FACULTEIT ECONOMIE
EN BEDRIJFSKUNDE

AN INTEGRATIVE APPROACH TO SCHEDULING THE PRE-TREATMENT AND TREATMENT PHASE IN RADIOTHERAPY

Aantal woorden / Word count: 21.781

Gilles Bauwens
Stamnummer / student number : 01802441

Promotor / supervisor: Prof. Dr. Broos Maenhout

Masterproef voorgedragen tot het bekomen van de graad van:
Master's Dissertation submitted to obtain the degree of:

Master in Business Engineering: Operations Management

Academiejaar / Academic year: 2022-2023

UNIVERSITEIT
GENT

Confidentiality and Scientific Integrity

I declare that the research was conducted in accordance with the rules governing scientific and academic integrity. I have read, and acted in accordance with, the Code of Ethics of the Faculty.

Author: Gilles Bauwens

Foreword

Selecting a topic for my master's dissertation was not an easy task and several interesting topics were waiting to be explored. Nevertheless, it was clear to me that I wanted a topic with societal relevance. This ultimately led me to the healthcare industry, more specifically radiation therapy. Learning more about radiotherapy and the different steps and actors involved was highly rewarding. Furthermore, I can confidently say that I am very pleased with the topic that I have chosen.

First and foremost, I would like to express my gratitude to my supervisor, Professor Maenhout, who was always ready to answer any questions that I had. His valuable feedback and guidance throughout this journey has been immensely helpful. Furthermore, I want to thank the people from the radiotherapy department at AZ Sint-Lucas, who allowed me to gain an increased understanding of radiation therapy. More specifically, I would like to thank Nathalie Deman and Dr. Wim Duthoy for their time and patience, and for providing real-world data. Additionally, I want to thank my friends for their support and for giving valuable advice. In particular, I would like to thank Pieter Dauwe, for proofreading a substantial part of this thesis and for providing feedback. Finally, I want to express my gratitude to my parents for proofreading my thesis, for their understanding and especially for their unconditional love and support.

Gilles Bauwens

Contents

Abbreviations

ADP	Approximate Dynamic Programming
ASAP	As Soon As Possible
CT	Computed Tomography
CTV	Clinical Target Volume
GA	Genetic Algorithm
GTV	Gross Tumor Volume
ILP	Integer Linear Programming
IMRT	Intensity Modulated Radiation Therapy
IP	Integer Programming
K-S	Kolmogorov-Smirnov
linac	Linear Accelerator
LS	Local Search
MDP	Markov Decision Process
MLC	Multi-Leaf Collimator
OAR	Organ at Risk
OF	Objective function
OS	Online Stochastic
OR	Operations Research
PTV	Planning Target Volume
SAA	Sample Average Approximation
S-W	Shapiro-Wilk
VMAT	Volumetric Modulated Arc Therapy

List of Tables

List of Figures

List of Algorithms

1 Introduction

Radiation therapy or radiotherapy in short aims to kill or shrink malignant tumors by applying high-energy radiation. It is commonly used as the main form of treatment for cancer. Furthermore, it can also be used together with other treatment forms, such as medical surgery or chemotherapy. Since the high-energy rays impact not only the tumor, but also the surrounding healthy cells, the total radiation dose is spread out over multiple fractions on consecutive days. If the treatment is successful, the tumor is destroyed or has shrunk. The surrounding tissue will eventually recover ("General information about radiotherapy", n.d.). Roughly 50% of all European cancer patients will at some point in their treatment be advised to undergo radiotherapy treatment. According to Lievens et al. (2020) however, more than one fourth of these patients don't actually receive the treatment they need. Even countries that generally have good access to resources would often experience a substantial gap in the utilisation of these resources, meaning that their actual utilisation is lower than the optimal utilisation (Lievens, 2017). Additionally, large discrepancies between European countries were discovered.

Lievens et al. (2020) identified multiple possible reasons as to why patients might not receive appropriate radiotherapy treatment. Appropriate radiotherapy treatment here means that the radiotherapy treatment has to be on time and effective. Next to patient-related factors such as age, education and a lack of awareness, geography-related factors can have an important impact. Furthermore, economic factors can restrict some population groups to receive the treatment they need. Finally and most relevant to the topic of this dissertation, resource shortages, extensive waiting lists and/or treatment delays may also play a substantial role.

Section 1.1 gives a brief literature overview on the importance of effective scheduling practices. Section 1.2 provides a detailed description of the various tasks that occur in radiotherapy.

1.1 Importance of effective scheduling practices

The impact of not receiving timely and effective treatment can be detrimental. If patients fail to receive the treatment they need, their chances of survival will decrease. Two related dimensions can be identified. First and most obvious, the quality of the radiotherapy treatment has to be (more than) adequate, i.e. the *effective* dimension. Furthermore, patients have to receive treatment in a timely manner, related to the *efficient* dimension. The latter is not only supported by indirect evidence, but also by direct academic evidence. A short overview of important academic readings is given below. Most of the academic research is done for breast and head, and neck cancers (Chen et al., 2008).

Jensen et al. (2007) conducted research on tumor progression in head and neck cancer. The authors found a negative impact of extended waiting times; the increase in tumor volume is significantly correlated to waiting times for radiotherapy treatment. However, they were not able to define a maximum on the waiting times that would avoid volume changes. Mackillop (2007) synthesized different types of evidence; theoretical and experimental, direct clinical and indirect clinical evidence. The author found evidence that delays in radiotherapy treatment increase the probability of local failure for head and neck cancers. More recently, this finding was supported by Liang et al. (2017), where the authors studied the direct impact of waiting times on survival rates in nasopharyngeal carcinoma, a specific type of head and neck cancer. They propose to adopt an 'as short as reasonably achievable' principle to guide radiotherapy waiting times. Additionally, Chen et al. (2008) also investigated the relationship between radiotherapy waiting times and local recurrence, i.e. the risk of cancer re-emerging on or close to the same place as the 'original' cancer. The authors studied the existing direct evidence that supports this relationship. They did not limit the systemic overview to specific cancer types, but instead included all cancer types. However, as most studies pertain to breast and head and neck cancers, conclusions are often limited to these particular cancer types. The main finding of their study is that a delayed radiotherapy treatment start leads to an increase in the risk of local recurrence. No evidence could be found for the negative relationship between waiting times and survival rates in breast cancers. For head and neck cancers however, they did find evidence and hence their findings are in line with Mackillop (2007) and Liang et al. (2017).

In conclusion, it is clear that the waiting times in radiotherapy treatment should be kept to a minimum. This way, the risk of local recurrence and loss of control can be decreased. Furthermore, the probabilities of tumor growth and a decrease in survival rates will not increase. These findings gave rise to a relatively new body of research that is further elaborated on in chapter 2.

1.2 Patientflow

In this section, a brief explanation of the patientflow in a radiotherapy department is given. The exact steps in the patient pathway may differ depending on the hospital and the country. The pathway presented here is based on AZ Sint-Lucas, a hospital in Ghent (Belgium). The concepts are explained using figure 1.

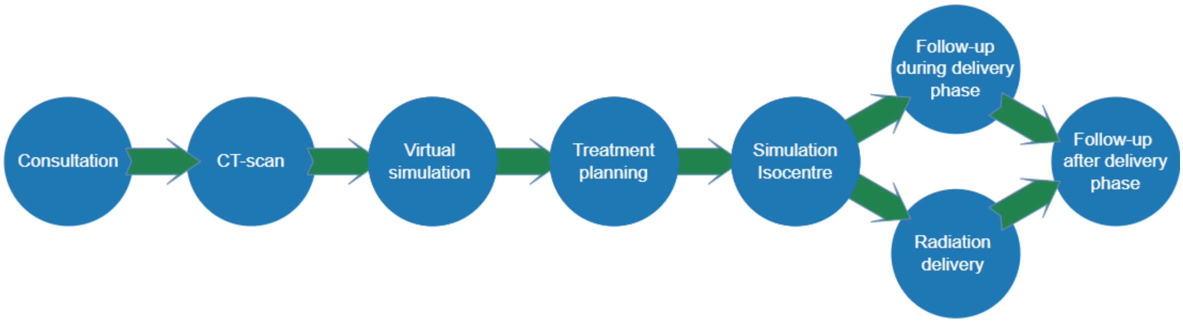


Figure 1: Patientflow in the radiotherapy department

New patient inflow can happen both internally and externally. In both cases, the respective patient case is usually already discussed on a multidisciplinary oncological consult and simultaneously, the decision to advance with radiotherapy has been made. During a first *consultation*, the radiation oncologist determines the treatment intent, e.g. curative or palliative. Radiotherapy can also co-exist next to chemotherapy. Furthermore, an anamnesis and a clinical examination are conducted and a decision on further steps and possibly additional examinations is being made. After the consultation, the patient moves forward to the CT/simulation unit, where appointments are being made for the CT-scan (computed tomography-scan), simulation of the isocentre and the start of the radiation delivery phase. Additionally, the patient receives more information on various practicalities and treatment preparations.

The next phase is the *CT-scan*. In this phase, a patient is positioned on the CT-table. For patients with a planning target volume (PTV) in the head and neck area, it is often necessary to construct a mask or beam direction shell that makes sure that the position of the patient remains stable. An advantage of using masks is that markings can be made on the mask, instead of directly on the skin. Also depending on PTV location is the presence of the radiation oncologist, e.g. she will mark the breast tissue in the case of breast treatment. Furthermore, points of reference are marked on the patient's body to ensure a consistent position. Preferably, all data regarding the positioning of the patient are stored in a central database. After the CT-scan, *virtual simulation* takes place. At this point in time, a nurse contours the organs at risk (OARs), if any, and these have to be checked by the radiation oncologist before the final treatment plan is being made. The radiation oncologist also contours the gross tumor volume (GTV), the clinical target volume (CTV) and the PTV. Based on these, the positioning of the isocentre is determined (see figure 2 for an understanding of the isocentre; the isocentre is the centre point along which the gantry, collimator and table rotate (Kotha et al., 2021)).

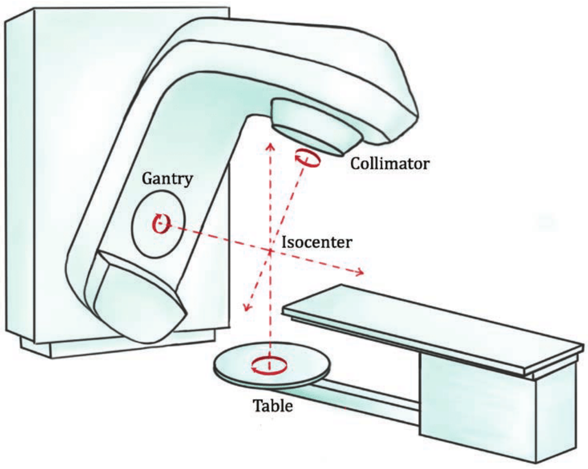


Figure 2: Linear accelerator, image taken from Kotha et al. (2021)

For the next phase, it is important to keep in mind that *planning* is here not related to the operations research meaning of the word, in the sense that it is unrelated to scheduling. During treatment planning, the radiation doses are determined and the radiation oncologist has to approve this dose prescription. Additionally, IMRT - intensity modulated radiation therapy - and VMAT - volumetric modulated arc therapy - optimisations have to be performed in order to complete the planning. More information on this topic is provided in section 2.1.1. Finally, every complete treatment plan has to be approved by the radiation oncologist.

The last phase before the delivery of the radiation can start is the *simulation of the isocentre*. This takes place a few days after the CT-scan. Various measurements regarding the position of the isocentre are done and compared to the measurements from the CT-scan. This comparison is to make sure that a deviation can be corrected in a timely manner. At the end of this phase, the patient receives a precise appointment time (i.e. day and hour) for the start of the treatment.

Prior to starting the actual *delivery of the radiation*, the patient receives some additional information on for example the number of fractions and potential side effects. Moreover, he or she is given the opportunity to ask questions. Next, radiation therapists (RTTs) complete a final checklist to guarantee successful delivery and they correctly position the patient, who is now ready to receive its first dose.

During and after the radiation delivery, regular follow-ups take place in order to ensure successful continuation of the treatment.

Based on the patientflow chart presented in figure 1, the pre-treatment stage involves several steps, including consultation, CT-scan, virtual simulation, treatment planning, and simulation of the isocentre. Afterwards, the treatment phase commences, which includes administering the treatment fractions, as well as monitoring and follow-up during and after the delivery phase. Scheduling the treatment phase has a specific feature that requires patients to undergo their treatment fractions on consecutive working days.

After stating the importance of effective scheduling practices and introducing the radiotherapy department, chapter 2 provides an overview of the literature related to radiation therapy and introduces two research questions. Chapter 3 then formulates the problem description in detail and presents a mathematical model. Three different methods to solve the scheduling problem are subsequently proposed in chapter 4. In addition, chapter 5 applies real-world data by executing several experiments and presents the main findings. Finally, chapter 6 presents the main conclusions and limitations, and provides directions for future research.

2 Literature

Operations research (OR) is defined as "the discipline of applying advanced analytical methods to help make better design and operational decisions for a system, usually under conditions requiring the allocation of scarce resources" (Maenhout, 2020, p. 1). Within radiotherapy treatment, several directions can be identified with regard to the OR literature. An overview is given in figure 3.

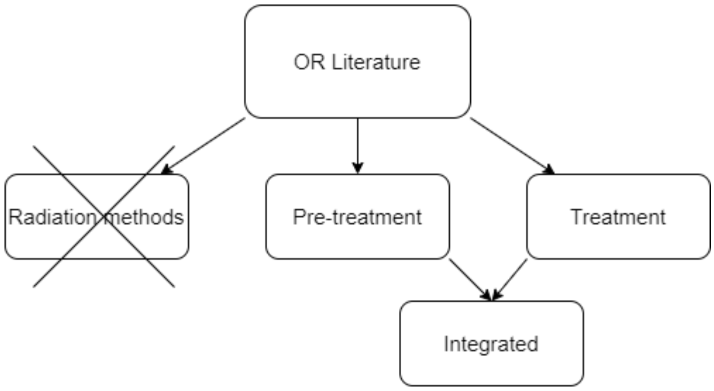


Figure 3: OR literature in radiotherapy

Section 2.1 examines the literature based on these directions. Section 2.2 discusses the literature with specific attention towards solution procedures. Finally, section 2.3 introduces the main contributions to the literature of this dissertation and introduces two research questions in order to materialise the contributions.

