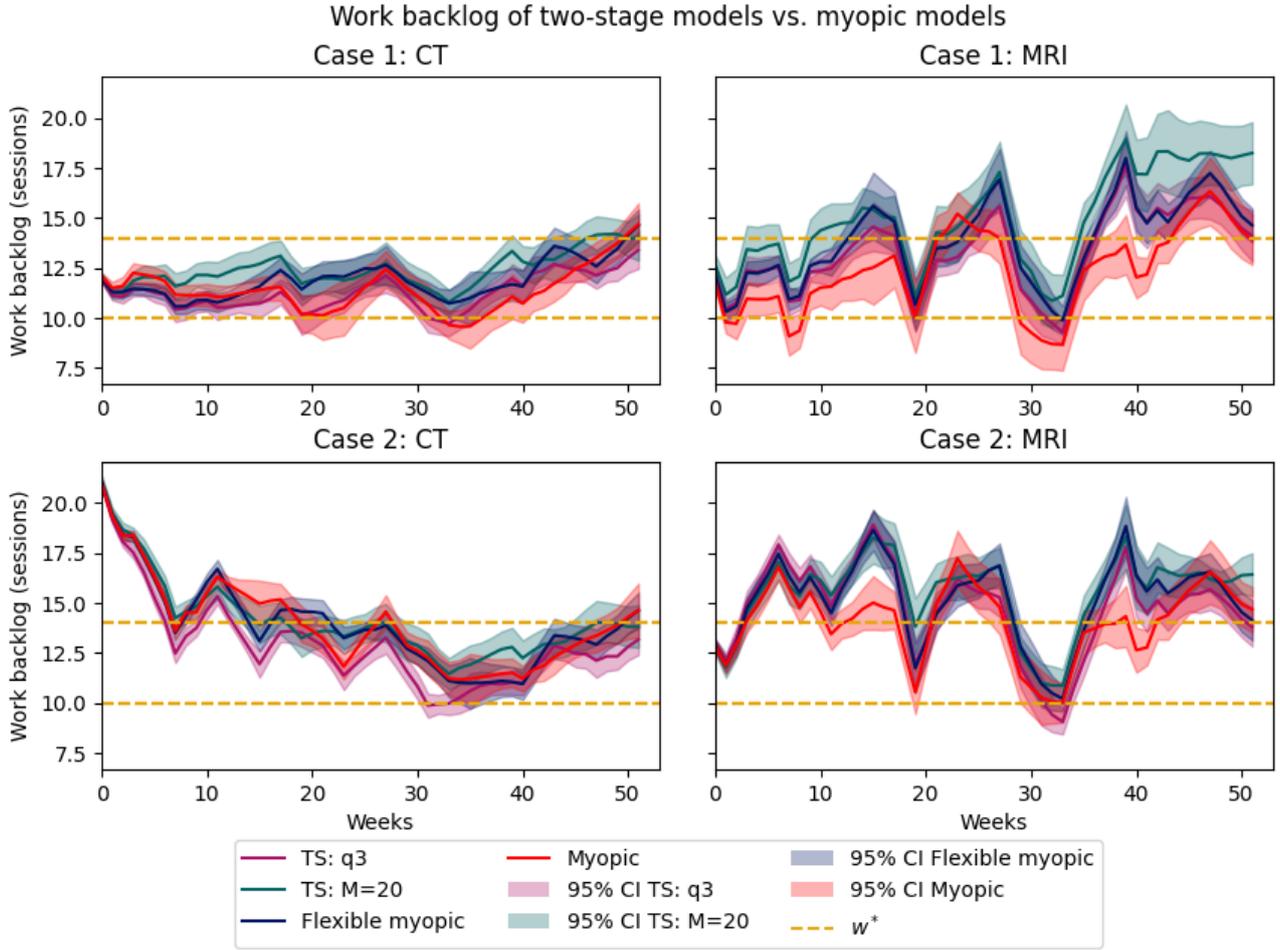
Figure 6.12: This figure depicts the work backlog over time for CT and MRI in cases 1 and 2 as generated by the two-stage (TS) methods using q3 and $M = 20$, the flexible myopic model and the myopic model. The mean work backlog over 20 episodes is plotted, and the 95% confidence interval (CI) interval is shaded. Yellow dashed lines indicate the desirable work backlog range.

When we compare the two-stage models to the myopic models, we see that the work backlog for CT is mostly within the desirable interval for both cases for all models. The two-stage model $M = 20$ has a higher work backlog for CT and MRI in case 1 than the other methods. As a result, the work backlog for CT of $M = 20$ is very close to w^* for the first 40 weeks. The work backlog of the myopic model and q3 is also within the desirable range, though. In case 2, q3 has a predominantly lower work backlog than the myopic models. Consequently, it decreases faster towards the desirable interval than the other methods. The work backlog for MRI varies significantly over time across all methods, with a similar trend observed for each. The performance of the flexible myopic model is very similar to the two-stage methods, indicating that this method is a decent heuristic for a the two-stage stochastic flexible framework.

The myopic model appears to align more closely with the desired interval compared to the proposed flexible methods. This suggests superior performance of the myopic approach, which bases its scheduling decisions on the mean. Our hypothesis is that when weekly orders and production exhibit minimal fluctuations, scheduling based on the mean yields favourable outcomes. In contrast, two-stage methods rely on a limited set of scenarios, which deviates from the mean. Additionally, monthly operational offline adjustments contribute to increased fluctuations in this test environment. Specifically, if the predicted work backlog falls within a different interval during these adjustments, the number of goal sessions changes, altering the scheduled decisions. While this adjustment theoretically aims to improve the session-demand match, overly aggressive corrections result in greater fluctuations. The myopic model experiences slightly less pronounced fluctuations due to its long-term (3-month) horizon, delaying model reactions and thereby smoothing out fluctuations.

We investigated our hypothesis by testing the models in two different environments. The first test was designed to decrease the quality of the information of the model about the orders, creating larger deviations from the model's mean. Specifically, we shifted the mean of the distributions, used to sample the datasets, of the elective

and emergency orders with a sample from a normal distribution with a mean 10% of the mean of the original distribution. Moreover, we added noise with probability 0.25 to the weekly orders by adding a sample of the normal distribution $\mathcal{N}(0,5)$ orders to either CT or MRI. The second test was designed to decrease the quality of the information at the tactical level specifically, which also creates larger deviations of the orders from the model's mean. Particularly, the mean of the deviation of planned and rostered sessions of the referring departments was increased from 1 to 2. Note that the distributions given to the models to find the expected work backlog, expected orders, and scenarios were not changed. The results of the two tests are given in Figure 6.13 and Figure 6.14.
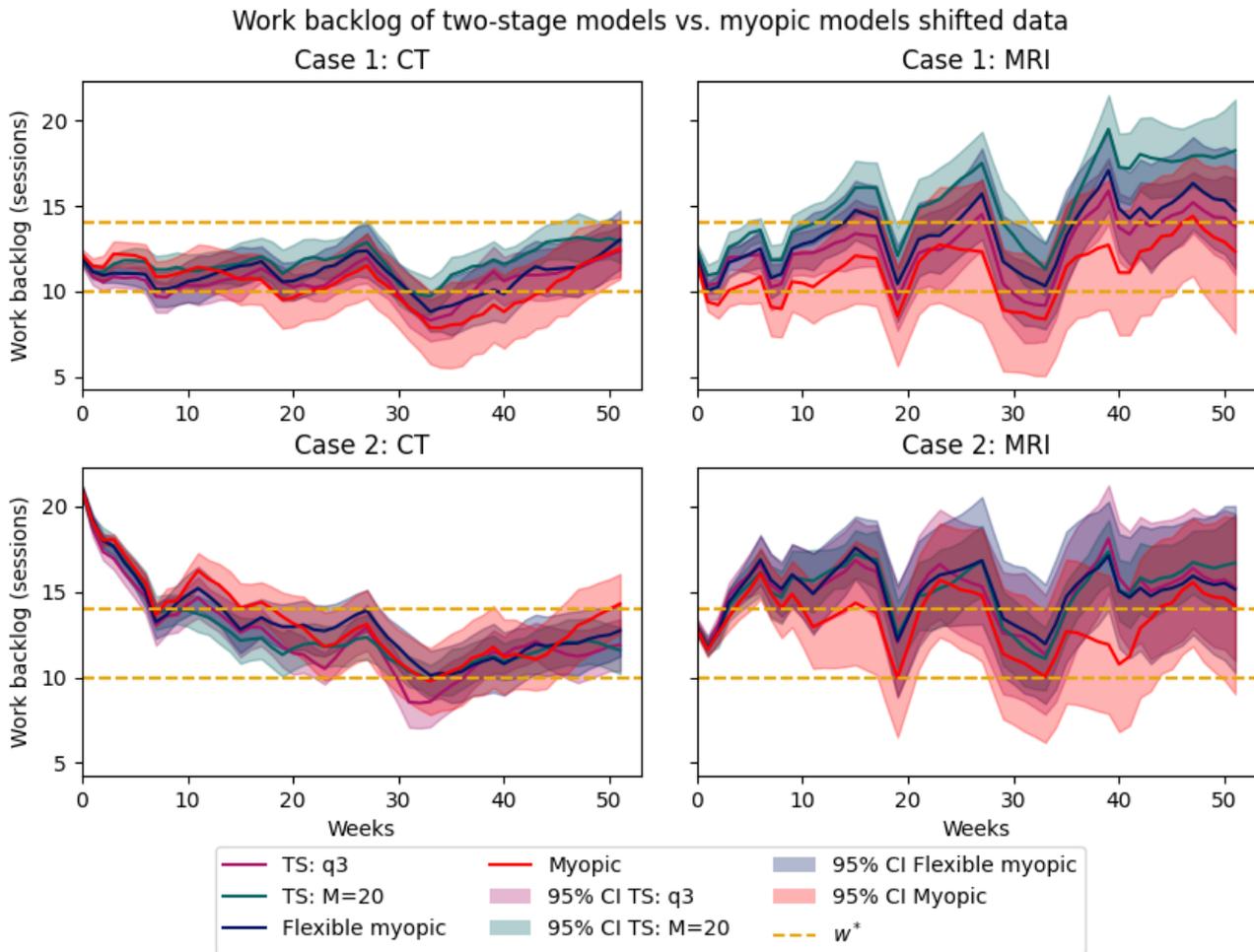
Figure 6.13: This figure depicts the work backlog over time for CT and MRI in cases 1 and 2 for a shifted distribution with added noise as generated by the two-stage (TS) methods using q3 and M = 20, the flexible myopic model and the myopic model. The mean work backlog over 20 episodes is plotted, and the 95% CI interval is shaded. Yellow dashed lines indicate the desirable work backlog range.

The general trend of the work backlog over time in Figure 6.13 is similar to that in Figure 6.12. The CI of the Myopic model is approximately 1.5 as wide for CT and 2-3 times as wide for MRI in comparison to the other models, which indicates that the results for the myopic model are less consistent than the other methods. This indicates that the flexible methods are less dependent on having an accurate distribution that is used for the prediction of the work backlog and orders. Nonetheless, the mean performance of the myopic model again outperforms the other models. The mean work backlog of the flexible myopic and TS q3 and $M = 20$ show very similar performance. Except for the work backlog for MRI in the first case, where the mean of $M = 20$ is consistently higher than the other models.
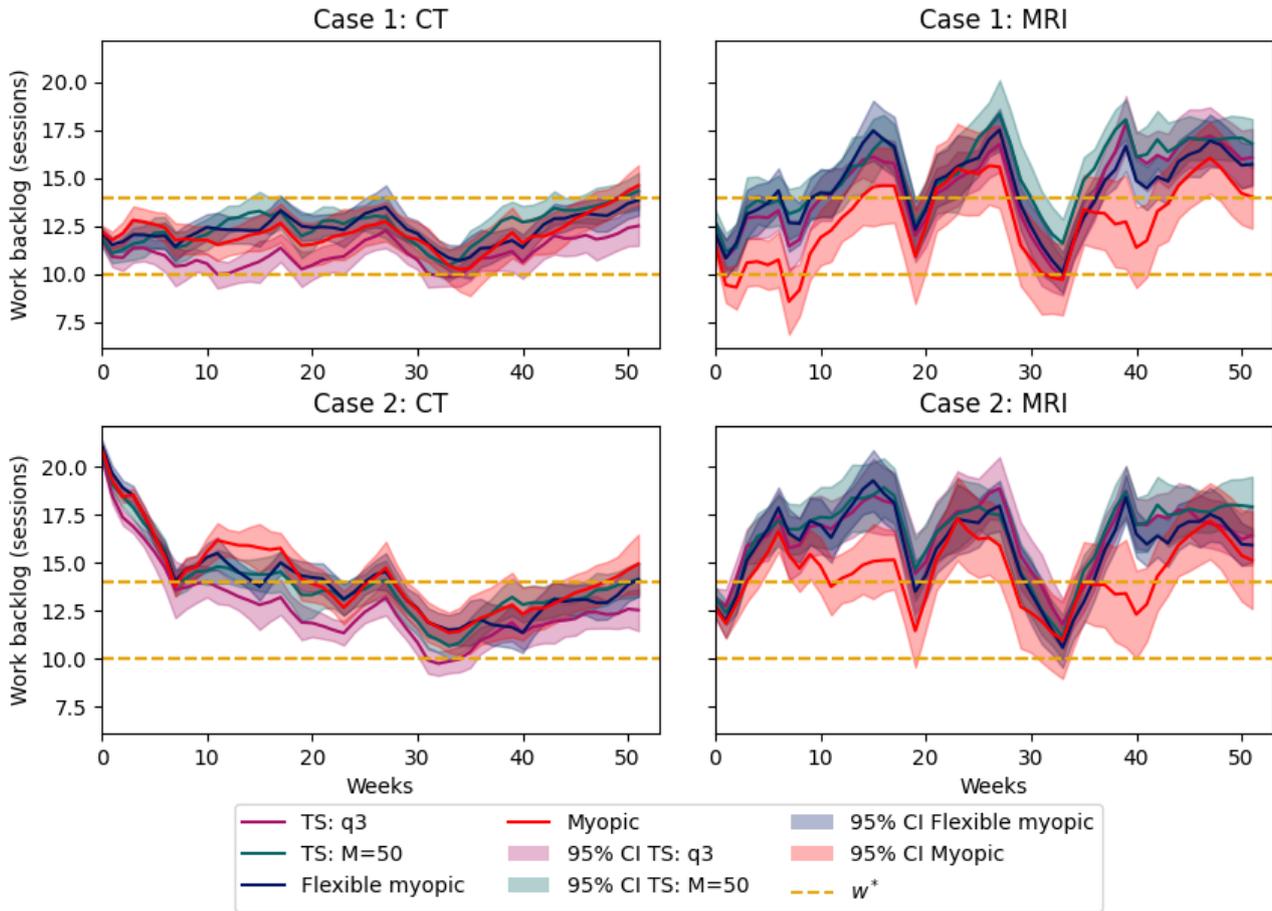
Figure 6.14: This figure depicts the work backlog over time for CT and MRI in cases 1 and 2 with worse information as generated by the two-stage (TS) methods using q3 and $M = 20$, the flexible myopic model and the myopic model. The mean work backlog over 20 episodes is plotted, and the 95% CI interval is shaded. Yellow dashed lines indicate the desirable work backlog range.

Figure 6.14 again shows the same results as we have seen before in Figure 6.12. The CI of all methods are wider, and the differences between TS q3 and Myopic in the first case for CT are larger. However, the differences are not consistent. We conclude that having worse information at the tactical stage, a standard deviation of 2 sessions instead of 1, does not significantly influence the results. The myopic model is, therefore, robust against deviations of this frequency and size. The deviations were applied every week, which supposedly still still smooth out on a monthly basis. Therefore, we tested the model for another set of datasets that had significant fluctuations in the number of orders for longer periods of time that became known at the operational offline stage.