2.1 Problem description

In this section, multiple directions in the literature on radiotherapy scheduling are discussed. A first direction is presented in 2.1.1. It briefly explains some of the technologies used in radiation treatment in order to give the reader a more accurate understanding of the topic. In the remainder of this dissertation, however, this body of research is not revisited. In section 2.1.2, a distinction is made between research that handles the pre-treatment phase and literature that is related to the treatment phase. In addition, research is identified that pertains to both phases simultaneously. Section 2.1.3 divides the literature in myopic and dynamic policies. Finally, offline and online scheduling are introduced in section 2.1.4 as two opposing practices.

2.1.1 Radiation methods

The most common form of radiotherapy is external beam radiation, in which the radiation is delivered via a machine that is placed outside of the body ("General information about radiotherapy", n.d.). Another method, called brachytherapy, is by placing the radiation source in the body. In this dissertation, planning of radiotherapy patients refers to the planning of patients that are assigned to an external beam radiation treatment, because the pathway for a brachytherapy patient is significantly different. The radiation is delivered through a linear accelerator (linac) that creates and directs high-energy radiation beams and that is typically placed on a rotating gantry, as illustrated in figure 2 in chapter 1 (Kotha et al., 2021). The goal of using a linac is to minimise the intensity of the beams going through the surrounding OARs, while at the same time making sure that the PTV receives a sufficient amount of radiation. Often, a multi-leaf collimator (MLC) is used to guide the beams in the right direction. An MLC exists, as the name suggests, of several separate leaves that can be positioned independently. The position of all leaves together forms an MLC aperture (Pudsey et al., 2021). An example of an MLC aperture can be found in figure 4. The MLC aperture plays an important role in the optimisation problems discussed in this section. Another major constituent is the intensity of the beams, called fluence weights in the literature (e.g. Akartunalı et al., 2015). The simultaneous optimisation of apertures, fluence weights and beam angles is the subject of most OR literature on radiation methods.

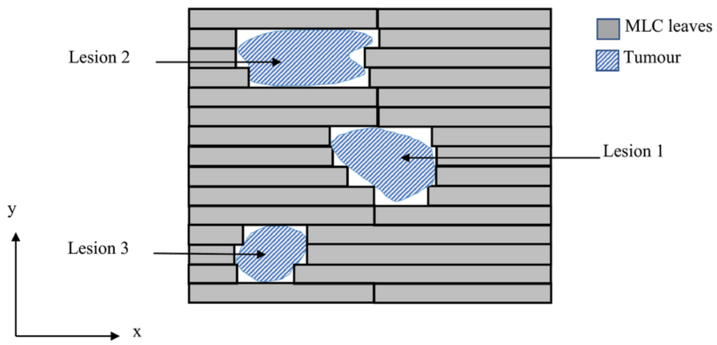


Figure 4: MLC aperture, image taken from Pudsey et al. (2021)

At least two distinctions can be made in the type of rotation of the gantry. Firstly, intensity modulated radiation therapy (IMRT) is a technique that modulates the intensity of the radiation beams and distributes them via a number of discrete directions or angles. IMRT was first introduced by Brahme (1984). Over time, the technique has experienced major developments in order to achieve a higher degree of conformality, i.e. reaching the prescribed dose for the PTV while respecting upper bounds on the surrounding tissue (Bortfeld, 2006). Bortfeld (2006) labels the optimisation of the non-uniform intensities based on the fractionation dose prescription in the PTV and the surrounding healthy structures as 'inverse planning' problems. Even though new methods like volumetric modulated arc therapy (VMAT) have become increasingly important, as of today, IMRT still has its place in radiotherapy centres. More recent research is done by for example Salari and Romeijn (2012) and Bertsimas et al. (2013). VMAT is the second important technique in the field. It builds on the IMRT method and can in fact be seen as a specific type of IMRT (Bedford & Warrington, 2009). Similarly, it modulates the intensities of the radiation beams for different angles or directions. Where IMRT discretely varies the MLC aperture and radiation according to the beam angle however, VMAT does this in a continuous manner; radiation is continuously delivered through one or more so called arcs. An arc is a continuous rotation of the gantry over a specified range, e.g. 360° (Akartunalı et al., 2015). Specialised literature on volumetric modulated arc therapy can be found in for example Peng et al. (2012), Dursun et al. (2016) and Dursun et al. (2019). In addition to these two widely-used techniques, other techniques include tomotherapy, cyberknife, etc. (Akartunalı et al., 2015).

2.1.2 Pre-treatment and treatment

Following the patientflow in figure 1 (chapter 1), the pre-treatment phase comprises of consultation, CT-scan, virtual simulation, treatment planning and simulation of the isocentre. Treatment then encompasses the actual delivery of the fractions and the follow-up during and after the delivery phase. In this section, the literature related to these two phases is discussed. An overview of all papers described in this section can be found at the end of the section, in table 1.

2.1.2.1 Pre-treatment

Kapamara and Petrovic (2009) were among the first to conduct research on scheduling problems in the pre-treatment phase of radiation therapy. The objective of their research was to minimise the total weighted lateness, with weights given to each patient based on his/her priority and lateness being defined as the "difference between the date the patient's details are referred to the centre and the targeted date for the start of their treatment" (Kapamara and Petrovic, 2009, p. 2). The authors developed a two-step solution procedure. First, four different heuristics were developed to construct an initial solution. In the next step, a steepest hill climbing technique was used to improve the initial solution. Technically, their research combines the pre-treatment and treatment phase, since patients are scheduled in the *treatment unit*, in addition to what they presented as the planning, physics and pre-treatment unit. However, scheduling in the treatment phase differs considerably from the pre-treatment phase (cf. section 2.1.2.2). The specific aspects of the latter are not explicitly modelled. Therefore, their research falls under the pre-treatment category in this dissertation. In addition to including lateness in the objective function, Castro and Petrovic (2011) consider a penalty for early idle time of the resources used in the pre-treatment phase. The idea behind this is to ensure sufficient capacity for future arrivals by penalising non-use of early available capacity.

Castro and Petrovic (2012) make use of combined mathematical programming and heuristics to solve the radiotherapy pre-treatment scheduling problem. First, some important terms have to be explained. Waiting times and target waiting times are defined as the final day of the pre-treatment pathway and the preferred final day of pre-treatment for each patient, respectively. Three objectives were considered and ranked according to their perceived importance. This way, the scheduling problem is solved as a series of single-objective optimisation problems. First, a minimisation of the weighted number of patients having waiting times that exceed the target waiting times is done (i.e. the relative amount of patients that finish pre-treatment too late). Then, the maximum lateness, defined as the difference between actual waiting times and target waiting times, is minimised. The final objective is subsequently to minimise the total sum of (weighted) lateness. Since computational issues were observed for a realistic dataset, dispatching rules were constructed to come up with an initial solution. The main advantage of the modelling approach by Castro and Petrovic (2012) is the flexibility to define patient pathways. They focused on a hospital in the United Kingdom that was interested in testing a 10% increase in patient inflow, but the scheduling problem can easily be adapted by other hospitals. Furthermore, recirculation and resource concurrence were included for the very first time in radiotherapy pre-treatment scheduling. Recirculation means that a patient can require the same resource multiple times, while resource concurrence means that an operation might require several resources simultaneously. A noteworthy disadvantage is the lack of integration with the treatment phase. Situations may occur in which a patient fails to meet its due date because of linear accelerator unavailability, despite finishing pre-treatment in time. At the same time, the exact opposite can happen; a patient who can start treatment on a linac, but has to wait for a pre-treatment task to finish. When simultaneously taking pre-treatment and treatment into account, such situations can be avoided (cf. 2.1.2.3).

2.1.2.2 Treatment

The treatment phase consists of the actual delivery of the radiation fractions, follow-up during delivery phase and follow-up after delivery phase. With regard to the scheduling of these tasks, usually, only the delivery of the fractions are included in the formulation of the optimisation problem. Conversely, the follow-up is done in a flexible and dynamic manner and is therefore not subject to modelling. A particular characteristic of treatment scheduling is that patients have to receive their fractions on consecutive working days.

S. Petrovic et al. (2006) use a two-step heuristic approach to assign appointments on linear accelerators to all patients. Patient arrivals are accumulated over the course of a day. At the end of the day, a scheduling decision is made for each of these patients. The first step in the scheduling procedure is to construct a priority list containing all the patients that arrived during the day. To do this, dispatching rules are used. Secondly, an 'as soon as possible' (ASAP) or a 'just-in-time' (JIT) algorithm is used to successively assign each patient in the priority list to its appointment(s).

Conforti et al. (2008) were the first to propose constraint based mathematical models, more specifically integer programming models (IP), to optimise radiotherapy treatment planning. The total pool of patients is divided into patients that are already scheduled and patients on a (prioritised) waiting list. Two models are then presented; a base model that keeps the appointments of already scheduled patients fixed and an extended model that allows for a revision of these appointments. The extended model does specify however that patients who already started treatment finish their remaining fractions, regardless of any revision. In both cases, the primary objective is to maximise the number of new patients that receive a schedule for the next week. A key assumption made in this paper is that the availability of linear accelerators is expressed as temporal blocks of equal size, referred to as a *block-scheduling* system. Despite a clear simplification of reality, a block system is still widely used in radiotherapy centres (Conforti et al., 2011). Furthermore, only one linac is available, although this assumption can easily be relaxed.

In 2011, Conforti et al. extended their 2008 paper by including some additional requirements, e.g. patient availability and patient preferences. Taking into account patient availability is explained to be important because of a potential co-existing therapy. Hereby, a patient can have a clear timeframe that has to be respected so that additional treatment (e.g. chemotherapy or surgery) can be given at the appropriate time. To accommodate for this, they introduced a release time slot and a due time slot for each day. These time slots have a different meaning as compared to what is typically called a release date and a due date (e.g. Castro and Petrovic, 2012). They denote the first and last time slot respectively on a specific day in the planning horizon during which the patient can be scheduled. The objective function is very similar to Conforti et al. (2008). However, an additional term is included; the total number of booked appointments over the scheduling horizon (one week) is also maximised. Following their 2008 paper, a base model and an extended model including revision of already scheduled appointments are proposed.

The mathematical models in the two previous papers by Conforti et al. were constructed based on a block-scheduling system, meaning that a working day is split into several equidistant time blocks, often referred to as *slots* or *timeslots*. Despite the relative popularity of the block system, some inefficiencies might occur as a consequence of the fact that sometimes only a limited portion of a timeslot is necessary to accommodate the treatment delivery. Therefore, the authors in Conforti et al. (2010) use a *non-block* system, which allows for different treatment times to be assigned to patients and is a more accurate representation of the real workload. According to the authors, it was the first time that a full mathematical representation of the non-block radiation therapy treatment scheduling problem is presented. The objective of the optimisation model is to maximise newly scheduled patients while minimising waiting times. Similarly to the extended models in Conforti et al. (2008) and Conforti et al. (2011), patients that are already scheduled and have a treatment plan in progress can be rescheduled, provided that they can continue their treatment. The extension of including additional requirements such as patient availability and patient preferences, as done in Conforti et al. (2011), is not included. However, it is mentioned that the model can be easily adapted to accommodate for this. The main disadvantage of these three papers is that after scheduling, not all patients waiting to start their treatment are guaranteed an appointment. It is possible that their scheduling decision is postponed to a later date.

Saure et al. (2012) used a vastly different approach to solve the problem under study. Instead of using an IP, they propose a discounted infinite-horizon Markov decision process. According to the authors, directly solving the Markov decision process (MDP) to optimality is impossible. Even on small instances, the model is computanionally intractable. Therefore, the discounted infinite-horizon MDP is transformed into a linear programming format. The resulting linear programming in itself does not provide a solution to the sizing problem. However, approximate dynamic programming (ADP) methods exist and such a technique is used by Saure et al. (2012). An important note made by the authors is that they do not schedule specific appointment times, but instead they only book a first day of treatment for each new patient, taking into account capacity requirements for each request. At a later stage, and when all patients are known for a particular day, the individual appointments will be scheduled using a lower-level scheduling technique. More recently, Gocgun (2018) extended the model of Saure et al. (2012) by including cancellation of treatments. The authors also used a slightly different solving procedure, namely a simulation-based approximate dynamic programming technique in contrast to an approximate linear programming approach to ADP. This change is necessitated because of the additional computational complexity by including cancellations of treatment.

Legrain, Fortin, et al. (2015) also researched the radiotherapy treatment scheduling problem. Their objective is to establish the first day of treatment and subsequently the slot. Similar to Conforti et al. (2008), Conforti et al. (2010) and Conforti et al. (2011), the authors propose to use an integer programming model. Another example of using en integer programming model to schedule to treatment phase can be found in Chang et al. (2020). They use an objective function that minimises the total sum of days of first treatment. An example to clarify: imagine two patients that can start treatment on day 5 and 6 respectively and that other constraints do not allow them to start earlier. The objective function value is then equal to $5 + 6 = 11$. The primary disadvantage of this objective is that patient priorities are not taken into consideration. Riff et al. (2016) formulated a model that is closely related to S. Petrovic et al. (2006). The authors aim to schedule appointments once a week and give every patient a fixed schedule of their entire course of treatment. They propose a new method, named 'radiotherapy scheduling with on-the-fly priorities'. The on-the-fly priorities refer to the dynamically updated priority list of patients to be scheduled. A patient's priority increases when the current day reaches his/her last day to begin treatment. The optimisation algorithm comprises of two steps. First, a greedy algorithm goes over each working day in the current week and schedules new patients and patients in the waiting list to determine an initial solution to the planning problem. Then, a local search algorithm is used to improve this initial solution. The local search algorithm applies three successive hill climbing approaches.

In Vogl et al. (2019), the scheduling of a more expensive form of radiation therapy is researched. The authors model an ion beam facility. Only one machine that serves several treatment rooms is used, instead of one machine per treatment room. Because of this particular configuration, the nature of the optimisation problem is different; the use of the particle beam machine has to be distributed effectively between the different treatment rooms, in order to maximise utilisation. Although their scheduling problem has some distinctively different characteristics compared to the one investigated in this master's dissertation, certain elements of the formulation can be useful nonetheless (cf. chapter 3). Regarding the solution methods, the authors propose three metaheuristics: a genetic algorithm, an iterated local search procedure and a combination of the two previous methods. The latter performs best on-real world instances.

Sharing our observation that only Legrain, Fortin, et al. (2015), Legrain, Widmer, et al. (2015), Saure et al. (2012) and Gocgun (2018) incorporate a stochastic mechanism in their approach (see section 2.1.3), Pham et al. (2021) identify that both the online stochastic (OS) optimisation models used in the former two and the MDP models in the latter two are difficult to scale up and hence inappropriate to use in large radiotherapy centres. Therefore, they propose a prediction-based approach to solve the treatment problem. By applying machine learning techniques on a large set of instances, the authors can utilize optimal offline solutions to predict when an arriving patient should receive their first fraction. Their proposed approach is easily scalable and allows for large instances of up to eight linear accelerators.

2.1.2.3 Pre-treatment and treatment integrated

When optimising the scheduling of either the pre-treatment phase or the treatment phase in isolation, the results might be too one-sighted. Failing to integrate both phases in decision making may lead to undesirable consequences. Two examples are given to illustrate this potential drawback. On the one hand, it can occur that a patient receives a schedule to start treatment on the first day of week 2 in the scheduling horizon. If there is not sufficient capacity to finish all the pre-treatment tasks before this date however, the patient will experience an unexpected delay. In addition to a potential decrease in survival rate, this can cause serious psychological distress, which is highly unwanted (Mackillop, 2007). On the other hand, the final step in the pre-treatment phase prior to starting the delivery of the radiation fractions is simulation of the isocentre, as described in section 1.2. As a general practice, it is best not to have too many days between this step and the start of treatment to ensure that all parameters such as points of reference, tumor size etc. are still valid. In AZ Sint-Lucas for example, the time window is a maximum of two days. Therefore, a close integration is needed to ensure that the time window constraints are not violated. The simultaneous optimisation of pre-treatment and treatment scheduling is referred to as 'radiotherapy scheduling' in the remainder of this section.

Kapamara et al. (2006) were among the first to focus on scheduling problems in radiotherapy. The goal of their research was to report on the main elements of radiotherapy scheduling and to provide directions for future research. They acknowledged the importance of combining both the pre-treatment and treatment phase into the problem formulation. Furthermore, the authors reviewed different methods that can be used to solve the scheduling problem. Among the reviewed methods are heuristics (e.g. dispatching rules, simulated annealing) and metaheuristics (e.g. tabu search, genetic algorithms).

In D. Petrovic et al. (2011), the authors propose genetic algorithms to solve the radiotherapy scheduling problem and included multiple objectives. Both the average of patient waiting times and the average of patient lateness are minimised. They defined waiting time as the time between the first day of delivery of treatment and the day that the decision to use radiation therapy as form of treatment was made. Lateness on the other hand, is calculated as the difference between actual waiting times and target waiting times, similar to Castro and Petrovic (2012). Resource recirculation is only possible on the treatment machines; the linear accelerators. The last operation, defined as "the servicing of a patient on a machine or facility" (D. Petrovic et al., 2011, p. 6995), for each patient is the delivery of the first fraction. It is assumed that the remaining fractions are given each working day after the day of the first fraction. The authors propose three different genetic algorithms; a standard genetic algorithm that does not take into account differences in patient priority, a knowledge-based genetic algorithm that utilises domain knowledge to prioritise emergency patients and a weighted genetic algorithm that gives different weights to the three patient categories (radical, palliative and emergency). Overall, the knowledge-based genetic algorithm performed best.

In Legrain, Widmer, et al. (2015), the authors extended the research from Legrain, Fortin, et al. (2015) by taking into account two dosimetry tasks that are performed prior to the treatment on linear accelerators. Dosimetry is defined as "pre-treatment steps that consist mainly in planning the shape, the intensity, and the direction of the beams of the linear accelerator" (Legrain, Widmer, et al., 2015, p. 1-2). It can be seen that their view on pre-treatment is slightly different as compared to the pre-treatment steps identified in this dissertation (see figure 1, chapter 1). Within dosimetry, they identified two tasks, performed by two different dosimetrists: treatment preparation and verification. The authors then modeled the problem as a flow-shop process with two different 'machines': dosimetry and linear accelerators. The dosimetry part is solved using a genetic algorithm.

An overview of the literature in relation to the main phases in radiotherapy scheduling reveals the scarcity of papers that discuss both the pre-treatment and treatment phase in an integrated manner, despite Kapamara et al. (2006) emphasising the importance of doing so (table 1). Upon closely examining the papers by D. Petrovic et al. (2011) and Legrain, Widmer, et al. (2015), a research gap was identified. Both research papers propose a very specific model for the pre-treatment phase and hereby overlook the differences in scheduling practises across different radiotherapy departments. The aim of our research is to fill this gap by developing a model that not only integrates the pre-treatment and treatment phase, but also allows for straightforward adaptations according to the needs of an individual radiotherapy department.

Paper	**Pre-treatment**	**Treatment**	**Integrated**
Kapamara and Petrovic (2009)	x		
Castro and Petrovic (2011)	x		
Castro and Petrovic (2012)	x		
S. Petrovic et al. (2006)		x	
Conforti et al. (2008, 2010, 2011)		x	
Saure et al. (2012)		x	
Gocgun (2018)		x	
Legrain, Fortin, et al. (2015)		x	
Riff et al. (2016)		x	
Vogl et al. (2019)		x	
Chang et al. (2020)		x	
Pham et al. (2021)		x	
Kapamara et al. (2006)			x
D. Petrovic et al. (2011)			x
Legrain, Widmer, et al. (2015)			x

Table 1: Literature - relating to the phases in radiotherapy scheduling

2.1.3 Myopic vs dynamic

Two contrasting perspectives on scheduling issues are myopic scheduling and stochastic or dynamic scheduling. In myopic scheduling, there is no consideration for how today's decisions will affect the future state of the system. In contrast, stochastic or dynamic, in case of a Markov decision process, scheduling typically generates better schedules by taking into account the impact of today's scheduling decisions on the future state of the system (Saure et al., 2012).

Castro and Petrovic (2012) don't take future patient arrivals into account and hence, their modelling approach can be seen as myopic, as opposed to dynamic or stochastic. This could result in a suboptimal situation, where inclusion of future patients would have resulted in a different, and more appropriate schedule. Additionally, the authors in Conforti et al. (2008), Conforti et al. (2010), Conforti et al. (2011) and Vogl et al. (2019) also used a myopic approach that ignores the impact on future arrivals of decisions made today. Saure et al. (2012) contributed to the literature by modelling the radiotherapy scheduling problem as a discounted infinite-horizon Markov decision process. This modelling technique is dynamic in nature. Similarly to Saure et al. (2012), Gocgun (2018) also takes a dynamic outlook on the scheduling decisions. Both Legrain, Fortin, et al. (2015) and Legrain, Widmer, et al. (2015) use stochastic information on patient arrivals to incorporate knowledge about the future state of the system. The method used in Pham et al. (2021) can also be classified as stochastic, although it uses a different method to incorporate stochastic information. Instead of building scenario's of future arrivals, as done in both Legrain, Fortin, et al. (2015) and Legrain, Widmer, et al. (2015), the authors use full offline optimal solutions for past, known arrivals to predict when new patients should receive their first fractionation dose, thereby implicitly taking future patients into account.

The overview in table 2 indicates that recent research articles consider a dynamic or stochastic approach. In these articles, considering the impact of today's scheduling decisions on future arrivals is shown to outperform the myopic paradigm. Therefore, one of the solution techniques proposed in this thesis follows the stochastic paradigm and consequently contributes to the body of literature on stochastic radiotherapy scheduling.

Paper	Myopic	Dynamic/stochastic
Kapamara and Petrovic (2009)	x	
Castro and Petrovic (2011)	x	
Castro and Petrovic (2012)	x	
Kapamara et al. (2006)	x	
S. Petrovic et al. (2006)	x	
Conforti et al. (2008, 2010, 2011)	x	
Saure et al. (2012)		x
Gocgun (2018)		x
Legrain, Fortin, et al. (2015)		x
Riff et al. (2016)	x	
Vogl et al. (2019)	x	
Chang et al. (2020)		x
Pham et al. (2021)		x
D. Petrovic et al. (2011)	x	
Legrain, Widmer, et al. (2015)		x

Table 2: Literature - relating to myopic vs dynamic scheduling

2.1.4 Offline vs online

An important difference when examining research on the radiotherapy scheduling problem is whether or not patients immediately receive their appointments upon issuing a treatment request. If not, we refer to the scheduling as offline scheduling. 'Full' offline scheduling means that all treatment requests are known in advance. In practice, some hospitals (e.g. AZ Sint-Lucas) want to immediately give the appointment dates and times to their patients. This necessitates online scheduling which indicates that upon receiving a treatment request, only the current state of the system, formed by previous requests, is known and that the patient is immediately assigned its booking(s). This makes online scheduling inherently inferior to offline scheduling. In some hospitals, patients have to wait a certain amount of time to get their appointments. Often, they receive a phone-call by the end of the day (e.g. Castro and Petrovic (2012)) or week (e.g. Riff et al. (2016)) with specific times for their appointments. In that case, incoming patient requests are collected and the optimisation model is being run at the end of the day or week, when all patient requests of that day or week are known. This is referred to as *batch scheduling*. In this dissertation, offline scheduling is used interchangeably for batch scheduling and 'full' offline scheduling. Solving the 'full' offline paradigm to optimality provides a lower (upper) bound for the minimisation (maximisation) problems.

In Castro and Petrovic (2012), patients are accumulated on a day-to-day basis (cf. batch scheduling). Additionally, Conforti et al. (2008), Conforti et al. (2010) and Conforti et al. (2011) also use a batch scheduling approach. Moreover, Saure et al. (2012), a booking agent has to wait for the end of the day, when all patient request are known, to do the scheduling exercise. Furthermore, the booking agent can postpone some of the decisions to the next day, leaving the principal drawback of Conforti et al. (2008), Conforti et al. (2010) and Conforti et al. (2011) intact. Riff et al. (2016) propose to accumulate requests on a weekly basis, hence also following the batch scheduling paradigm. Vogl et al. (2019) use metaheuristics to solve the offline problem instance.

The approach from Legrain, Fortin, et al. (2015) is different to what has been done before in at least one way; patients are scheduled in an online manner. This means that upon (sequential) arrival of each patient, (s)he is almost immediately given a schedule. The schedule here refers to a first day of treatment and a timeslot. In addition, all patients are guaranteed a schedule and no scheduling decisions are postponed. Legrain, Widmer, et al. (2015) build on the same online paradigm and include the integration of the pre-treatment phase in the scheduling decision.

The overview regarding the offline or online approaches in table 3 indicates that relatively few papers incorporate an online paradigm into their scheduling approaches. Consequently, the research in this dissertation contributes to the literature by considering online techniques, in addition to the offline paradigm.

Paper	Offline	Online
Kapamara and Petrovic (2009)	x	
Castro and Petrovic (2011)	x	
Castro and Petrovic (2012)	x	
Kapamara et al. (2006)	x	
S. Petrovic et al. (2006)	x	
Conforti et al. (2008, 2010, 2011)	x	
Saure et al. (2012)	x	
Gocgun (2018)	x	
Legrain, Fortin, et al. (2015)	x	x
Riff et al. (2016)	x	
Vogl et al. (2019)	x	
Chang et al. (2020)	x	x
Pham et al. (2021)	x	x
D. Petrovic et al. (2011)	x	
Legrain, Widmer, et al. (2015)		x

Table 3: Literature - relating to offline vs online scheduling

2.2 Solution methods

In this section, the literature is discussed in relation to solution procedures that can be used to solve the scheduling problem. In section 2.2.1, some of the techniques that are commonly used in offline radiotherapy scheduling are described. The solution procedures for online scheduling are similar to those used in offline scheduling. The difference is primarily related to the input; in addition to the patient requesting treatment, scenarios of future patient arrivals are generated and given as input to the model. Section 2.2.2 consequently briefly elaborates on the techniques used in online radiotherapy scheduling.

2.2.1 Offline

Various approaches have been taken in the existing body of literature on radiotherapy scheduling in an offline manner. These studies share the common assumption that all treatment requests are known in advance. Upon assuming this, the knowledge on future patient arrivals can be taken into account when making scheduling decisions. It is worth noting that it is unrealistic in the setting of a radiotherapy department to assume that all patient arrivals can be anticipated with 100% certainty. Nevertheless, finding the offline solution is proven to be useful. In many research articles, the offline solution is used as a benchmark to test the performance of more realistic approaches (e.g. Pham et al., 2021; Legrain, Fortin, et al., 2015).

2.2.1.1 I(L)P

In the literature, a common method is solving the optimisation problems via their integer (linear) programming formulation directly. Various optimisation software tools exist that provide efficient ways to run a mathematical programming model, e.g. CPLEX by IBM or Gurobi. The IP solver by CPLEX is used in for example Conforti et al. (2011), Castro and Petrovic (2012) and Legrain, Fortin, et al. (2015). In contrast to CPLEX, Gurobi does not offer a dedicated user interface. Instead, an API that can be seamlessly integrated with various programming languages is available. In this master's dissertation, the Gurobi API is used and integrated in Python, a widely used programming language.

When dealing with moderate to large problem instances, a drawback of using the I(L)P approach arises. In small problem instances, the problem can be solved in a timely manner by using all available data as input to the ILP formulation. This perk disappears when more realistic data instances are used, as a consequence of rapid growth in problem size. As a result, the time constraint issue, as described in El-Omari (2021), is present. In their research, the authors identified several issues that instantiate the need for alternative solution procedures. Among these are *time constraint*, indicating that there is a limited time window to solve the problem under study, *resource constraint* or being restricted in the computational resources and *problem difficulty* i.e. solving a problem that is rather complex in nature. In the following sections (2.2.1.2, 2.2.1.3 and 2.2.1.4), three alternative solution techniques are explained.