The fluctuations were defined as follows: the referral departments have eight planned outpatient appointment sessions for MRI and CT every week. The rostered sessions of the referring departments shift every four weeks between six and ten sessions for CT and MRI. For instance, in the first four weeks, the referring departments for CT have six outpatient sessions and MRI ten, and in the next 4 weeks, CT has ten, and MRI has six. The results are given in Figure 6.15. Note that we removed the reductions for the holiday weeks.
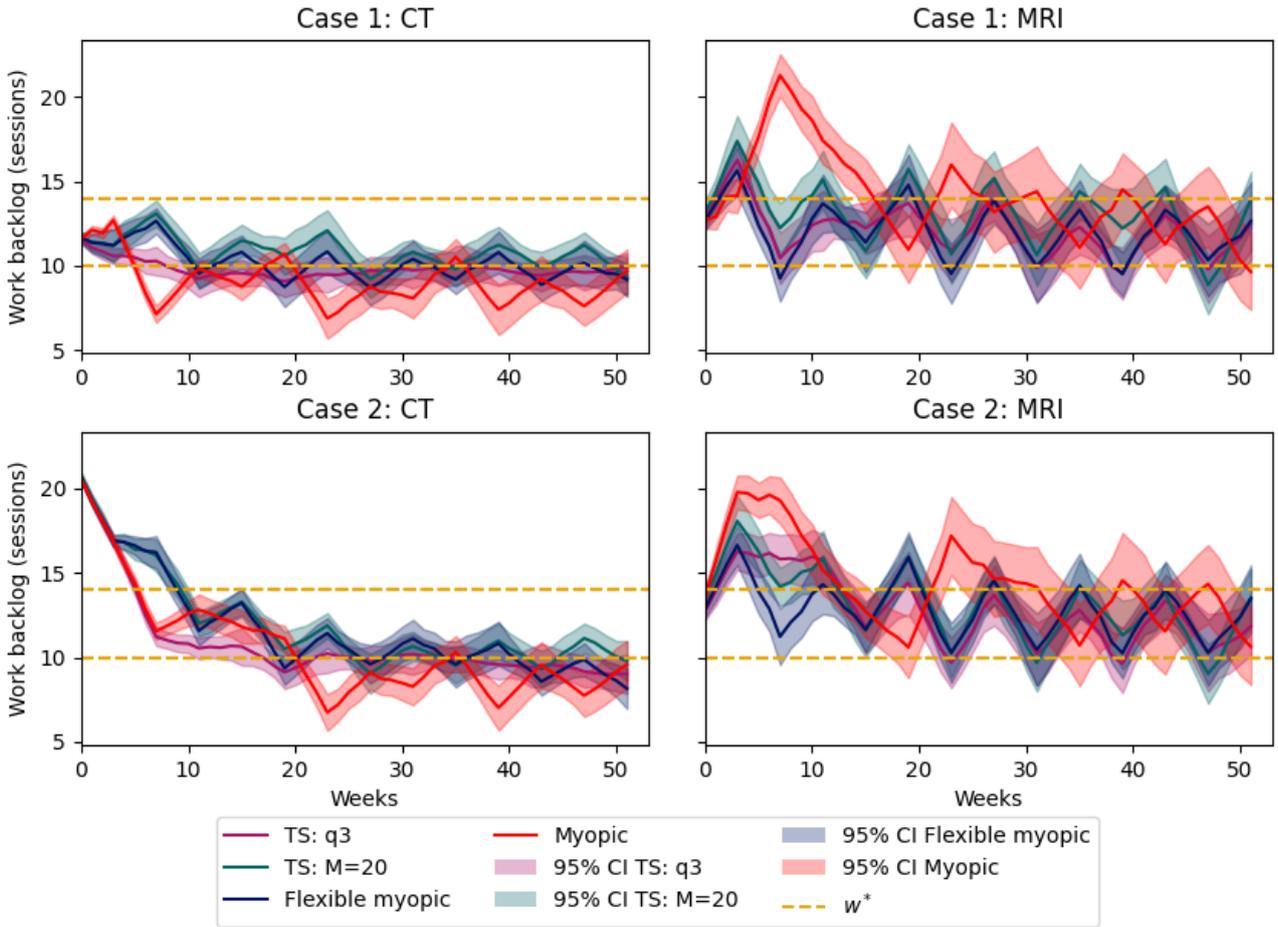
Figure 6.15: This figure depicts the work backlog over time for CT and MRI in cases 1 and 2 with 4-week shifts in demand as generated by the two-stage (TS) methods using q3 and M = 20, the flexible myopic model and the myopic model. The mean work backlog over 20 episodes is plotted, and the 95% CI interval is shaded. Yellow dashed lines indicate the desirable work backlog range.

In Figure 6.15, we see that the work backlog fluctuates for each method. The methods using a flexible framework all show similar results: the work backlog for CT fluctuates around the lower bound of the desirable interval, and the work backlog for MRI fluctuates from the upper bound to the lower bound. In contrast, the work backlog for CT of the myopic model fluctuates mostly below the desirable interval, thus showing worse performance. The work backlog for MRI of the myopic model first increases in both cases, decreases and increases again above the interval, after which it starts to fluctuate within the desirable interval. Its fluctuations are in anti-phase with the work backlog generated by the flexible method. These results show that when the deviations of the demand are for a prolonged period and significant in size, a flexible framework can adapt to the demand slightly better.

6.5   Case study

We conducted a case study about VMC to assess the efficacy of our developed model in a real-world context. This section begins with a description of the case and the scenario generation. We used data of VMC of 2023. While older data was available, changes in data management and referral patterns rendered it less representative for our analysis. The data from the first part of 2023 was used to establish parameters for scenario generation and such, leaving us with a 6-month period for testing. A detailed breakdown of the available and utilised data can be found in Appendix A.2.

Following this, we describe the validation and verification process in short. Next, we compare the model's performance across a month-based and week-based timeline and different scenario generation methods to find the best model configuration. Subsequently, we compare this model's performance with historical data from VMC. We evaluate multiple KPIs: the work backlog, the number of sessions, the scheduled FTE, the resulting RT schedule and the difference between tactical and operational decisions. Furthermore, we compare the model with two myopic models, as we have done previously for the test case.

Finally, we subject the model to two additional scenarios. In the first scenario, the modality composition was extended. Particularly, the composition of modalities is changed to VMC's future modality composition in 2024. In the second scenario, the flexibility of the model was reduced. This is of practical interest because transitioning from no flexibility to a little flexibility would be easier in practice.

6.5.1   Case description

The radiology department of VMC operates eleven different imaging modalities: CT,MRI, US, Mammography (often abbreviated to 'Mammo'), X-treme CT, dual X-ray absorptiometry (Dexa), Angiography (often abbreviated to 'Angio'), Blok, fluoroscopy lab (FL), and Bucky. The last three modality names refer to the location rather than the imaging technique. For example, X-rays are made in Bucky rooms in the diagnostic centre and in the "Blok". The Blok encompasses X-rays for inpatients and fluoroscopy in the surgery centre. Fluoroscopy is also done in the FL in the radiology department. Additionally, RTs are scheduled on administrative tasks: protocol, order, application, and guidance. Guidance is not scheduled during the summer holidays due to the absence of interns. Furthermore, two RTs are scheduled on 'flex' daily to cover absences. VMC has three hospital locations: Venlo, Venray, and Panningen.

The department employs 75 RTs. The composition of RT qualifications is visualised in Figure 6.16. The correct interpretation of the right diagram is as follows: upon examination of the initial column, it becomes evident that VMC employs 15.97 FTE of RTs that only have the qualification CT, besides being able to do general work. Moreover, we see that one RT working 0.78 FTE has CT and MRI qualifications, and 2-3 (1.89 FTE) RTs have CT and Angio qualifications. In general, we see that some qualification combinations are more popular than others. Such imbalance in the composition can constrain scheduling options and limit the ability to respond effectively to changing demands. For instance, relatively many RTs are qualified for Mammography and MRI, then if the demand for Mammography and MRI is high, the capacity staff-wise needs to be divided over Mammography and MRI.

The left diagram describes the available FTE per qualification: we count the total FTE of the RTs with this skill. Therefore, having 51.2 available FTE for general work, means there are 51.2 FTE employed that are qualified for general work, not that all 51.2 FTE can simultaneously be scheduled for general work. In practice, the hours of RTs are divided over all the different qualifications they have and indirect work. Consequently, from Figure 6.16, we can conclude that RTs need to be scheduled for approximately 40% of their shifts on their qualifications. This percentage is lower for angiography and mammography and higher for US and Xtreme-CT. The total FTE employed by VMC is 53.0, while the total FTE needed is 46.6, based on budgeted sessions and irregular shifts, but excluding indirect work. In our case study, we exclude all irregular shifts and indirect work. Therefore, we reduced the FTE of each employee with their indirect work as reported by VMC and a reservation for irregular shifts. In Section 6.5.5, we elaborate on the realised FTE by VMC and our model.

VMC determines the number of necessary sessions for CT, MRI, Dexa, US, and Mammography based on the demand. Blok and Bucky are omitted because these modalities are not scheduled with appointments but are on a walk-in basis. Moreover, Angiography is not scheduled demand-based because this involves almost solely emergency orders. X-treme CT is omitted because the demand is very low and mostly research-based. Lastly, FL sessions are mostly in cooperation with other departments, which makes flexible scheduling impractical. Therefore, we do not include FL either. The orders over 2023 for the demand-based scheduled modalities are visualised in Figure 6.17.
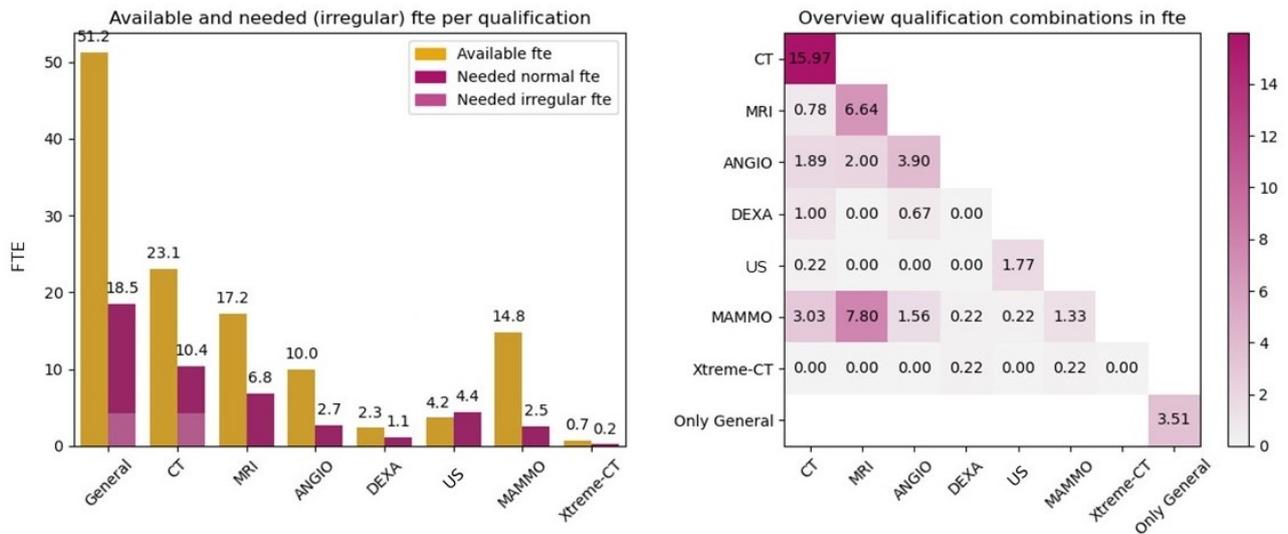
Figure 6.16: *The left figure depicts the total needed FTE and the total available FTE per modality. The FTE of an RT with multiple qualifications is included for all the modalities they are qualified for. The right figure shows a heatmap of the composition of the qualifications of all RTs.*
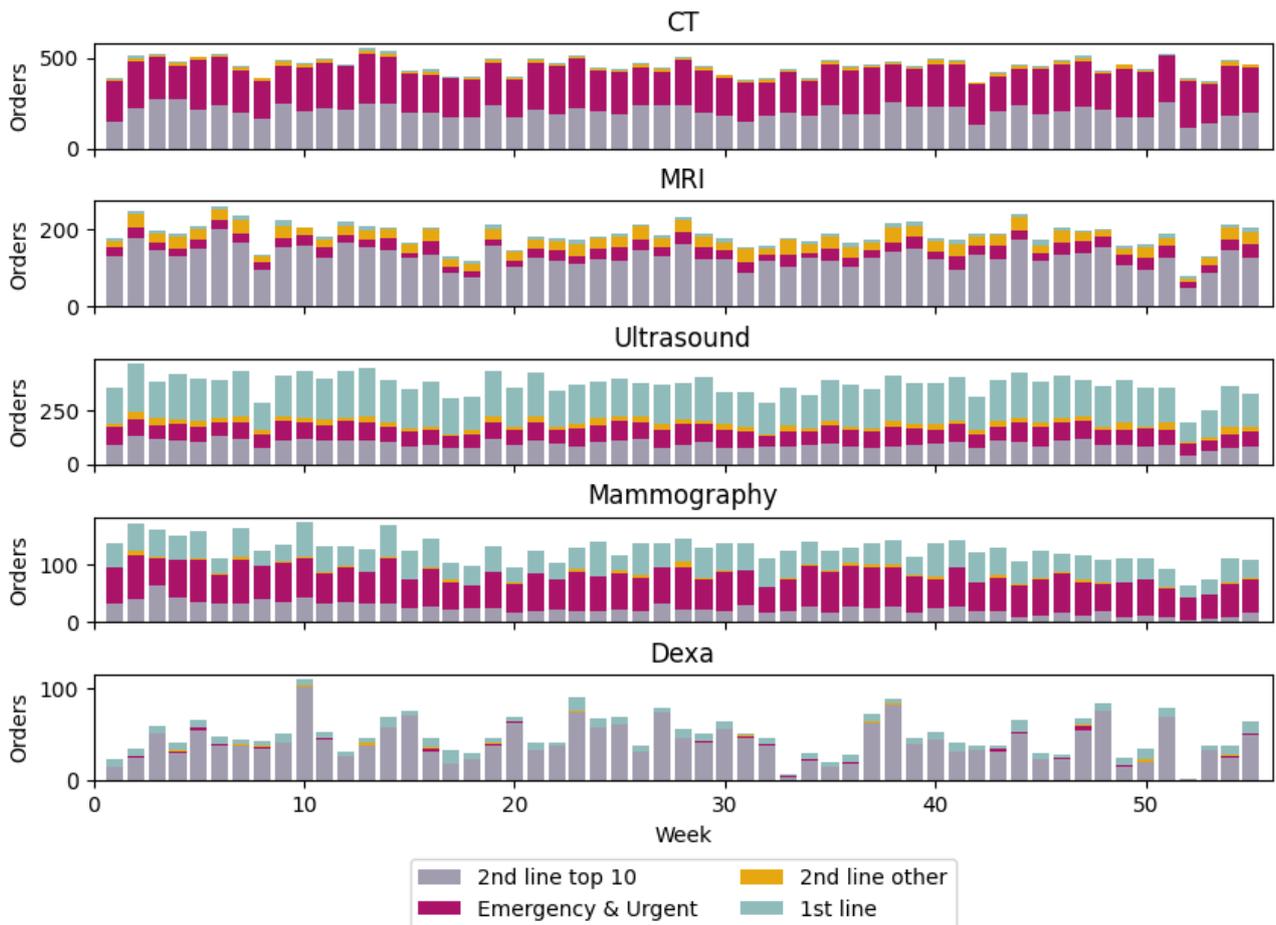

Figure 6.17: *This figure visualizes the number of orders per demand-based scheduled modality plotted per week of 2023. The different sources of the orders are printed in different colours.*

We see in Figure 6.17, that the orders fluctuate for all modalities. The ratio of the sources of the orders differs per department: approximately half of the orders for CT and mammography are emergency and urgent orders, whereas the orders for Dexa and MRI are mostly generated by the top 10 referring departments. The total number of orders fluctuates for all modalities. At the end of the year, we see a decrease in the number of orders for all modalities.

VMC has determined that sessions covering at least 40% of the expected demand for a demand-based scheduled modality during one week should be fixed three months in advance. Therefore, we set $u(i) = 0.4$ for all demand-based scheduled modalities. Consequently, allowing approximately 60% of the sessions of the demand-based scheduled modalities to be flexible. Moreover, the desired work backlog of a modality i lies between one week and three weeks' worth of the average production of this modality.

6.5.2 Scenario generation

Recall, from Section 6.1 that we need different distributions to generate scenarios. This subsection describes how we determined these distributions for the case study. Firstly, we constructed a histogram with ten bins from the historical data of the orders generated by the first line, emergency orders and the non-top 10 referring departments. We chose 10 bins because they capture the distribution sufficiently well without adding unnecessary complexity. From this histogram, we constructed empirical distributions for the weekly orders for each modality i from the first line F_i^{1st}, emergency department F_i^e and the non-top ten hospital departments F_i^{nt}. The emergency orders for CT are partly executed outside of normal shift times: 58% in 2023; therefore, we corrected the number of CT emergency orders with a factor of 0.42.

Similarly, we found an empirical distribution for the elective orders generated by the top 10 referring departments. Unfortunately, data on the referrals per appointment was not available; only aggregated data of the number of outpatient sessions and number of referrals per week was available. Therefore, we constructed an empirical distribution $F_{r,i}^{top}$ per modality i per referring department r for the success fraction of an outpatient session: how many elective orders a session produced. We classified long-term orders as elective orders for simplification purposes. The distributions for the referring departments for CT are plotted in Figure 6.18 per illustration.
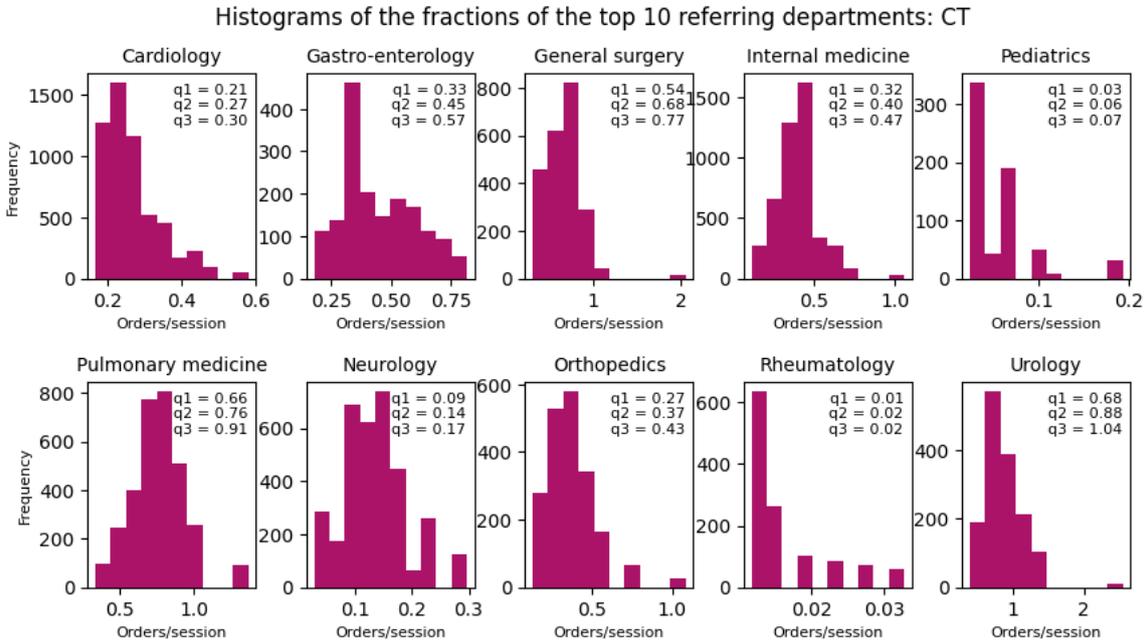

Figure 6.18: This figure shows the histograms of the fraction, the number of CT orders per outpatient session, for each referring department.

We used historical data to assess the deviation between realised and planned sessions A_r^2 of referring specialists. We analysed the histograms and saw a bell-shaped curve. After removing outliers using the interquartile range method, we conducted the Shapiro-Wilks test to confirm normality. This test showed that the deviations of all departments are normally distributed under $p = 0.05$. Consequently, we fitted normal distributions to the deviations for each department r: N^r. The number of outpatient sessions per referring department

can, therefore, be sampled by summing the planned sessions with a sample of the found normal distribution. Approximately one to two months in advance the scheduled outpatient sessions A_r^3 are available. Unfortunately, the data regarding the scheduled outpatient sessions was not available, therefore we used the realized sessions. The desirable access time of VMC is two weeks; therefore, as argued in Section 6.1, the elective orders relevant for the week θ are generated two weeks earlier.

We generated M scenarios using the distributions that were found. A scenario includes the number of orders of weeks θ and a work backlog prediction. Similarly, as in Section 6.1, we sampled the number of orders f for week θ for modality i for the scenarios in the tactical decision as:

$$f_i^{2,\theta} = F_i^{1st}(k=1) + F_i^{nt}(k=1) + F_i^e(k=1) + \sum_{r \in \mathcal{R}} \left(F_{r,i}^{top}(k = A_r^{2,\theta-2} + N^r(k=1)) \right), \tag{5}$$

and for the operational decision as:

$$f_i^{3,\theta} = F_i^{1st}(k=1) + F_i^{nt}(k=1) + F_i^e(k=1) + \sum_{r \in \mathcal{R}} \left(F_{r,i}^{top}(k = A_r^{3,\theta-2}) \right), \tag{6}$$

where k is the number of samples drawn from the respective distributions.

A prediction of the future work backlog was sampled by summing the current work backlog with the sampled incoming orders in the weeks within the horizon and subtracting the outflow of orders within the horizon, as described in Section 6.1. Furthermore, we did not generate scenarios regarding supply as highlighted in Section 6.1. However, the availability of the employees was updated from the tactical decision to the operational decision. VMC saves a version of the employee roster every three hours, which allowed us to construct their availability retrospectively.

6.5.3 Model validation and verification
The model was put through a short verification and validation process to ensure that it translated well to the VMC case. The demand penalties and sessions costs were chosen the same as in the test case and the coefficients of the fairness penalties and feasibility penalties were scaled according to the increased number of sessions: multiplied by $\frac{10}{4}$. The relative values of the fairness penalty coefficients were determined using an AHP-analysis as suggested in literature [11]. The AHP analysis was conducted as follows: the six fairness aspects $(\kappa_h, \kappa_r, \rho, \gamma, \lambda)$ were compared pairwise in a questionnaire and rated on a scale from 1 to 9. The questionnaire was filled out by the people who currently determine the RT schedule: the capacity manager, the two RT-schedulers, and the two team leaders. The average score was computed, and the AHP-matrix was constructed. Subsequently, the AHP scores were computed by dividing each element by the sum of its column and then taking the average of each row. The results of the AHP method can be found in Appendix A.1.

Lastly, the model was validated using a peer review: the results were discussed with the capacity manager and consultants of Rhythm to validate the model's behaviour. The results of a tactical and operational decision are given in Appendix A.3.

6.5.4 Model comparison
The performance of the model using different scenario methods and timelines is compared in this section. We evaluate five different scenario generation methods: sampling 20 and 50 scenarios and using a single scenario based on the mean, first quartile (q1), and third quartile (q3), and two different timelines: month-based and week-based. The models with different configurations were tested over a test period of 6 months from August 2023 until January 2024 using algorithm 2. This test period included 6 tactical and 6 operational offline decisions. Therefore, we tested over a horizon of 9 months because the tactical decision taken regarding August was taken in May. Due to the significantly longer computation time of AG in comparison to PI, and given the comparable performance of PI and AG in the test case, we opted to test only PI for this case study, as highlighted in Section 6.5.5.

At every tactical and operational decision, the most up-to-date data regarding the availability of RTs was used. This did not lead to infeasibility, therefore the implemented feasibility-relax functionality was not utilised. During the first three months, the recorded work backlog and sessions of VMC were used. Starting from August, the work backlog was updated using the actual inflow of orders and the realised average production per session multiplied by the sessions scheduled by the model.

In Table 6.6, three results of the tests are printed: the number of modalities with a work backlog within the desirable interval $w \in w^*$, the mean deviation of the work backlog $\Delta \bar{w}$, and the maximum deviation $\Delta \max w$ from the desired work backlog. These deviations were normalised to the number of weeks before calculating the mean and standard deviation across modalities to ensure comparability across different modalities. Note that the means are simply the averages over the test period.

Scenario method	$w \in w^*$		$\Delta \bar{w}$		$\Delta \max\{w\}$	
	Month	Week	Month	Week	Month	Week
q1	(2.42,1.36)	(1.08,1.35)	(0.99,0.34)	(1.59,0.55)	(1.91,0.63)	(2.38,0.48)
mean	(3.04,0.82)	(**3.12**,0.77)	(0.88,0.21)	(**1.06**,0.21)	(1.71,0.23)	(2.78,0.64)
q3	(2.31,1.01)	(2.35,1.16)	(1.26,0.24)	(1.24,0.28)	(2.42,0.38)	(**2.34**,0.26)
$S = 20$	(3.04,0.72)	(2.65,1.02)	(0.85,0.21)	(1.33,0.34)	(1.58,0.25)	(3.39,0.84)
$S = 50$	(**3.15**,0.83)	(2.88,0.86)	(**0.82**,0.2)	(1.23,0.3)	(**1.55**,0.23)	(3.04,0.87)

Table 6.6: This table offers an overview of test results comparing different scenario methods and timelines. Column $w \in w^*$ contains the number of modalities out of five with a work backlog in the desired range, column $\Delta \bar{w}$ contains the mean deviation of the work backlog from the desired work backlog in weeks, and column $\Delta \max\{w\}$ contains the maximum deviation of the work backlog from the desired work backlog in weeks. The values are presented in the format: $(\bar{x}, s)$, with $\bar{x}$ the mean and s the standard deviation over the test period. The highest value in the first column and the lowest value in the second and third columns are printed in bold font.

The sample standard deviation of $w \in w^*$ is between 0.8 and 1.4, indicating that for each scenario method and timeline, the number of modalities with a work backlog in the desirable range fluctuates approximately with one modality. Moreover, the mean scenario, $M = 20$ and $M = 50$, have, on average, at least three modalities within the desirable range using the month-based method. Using the week-based timeline, only the mean scenario has a mean of $w \in w^*$ above three. Scenario methods q1 and q3 score between 0.6 and 2 lower than the other scenario methods. The sample standard deviation of $\Delta \bar{w}$ and $\Delta \max\{w\}$ is around 0.2-0.3 for both timelines. In contrast, the standard deviations of the week-based timeline in the mean scenario, $M = 20$, $M = 50$, are higher: 06-0.9. In general, the average values of $\Delta \bar{w}$ and $\Delta \max\{w\}$ are higher using the week-based timeline.

When using a month-based timeline the value of $\Delta \bar{w}$ and $\Delta \max\{w\}$ is on average the lowest for $M = 50$, followed by $M = 20$. The single scenario methods q1 and q3 score relatively worse: the $\Delta \max\{w\}$ of q1 is up to 0.5 higher, and the $\Delta \bar{w}$ of q3 is 0.5 higher. Combining all the results, we conclude that $M = 20$ and $M = 50$ are the best-performing methods using a monthly timeline.

Furthermore, looking at the week-based timeline, we see that the mean scenario has the lowest $\Delta \bar{w}$ and q3 has the lowest $\Delta \max\{w\}$. The value of $\Delta \bar{w}$ of q3 and $M = 50$ only differ 0.01. Based on $M = 50$ having a higher $w \in w^*$ than q3, we consider $M = 50$ to be better. Therefore, the best-performing scenario methods for the week-based method are the mean scenario and $M = 50$. The best methods of both timelines are investigated in more detail in the next subsections

Month-based

The work backlog and scheduled sessions for the demand-based scheduling modalities using a month-based timeline for $M = 20$ and $M = 50$ are plotted in Figure 6.19.

The differences in work backlogs and scheduled sessions between the scenario methods are generally small. The work backlog for MRI starts out too high but slowly decreases to the desirable interval for both methods. The strong decrease at the end of the year may be explained by the decrease in demand in week 51 as seen in Figure 6.17. The work backlog of US is stable at first, then increases above the desirable interval and decreases again. The work backlog for Dexa decreases slightly faster for $M = 20$. Starting from week 50 the work backlog for Dexa for both methods decreases to zero. The number of orders for Dexa decreases heavily at the end of the year and the new year as seen in Figure 6.17.
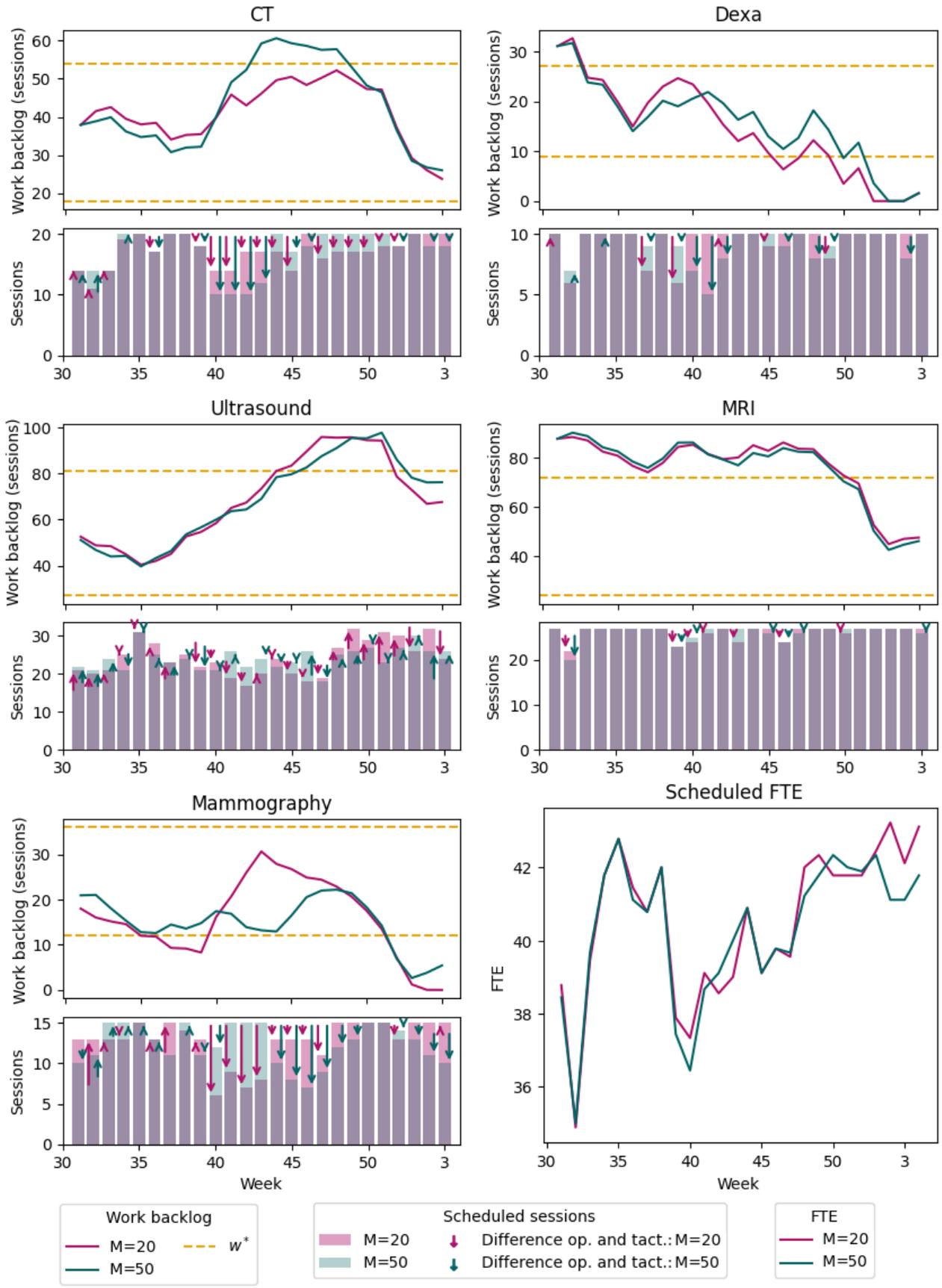
Figure 6.19: *This figure shows the work backlog, sessions and scheduled FTE over time for the month-based timeline. Each modality is represented by two figures. The upper figure plots the work backlog per scenario method. Yellow dashed lines indicate the desirable work backlog. The lower figure depicts the sessions scheduled per scenario method with overlapping bar graphs. The difference between the tactical and operational decisions is illustrated with arrows. The lower right figure plots the scheduled FTE over the test period per scenario method.*

The differences between both scenario methods are most noticeable in CT and mammography. The work backlog for CT of $M = 20$ is consistently within the desirable range, whereas $M = 50$ is slightly above, approximately 7 sessions, for 6 weeks. The work backlog for mammography fluctuates for both methods: $M = 20$ dips below the desirable interval at week 38, afterwards it shows a strong increase due to fewer scheduled sessions. In contrast, $M = 50$ shows smaller fluctuations. The strong peak of $M = 20$ can be explained by the work backlog dipping below the desirable interval, which triggers the model to schedule fewer sessions. Both scenario methods decrease towards zero around week 52, but $M = 50$ remains higher.