2.2.1.2 Heuristics

Heuristics are, following the explanation in El-Omari (2021), alternative solution techniques that aim at finding a (near-)optimal solution to the scheduling problem within a context dependent reasonable time frame. Compared to the class of metaheuristic algorithms described in the next section, heuristics are deemed to be more context specific (Abdel-Basset et al., 2018).

Castro and Petrovic (2012) use a combination of mathematical programming and heuristics to solve the pre-treatment scheduling problem; the heuristics are used to provide an initial solution that can be used as input to the mathematical programming formulation. Without using heuristics to come up with a good initial solution, the authors experienced that no feasible solution could be found within at least three hours by using the IP solver. Even in the rare case that an initial solution was obtained within a reasonable time frame, acquiring optimality in a timely manner became the new critical issue. In total, six so called dispatching rules were developed. These dispatching rules are used to construct a priority list containing all patients, where the position of each patient in the list denotes its rank when making the scheduling exercise. Upon iterating over the priority list, each selected patient has its operations booked in the earliest available timeslots that meet the constraints, hence following an ASAP procedure. Interestingly, S. Petrovic et al. (2006) used a similar approach to solve the treatment problem. An overview of the dispatching rules is presented in table 4. The description of these rules is adapted to the context of this dissertation.

Dispatching rule	**Abbreviation**	**Description**
Random	/	The priority list is constructed in a random manner
Waiting list status - earliest due date first	WLS-EDD	Patients are sorted based on their priority weight w_j, that in turn is based on their treatment intent. In case of indecisiveness, earliest due date has priority.
Waiting list status - latest doctor appointment	WLS-LD	Patients are ranked based on w_j. Ties are broken based on latest *estimated** pre-treatment operation.
Waiting list status - latest completion day	WLS-LC	Patients are sorted based on w_j. Ties are broken based on latest *estimated** treatment completion day.
Waiting list status - earliest doctor appointment	WLS-ED	Patients are ranked based on w_j. Ties are broken based on earliest *estimated** pre-treatment operation.
Waiting list status - earliest completion day	WLS-EC	Patients are sorted based on w_j. Ties are broken based on earliest *estimated** treatment completion day.

Table 4: Dispatching rules (adopted from Castro and Petrovic (2012))

*Estimations are made by taking into account current capacity on the resources and patient characteristics.

In the research by Castro and Petrovic (2012), the authors used data from the Nottingham City Hospital to construct both realistic and non-realistic instances and scenarios. Based on these instances, it was found that the WLS-LD rule performed best, compared to the other dispatching rules. To tackle the fact that optimality could not be achieved in a reasonable time frame, even after constructing an initial solution, a time limit and gap tolerance (120 seconds and 5%, respectively) were set.

Similarly to Castro and Petrovic (2012), Kapamara and Petrovic (2009) used heuristics to construct an initial solution. A total of four heuristics were defined, combining dispatching rules and other strategies to construct a priority list of all patients waiting to start treatment.

When the priority list is established, a greedy approach is used to schedule the operations. The authors have identified four phases or units in the radiotherapy department; planning unit, physics unit, pre-treatment unit and treatment unit. Each of these units has a specific heuristic procedure. Instead of using the initial solution as input to an integer programming model however, the authors used a hill climbing approach which can be classified as a meta-heuristic, more specifically a local search algorithm, to further improve the solution. This iterative improvement method is explained in the section on metaheuristics (2.2.1.3).

A third radiotherapy scheduling usecase of heuristics can be found in Riff et al. (2016). The authors propose an innovative algorithm, named 'radiotherapy scheduling with on-the-fly properties'. This method uses, like Kapamara and Petrovic (2009), a combination of heuristics and metaheuristics. Similarly to both Castro and Petrovic (2012) and Kapamara and Petrovic (2009), the patient waiting list is prioritised based on a dispatching rule; the priority increases when the current day gets closer to the last possible day to start treatment. The heuristic part of the algorithm applies to this dispatching rule and the greedy manner in which requests are handled on each day. The greedy approach works as follows. When iterating through the scheduling horizon, each patient in the current day's waiting list gets scheduled, if it adheres to the constraints of the model. If not, he or she is put in the waiting list for the next day. After the application of the greedy algorithm, a local search method is implemented. The local search method and its difference compared to the local search method in Kapamara and Petrovic (2009) are presented in the next section.

2.2.1.3 Metaheuristics

Metaheuristics are, like heuristics, methods that search for a close-to-optimal solution within a reasonable timeframe. According to Abdel-Basset et al. (2018), the main difference between the two is that metaheuristics are more problem independent. Hence, they can be applied to various different optimisation problems with only a few adaptations. In fact, the authors argue that heuristics can be used by metaheuristics as domain specific building blocks in the overarching algorithm. Two strategies are often used to describe metaheuristic algorithms. On the one hand, the intensification strategy has the goal of improving the solution as much as possible. On the other hand, the diversification strategy searches for different solutions that are not necessarily better. The need for the latter arises when the possibility exists for the algorithm to get stuck in a local optimum. By accepting other, possibly worse solutions, a different area of the search space can be explored and better solutions might be found. An illustration can be found in figure 5. If a diversification strategy is applied, point 2 might be accepted, in spite of it being a worse solution than the local optimum (point 1). By accepting point 2 as the new incumbent solution, the global optimum (point 3) might be reached.

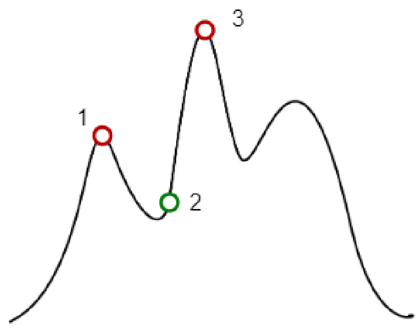


Figure 5: Local & global optima

Within the broad field of metaheuristics, a first classification can be made based upon whether or not the algorithm mimics behaviour pertaining to a certain discipline (e.g. nature, physics, psychology etc.). If so, the metaheuristics are metaphor based. Within the metaphor based metaheuristics, the techniques can be further categorised based upon the related discipline that serves as inspiration for the metaheuristic. The most common ones are inspired by nature (or sometimes named biology). Examples of this last category include genetic algorithms (GA) and bat algorithms. Non-metaphor based examples are local search (LS), tabu search and variable neighbourhood search (El-Omari, 2021; Abdel-Basset et al., 2018).

Local search

Both Kapamara and Petrovic (2009) and Riff et al. (2016) incorporate hill climbing, a local search technique, to improve the performance of their scheduling algorithm after generating an initial solution using dispatching rules (i.e. heuristics). The basic idea behind hill climbing is to generate neighbour solutions through for example swapping the appointments of two patients. Each neighbouring solution is evaluated based on the objective function. If the evaluation is positive, this solution is now the current or incumbent solution and the previous steps are repeated until a certain stop criterion is met. Examples of stop criteria include a maximum number of iterations or an appropriate objective function threshold.

Riff et al. (2016) make use of three different hill climbing methods in their local search algorithm. First, they propose to try swaps between scheduled patients and patients on the waiting list, or simply to try inserting patients currently on the waiting list into the schedule.

Since the scheduling approach in this master's thesis does not use waiting lists, the first hill climbing method is of little relevance. A second step is to swap the appointments pairwise between scheduled patients, if possible. It is possible that free timeslots are still present. Therefore, the third and final step is to try to reschedule appointments into empty timeslots that occur sooner. In conclusion, their algorithm outperforms the well-known 'as soon as possible' and 'just-in-time' approaches based on their testcase data. The hill climbing technique used in Kapamara and Petrovic (2009) differs slightly for each of the four radiotherapy units (planning, physics, pre-treatment and treatment), since each unit has some contextual particularities. The general approach, however, is similar to the third hill climbing algorithm in Riff et al. (2016); a neighbouring solution is obtained by trying to reschedule an appointment to an empty timeslot on a previous date. The empirical results further confirm the effectiveness of a hill climbing algorithm in improving performance.

In Vogl et al. (2019), the authors also use a local search algorithm, an iterated local search to be precise. In addition, a genetic algorithm and a combination of the (I)LS and GA methods are presented. The context of the scheduling problem is different, since an ion beam facility does not use linear accelerators. Instead, one particle accelerator is used. The accelerator is located in a central room and is connected to several treatment rooms. However, as pointed out by Abdel-Basset et al. (2018) and El-Omari (2021), metaheuristics can be applied across various scheduling problems. Therefore, their approach is further elaborated on. In order to successfully apply each of the three proposed metaheuristic algorithms (ILS, GA and the combination of both), an initial solution has to be constructed. This is done in a partly greedy and partly random manner. Because of the specific and different context of the scheduling problem, the details are here not discussed. Regarding the local search step, a different approach is taken compared to Kapamara and Petrovic (2009) and Riff et al. (2016). A variable neighbourhood descent technique is applied in the local search algorithm, which is a variation of the hill climbing method.

Genetic algorithm

A genetic algorithm is a metaphor based evolutionary metaheuristic algorithm. The complete set of evolutionary algorithms "mimics or simulates the biological progression of evolution at the cellular level employing selection, crossover, mutation, and reproduction operators to generate increasingly better candidate solutions (chromosomes)" (Abdel-Basset et al., 2018, p. 188). Genetic algorithms are the most popular of all evolutionary algorithms. The notion of population, chromosomes and genes is critical in understanding the algorithm. A population is formed by chromosomes and represents the solution to the problem under study. An example in the context of radiotherapy scheduling could be that each chromosome is constituted of the operations for all patients, i.e. operation-based representation (D. Petrovic et al., 2009). The chromosome itself is comprised of genes, which then relate to operations separately in the radiotherapy setting. In order to create several different solution possibilities to select from, the following operators are typically defined in genetic algorithms: selection, crossover and mutation. Crossover mechanisms allow certain genes from two selected chromosomes to be swapped. Mutation, on the other hand, allows for the changing of genes within a particular chromosome. Genetic algorithms are widely used within the radiotherapy scheduling domain, both for offline (D. Petrovic et al., 2011; Castro and Petrovic, 2011; Vogl et al., 2019) and online (Legrain, Widmer, et al., 2015; Chang et al., 2020) procedures.

2.2.1.4 Batch

Another possibility to alleviate the time and resource burden is to make use of the available information in an iterative manner. This approach is an extension on the batch scheduling method presented in Pham et al. (2021). Similarly to that approach, treatment requests are accumulated over a specific time period. Then, the scheduling exercise is made for this batch of requests. The difference between the method used in this master's dissertation and the one described by Pham et al. (2021) is the utilisation of (part of) the available information on future patient arrivals. In Pham et al. (2021), no such information is used and the scheduling approach is essentially oblivious to future arrivals. In this manuscript however, arrivals beyond the scheduling horizon are taken into account. In each iteration, it is subsequently not only the scheduling horizon that moves forward, but also the period of future arrivals, named the *future horizon*. As an illustration (see figure 6), take a scheduling horizon of 5 working days and a future horizon of 15 working days. Upon moving forward from $t = 1$ to $t = 2$, the scheduling horizon changes from days 1-5 to days 6-10 and the future horizon advances from days 6-20 to 11-25. Consequently, the total horizon is 20 working days each time. The batch scheduling algorithm used in this dissertation is further explained in section 4.1.

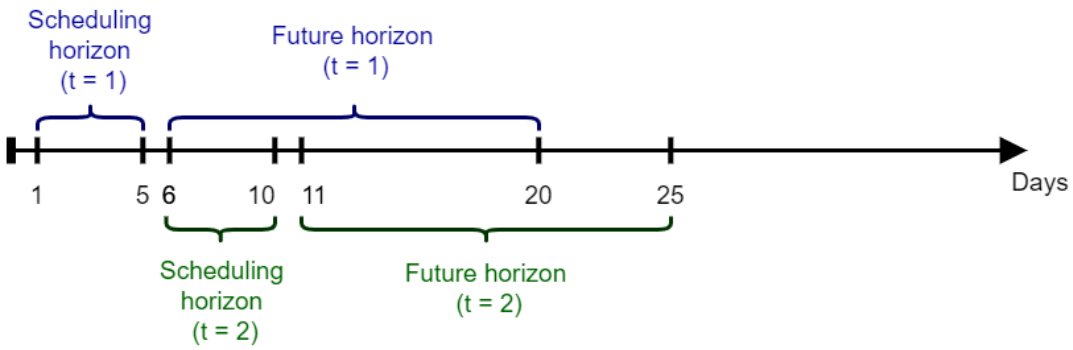


Figure 6: Scheduling horizon & future horizon

2.2.2 Online and online stochastic

As explained in section 2.1.4, online scheduling indicates that a patient immediately receives the bookings of its (pre-)treatment sessions upon request. In contrast to offline scheduling, no information about patients arriving afterwards is available. As a consequence, online scheduling is inherently inferior to offline scheduling in terms of performance on important objectives, such as minimising total waiting time for all patients over a period of time. In order to overcome at least part of this drawback, stochastic information on patient arrivals can be used. Online stochastic scheduling is not exclusively used in radiotherapy scheduling problems, but also in nuclear medicine (Pérez et al., 2013), packet scheduling (Bent & Van Hentenryck, 2005), vehicle routing (Van Hentenryck et al., 2010), kidney transplantation (Awasthi & Sandholm, 2009) and other disciplines.

The solution algorithms for online scheduling are similar to those used in offline scheduling. Therefore, the division in subsections based on the different methods is no longer explicitly made. Instead, the distinctive features of online and online stochastic solution procedures are emphasised. It can be seen throughout this section that some authors directly apply the offline techniques. In that case, in addition to the patient requesting treatment, different scenarios are used to generate potential future patient arrivals.