The arrows in Figure 6.19 indicate the change in scheduled sessions between the operational decision and the tactical decision. Both methods show almost no changes for MRI. The changes are the largest and most frequent for mammography and US. The corrections regarding US are both upwards and downwards. The corrections for mammography are biggest around weeks 40-43 for $M = 20$ and weeks 44-47 for $M = 50$. We notice that the changes are made for an entire month. Thus, it seems that the prediction of the mean scenario was too high at the tactical level for mammography, leading to a decrease in scheduled sessions during the operational decision. Lastly, $M = 50$ cancels more CT sessions in weeks 40-44 than $M = 20$, explaining the increase in work backlog.

The scheduled FTE of both methods is very similar. We see a dip during the first two weeks due to the holidays. Moreover, we see a dip around week 40, the autumn break. The FTE increases slowly back to 42. This slow increase might be because a new CT arrived these weeks which decreased the available FTE due to additional indirect work.

Based on these results, we see that $M = 50$ shows better performance on Dexa and Mammography, whereas the performance of $M = 20$ is slightly better for CT. Since the work backlog of CT is only 7 sessions above the desired interval, we prefer $M = 50$.

Week-based

The results using a week-based timeline for the mean scenario and $M = 50$ are plotted in Figure 6.20. In general, the results are comparable to the results of the month-based timeline. Especially, the work backlogs for US and Dexa are very similar. In contrast, the work backlog for MRI is much higher, showing an increase instead of the decrease we saw in Figure 6.19. This difference can be explained by the decrease in scheduled MRI sessions during weeks 37-39.

The mean scenario and $M = 50$ show comparable results. We see that $M = 50$ has a higher work backlog for CT than the mean scenario and than desirable. The work backlog for mammography of $M = 50$ shortly dips above the desirable interval. In contrast, the work backlog for mammography of the mean scenario dips 5 sessions below the desirable interval during week 52 of 2023 and 1-3 of 2024.

The changes between the operational and tactical decisions indicated by the arrows are most frequent for mammography, Dexa and CT for both scenario methods. In comparison to the month-based method, we see that the corrections regarding CT of $M = 50$ are smaller and more frequent. The changes are not consistent for one month anymore, showcasing the ability to adjust weekly.

The scheduled FTE is similar for both scenario methods, except that the mean scenario schedules more RTs during weeks 38 and 39. We see that the scheduled FTE dips around the summer holidays (weeks 31 and 32) and weeks 37-40. Based on the described results, we conclude that the mean scenario method shows the best performance for a week-based scheduling timeline.
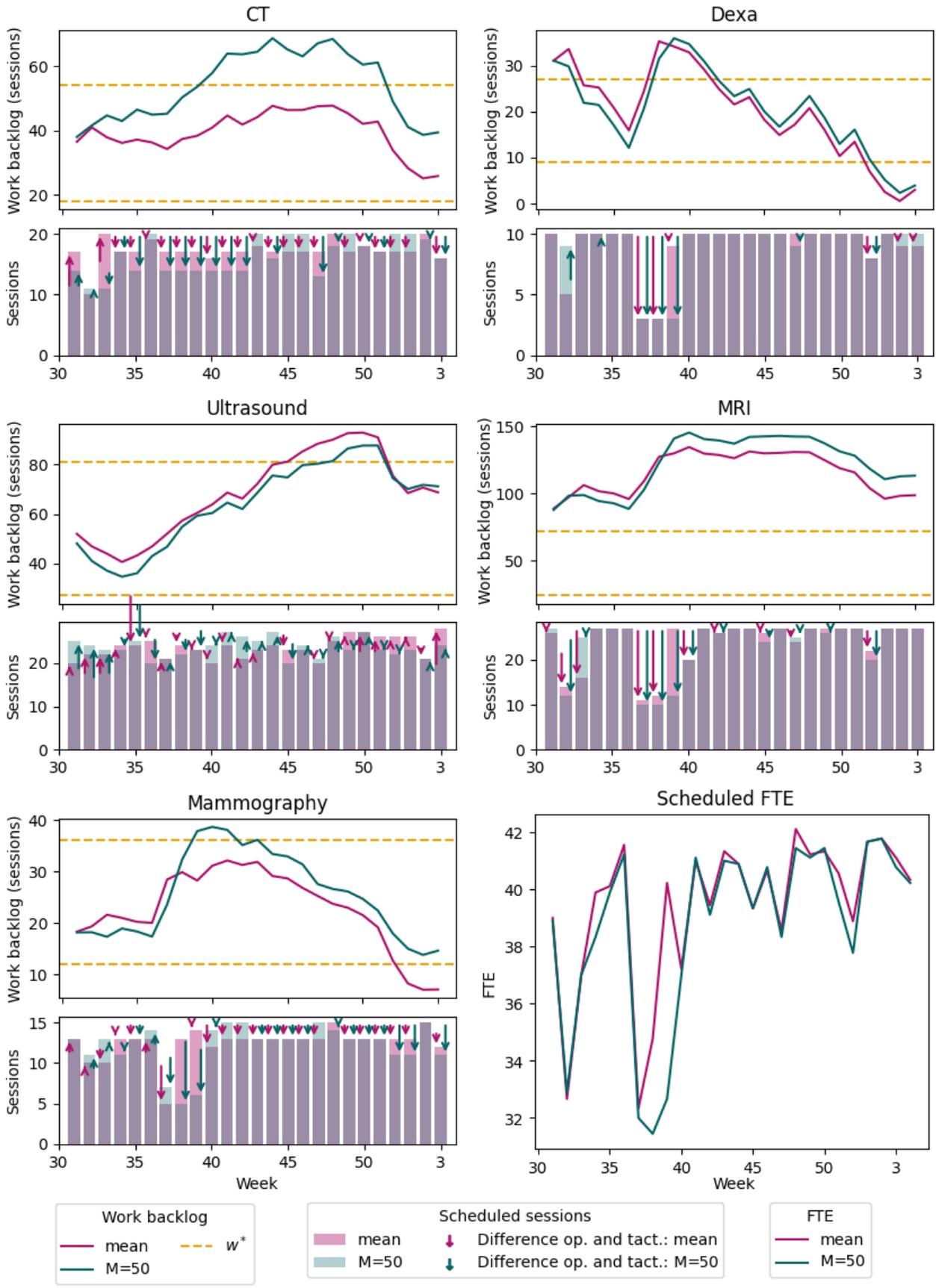
Figure 6.20: *This figure shows the work backlog, sessions and scheduled FTE over time for the week-based timeline. Each modality is represented by two figures. The upper figure plots the work backlog per scenario method. Yellow dashed lines indicate the desirable work backlog. The lower figure depicts the sessions scheduled per scenario method with overlapping bar graphs. The difference between the tactical and operational decisions is illustrated with arrows. The lower right figure plots the scheduled FTE over the test period per scenario method.*

The fairness score for the month-based and week-based timelines for the chosen scenario methods in Figure 6.21. The fairness score of all models lies within the range of 35-42. Therefore, these scores suggest that the models generate comparable schedules in terms of fairness. The month-based methods score predominantly lower than the week-based methods. During weeks 43-48 and 51-3 $M = 20$ for the month-based method scores a lower than the other methods.

The coefficients of the fairness penalties were in the range of 0-10. A fairness penalty of 40, would be caused by for example 5 out of the 75 RTs not being scheduled for 2 sessions on their qualifications that require postgraduate professional education (CT, MRI, US, and angiography) and 5 RTs being scheduled on one of their roster-free shifts. Therefore, we suppose the schedules are generally quite fair.
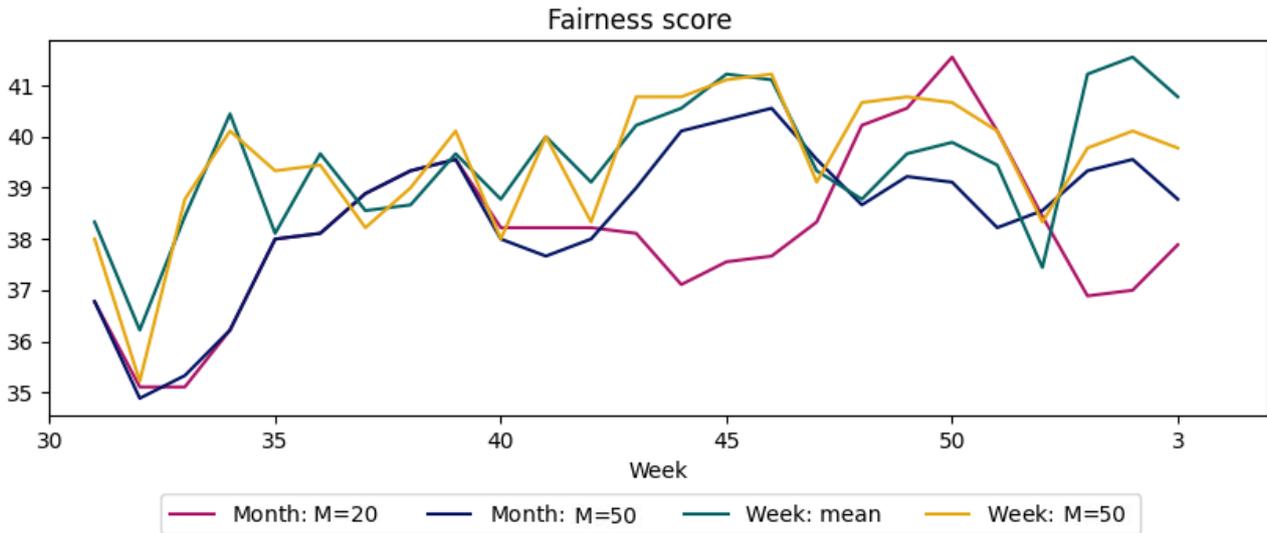
Figure 6.21: *This figure depicts the fairness score of the month-based and week-based timelines for different scenario methods over the tested horizon.*

The results for the month and week-based timeline are not completely comparable, because the availability of employees differed: the week-based method used more up-to-date information due to it taking decisions every week. Particularly, we see a strong decrease in scheduled FTE during weeks 37-39, which we did not see in the month-based scenario.

When comparing the results, regardless, we see that the month-based timeline performed slightly better in terms of work backlog. In more detail, the work backlog for mammography was more stable, the work backlog for Dexa remained within the desirable interval longer and the work backlog for MRI was much lower. In contrast, the work backlog for CT of the week-based timeline was more consistent and always within the desirable interval. Taking into account that a month-based method is preferred in practice, we decided to explore the month-based scheduling method using 50 scenarios further.

6.5.5 Model performance
We compare the performance of two-stage stochastic program (TS) with a month-based timeline using 50 scenarios to VMC in Figure 6.22 and to a flexible myopic model and two myopic models in Figure 6.26.

Model vs.VMC

In Figure 6.22 we plotted the results of the model as seen in Figure 6.19, the recorded work backlog of VMC, the theoretical work backlog of VMC and the realised sessions of VMC. The theoretical work backlog is the work backlog VMC would have if you calculate the work backlog by adding the incoming orders to and subtracting the production (orders/session) from the current work backlog. Firstly, we notice that the recorded work backlog shows larger fluctuations than the recorded work backlog. These deviations are caused by for example canceled orders. Our model's results are most comparable to the theoretical work backlog of VMC as both were computed using identical methodologies.
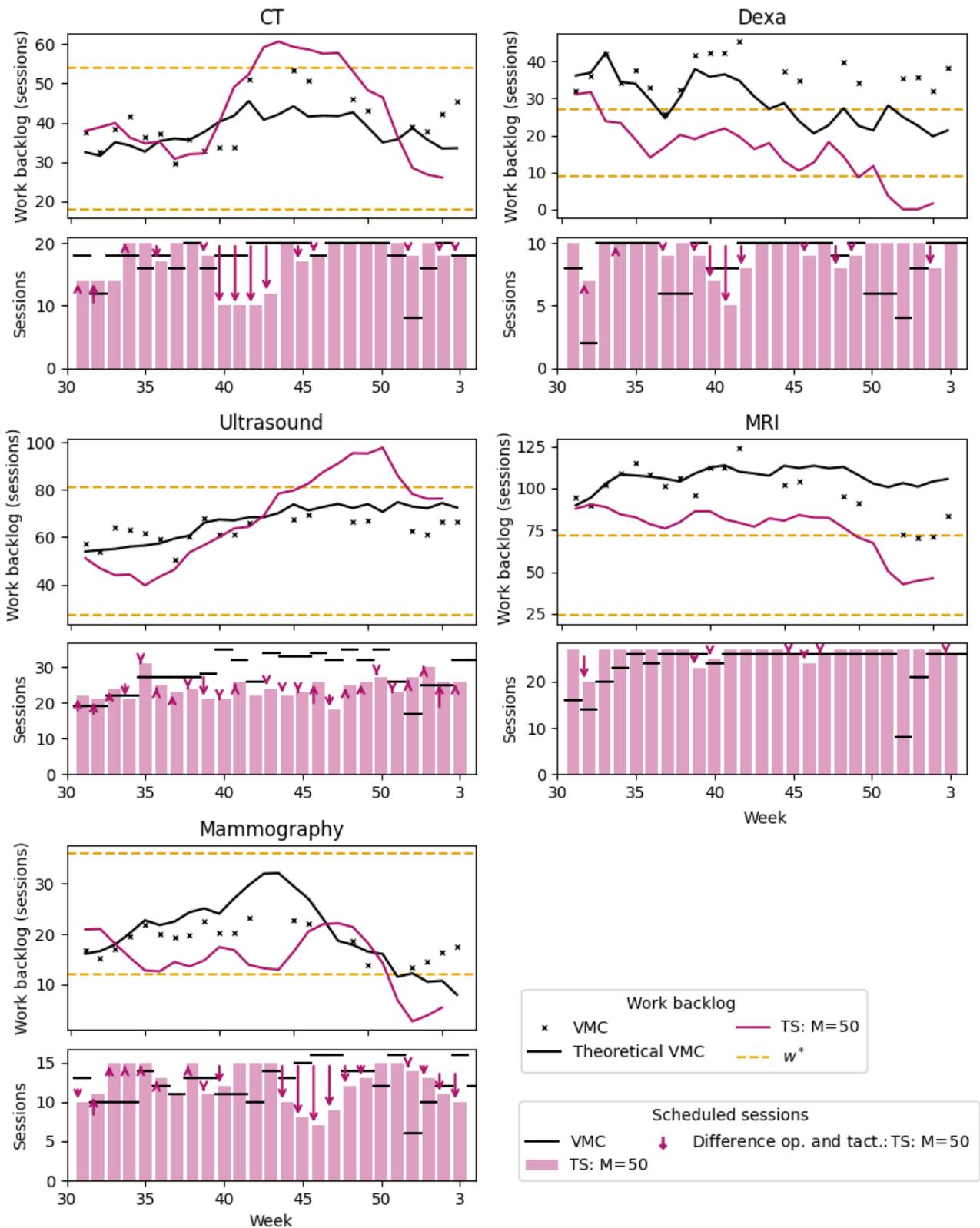
Figure 6.22: *This figure compares the work backlog, sessions and scheduled FTE over time for the model and VMC. Each modality is represented by two figures. The upper figure illustrates the work backlog of the model and the theoretical work backlog of VMC, depicted as lines. The realised work backlog of VMC is depicted as a scatter plot, due to some missing data points. Yellow dashed lines indicate the desirable work backlog. The lower figure depicts the sessions scheduled by the model with bars and the realised sessions of VMC are represented by black lines. The difference between the tactical and operational decisions of the model are illustrated with arrows.*

We see in Figure 6.22 that the work backlog for MRI of the model is lower than VMC and is within the desirable interval from week 50. The model schedules more MRI sessions in the summer holidays and Christmas holidays. For instance, looking at MRI, VMC scales down to 14-16 sessions while the model schedules 21-24 sessions during weeks 31-33.