In Legrain, Fortin, et al. (2015), patients are given a schedule for their appointments quickly after they arrive at the radiotherapy department. The schedule refers to a first day of treatment and a timeslot. Instead of directly developing an online stochastic solution method, the authors start from a version where the complete patient set is known, i.e. offline. They then successively adapt the algorithm to perform an online and online stochastic optimisation. Consequently, Legrain, Fortin, et al. (2015) provide an excellent view of how offline, online and online stochastic methods are related. To avoid computational issues during execution of the optimisation in the offline procedure, the set of feasible schedules for a patient is updated by using column generation. A schedule is feasible when it falls on or after the release date (end of pre-treatment phase) and before the deadline of first treatment session plus some constant parameter that denotes how much the actual first day of treatment can deviate from the deadline. With regard to the online algorithm, both a tailored greedy and a primal-dual algorithm are proposed. The greedy algorithm is the one used by their partnering hospital. It gives results that are very similar to an ASAP procedure. The online stochastic solution technique is based on the formulation of the offline model but includes scenarios of future patient arrivals and also uses a primal-dual algorithm. A larger set of future scenarios will naturally lead to better results while negatively impacting the time needed to solve the problem. The procedure is split up into two variations; one variation assumes that all lower-priority patients are known in advance, the other one does not make this assumption. Unsurprisingly, the former produces better results. The stochasticity in these approaches refers to uncertainty in patient arrivals, their priority and their treatment duration.

Legrain, Widmer, et al. (2015) build upon the research of Legrain, Fortin, et al. (2015). They include the pre-treatment phase and adapt the online stochastic procedure in order to fit the problem description. In addition, a genetic algorithm is proposed. Chang et al. (2020) also make use of a genetic algorithm and they use scenarios to take into account stochastic information.

Similarly to Legrain, Fortin, et al. (2015), the authors in Pham et al. (2021) present offline, online and online stochastic procedures. The online procedure is based on a greedy heuristic that produces results similar to an ASAP algorithm. With respect to the online stochastic procedure, the authors utilise a different approach compared to what has been done before. Instead of creating scenarios of future arrivals by sampling, the authors build and train machine learning models by solving an IP formulation with complete historical data sets as input. These models are consequently used to predict the preferred first day of treatment.

An overview of the methods used in the literature on scheduling in radiotherapy is given in table 5. In chapter 4, three solution algorithms are developed; an offline (batch), online (heuristic) and an online stochastic procedure.

Paper	**Heuristics**	**GA**	**LS**	**Batch**
Kapamara and Petrovic (2009)	x		x	
D. Petrovic et al. (2011)		x		
Castro and Petrovic (2011)		x		
Castro and Petrovic (2012)	x			
Riff et al. (2016)	x		x	
Vogl et al. (2019)		x	x	
Pham et al. (2021)	x			x
Legrain, Fortin, et al. (2015)	x			
Legrain, Widmer, et al. (2015)		x		
Chang et al. (2020)		x		

Table 5: Offline solution methods - overview

GA = genetic algorithm; LS = local search

2.3 Contributions and research questions

Regarding the pre-treatment and treatment phase in radiotherapy scheduling, Kapamara et al. (2006) stated that it is preferable to consider both phases simultaneously. However, the existing literature is primarily focused on scheduling problems in the treatment phase. In fact, only D. Petrovic et al. (2011) and Legrain, Widmer, et al. (2015) successfully integrate both phases. In relation to offline and online (pre-)treatment scheduling, few papers consider the online paradigm. Furthermore, the literature study has shown that a dynamic or stochastic approach, whereby the impact of today's scheduling decisions on future arrivals is incorporated, provides better results in comparison to a myopic approach. Most recent studies on the topic have incorporated this view. Synthesizing these findings indicated the direction of this dissertation. To the best of our knowledge, Legrain, Widmer, et al. (2015) is the only existing study that effectively solves the pre-treatment and treatment simultaneously and in an online stochastic manner. The main disadvantage of the model proposed in their study is that their model is tailored specifically to the Centre Intégré de Cancérologie de Laval (CICL), a Canadian health centre. For the pre-treatment phase for example, it is sufficient to compare the pre-treatment phase in Legrain, Widmer, et al. (2015) to the one in Castro and Petrovic (2012) to understand that various hospitals organise the radiotherapy department differently.

Therefore, in relation to the radiotherapy scheduling literature, the contributions of our study are threefold:

1. A general purpose mathematical model is developed that integrates the pre-treatment and treatment phase in radiotherapy. Furthermore, the model is tested in a variety of experiments using real-world data.
2. A contribution to the growing literature on stochastic radiotherapy scheduling is made by developing a stochastic algorithm.
3. In addition to an offline technique, online solution techniques are presented, hereby contributing to the scarce literature on online radiotherapy scheduling.

In order to successfully materialise these contributions, the following research questions are defined:

RQ 1: Can we formulate a general purpose model that includes characteristics of both the pre-treatment and treatment phase?

RQ 2: Can we subsequently solve this model in an online stochastic manner?

In chapter 3 of this dissertation, a mathematical model formulation is proposed as a response to *RQ 1*. In response to *RQ 2*, an offline, online and online stochastic procedure are introduced in chapter 4. The mathematical model and solution techniques are tested in chapter 5.

3 Problem definition and formulation

The principal goal of scheduling practices in radiotherapy is to aid in the provision of effective and efficient healthcare treatment to patients. In order to achieve this objective, patients have to receive timely appointments for their (pre-)treatment operations. The literature suggests that the best way to guarantee the performance of the scheduling practices is to include the minimisation of waiting times or related concepts (S. Petrovic et al., 2006). A distinction is often made between allocation scheduling and advance scheduling. In this study, the definitions of Saure et al. (2012) are used. The authors state that *allocation scheduling* involves the allocation of resources and time slots to patients on the day of service, once all patients to be served on that day are known. This is in contrast to *advance scheduling*, where patients receive appointments prior to the service date. In this dissertation, an advance scheduling problem is studied. Section 3.1 defines the problem description and some key assumptions of the scheduling problem under study. This allows to create a common understanding of the problem under study and to set the stage for the mathematical model formulation presented in section 3.2.

3.1 Problem description

As mentioned in section 2.3, the literature indicates that scheduling procedures in radiotherapy departments can differ substantially. With regard to the problem description, the aim of this dissertation is hence to present a model that is flexible and able to cater to the needs of individual medical centres. As described in section 2.1.2, a distinction can be made between the pre-treatment and treatment phase. The pre-treatment phase includes all preparatory operations that are necessary to successfully commence the actual administration of the radiation doses, i.e. the treatment phase. To accurately model the pre-treatment pathway, the research from Castro and Petrovic (2012) is primarily used. Regarding the treatment phase, the main source of inspiration is the problem description presented by Pham et al. (2021). In the following paragraphs, fundamental concepts underlying the problem description are explained.

Patients in the radiotherapy department are typically classified into various categories, depending on factors such as treatment intent, waiting list status, etc. Each of these categories is paired with a set of characteristics, e.g. waiting time target and patient priority. In the literature, several classifications exist, reflecting the differences in radiotherapy departments.

In this study, patient categories are defined based on their treatment intent, following the guidelines in AZ Sint-Lucas. Three groups exist: palliative patients, curative patients and patients with definitive tumor control. The last two groups share the same characteristics. Therefore, they are sometimes referred to as *non-palliatives*, as opposed to *palliatives*. Additionally, emergency patients are treated on an ad-hoc basis, meaning that they will always receive timely treatment by using overtime capacity if needed. These patients are therefore not explicitly included in the model.

The literature study revealed that there is no singular correct way to model the pre-treatment phase. As a result, the problem is described in broad terms, in order to allow for a straightforward tailoring to specific hospitals. First, figure 7 illustrates that for each pre-treatment operation, various resource requirements can exist. Every resource requirement can be fulfilled by one or more resources. This flexible approach leaves room for pools of resources consisting of multiple eligible resources for a certain resource requirement (e.g. a pool of eligible nurses for the human resource requirement). At the same time, it is possible that only 1 resource meets the qualifications to serve a certain task, e.g. a scanner machine. In that case, that scanner is the only resource in the eligible machines pool. Both situations are effectively conceptualised in the model. The constructs of resource requirements and eligible resources are derived from Vogl et al. (2019).

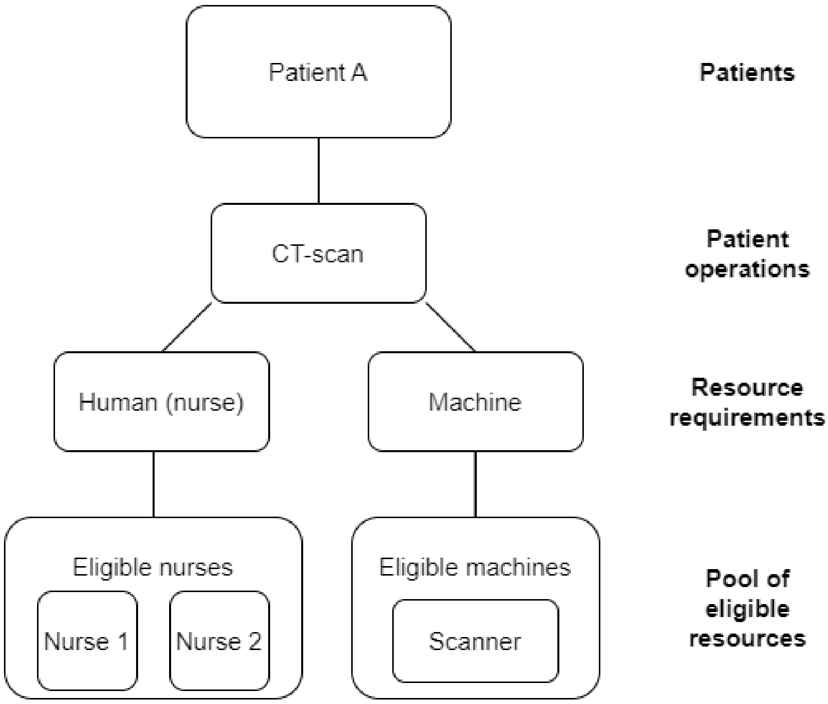


Figure 7: Resource levels

Secondly, a hybrid form of block-scheduling is used, following the method of Castro and Petrovic (2012). This indicates that timeslots are defined for each resource. Each pre-treatment task has to be scheduled in one of these timeslots. However, the difference with a more conventional block-scheduling system is that the actual starting and end time of a pre-treatment operation need not be exactly equal to the start and end time of the slot. This ensures that operations of different duration but with the same resource requirements are effectively modelled. As an example, figure 8 shows the available morning slots on Monday for a certain resource and two tasks with a different duration to be planned. Task A has a duration of 45 minutes and task B takes 1 hour and 15 minutes to be completed. By defining multiple slots with a duration that is a multiple of some base duration, every task can be booked. It is worth noting that the base duration should be smaller than or equal to the shortest task to be scheduled on a resource, in order to make sure that enough slots are defined. After scheduling task A in slot 5 and task B in slot 9, the resource has no capacity left. This can be seen in the second part of figure 8. When the resource is fully utilised, a mechanism has to be in place to ensure that no additional tasks are booked on the remaining slots. In the mathematical formulation (section 3.2), this is accomplished by equations 6.

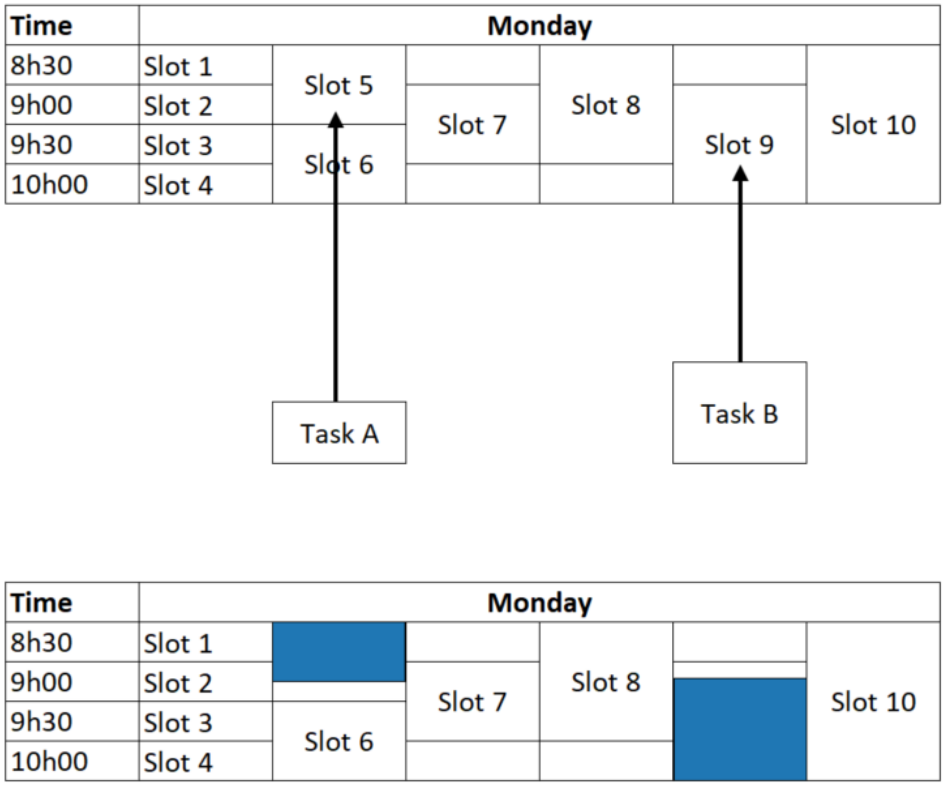


Figure 8: Hybrid block-scheduling

Additionally, time window constraints are defined. First, it is assumed that pre-treatment operations can have a certain lead time, in addition to the duration of the operation. This concept is taken directly from Castro and Petrovic (2012). In their research, the example is given for the beam direction shell operation. The duration of this operation is only one hour. However, it takes an additional 24 hours for the device to be hardened and ready to use. A second time window constraint relates to the connection between the pre-treatment and treatment phase. The assumption is made that the treatment phase has to start within a certain amount of time after completing the final pre-treatment task. The idea behind this is to ensure the validity of the parameters when starting the treatment.