The theoretical work backlog of VMC is more stable for CT and US. VMC schedules more sessions for CT during weeks 40-44, explaining why VMC's work backlog remains stable while the model's increases. The model might have scaled down the CT sessions due to the arrival of a new CT in weeks 42-43. Namely, during these weeks a lot of 'indirect' shifts were planned for RTs with a CT qualification. The indirect shifts were to help familiarize the RTs with the new CT. Consequently, fewer RTs were available, which might have influenced the model to scale down.

VMC schedules up to 10 sessions more for US starting from week 40. This difference is due to a change in US management: having RTs making ultrasounds instead of radiologists. The RTs need more time for the same appointment, therefore the production was reduced. However, the production used by the model to determine the number of needed sessions was not updated. Explaining, the increase in the model's work backlog.

The work backlog for mammography is predominantly lower for the model. Both the model's and VMC's work backlog for mammography decrease in the beginning of 2024. However, the work backlog of VMC only slightly dips below the desirable interval, where the model decreases towards zero. Lastly, the work backlog for Dexa fluctuates for both VMC and the models. The work backlog of VMC fluctuates consistently around the upper bound of the desirable interval. In contrast, the work backlog of the model fluctuates within the desirable interval and dips below the interval around week 50. This dip can be explained by the decreased number of Dexa orders at the end of the year. VMC schedules 4-5 sessions less at the end of the year, while the model does not scale down.

The differences between the tactical decision and operational decision are visualised with arrows in Figure 6.22 and the proportion of sessions that were not changed from the tactical to the operational session schedule is given per modality in 6.7. We see that MRI is changed less than 3%. The mean fractional change is approximately 0.86 across all methods.

	Modality					
	CT	Dexa	Ultrasound	MRI	Mammo	mean
Sessions Reg.	(17.7,3.3)	(9.3,1.2)	(24.4,3.0)	(26.5,1.1)	(12.5,2.6)	(18.1,7.1)
Sessions Hol.	(14,3.3)	(8.5,1.7)	(22.8,2.9)	(24.8,3.3)	(11.8,1.7)	(16.4,6.9)
Fixed	(0.8,0.33)	(0.89,0.22)	(0.86,0.1)	(0.97,0.07)	(0.77,0.29)	(0.86,0.23)

Table 6.7: This table offers an overview of the number of sessions in regular weeks and holiday weeks and the proportion of sessions that did not change from the operational decision from the tactical decision, categorized by modality and presented for $M = 50$ using the month-based timeline model, and are presented in the format: $(\bar{x}, s)$, with $\bar{x}$ the sample mean and s the sample standard deviation.

The differences between the planned session schedule in the tactical decision and the realised sessions of VMC are plotted in Figure 6.23 for weeks 31-52. Looking at Figure 6.23, in general, more sessions are closed than opened. US, CT, and Dexa sessions are closed most frequently. The model also changes the sessions of these modalities most frequently, as seen in Table 6.7 and Figure 6.22. Some changes overlap, such as closing Dexa sessions during weeks 40-41. Moreover, both VMC and the model frequently close some US sessions. The large peaks in closed sessions for the US of VMC during weeks 36-39 are due to a delayed implemented change in US management. VMC scheduled more US sessions due to the planned change in management, but the change was not implemented until week 40. However, not all changes are similar, the model closes mammography sessions during weeks 44-49, while VMC opens sessions. Furthermore, VMC closes up to 4 CT sessions during weeks 34-40, in Figure 6.22, where the model closes 10 sessions during weeks 40-44. Both large corrections in sessions by the model can be related to the work backlog of the modalities being on the border of the desirable interval, causing a strong reaction due to the nature of the demand constraints.
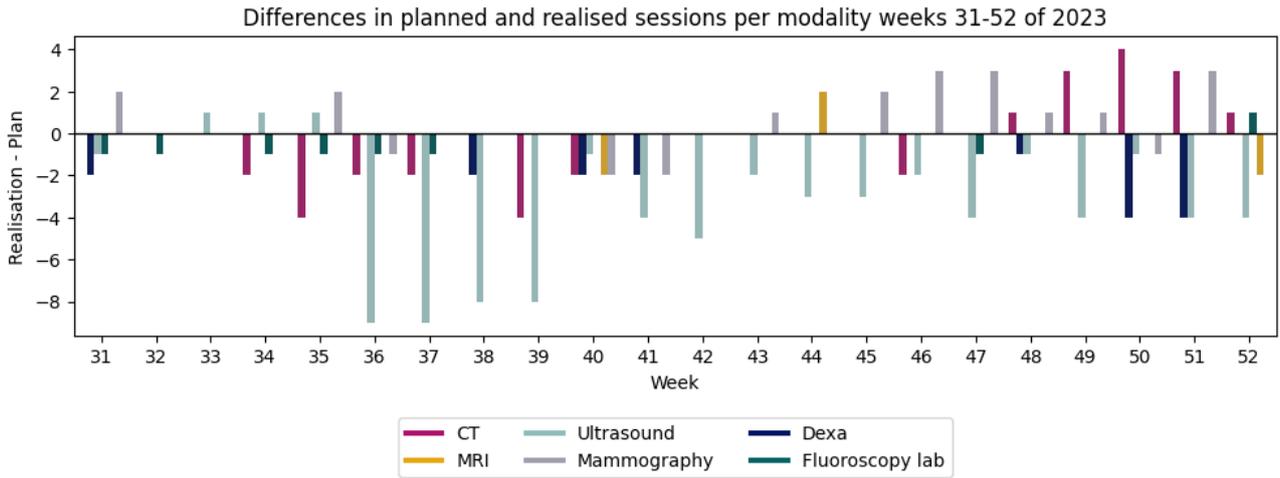
Figure 6.23: This figure depicts the differences in planned (tactical decisions) and realised sessions for weeks 31-52 of 2023 of VMC. Positive bars indicate additional sessions and negative bars indicate canceled sessions.

Besides, these large changes, we see that the order of magnitude of the changes is similar, even though the changes by VMC are made at any point between the tactical decision up to the day of execution and the model's changes are strictly made during the operational decision. This indicates that our model achieves a decent work backlog with less ability to change the session schedule.

The scheduled FTE by $M = 50$, the reported FTE by VMC and the FTE calculated based on the realised sessions and $e(i, t, l)$ are plotted in Figure 6.24. The scheduled FTE by VMC was corrected not to include indirect work, such as education, onboarding, and learning how to operate a new machine. We see that the calculated FTE is approximately equal to the reported FTE.
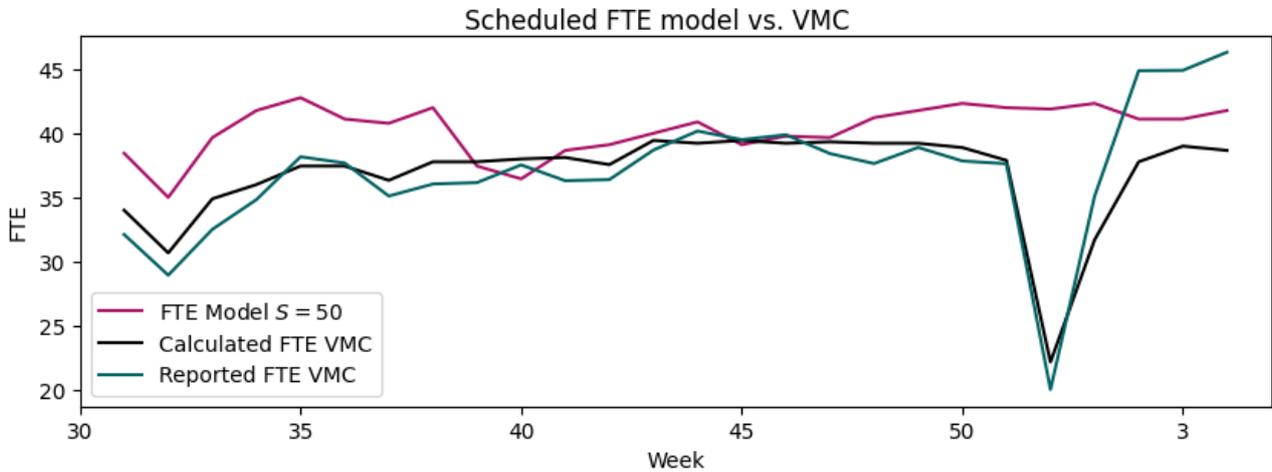

Figure 6.24: This figure depicts the scheduled FTE by the model using $M = 50$, the calculated FTE of VMC based on sessions and reported FTE by VMC over the test period.

Figure 6.24 shows that the model schedules approximately equal FTE as VMC in the middle of the tested horizon. However, in the first weeks and the last weeks of the tested interval, the model schedules approximately 7 FTE more than VMC. Recall that the model had the flexibility to schedule one additional session for an RT each week beyond their availability, albeit with a penalty for doing so. Since the model consistently schedules more FTE than is available, this penalty might have been too low.

The differences in scheduled FTE are largest in holiday weeks. The average number of sessions in regular weeks and holiday weeks scheduled by the model is given in 6.7. We see that the difference between holiday weeks and regular weeks is 2-3 sessions, whereas we see in Figure 6.22 that VMC closes many more sessions. Some of the additionally scheduled FTE in weeks 31 and weeks 32 are used to facilitate more MRI sessions. This causes the model's work backlog for MRI to decrease while VMC's work backlog increases. While these additional sessions

may not reflect real-world constraints, they offer practical insight. Specifically, optimizing work backlogs across all modalities could involve scaling down less for modalities with high backlogs during holiday periods compared to those with lower backlogs.

The RT schedule of VMC before publication and of the model is analyzed in Figure 6.25. This figure depicts the distribution of scheduled hours per modality and the number of qualifications of RTs. Therefore, it provides insight into how the shifts are distributed. The largest difference is observed for CT. Namely, RTs with CT qualification are scheduled less for CT by the model, which may be explained by the fact that the model does not take into account irregular shifts, while VMC does. Moreover, the model schedules RTs with four qualifications less for Application and Dexa and more for MRI. However, in general, the distribution of scheduled hours per qualification of VMC and the model are quite similar. This suggests that the schedules are constructed similarly, implying that the model's schedules are realistic.
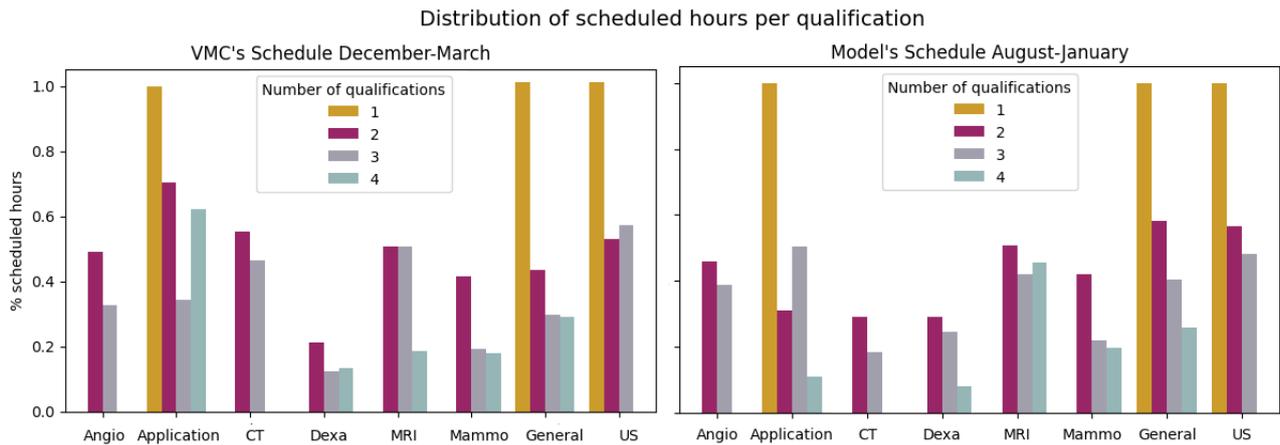

Figure 6.25: This figure depicts the distribution of scheduled hours per modality and number of qualifications of RTs by VMC and the model.

Model vs. myopic methods

In Figure 6.26, we see that the myopic model schedules mostly the same number of sessions for each modality over the entire 6 months. It achieves comparable performance to the flexible methods in doing so. The myopic method generates a lower work backlog for MRI, due to consistently scheduling the maximum number of sessions. Moreover, the work backlog for CT is more constant, due to not scaling down in the weeks 40-43. The myopic method performs worse for US and Dexa, having the work backlog for US increase further above the desirable interval and having the work backlog for Dexa dip below to desirable interval around week 45.

We notice some differences between the two flexible methods. The flexible myopic model, for instance, achieves a more stable work backlog for CT overall, albeit experiencing a dip below the desirable interval at the beginning of 2024. This decrease in work backlog across all models can be attributed to a reduced number of orders for CT in this period. Furthermore, while the work backlog for Dexa in the flexible myopic model dips below the desirable interval during weeks 40-47, the TS maintains its position within the interval. Lastly, the work backlog for mammography exhibits a sharp peak around weeks 40-45, which can be traced back to session cancellations during weeks 40-43. Overall, despite these variations, the performance of the two flexible methods remains quite similar, demonstrating that the flexible method serves as a robust heuristic for the TS.

The scheduled FTE of the flexible methods is approximately the same over the entire test period. The myopic method seems to schedule fewer RTs, except in weeks 40-47 when the flexible models schedule fewer FTE. The reason for this difference is not clear.
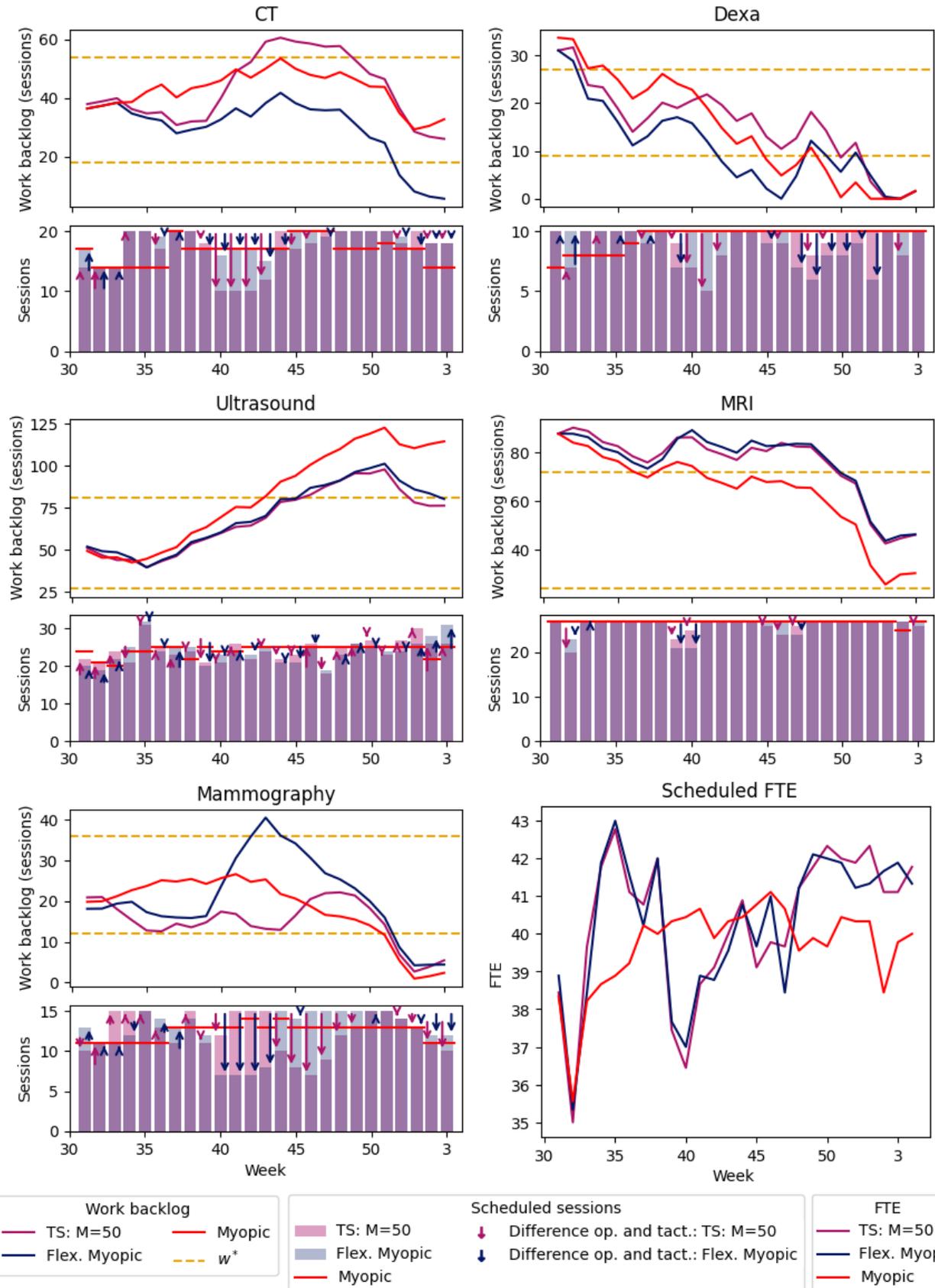
Figure 6.26: *This figure compares the work backlog, sessions and scheduled FTE over time of the two-stage model, the flexible myopic model and the myopic model. Each modality is represented by two figures. The upper figure plots the work backlog per model. Yellow dashed lines indicate the desirable work backlog. The lower figure depicts the sessions scheduled with overlapping bar graphs for the TS and flexible myopic method. The sessions of the myopic method are visualised with horizontal lines. The differences between the tactical and operational decisions of the flexible models are illustrated with arrows. The lower right figure plots the scheduled FTE over the test period per model.*

Computation time

The computation time was measured for six tactical decisions for the five previously used different scenario generation methods using AG and PI. The computation time is plotted in Figure 6.27. The computation time of PI is much lower than the computation time of AG using single scenario methods. If a scenario method with $M > 1$ was employed, the computation time of AG exceeded 3 hours for a single tactical decision. Therefore, these tests were stopped and not included in Figure 6.27. Note, that the model needs to make 6 tactical and operational decisions for one evaluation of the test period. Thus, the computation time of AG for multiple scenario methods would likely exceed 36 hours. The large difference in computation time shows that solving multiple smaller mixed integer linear programs has its benefits over solving one larger mixed integer linear program as soon as the problem size is large enough. Using PI, we see that the computation times increase significantly as the number of scenarios increases.
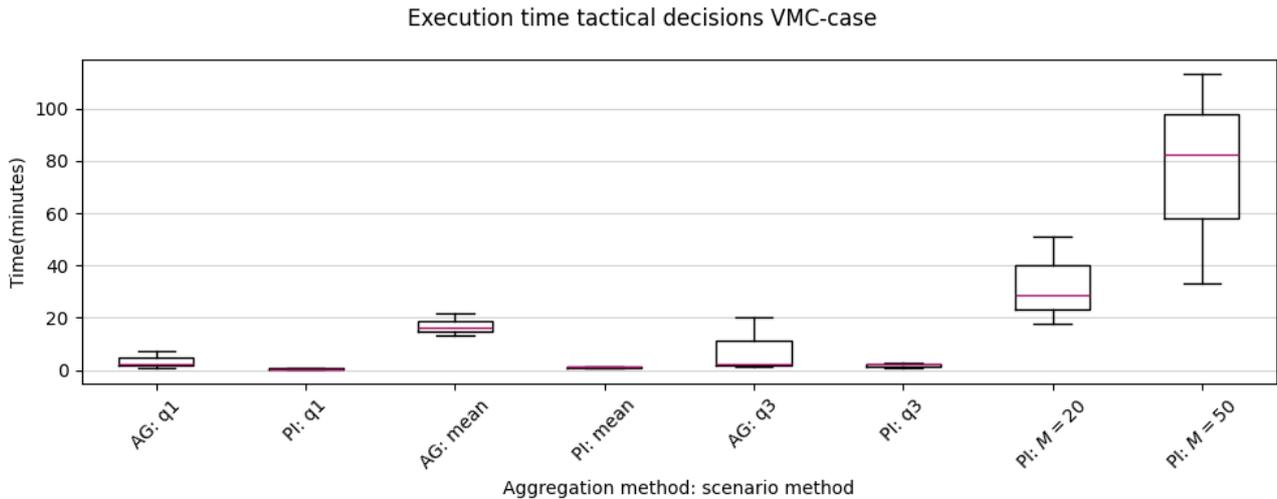

Figure 6.27: *This figure depicts the computation time for different scenario methods for AG and PI. The execution time of the multiple scenario methods using AG is left out because the model did not finish within 3 hours.*

6.5.6   Less flexible model

We investigate the effects of fixing 80% of the mean sessions as given in 6.7. This means that the model must schedule at least 14 CT sessions, 7 Dexa sessions, 19 US sessions, 10 mammography sessions and 20 MRI sessions. We lowered these obligatory sessions with 4 sessions for the holiday weeks 31, 32, 40 and 52. The results are compared to the original model in Figure 6.28.

As a result of the reduced flexibility, we see fewer large changes between the operational and tactical decisions for CT and Dexa in comparison to the original model. Consequently, the peak of the work backlog for CT outside the desirable range of the original model does not occur for the less flexible model. The work backlog for Dexa is the same for the first 8 weeks, but afterwards, the original model scales down, whereas the less flexible model does not. It schedules 10 Dexa sessions, in weeks 41 and 42, even though only 7 sessions are mandatory. As a consequence, the work backlog reduces to below the desirable range.

The sessions and work backlog for MRI are more or less the same. The work backlog for mammography of both models fluctuates around the lower bound of the desirable interval. The less flexible model reduces the sessions in weeks 40-43, while the normal model reduces in the weeks after. The work backlog of the model with reduced flexibility decreases towards zero starting from week 50. Yet, it schedules more sessions than the 10 fixed sessions. The scheduled FTE is approximately the same for both models. We conclude that the performance of a model with reduced flexibility is rather similar and could therefore be applied instead of the original model if preferred in practice.
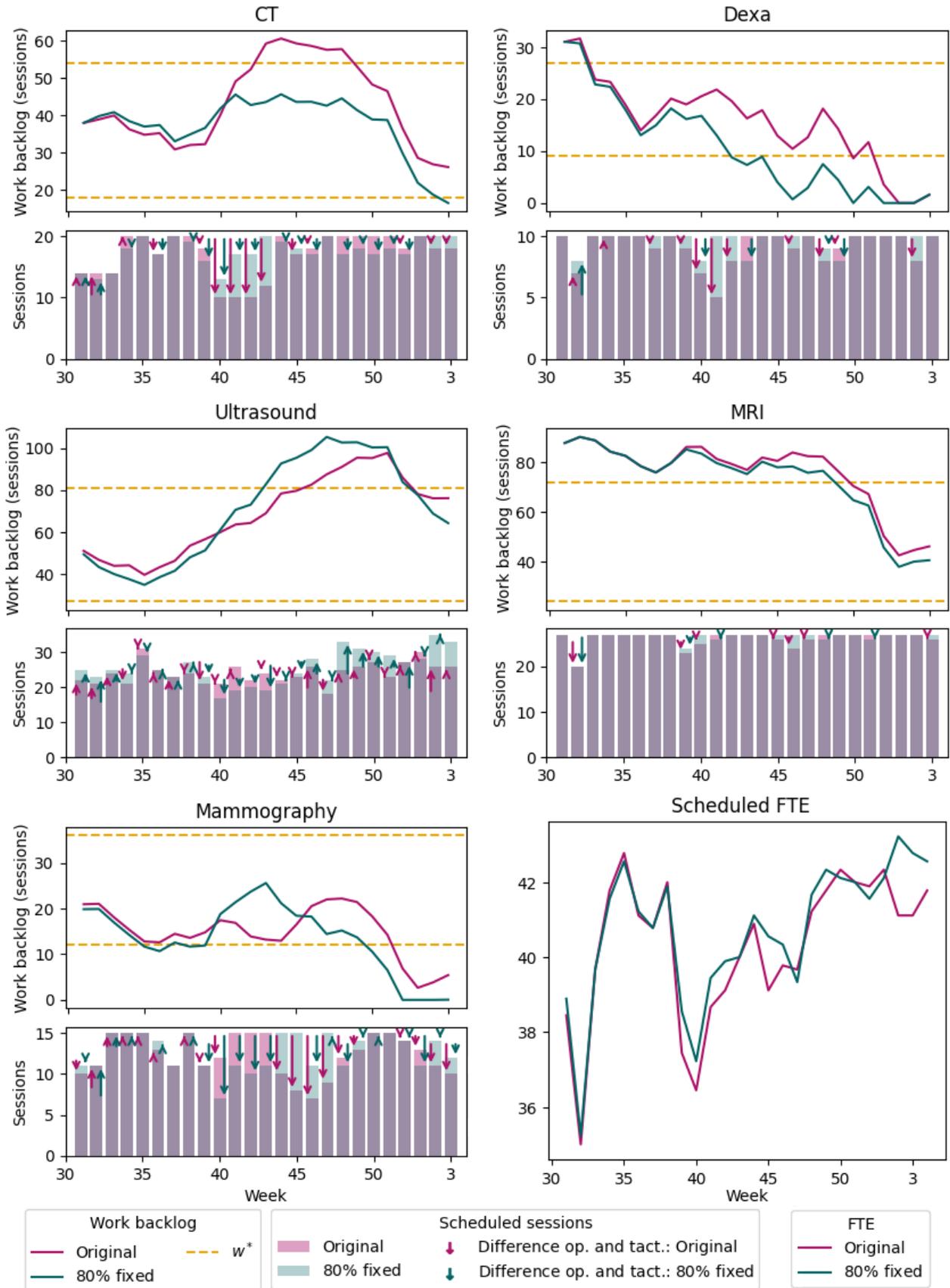
Figure 6.28: This figure shows the work backlog, sessions and scheduled FTE over time for the original model and the model with the flexibility reduced to 80%. Each modality is represented by two figures. The upper figure plots the work backlog per model. Yellow dashed lines indicate the desirable work backlog. The lower figure depicts the sessions scheduled with overlapping bar graphs. The differences between the tactical and operational decisions are illustrated with arrows. The lower right figure plots the scheduled FTE over the test period per model.

6.5.7 Modality capacity

We test the model on the extended modality composition that VMC will acquire in 2024. The extension of the modality capacity includes an additional CT and MRI scanner in Venlo. The productivity of the CT is historically higher in Venlo than in Venray due to a different case mix. The case mix is expected to be different when the new CT arrives, therefore we used a different production for the CTs than currently achieved: the current average is 18 orders per session, whereas the new production is expected to be 14.5. As a result, more sessions need to be scheduled to achieve the same production. The results of the new modality composition are compared to the previous results in Figure 6.29.

We see in Figure 6.29 that approximately 5-10 additional CT sessions are scheduled without drastically decreasing the work backlog, aligning with the decreased production. Furthermore, the extended capacity model schedules more MRI sessions. Consequently, the work backlog for MRI is consistently lower than the original model, and after week 52 it dips below the desirable interval. Thus, we see that the model schedules more CT and MRI sessions, helping to decrease the work backlog of these modalities. Consequently, the model with extended capacity schedules more FTE than the original model starting from week 36. It indicates that to benefit from the extended modality capacity, VMC would need more FTE or change the number of RTs necessary to operate a modality.

The capacity for US, Dexa, and mammography was not changed, however, there are some changes visible. The work backlog for US shows an earlier and stronger peak than the original model. Moreover, the work backlog of mammography shows a sharper peak towards the upper bound of the desirable interval. Lastly, the extended model schedules more Dexa sessions in weeks 41 and 42 and consequently, it has a lower work backlog fluctuating just above zero. We attribute these changes to variations in work backlog intervals, resulting in a different number of goal sessions, as well as alterations in RT availability due to the scheduling of additional CT and MRI sessions.
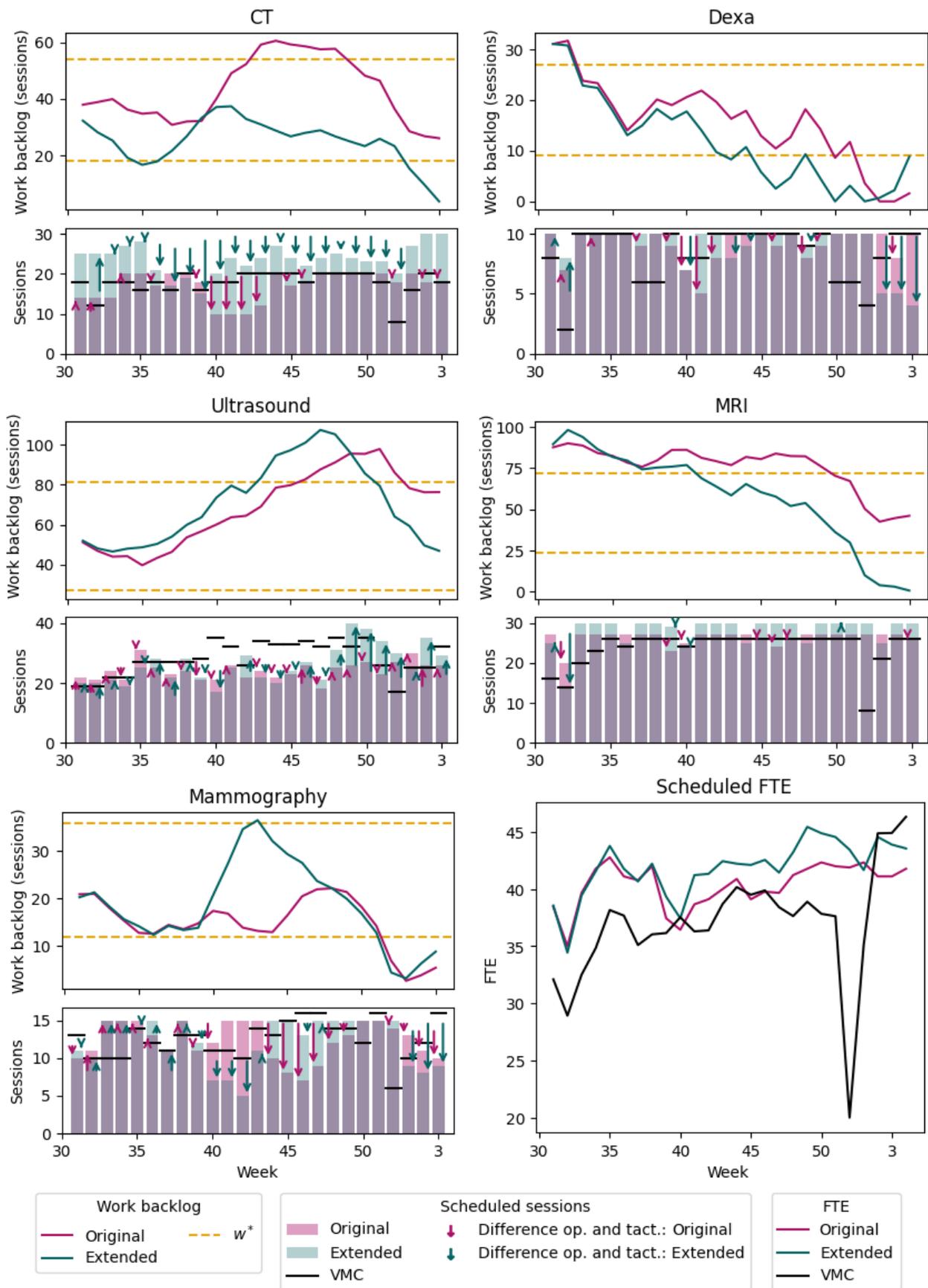
Figure 6.29: *This figure shows the work backlog, sessions and scheduled FTE over time for the original (current) and extended modality capacity. Each modality is represented by two figures. The upper figure plots the work backlog per modality capacity. Yellow dashed lines indicate the desirable work backlog. The lower figure depicts the sessions scheduled with overlapping bar graphs. The difference between the tactical and operational decisions is illustrated with arrows. The lower right figure plots the scheduled FTE over the test period per modality capacity.*

Chapter 7

Conclusions and discussion

7.1 Discussion

In this thesis, we researched the effects of a flexible scheduling framework for a radiology department. First, a mathematical model was developed that describes the different decision levels involved in the scheduling of a radiology department: strategic, tactical, operational online and operational offline. The decision levels and their definitions were taken from Boucherie et al. [3]. Our model links all decision levels. We realize that it is not necessary in practice to take all decision levels into account in for example a strategic decision. However, in doing so the model does provide additional insights into the different components of the scheduling framework, their frequencies, their relations and the involved random variables.