The elements described above serve as input to the pre-treatment part of the scheduling problem. As output, each patient receives precise timestamps that indicate when (s)he has to undergo his or her pre-treatment operations. Furthermore, for each operation and each resource requirement, one eligible resource is chosen to assist in the execution of that operation. In the following paragraphs, the input for the treatment phase is presented. The goal of that part of the scheduling problem is to provide each patient with a day for its first treatment session. In addition, the linear accelerator that will be used throughout the entire treatment phase of that patient is determined.

In this dissertation, the approach used for the treatment phase largely follows the perspective of Pham et al. (2021). The output relating to patient scheduling for the treatment phase only involves the start date of treatment, with each patient receiving a day for their first fraction administration. The exact starting time is provided later, when linear accelerator appointments are booked for all patients for a given day or week (cf. allocation scheduling). Unlike the approach taken by the authors in Pham et al. (2021), this dissertation does not reserve specific slots for palliative patients, as AZ Sint-Lucas currently does not follow this policy. Nonetheless, the model presented in chapter 3.2 can be extended to accommodate a similar policy if necessary. To give a patient an appointment on a linear accelerator for their first fraction, it has to be ensured that this patient can continue to receive the remainder of its treatment on the same linear accelerator. Therefore, linac capacities form a critical component of the input. In figure 9, an illustration for a situation with one linac is given. The treatment duration for a certain patient is 4 days. This means that the earliest feasible day to start treatment is day 4. Starting earlier would mean that the patient's treatment is interrupted on day 3, as there is no capacity left on that day.

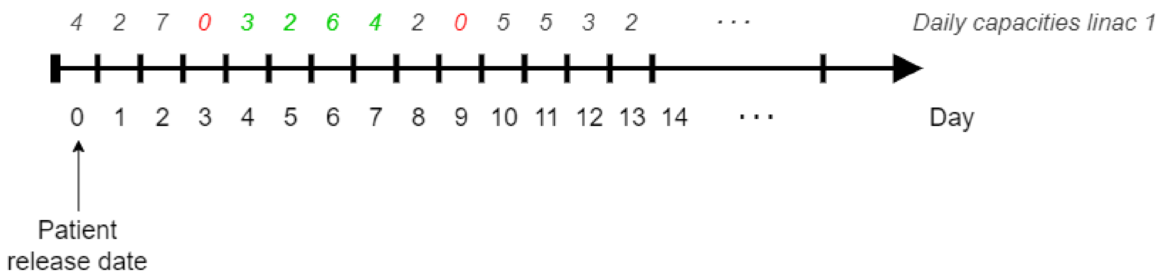


Figure 9: Linac capacities, based on Pham et al. (2021)

Furthermore, it is common practice in the literature to double the fraction duration of the first fraction (e.g. Conforti et al., 2011; Saure et al., 2012). The motivation behind this is that it allows for a set-up of the linear accelerator and parameter verification.

Finally, the term 'no-show' is used to refer to a situation where a patient fails to attend their scheduled appointment (Diamant et al., 2018). No-shows have the possibility to negatively impact utilisation rates of the resources. Fortunately, consultation with a health practitioner learned us that no-shows are almost non-existent in the radiotherapy department. Consequently, it is assumed that patients always arrive at their assigned appointments.

3.2 Mathematical formulation

In this section, the mathematical formulation of the problem description given in 3.1 is presented. The problem is formulated as an integer linear programming (ILP) model. This means that all variables are integers, either binary or positive and without upper bound. In addition, the linearity assumption states that the objective function and all constraints are linear. As described in the literature overview in chapter 2, research was conducted into both the pre-treatment and the treatment phase as well as into their integration. To accurately model the pre-treatment pathway, Castro and Petrovic (2012) served as the main source of inspiration. The treatment phase on the other hand is based upon the mathematical formulation in Pham et al. (2021). Additionally, extra constraints are added to reinforce the integration of both. It is worth noting that extensive adaptations have been made to the models by Castro and Petrovic (2012) and Pham et al. (2021), in order to match the problem description. For example, to add a level of generality to the pre-treatment phase that is not present in Castro and Petrovic (2012), extra notation based on Vogl et al. (2019) has been included.

3.2.1 Sets

First, the notation of all sets is given. $\mathcal{P}$, $\mathcal{E}$ and $\mathcal{O}$ are taken integrally from Castro and Petrovic (2012), while $\mathcal{K}$ and $\mathcal{V}$ stem from Pham et al. (2021), albeit having different notation. The sets of pre-treatment resource requirements $\mathcal{M}$ and pre-treatment resources $\mathcal{I}$ relate to the concept of resource requirements and individual resources stated in the problem description and are originally deducted from Vogl et al. (2019). $\mathcal{Q}$ consists of elements that each exist of two pre-treatment operations l and l' from a certain patient $j \in \mathcal{P}$, indexed by a tuple $((j,l),(j,l'))$ and where task l has to be finished before l' can start. Hence, set $\mathcal{Q}$ denotes the existing precedence relations between pre-treatment operations from all patients.

$\mathcal{K}$: set of working days in the planning horizon.

$\mathcal{P}$: set of patients.

$\mathcal{E}$: set of all free time slots.

$\mathcal{M}$: set of pre-treatment resource requirements

$\mathcal{I}$: set of pre-treatment resources

$\mathcal{O}$: set of pre-treatment operations.

$\mathcal{Q}$: set of precedence relations between pre-treatment operations.

$\mathcal{V}$: set of linear accelerators (linacs).

3.2.2 Parameters

Secondly, an overview of the different parameters used in the model is given. They are grouped in smaller subclasses that relate to a particular modelling aspect.

3.2.2.1 General parameters

A general parameter that is not specifically related to any modeling concept is M.

M: big M parameter, used in constraints 6, 7, 8 and 11a. The value of this parameter is arbitrarily large.

3.2.2.2 Parameters regarding patient information

The due date d_j of a patient is the day by which the first radiation fraction has to be delivered. Some countries have strong guidelines on the target waiting times or correspondingly the due dates. For example in the UK, the Joint Collegiate Council for Oncology sets waiting time targets. In Belgium however, the guidelines are less strict and the treating hospital is the primary decision maker. Weights w_j are assigned to patients on a similar basis as due dates, i.e. based on treatment intent. The release date of a patient a_j is the earliest day on which he or she can start the pre-treatment pathway. TW_j represents the fact that the last pre-treatment phase, simulation of the isocentre, has to be followed by the first radiation fraction in a timely manner, to guarantee parameter validity.

d_j: due date of patient $j \in \mathcal{P}$ (set as day).

w_j: priority weight of patient j.

a_j: release date of patient j.

TW_j: maximum time (in days) between the final pre-treatment step and the day of first fraction.

3.2.2.3 Parameters regarding pre-treatment operations

s_j represents the amount of tasks or operations in the pre-treatment phase that are subject to scheduling, i.e. those tasks that require the assignment of a definitive timeslot, starting time and one or more resources. O_j then indicates the operations or tasks, subject to scheduling, of a patient j. Each operation has certain resource requirement(s), denoted by M_{jl}. For a given resource requirement, I_{jlm} provides a pool of resources that meet the eligibility criteria (cf. figure 7). If only one resource is eligible, then $|I_{jlm}| = 1$. p_{jl} expresses the duration of each operation, in minutes. The lead time of an operation is the minimum time that has to elapse between this operation and the next. $W1_{jl}$ and $W2_{jl}$ both represent the same concept, but with a different granularity.

s_j: number of pre-treatment operations of patient j.

O_j: set of pre-treatment operations of patient j. $O_j \in \mathcal{O}$

M_{jl}: set of resource requirements for processing operation $l \in O_j$ of patient j. $M_{jl} \in \mathcal{M}$

I_{jlm}: set of eligible resources for the l'th operation of patient j and resource requirement $m \in M_{jl}$. $I_{jlm} \in \mathcal{I}$

p_{jl}: processing time of operation l of patient j (in minutes).

$W1_{jl}$: lead time of operation l of patient j (in minutes).

$W2_{jl}$: lead time of operation l of patient j (in days).

3.2.2.4 Parameters regarding treatment operations

The capacity of each linear accelerator on a given day is represented by C_v^k and it is expressed as an integer that relates to the number of base timeslots that are available. Each fraction of a patient has a duration expressed as a number of base timeslots: z_j. The total amount of fractions that a patient has to receive, i.e. the number of successive days that a patient has to be treated on a linac is represented by n_j.

C_v^k: capacity of linac $v \in \mathcal{V}$ on day $k \in \mathcal{K}$ (in number of base timeslots).

z_j: duration of each fraction of patient j (in number of base timeslots).

n_j: number of treatment operations of patient j, i.e. the number of fractions each patient has to receive.

3.2.2.5 Parameters regarding timeslots

The concept of timeslots is heavily based on Castro and Petrovic (2012). The authors defined timeslots for every resource. In this master's dissertation, an index to denote the day was added with the purpose of making the implementation in software programs more straightforward and to make the formulation ever so slightly easier to comprehend.

$\mathcal{E}_{ijl}$ represents the set of timeslots of a resource that are feasible to perform a particular pre-treatment operation. Slot $S_{iek} \in \mathcal{E}_{ijl}$ is feasible for pre-treatment operation l of patient j, if the duration of the operation does not exceed the duration of the timeslot and if the timeslot is on or after the patient's release date a_j.

> S_{iek}: slot e of day k of resource i, $S_{iek} \in \mathcal{E}$.
>
> $\mathcal{E}_{ijl}$: set of feasible slots for (pre-treatment) operation l of patient j and resource i.
>
> $\underline{\tau}(S_{iek})$: the time elapsed from the beginning of the scheduling horizon to the start of slot S_{iek} (in minutes).
>
> $\overline{\tau}(S_{iek})$: the time elapsed from the beginning of the scheduling horizon to the end of slot S_{iek} (in minutes).
>
> $\tau(S_{iek})$: duration of slot (S_{iek}); $\tau(S_{iek}) = \overline{\tau}(S_{iek}) - \underline{\tau}(S_{iek})$ (in minutes).

3.2.3 Decision variables

Thirdly, the decision variables are presented. A useful distinction can be made between primary and secondary decision variables. The former has meaning in the problem context, whereas the latter serves the sole purpose of modelling the desired structure in a set of equations.

3.2.3.1 Primary decision variables

It is important to understand the interaction between x_{jlmiek} and t_{jlmi}. If $x_{jlmiek} = 1$, operation l of patient j will be performed in timeslot $S_{iek} \in \mathcal{E}_{ijl}$ for a resource requirement m. Subsequently, t_{jlmi} will lie between $\underline{\tau}(S_{iek})$ and $\overline{\tau}(S_{iek})$ and it has to be made sure that the completion time of the operation (i.e. $t_{jlmi} + p_{jl}$) does not exceed $\overline{\tau}(S_{iek})$, which is the end time of the timeslot. In section 3.2.5, the necessary constraints to model this behaviour are explained. Both variables relate to scheduling of the pre-treatment tasks. The tardiness (i.e. positive lateness) of a patient, L_j, is the positive difference between the day of the first fraction dose and the due date of the patient. Hereby, a deviation is made from Castro and Petrovic (2012) who use the lateness, both positive and negative, of a patient. In the remainder of this dissertation, the terms lateness and tardiness will be used interchangeably, but always referring to the positive lateness of a patient. x^j_{kv} denotes whether or not a patient receives its first radiation treatment session on day k on linac v. If so, the patient will be scheduled on that linac for the $n_j - 1$ days thereafter as well. This is not shown by means of a decision variable. Instead, this behaviour is modelled by using a specific set of constraints (cf. section 3.2.5).

$$x_{jlmiek} = \begin{cases} 1, & \text{if the } l\text{'th operation of patient } j \text{ is processed in timeslot} \\ & S_{iek} \in \mathcal{E}_{ijl} \text{ for resource requirement } m \in M_{jl} \\ 0, & \text{otherwise} \end{cases}$$

t_{jlmi}: the time elapsed from the beginning of the scheduling horizon to the start of operation l from patient j on resource i for resource requirement m (in minutes). $t_{jlmi} \geq 0$

L_j: tardiness of patient j (in days) $L_j \in Z$. $L_j \geq 0$

$$U_j = \begin{cases} 1, & \text{if patient } j \text{ exceeds the waiting time target and hence has } L_j > 0 \\ 0, & \text{otherwise} \end{cases}$$

$$x^j_{kv} = \begin{cases} 1, & \text{if patient } j \text{ receives their first fraction on day } k \text{ on linac } v \\ 0, & \text{otherwise} \end{cases}$$

3.2.3.2 Secondary decision variables

Both secondary decision variables are used to model an either-or structure on a set of two constraints. This means that either the first constraint will always be satisfied and the second has to be adhered to, or the other way around, depending on the value of the decision variable. These variables are to be combined with a big M parameter to ensure the correct functioning.

$$b_{jflgmni}\colon \begin{cases} 1, & \text{if constraint 6b has to be adhered to} \\ 0, & \text{if constraint 6a has to be adhered to} \end{cases}$$

$$b_j\colon \begin{cases} 1, & \text{if constraint 7d has to be adhered to} \\ 0, & \text{if constraint 7c has to be adhered to} \end{cases}$$

3.2.4 Objective function

In this dissertation, an optimisation problem with several objectives is considered. Inspiration is taken from Castro and Petrovic (2012) and the objectives have been proposed to the head of quality at the radiotherapy department in AZ Sint-Lucas to determine the relative importance of each objective. It was asked whether it is preferred to have for example one patient with a delay of three days or three patients with a delay of one day each. In practice, neither of these situations are desired. Therefore, it is decided to assign weights G_1 and G_2 to each part of the objective function respectively. The first objective, the first term in (1), is to minimise the weighted amount of patients that exceed the waiting time targets. If only this part of the objective would be included in the model, it is theoretically possible that an exorbitant situation like the following will occur. Imagine that the capacity in the radiotherapy department is not sufficient to schedule all patients on time, or that the patient inflow is too high for a given capacity. In that case, a model including the first term of the objective as the sole objective, tries to minimise the amount of patients that experience a delay, independent of the size of the delay. Subsequently, it is possible that one patient has a tardiness of 10 days, while all other patients can start treatment on time. For the same data instance, scheduling two patients with a delay of 1 day each was also possible and arguably better. However, this outcome is disregarded by not taking into account the second term. Therefore, it is critical to include both terms in the objective. The second term of the objective function minimises the total weighted tardiness of all patients. Furthermore, the denominator (the weighted sum over all patients of the maximum tardiness of each patient) in this term is included to normalise the term. This is necessary to determine the values of G_1 and G_2 (cf. section 5.1.3).

$$\begin{aligned} minimise \quad & G_1 \cdot (\sum_{j \in \mathcal{P}} w_j \cdot U_j) / \sum_{j \in \mathcal{P}} w_j \\ & + G_2 \cdot \sum_{j \in \mathcal{P}} (w_j \cdot L_j) / \sum_{j \in \mathcal{P}} ((max(\mathcal{K}) - d_j) \cdot w_j) \end{aligned} \tag{1}$$

3.2.5 Constraints

In this section, the constraints of the optimisation model are given. The combination of all constraints ensures that the model follows the principles outlined in section 3.1.