To investigate the effects of introducing flexibility into this framework, we narrowed our scope to the tactical and operational offline level. Moreover, we approximated the resulting multi-stage model with a two-stage model using stage aggregation on a rolling horizon (RH). We opted for this approximation technique because multi-stage stochastic models are virtually intractable for large instances [22], especially if the feasible region of the underlying optimisation problem is a mixed integer set [11]. We did not investigate a bound and can, therefore, not draw any conclusions on how well our aggregate model performs in comparison to a multi-stage stochastic program. Birge and Louveaux [42] show that in most cases, a lower and upper bound can be found for an aggregate form of a multi-stage stochastic program. However, the tightness of this bound can vary, thus significantly affecting the bounding results. Cuisinier et al. [36] found that aggregating long-term decisions in pursuit of balancing long-term and short-term decisions on a RH leads to low and stable computation times. Moreover, they showed robust performance under a sensitivity analysis in comparison to a myopic and retrospective method.

Furthermore, we extended this approximation with the assumption of independent second stages (PI), allowing us to split the second stage into multiple minimisation problems that could be solved in parallel. Numerical experiments showed that the computation time of PI was larger for small instances, likely due to the increased computation overhead. However, for the case study, a real-world problem, PI was much faster as a result of the computation time for the smaller subproblems being much shorter. Therefore, this heuristic offers a large computational advantage. The performance of PI and AG were very similar in the numerical experiments on the test set. A limitation of this comparison was that the specifications of the test case might have eliminated some possible differences between the methods. The differences could be explored by designing test cases that exploit the differences caused by the independence assumption. More interestingly, since PI offers a large computational advantage, it would be of interest to investigate its theoretical properties further.

We compared models with different scenario methods and timelines. Our findings indicated that utilising a single scenario method based on a worst-case scenario, specifically the third quartile (q3), yielded the most favourable outcomes for the test case. Conversely, q3 demonstrated inferior performance compared to the multiple scenario methods in the case study. The method q3 performing well on the test case might be due to problem-specific characteristics. Notably, the case study achieved optimal results with 50 scenarios, the highest number tested. Furthermore, our analysis revealed that myopic flexible scheduling, employing a regular mixed integer linear program, performed comparably well to its stochastic counterpart in most instances. Consequently, this heuristic emerges as a more easily implementable approach with satisfactory performance in practical applications. Nonetheless, our experiments were limited to 50 scenarios, and given that the quality of a stochastic model

typically improves with an increasing number of scenarios [22], further studies incorporating more scenarios would be valuable for further comparison of this heuristic method against the two-stage stochastic model.

Our results showed that implementing a monthly timeline is favourable, which may be explained by the fact that the underlying probabilistic information is rather stationary [26]. Therefore, the effects of having more recent information, as the model using a week-based method has, are minimised. Moreover, the week-based timeline considers only one week in the operational offline decision. As a result, the model corrects slight fluctuations without taking the rest of the month into account, creating more fluctuations in the work backlog. It would be interesting to evaluate the effect of applying the week-based method on a RH that considers a month. In more detail, to take a month of operational offline decisions into account but only fix the decisions regarding one week. Chang et al. [26] compare such a method to a deterministic method and observe superior performance. The theoretical difference in performance between an (approximated) multi-stage stochastic program and an online approach such as a RH has not been settled [26].

The performance of the month-based flexible model was compared to a myopic model, fixating the schedule three months in advance. The results showed that such a myopic model either outperforms the flexible models or shows similar performance. The case study was performed on just one 6-month test period. More results on larger cases are needed to confirm these results in a real-life setting. However, the test case was tested for multiple data instances. The myopic model showed consistently superior performance even when we added noise to the data or decreased the quality of information three months in advance. The long-term scheduling based on the mean demand that the myopic model applies supposedly dampers out the weekly fluctuations in demand. In contrast, the introduced flexible models respond to weekly fluctuations, resulting in overcompensation, which results in larger fluctuations in the work backlog. This effect is commonly known as the bullwhip effect [43]. However, experiments with a demand that deviated from the mean for prolonged periods showed that, in such cases, the introduction of flexibility improved the work backlog over time.

The performance of the proposed flexible method is heavily related to the chosen formulation of the objective and constraints. The model relied on demand constraints based on practical arguments, using fixed intervals for the work backlog and corresponding goals. In addition, these goals were strengthened with the introduction of scaling coefficients. The coefficients and goals were established through iterative experimentation, indicating room for improvement in the optimization process. Moreover, the results showed that the nature of these constraints caused unwanted side effects: boundary effects and strong fluctuations. In further research, demand constraints depending on a continuous function could be investigated to see if this would improve the performance. Additionally, the number of orders, referrals, and production is not constant over time. Consequently, it would be interesting to investigate the merits of changing this framework to an adaptive framework that learns over time. Furthermore, the level of flexibility could be included in the optimisation model: assigning a cost to flexibility would allow the level of flexibility to be optimised. In this report, only the maximal flexibility and 80% of flexibility were considered. The results of reducing the flexibility of our flexible model showed poorer performance. However, these deviations could be explained by the boundary effects of the demand constraints.

The result section of this report focused largely on the effect of the formulated models on the work backlog over time. However, another important effect of adopting such a model is the efficacy of workforce management. Our model predominantly scheduled more FTE than VMC has available, which reduces the comparability of our results to VMC. However, it did provide some valuable practical insights. Firstly, prioritising modalities with a high work backlog in holiday periods can help to get better work backlog levels over all modalities. Secondly, the experiments using the future extended modality composition of VMC indicated that either the available FTE should be increased or the number of required RTs per modality should be adjusted to have the work backlog levels benefit from the extended capacity.

Moreover, we integrated workforce scheduling with session scheduling and included fairness constraints that ensured that the resulting RT schedules were suitable. The results showed that the distribution of shifts over RTs was similar to the current scheduling methods, indicating realistic schedules. Such models are not prevalent in literature [10]. Our AHP analysis and fairness constraints contribute to the understanding of different fairness concepts regarding activity scheduling of RTs. The integration of workforce scheduling and session scheduling meant that the session schedule was adjusted to the available employees. Bentayeb et al. [9] previously showed improved performance in comparison to a real hospital department when integrating patient appointment scheduling and RT scheduling. Our model did not outperform VMC based on work backlog management. This is largely explained by our model's limitations, as explained above, but also due to data anomalies, such as the arrival of a new CT, an unexpected decrease in orders at the end of 2023, and a changed US management, and operational online decisions that were not considered within our model.

However, in terms of workforce scheduling, our results show that it is possible to fix the sessions and employee schedules one month in advance and maintain a decent work backlog over time. Namely, the changes made between the tactical and operational offline decisions were in the same order of magnitude VMC does now, but one month in advance. In practice, the RT schedule will still need to be adjusted due to sickness and such. Nevertheless, our results do show proof of concept that adopting a flexible scheduling framework is possible while maintaining an equally or more stable schedule for the RTs and a decent work backlog.

Therefore, the proposed flexible framework is suitable for implementation in organisations where some of the resources, such as employees, need to be fixed earlier than other resources; it facilitates flexibility in the scheduling framework where possible. Our results do not show an improved performance based on work backlog management, however, they do show that the introduction of flexibility is possible while still maintaining a stable staff schedule. As a result, the organisation can become more flexible which can be beneficial in an environment with changes that are hard to predict.

7.2   Conclusion

In this report, we developed a multi-stage stochastic model to describe the scheduling framework of a radiology department. We proposed a more flexible approach, allowing for the determination of up to 60% of the session schedule one month in advance. This adaptation integrated employee scheduling with flexible session scheduling, dependent on demand predictions and current employee availability, whilst taking into account the fairness of the schedule. It was approximated using stage aggregation on a RH, with independent operational offline decisions. Future research could delve into the theoretical implications of the quality of these approximation methods.

Numerical experiments on a test case showed that the flexible framework using a scenario based on the third quartile is able to maintain work backlogs mostly within the desirable range. A month-based operational decision showed more constant behaviour than a week-based decision that caused stronger fluctuations. An alternative RH method that takes decisions weekly but considers a monthly horizon would be interesting to investigate further. We found that a myopic method that fixes the schedule three months in advance outperforms the flexible framework unless the demand varies significantly from the mean for a prolonged time. The flexible framework was seen to introduce more fluctuations in the work backlog, supposedly caused by the heuristic nature of the demand constraints. Future research could explore alternative demand constraints and objectives to reduce these effects.

Moreover, a case study regarding VieCuri Medical Centre showed that in a real-life example, a myopic method performs comparably to the two-stage stochastic method. We found that using 50 scenarios on a month-based timeline performed best out of the tested two-stage configurations. The heuristic using a regular mixed integer linear program and the flexible scheduling framework showed similar performance to the stochastic two-stage method. VMC's current management maintains a more stable work backlog for most modalities than the proposed flexible model. However, our model cannot make operational online adjustments where VMC can. Therefore, our results show a proof of concept for a flexible method that is able to make operational adjustments earlier and offers the opportunity to adjust the session schedule while maintaining a decent work backlog and an equally or more stable RT schedule.

Bibliography

[1] ABF Reseach. Arbeidsmarktprognoses zorg en welzijn 2021-2031, 2022. URL https://abfresearch.nl/publicaties/arbeidsmarktprognoses-zorg-en-welzijn-2021/.

[2] Hans Langenberg, Chantal Melser, and Marjolein J. Peters. Arbeidsmarktprofiel van zorg en welzijn in 2022. Technical report, CBS, 2023. URL https://www.cbs.nl/nl-nl/longread/statistische-trends/2023/arbeidsmarktprofiel-van-zorg-en-welzijn-in-2022.

[3] Richard Boucherie, Aleida Braaksma, and Gréanne Leeftink. Van strategische visie tot operationele afspraken. *CHOIR Netwerk Zorglogistiek*, 2022.

[4] Nederlandse vereniging ziekenhuizen. Collectieve-arbeidsovereenkomst-ziekenhuizen-2023-2025, 2023. URL https://cao-ziekenhuizen.nl/.

[5] Maartje E. Zonderland, Richard J. Boucherie, Erwin W. Hans, and Nikky Kortbeek. *Handbook of Healthcare logistics*. Springer Charm, 1 edition, 2021. ISBN 978-3-030-60212-3. doi: 10.1007/978-3-030-60212-3.

[6] Michael Pinedo, Christos Zacharias, and Ning Zhu. Scheduling in the service industries: An overview. *Journal of Systems Science and Systems Engineering*, 24(1):1–48, 2015. ISSN 18619576. doi: 10.1007/s11518-015-5266-0.

[7] Edmund Burke, Patrick De Causmaecker, Greet Vanden Berghe, and Hendrik Van Landeghem. The State of the Art of Nurse Rostering. *Journal of Scheduling*, 7:441–499, 11 2004. doi: 10.1023/B:JOSH.0000046076.75950.0b.

[8] Melanie Erhard, Jan Schoenfelder, Andreas Fügener, and Jens O Brunner. State of the art in physician scheduling. *European Journal of Operational Research*, 265(1):1–18, 2018. doi: 10.1016/j.ejor.2017.06.037.

[9] Dina Bentayeb, Nadia Lahrichi, and Louis Martin Rousseau. On integrating patient appointment grids and technologist schedules in a radiology center. *Health Care Management Science*, 26(1):62–78, 2023. ISSN 15729389. doi: 10.1007/s10729-022-09618-z.

[10] Elín Björk Bövarsdóttir, Niels Christian Fink Bagger, Laura Elise Høffner, and Thomas J.R. Stidsen. A flexible mixed integer programming-based system for real-world nurse rostering. *Journal of Scheduling*, 25(1):59–88, 2022. ISSN 10991425. doi: 10.1007/s10951-021-00705-7.

[11] Ping Shun Chen, Ying Jie Lin, and Nai Chun Peng. A two-stage method to determine the allocation and scheduling of medical staff in uncertain environments. *Computers and Industrial Engineering*, 99:174–188, 2016. ISSN 03608352. doi: 10.1016/j.cie.2016.07.018.

[12] Xin Pan, Jie Song, and Bo Zhang. Resource allocation via dynamic admission control in healthcare system. In *IEEE International Conference on Automation Science and Engineering*, volume 2017-Augus, pages 560–561, 2017. doi: 10.1109/COASE.2017.8256163.

[13] Jonathan Patrick, Martin L Puterman, and Maurice Queyranne. Dynamic Multipriority Patient Scheduling for a Diagnostic Resource. *Operations Research*, 56(6):1507–1525, 2 2008. ISSN 0030364X, 15265463.

[14] Rodolfo Benedito Zattar da Silva, Flávio Sanson Fogliatto, André Krindges, and Moiseis dos Santos Cecconello. Dynamic capacity allocation in a radiology service considering different types of patients, individual no-show probabilities, and overbooking. *BMC Health Services Research*, 21(1):968, 2021. ISSN 1472-6963. doi: 10.1186/s12913-021-06918-y.

[15] Z. Abtahi, R. Sahraeian, and D. Rahmani. A Stochastic Model for Prioritized Outpatient Scheduling in a Radiology Center. *International Journal of Engineering, Transactions A: Basics*, 33(4):598–606, 2020. ISSN 17281431. doi: 10.5829/IJE.2020.33.04A.11.

[16] Wineke A M Van Lent, Joost W Deetman, H Jelle Teertstra, Sara H Muller, Erwin W Hans, and Wim H van Harten. Reducing the throughput time of the diagnostic track involving CT scanning with computer simulation. *European journal of radiology*, 81(11):3131–3140, 11 2012. ISSN 1872-7727 (Electronic). doi: 10.1016/j.ejrad.2012.03.012.

[17] Bruno Vieira, Derya Demirtas, Jeroen B van de Kamer, Erwin W Hans, and Wim van Harten. A mathematical programming model for optimizing the staff allocation in radiotherapy under uncertain demand. *European Journal of Operational Research*, 270(2):709–722, 2018. ISSN 0377-2217. doi: 10.1016/j.ejor.2018.03.040.

[18] Ivan B Vermeulen, Sander M Bohte, Sylvia G Elkhuizen, Han Lameris, Piet J M Bakker, and Han La Poutré. Adaptive resource allocation for efficient patient scheduling. *Artificial Intelligence in Medicine*, 46 (1):67–80, 2009. ISSN 0933-3657. doi: 10.1016/j.artmed.2008.07.019.

[19] Thomas R Rohleder and Kenneth J Klassen. Rolling horizon appointment scheduling: a simulation study. *Health care management sciencecare management science*, 5(3):201–209, 8 2002. ISSN 1386-9620 (Print). doi: 10.1023/a:1019748703353.

[20] Osman Aydas, Anthony Ross, Matthew Scanlon, and Buket Aydas. Short-Term nurse schedule adjustments under dynamic patient demand. *Journal of the Operational Research Society*, 74:1–20, 2 2022. doi: 10.1080/01605682.2022.2039566.

[21] Wen-Tso Huang, Ping-Shun Chen, John J. Liu, Yi-Ru Chen, and Yen-Hsin Chen. Dynamic configuration scheduling problem for stochastic medical resources. *Journal of Biomedical Informatics*, 80:96–105, 2018. doi: 10.1016/j.jbi.2018.03.005.

[22] Warren B. Powell. *Reinforcement learning and stochastic optimization A unified framework for sequential decisions*. A JOHN WILEY & SONS, INC., 2019. ISBN 9781119815051.

[23] Pierre Carpentier, Jean Philippe Chancelier, Guy Cohen, and Michel De Lara. *Stochastic Multi-Stage Optimization: At the Crossroads between Discrete Time Stochastic Control and Stochastic Programming*, volume 75. 2015. ISBN 9783319181370.

[24] Willem K. Klein Haneveld, Maarten H. van der Vlerk, and Ward Romeijnders. *Recourse Models*. Springer, 2020. ISBN 9783030292188. doi: 10.1007/978-3-030-29219-5.