Equation 2 ensures that each operation is assigned exactly one slot. Note that the equation is an equality. This assumes that the horizon is large enough to ensure that every patient can and will receive appointments for its pre-treatment operations. Furthermore, the assumption is made that for each resource requirement m, one and only one of the eligible resources is used. Hence the summation of i over $\mathcal{I}_{jlm}$.

$$\sum_{i \in I_{jlm}} \sum_{S_{iek} \in \mathcal{E}_{ijl}} x_{jlmiek} = 1 \qquad \forall j \in \mathcal{P}; l \in O_j; m \in M_{jl} \tag{2}$$

By using equation 3, it is ensured that the limit of each timeslot's capacity is not exceeded.

$$\sum_{j \in \mathcal{P}} \sum_{l \in O_j} \sum_{m \in M_{jl}} [p_{jl} \cdot x_{jlmiek}] \leq \tau(S_{iek}) \qquad \forall S_{iek} \in \mathcal{E} \tag{3}$$

Equations 4a and 4b make sure that the starting time of each operation is constrained within the allocated feasible timeslot's interval. If resource i is not used for a certain resource requirement m for operation l of patient j, t_{jlmi} will be equal to 0 according to these equations. It is important to remember this, since it will be used in some of the following constraints.

$$\sum_{S_{iek} \in \mathcal{E}_{ijl}} [\underline{\tau}(S_{iek}) \cdot x_{jlmiek}] \leq t_{jlmi} \qquad \forall j \in \mathcal{P}; l \in O_j; m \in M_{jl}; i \in I_{jlm} \tag{4a}$$

$$t_{jlmi} \leq \sum_{S_{iek} \in \mathcal{E}_{ijl}} [(\overline{\tau}(S_{iek}) - p_{jl}) \cdot x_{jlmiek}] \qquad \forall j \in \mathcal{P}; l \in O_j; m \in M_{jl}; i \in I_{jlm} \tag{4b}$$

Equation 5 guarantees that the pre-treatment pathway of each patient is carried out in the correct order and with an accurate time window structure. The summation of i over $\mathcal{I}_{jlm}$ has to be included because one and only one resource will be used for each resource requirement m and t_{jlmi} will be equal to zero for the other resources.

$$\sum_{i \in I_{jlm}} t_{jlmi} + p_{jl} + W1_{jl} \leq \sum_{i' \in I_{jl'm'}} t_{jl'm'i'} \qquad (5)$$
$$\forall j \in \mathcal{P}; ((j,l),(j,l')) \in \mathcal{Q}; m \in M_{jl}; m' \in M_{jl'}$$

Equations 6 prevent a given resource from processing more than one operation at a time. These constraints are rather complex and it might be difficult to grasp the logic behind them. Therefore, a brief explanation is given. For now, imagine that the last term in both equations is not included. In equation 6a, it is then stated that the starting time of operation g from a patient f on resource i has to be larger than or equal to the starting time of an operation l from patient j on the same resource i plus the processing time of operation l from patient j. However, when operation g is not processed on resource i, t_{fgni} will be zero and the equation might cause infeasibility without the inclusion of the term $-(1 - \sum_{S_{iek} \in \mathcal{E}_{ifg}} x_{fgniek}) \cdot M$. For equation 6b, the same reasoning applies. It is critical to realise that jlm cannot be equal to fgn, since it that case, exactly the same operations are concerned and t_{fgni} has to be equal to t_{jlmi}. A final note is the inclusion of the last term in both equations. They are necessary to have the correct either-or structure.

$$t_{fgni} \geq t_{jlmi} + (p_{jl} \cdot \sum_{S_{iek} \in \mathcal{E}_{ijl}} x_{jlmiek}) - (1 - \sum_{S_{iek} \in \mathcal{E}_{ifg}} x_{fgniek}) \cdot M - b_{jflgmni} \cdot M \qquad (6a)$$
$$\forall j, f \in \mathcal{P}; l \in O_j; g \in O_f; m \in M_{jl}; n \in M_{fg}; jlm \neq fgn; i \in \{I_{jlm} \cap I_{fgn}\}$$

$$t_{jlmi} \geq t_{fgni} + (p_{fg} \cdot \sum_{S_{iek} \in \mathcal{E}_{ifg}} x_{fgniek}) - (1 - \sum_{S_{iek} \in \mathcal{E}_{ijl}} x_{jlmiek}) \cdot M - (1 - b_{jflgmni}) \cdot M \qquad (6b)$$
$$\forall j, f \in \mathcal{P}; l \in O_j; g \in O_f; m \in M_{jl}; n \in M_{fg}; jlm \neq fgn; i \in \{I_{jlm} \cap I_{fgn}\}$$

Constraints 7 specify the tardiness of each patient. Tardiness is defined as the positive difference between the actual waiting time for administration of the first radiation fraction and the waiting time target. The due date d_j of a patient is based on the waiting time target and the release date a_j. Equation 7a ensures that the tardiness of each patient is larger than or equal to the difference between the day of the first treatment session and the due date. According to this equation in isolation, the tardiness can be negative, if the patient is scheduled before the due date. Therefore, equation 7b is included. At this point, there is no upper bound on the value of L_j. Inequalities 7c and 7d ensure that the tardiness of a patient is either smaller than or equal to the difference between the day of the first treatment session and the due date, or smaller than or equal to zero. Combining all four constraints results in the correct behaviour; if a patient is scheduled before or on the due date, the tardiness will be zero and otherwise, it will be exactly equal to the difference between the actual waiting time and the waiting time target.

$$L_j \geq \sum_{k \in \mathcal{K}} \sum_{v \in \mathcal{V}} [x^j_{kv} \cdot (k - d_j)] \qquad \forall j \in \mathcal{P} \tag{7a}$$

$$L_j \geq 0 \qquad \forall j \in \mathcal{P} \tag{7b}$$

$$L_j \leq \sum_{k \in \mathcal{K}} \sum_{v \in \mathcal{V}} [x^j_{kv} \cdot (k - d_j)] + M \cdot b_j \qquad \forall j \in \mathcal{P} \tag{7c}$$

$$L_j \leq 0 + M \cdot (1 - b_j) \qquad \forall j \in \mathcal{P} \tag{7d}$$

In order to keep record of how many patients exceed their respective waiting time targets, the binary decision variable U_j has been introduced. Constraints 8 set the value of this variable accordingly. If L_j is equal to zero, U_j has to be equal to zero as well. If, on the other hand L_j is larger than zero, U_j should be equal to 1.

$$U_j \leq L_j \qquad \forall j \in \mathcal{P} \tag{8a}$$

$$L_j \leq U_j \cdot M \qquad \forall j \in \mathcal{P} \tag{8b}$$

When two or more resources are needed to process an operation, the starting time of the operation has to be the same on all resources. This is accomplished by equation 9.

$$\sum_{i \in I_{jlm}} t_{jlmi} = \sum_{i' \in I_{jlm'}} t_{jlm'i'} \qquad \forall j \in \mathcal{P}; l \in O_j; m, m' \in M_{jl} \tag{9}$$

Equation 10 makes sure that all patients receive a date for their first treatment session. At this stage, this appointment is merely a day. A specified appointment hour will be decided and communicated by the radiotherapy department later, when a lower level allocation scheduling practice is performed.

$$\sum_{k \in \mathcal{K}} \sum_{v \in \mathcal{V}} x^j_{kv} = 1 \qquad \forall j \in \mathcal{P} \tag{10}$$

The lead time, in days, of the last pre-treatment operation is presented by $W2_{js_j}$. Furthermore, the maximum time between this operation and the start of the treatment phase is given by TW_j. Equations 11 ensure that the first radiotherapy session is planned within the corresponding timeframe related to the final pre-treatment session. Inequality 11a models the minimum time that has to elapse between pre-treatment completion and start of treatment. Usually, a minimum amount of days is necessary to allow the medical staff at the radiotherapy centre to prepare the treatment and validate the patient's parameters. Equation 11b, on the other hand is included to prevent the model from scheduling the first fraction delivery at an unacceptably late date after the pre-treatment completion. Including both equations for only one resource requirement is sufficient, because the other constraints in the model ensure the correct functioning for the other resource requirements.

$$\sum_{i \in I_{js_j 1}} \sum_{S_{iek} \in \mathcal{E}_{ijs_j};(k \leq k')} [k \cdot x_{js_j 1iek}] + W2_{js_j} - (1 - x^j_{k'v}) \cdot M \leq k' \cdot x^j_{k'v} \qquad \forall j \in \mathcal{P}; k' \in \mathcal{K}; v \in \mathcal{V} \tag{11a}$$

$$\sum_{i \in I_{js_j 1}} \sum_{S_{iek} \in \mathcal{E}_{ijs_j};(k \leq k')} [k \cdot x_{js_j 1iek}] + TW_j \geq k' \cdot x^j_{k'v} \qquad \forall j \in \mathcal{P}; k' \in \mathcal{K}; v \in \mathcal{V} \tag{11b}$$

Although only the first day of the treatment phase is being explicitly scheduled, the subsequent days have to be taken into account as well. To accommodate for this, constraints 12a and 12b ensure that sufficient capacity is available on the designated linear accelerators over

the full length of each patient's treatment phase. In Pham et al. (2021), the constraint was originally modelled (using the notation from this dissertation) as:
$\sum_{j\in\mathcal{P}}\sum_{k'=max(0,k-n_j+1)}^{k}[z_j \cdot x_{k'v}^{j}] \leq C_v^k \quad \forall k \in \mathcal{K}, v \in \mathcal{V}$
This equation functions perfectly when every fraction in a patient's treatment pathway has the same duration equal to z_j. However, the first session requires twice that duration, in order to perform a set-up and additional validation of critical parameters. Accordingly, the equation has been altered to include this functionality. The result can be seen in constraints 12. In (12a), summation in the first term is done with an upper bound $k-1$ instead of k. Moreover, a second term has been added to model the double duration of the first fraction delivery. Using only this constraint and over all days in the horizon would result in infeasibility as k' must not be equal to -1. Subsequently, equation 12a is only valid for $k \in \mathcal{K}\backslash\{0\}$ and equation 12b has been added for the special case where $k=0$. The attentive reader will notice that when $n_j = 1$, $max(0, k-n_j+1)$ is smaller than $k-1$ and thus the lower bound of the summation is larger than the upper bound in that case. Although not visually pleasing, this does not provide any issues as the resulting sum is referred to as a 'nullary sum' and is equal to zero by definition (Harper, 2016, p. 119).

$$\sum_{j\in\mathcal{P}}\sum_{k'=max(0,k-n_j+1)}^{k-1}[z_j \cdot x_{k'v}^{j}] + \sum_{j\in\mathcal{P}}[(2\cdot z_j)\cdot x_{kv}^{j}] \leq C_v^k \\ \forall k \in \mathcal{K}\backslash\{0\}, v \in \mathcal{V} \tag{12a}$$

$$\sum_{j\in\mathcal{P}}[(2\cdot z_j)\cdot x_{0v}^{j}] \leq C_v^0 \qquad \forall v \in \mathcal{V} \tag{12b}$$

Finally, equations 13 ensure that the decision variables x_{jlmiek}, U_j and x^j_{kv} are binary and that decision variables t_{jlmi} and L_j are larger than or equal to zero.

$$x_{jlmiek} \in \{0,1\} \qquad \forall j \in \mathcal{P}; l \in O_j; m \in M_{jl}; i \in I_{jlm}; S_{iek} \in \mathcal{E}_{ijl} \tag{13a}$$

$$t_{jlmi} \geq 0 \qquad \forall j \in \mathcal{P}; l \in O_j; m \in M_{jl}; i \in I_{jlm} \tag{13b}$$

$$L_j \geq 0 \qquad \forall j \in \mathcal{P} \tag{13c}$$

$$U_j \in \{0,1\} \qquad \forall j \in \mathcal{P} \tag{13d}$$

$$x^j_{kv} \in \{0,1\} \qquad \forall j \in \mathcal{P}; k \in \mathcal{K}; v \in \mathcal{V} \tag{13e}$$

4 Solution techniques

In this chapter, three solution procedures are presented; an offline, online and online stochastic procedure. The offline procedure is a batch algorithm and is presented in 4.1. Additionally, the developed online algorithm follows an ASAP paradigm and is proposed in section 4.2. Finally, section 4.3 discusses the online stochastic procedure. It is worth mentioning that these techniques are presented on a high level in this chapter. In chapter 5, the procedures are tailored to AZ Sint-Lucas and detailed pseudocode representations are given.

4.1 Batch procedure

Building upon the batch scheduling paradigm discussed in the section 2.2.1.4, the batch algorithm developed in this dissertation is hereby presented. As stated in chapter 2, the full offline method for scheduling all patients over the course of a year is intractable. Therefore, an alternative solution that is able to alleviate the time and resource burden had to be constructed. In this dissertation, the batch procedure is chosen to accomplish this purpose. The proposed algorithm is straightforward to implement and results can be achieved within a reasonable timeframe. Moreover, it is expected to provide a better approximation of the full offline solution than the other alternatives (heuristics and metaheuristics). Consequently, this technique serves as a lower bound when comparing with the online and online stochastic method. A graphical illustration of the procedure is given in figure 10.