[25] Fei Xie and Yongxi Huang. A multistage stochastic programming model for a multi-period strategic expansion of biofuel supply chain under evolving uncertainties. *Transportation Research Part E: Logistics and Transportation Review*, 111:130–148, 2018. ISSN 1366-5545. doi: 10.1016/j.tre.2018.01.015.

[26] Yanbin Chang, Yongjia Song, and Burak Eksioglu. A stochastic look-ahead approach for hurricane relief logistics operations planning under uncertainty. *Annals of Operations Research*, 319(1):1231–1263, 2022. ISSN 15729338. doi: 10.1007/s10479-021-04025-z.

[27] Ahmed Abdelmoumene Kadri, Romain Perrouault, Mouna Kchaou Boujelben, and Céline Gicquel. A multi-stage stochastic integer programming approach for locating electric vehicle charging stations. *Computers & Operations Research*, 117:104888, 2020. ISSN 0305-0548. doi: 10.1016/j.cor.2020.104888.

[28] Alexander Shapiro and A Philpott. A tutorial on stochastic programming. 2007.

[29] Sophie N Parragh, Fabien Tricoire, and Walter J Gutjahr. A branch-and-Benders-cut algorithm for a bi-objective stochastic facility location problem. *OR Spectrum*, 44(2):419–459, 2022. ISSN 1436-6304. doi: 10.1007/s00291-020-00616-7.

[30] Linxuan Shi, Zhengtian Xu, and Miguel Lejeune. An Integer L-shaped Method for Dynamic Order Dispatching in Autonomous Last-Mile Delivery with Demand Uncertainty. pages 1–22, 2022. doi: 10.48550/arXiv.2208.09067.

[31] Julien Keutchayan, Janosch Ortmann, and Walter Rei. Problem-driven scenario clustering in stochastic optimization. *Computational Management Science*, 20(1):13, 2023. ISSN 1619-6988. doi: 10.1007/s10287-023-00446-2.

[32] Pavlo Glushko, Csaba I. Fábián, and Achim Koberstein. An L-shaped method with strengthened lift-and-project cuts. *Computational Management Science*, 19(4):539–565, 2022. ISSN 16196988. doi: 10.1007/s10287-022-00426-y.

[33] Martin Biel and Mikael Johansson. Dynamic cut aggregation in L-shaped algorithms. *arXiv: Optimization and Control*, 2020. doi: 10.48550/arXiv.1910.13752.

[34] Emilia Grass, Kathrin Fischer, and Antonia Rams. An accelerated L-shaped method for solving two-stage stochastic programs in disaster management. *Annals of Operations Research*, 284(2):557–582, 2020. ISSN 15729338. doi: 10.1007/s10479-018-2880-5.

[35] Felipe Silva Placido dos Santos and Fabricio Oliveira. An enhanced L-Shaped method for optimizing periodic-review inventory control problems modeled via two-stage stochastic programming. *European Journal of Operational Research*, 275(2):677–693, 2019. ISSN 0377-2217. doi: 10.1016/j.ejor.2018.11.053.

[36] Étienne Cuisinier, Pierre Lemaire, Bernard Penz, Alain Ruby, and Cyril Bourasseau. New rolling horizon optimization approaches to balance short-term and long-term decisions: An application to energy planning. *Energy*, 245, 2022. ISSN 03605442. doi: 10.1016/j.energy.2021.122773.

[37] Rüdiger Schultz. Continuity and stability in two-stage stochastic integer programming. *K. Marti, editor, Lecture Notes in Economics and Mathematical Systems*, 379:81–92, 1992.

[38] J. Dupačová and K. Sladký. Comparison of multistage stochastic programs with recourse and stochastic dynamic programs with discrete time. *ZAMM Zeitschrift fur Angewandte Mathematik und Mechanik*, 82 (11-12):753–765, 2002. ISSN 00442267. doi: 10.1002/1521-4001(200211)82:11/12⟨753::AID-ZAMM753⟩3. 0.CO;2-5.

[39] Gilbert Laporte and François V Louveaux. The integer L-shaped method for stochastic integer programs with complete recourse. *Operations Research Letters*, 13(3):133–142, 1993. ISSN 0167-6377. doi: 10.1016/ 0167-6377(93)90002-X.

[40] John R Birge and François V Louveaux. A multicut algorithm for two-stage stochastic linear programs. *European Journal of Operational Research*, 34(3):384–392, 1988. ISSN 0377-2217. doi: 10.1016/0377-2217(88) 90159-2.

[41] Gurobi Optimization. Gurobi Optimizer Reference Manual, 2024. URL https://www.gurobi.com.

[42] John R. Birge and François Louveaux. *The Value of Information and the Stochastic Solution*. 2011. ISBN 9781461402367. doi: 10.1007/978-1-4614-0237-4{_}4.

[43] Xun Wang and Stephen M Disney. The bullwhip effect: Progress, trends and directions. *European Journal of Operational Research*, 250(3):691–701, 2016. ISSN 0377-2217. doi: 10.1016/j.ejor.2015.07.022.

Appendix A   Supplements case study

A.1   Results AHP-analysis

Based on the results of the questionnaire the following AHP-matrix was constructed:

$$
\begin{bmatrix}
 & \eta & \kappa & \psi & \gamma & \rho \\
\eta & 1.00 & 0.14 & 0.11 & 0.13 & 0.11 \\
\kappa & 7.00 & 1.00 & 0.12 & 0.12 & 0.11 \\
\psi & 8.80 & 8.40 & 1.00 & 7.00 & 1.00 \\
\gamma & 7.80 & 8.20 & 0.14 & 1.00 & 0.14 \\
\rho & 9.00 & 8.80 & 1.00 & 7.40 & 1.00
\end{bmatrix}
$$

The scores per category were calculated as $\eta = 0.38, \kappa = 0.14, \psi = 0.07, \gamma = 0.03$ and $\rho = 0.37$.

A.2   Data overview

Table A.1 provides an overview of the used datasets, any processing that was done, and the availability of the data.

Dataset	Contains	Preprocessing	Availability
Modality specifications	Location, $\bar{x}, \underline{x}$, $e(i,t,l), k(i)$, and $m(i)$	-	-
RT specifications	FTE and qualifications.	-	-
RT availability	The sick days, holidays, and roster free days and indirect hours.	-	Weeks 31-52 of 2023, weeks 1-3 of 2024. A snapshot of the current status, each Monday at 8:00.
Sessions referring departments	Planned sessions and realized sessions per department per week.	The normal distributions of planned versus realised sessions were determined using data from weeks 1-17 of 2023.	Weeks 1-52 of 2023 and 1-3 of 2024. The rostered sessions were not available, the realized sessions were used instead.
Demand	Work backlog per modality per week, orders per modality per department week.	The orders of weeks 1-17 of 2023 were used to find empirical distributions of the first line and emergency orders. In combination with the realized sessions of referring departments the fraction: orders/session was determined.	Weeks 1-52 of 2023 and weeks 1-3 of 2024. The work backlog of weeks 40, 42, 43, 47, 50 en 51 were missing.
Production	The production per modality per location $q^p(i,l)$ and the average production $\bar{q}(i)$.	The production of 2022 and 2023 until week 18 were used to calculate averages. The realised production of CT and mammography were corrected for emergency orders that misrepresented the true production of sessions.	-
Sessions VMC	Planned sessions and realised sessions	-	Weeks 1-52 of 2023 and weeks 1-3 of 2024.

Table A.1: This Table provides an overview of the used datasets.

A.3   Example schedules case study

In Figure A.1 the resulting schedule from a tactical decision regarding one month consisting of 4 weeks is given. The values for $e(i,t,l), \bar{x}, \underline{x}, V_{i,t}, m(i)$ and $k(i)$ are not disclosed specifically, because of confidentiality.
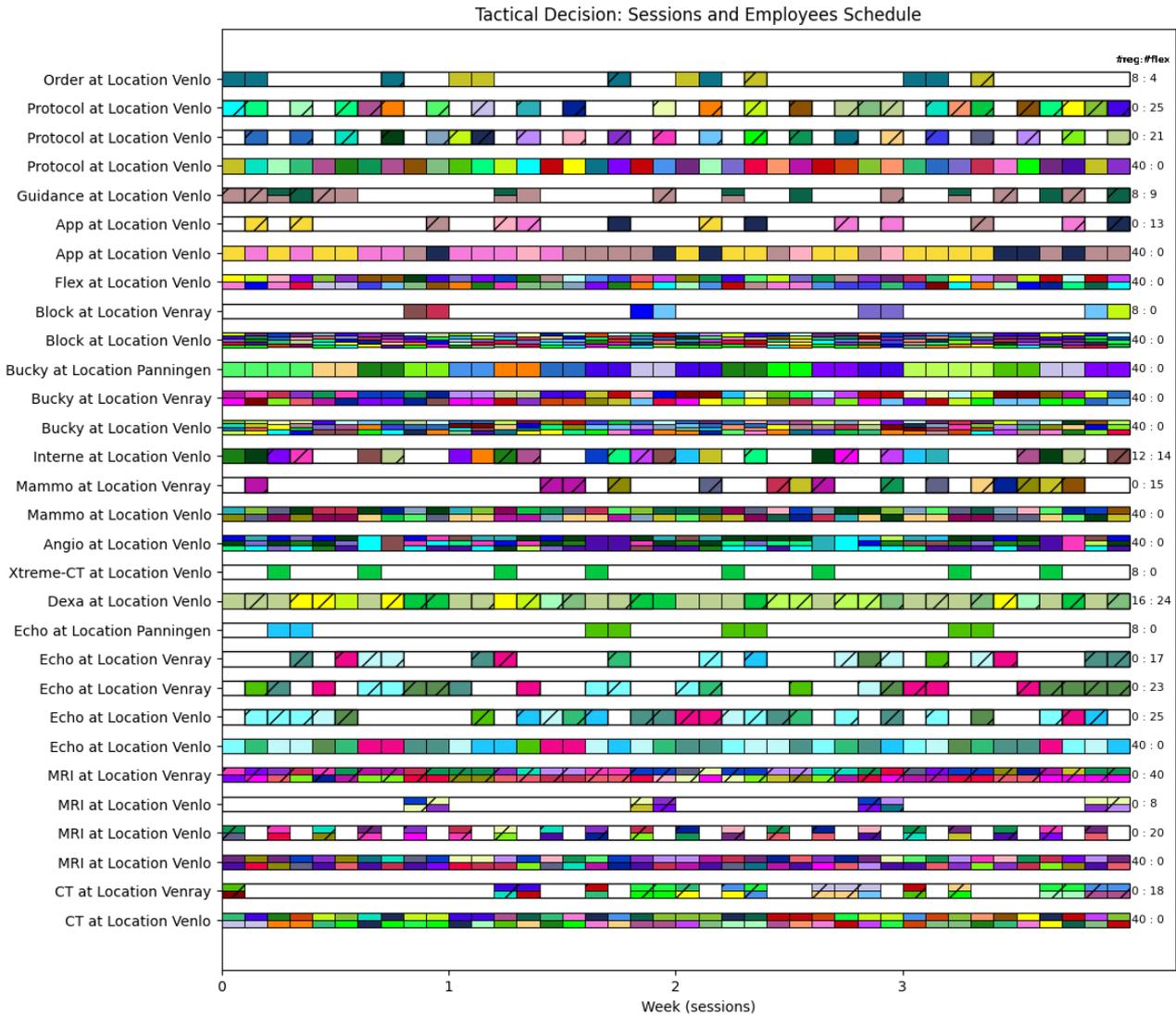

Figure A.1: *This figure depicts the schedule resulting from a tactical decision. Each horizontal bar represents the schedule of a modality at a certain location over time. The sessions are indicated by colour, where each unique colour represents an RT working that modality during that scheduling block. Multiple colours in one scheduling block indicate that multiple RTs are scheduled on that modality. On the right of the figure, the number of fixed (regular) and flexible sessions is printed. The flexible sessions are hatched within the schedule.*

In Figure A.2 the resulting schedule from an operational decision regarding one month consisting of 4 weeks is given. Figure A.1 and Figure A.2 do not correspond to the same decision, but merely show an example schedule of arbitrary months.
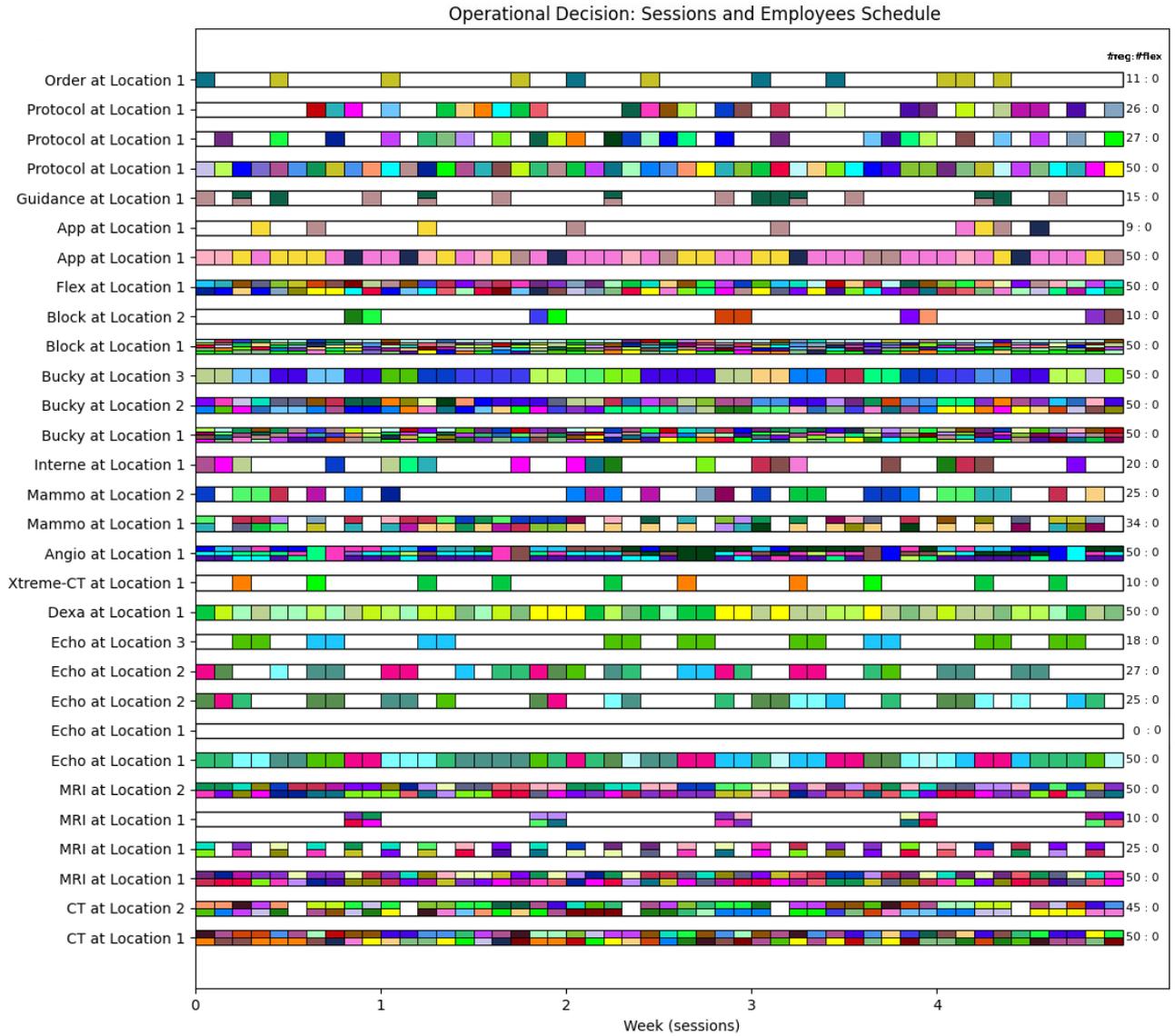

Figure A.2: This figure depicts the schedule resulting from an operational decision. Each horizontal bar represents the schedule of a modality at a certain location over time. The sessions are indicated by colour, where each unique colour represents an RT working that modality during that scheduling block. Multiple colours in one scheduling block indicate that multiple RTs are scheduled on that modality. On the right of the figure, the number of fixed (regular) is printed.