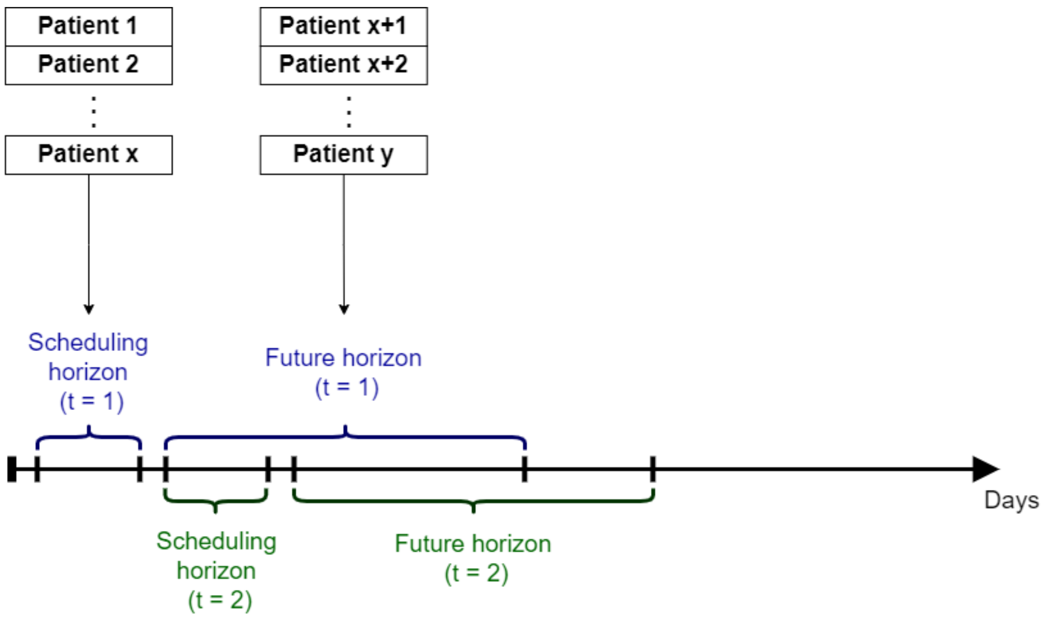


Figure 10: Batch procedure

Treatment requests are accumulated over a specific time period, conveniently named the *scheduling horizon*. In addition to this, arrivals beyond the scheduling horizon are also taken into account. We call this period the *future horizon*. Consequently, the batch consists of two parts; those treatment requests arriving during the scheduling horizon and those that arrive during the future horizon. Then, the scheduling exercise is made for the entire batch of requests. Instead of fixing the schedule for each of these requests, only patients requesting treatment in the scheduling horizon receive a fixed schedule (patients 1 to x in the figure). By including the second part of the batch as input to the ILP solver, the solver schedules patients from the scheduling horizon by explicitly considering the impact on patients that request treatment during the future horizon (patients x+1 to y). In a subsequent iteration, both the scheduling and future horizon move forward with a period equal to the length of the scheduling horizon.

4.2 ASAP procedure

In this section, a myopic online procedure is proposed. The method is essentially oblivious since it does not take into account any stochastic information on patient arrivals. Upon scheduling a newly arriving patient, the only available information is the patient's characteristics and previously scheduled patients and their booked appointments. The chosen procedure is based on the as soon as possible paradigm. ASAP algorithms or greedy procedures that give almost identical results are commonly used in the radiotherapy scheduling literature to reflect the current manual scheduling practice (e.g. S. Petrovic et al., 2006; Legrain, Fortin, et al., 2015; Pham et al., 2021). The ASAP algorithm is also the scheduling heuristic that most closely reflects the current scheduling practice at AZ Sint-Lucas. Furthermore, the method will serve as an upper bound when comparing with the online stochastic procedure. Figure 11 provides an illustration of the algorithm.

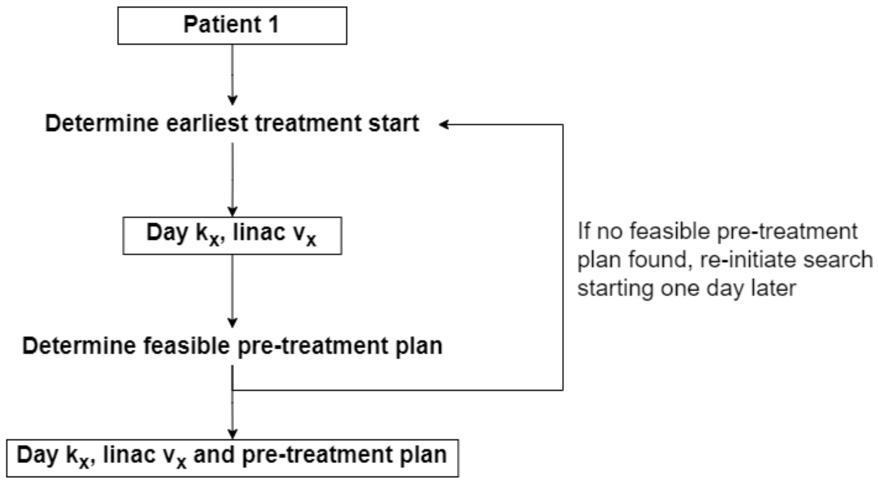


Figure 11: ASAP procedure

When a request for treatment is issued upon arrival of a patient (patient 1 in the figure) at the radiotherapy department, the goal of the ASAP procedure is to assign a linac appointment on the earliest feasible day for that patient. To ensure feasibility, it is also necessary that the pre-treatment operations can be scheduled without violating any constraints. If no feasible pre-treatment appointments can be generated, the algorithm re-initiates the search for the earliest treatment start, but starting one day later. This process is continued until the patient has a feasible pre-treatment plan and start date for treatment.

4.3 Online stochastic procedure

In this section, an online stochastic procedure is presented. Contrary to the oblivious method in the previous section, this method factors for future patient arrivals, by taking into consideration historical information. The proposed method is based on the sample average approximation paradigm (Verweij et al., 2003). The SAA method uses the concept of *scenarios*, which is commonly used in online stochastic modelling related to the radiotherapy department as a way to predict the future state of a system and successfully anticipate future capacities (e.g. Legrain, Fortin, et al., 2015; Chang et al., 2020). In this dissertation, sampling is used to construct several scenarios of future patient arrivals. These arrivals are subsequently used as input to the ILP model. Solving the model for every scenario results in multiple candidate solutions to the problem. Figure 12 provides on overview of this technique.

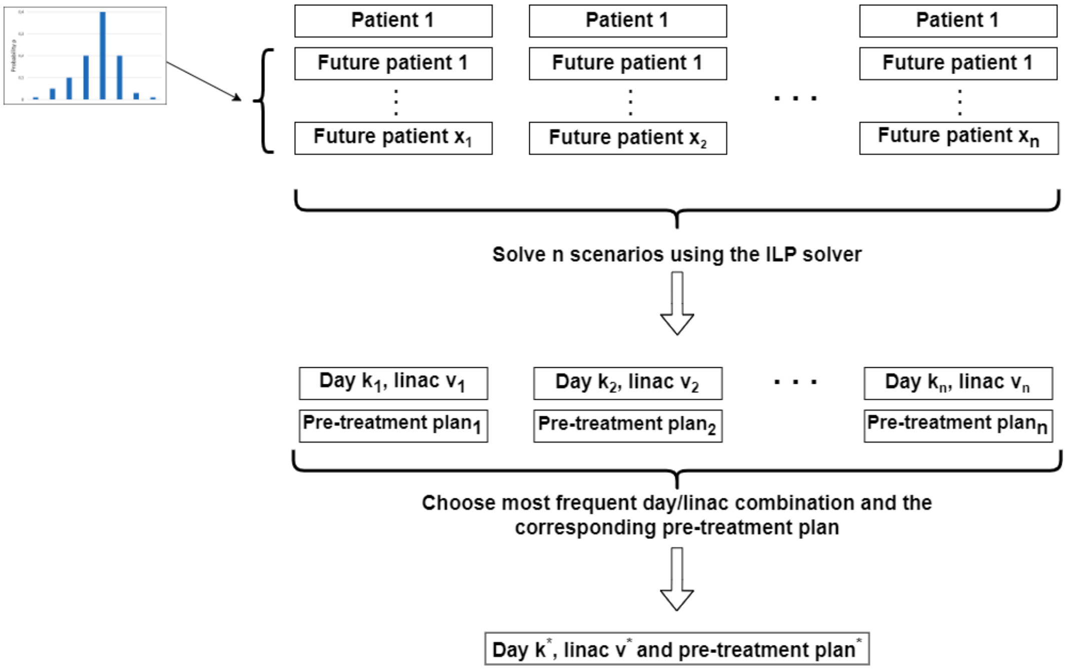


Figure 12: Online stochastic procedure

When a patient (patient 1 in the figure) arrives at the medical centre and a treatment request is issued, the OS algorithm uses empirical distributions based on historical data to sample future patients and their characteristics. In accordance with the SAA method, n scenarios are solved using the ILP solver (Verweij et al., 2003). From the n outputs, the most frequent day and linear accelerator combination are determined. The output of the algorithm is the chosen start date for treatment and the corresponding pre-treatment plan.

5 Numerical results

In this chapter, the mathematical model and solution procedures are tested on real-world data provided by AZ Sint-Lucas. Section 5.1 tailors the problem description to the specific case of AZ Sint-Lucas and introduces the input for the ILP model. Section 5.2 presents the implementation of the solution procedures. Finally, section 5.3 presents various experiments that are conducted and their results.

5.1 Test design

Section 5.1.1 presents the problem description tailored to AZ Sint-Lucas. In section 5.1.2, an overview of the collected data is presented. Furthermore, the values of the weights used in the objective function in chapter 3 are determined in section 5.1.3.

5.1.1 Assumptions

Tailoring the problem description (cf. chapter 3) to the radiotherapy department of AZ Sint-Lucas allows us to reduce the problem size. Figure 13 illustrates the patient pathway up until (s)he receives the necessary appointments to start treatment. Exceptions to this pathway are possible, but not included in the scope of this dissertation; only the general or average pathway of a patient is discussed.

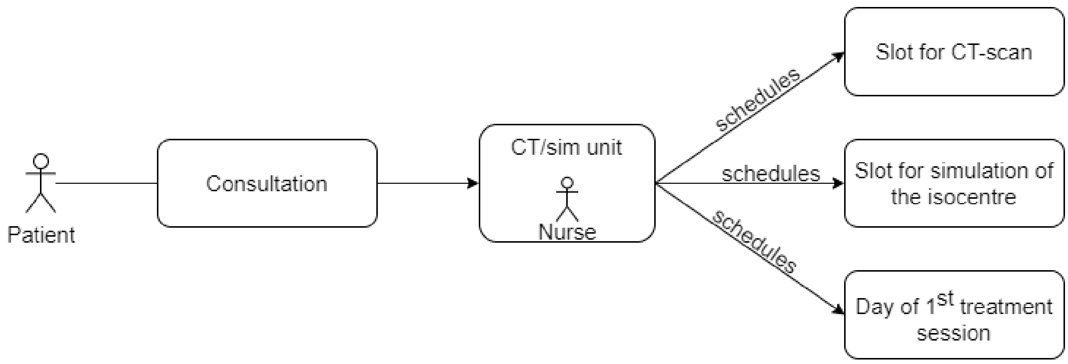


Figure 13: Patientflow at AZ Sint-Lucas until booking of relevant sessions

First, a patient arrives at the hospital at the time of the consultation appointment with its radiation oncologist, which has been booked in advance. Afterwards, the radiation oncologist guides the patient to the CT/simulation unit, where a nurse schedules appointments for the CT-scan, the simulation of the isocentre and the start of the radiation delivery phase. The nurse manually schedules the appointments for each patient and uses his or her experience to ensure that all patients can start their treatment in a timely manner. To test the methodologies proposed in this master's dissertation, the same scheduling decisions that the nurse makes in the CT/sim unit are made. This implies that two operations from the pre-treatment phase are scheduled (CT-scan and simulation of the isocentre). In addition, the patient receives a date that stipulates the first day on which a radiation fraction will be delivered.

Patients are categorised based on their treatment intent. As a result, three patient categories are defined: palliative patients, curative patients and patients with the goal of definitive tumor control. The last two groups are also described as non-palliative patients (cf. section 3.1).

Figure 14 illustrates the time window rules between operations. Both the CT-scan and the simulation of the isocentre take 30 minutes to complete and each patient has to undergo both pre-treatment tasks. A minimum of 5 days has to elapse between the two pre-treatment operations. Once the simulation has been completed, the patient is ready to receive its first fraction. This fraction has to be delivered within 2 days. Both time constraints in figure 14 are treated as hard constraints in the model, indicating that no deviation is allowed.

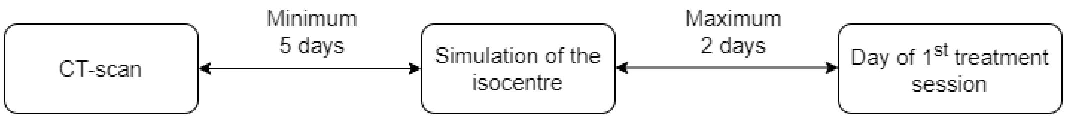


Figure 14: Time windows between operations

Resources are the same for both pre-treatment operations. A flexible pool of three nurses is always available, of which two are needed to perform a pre-treatment operation. The other nurse is available to perform other tasks in the department. Moreover, one machine is available and necessary to perform both the CT-scan and the simulation. Since the machine is the bottleneck in these operations, the nurse pool is not considered as input to the model (figure 15). The CT-sim machine can be operated from 8h30 to 17h30 on weekdays, resulting in 18 slots of 30 minutes per day. Since the duration of the pre-treatment operations is always 30 minutes, only having timeslots with a duration of 30 minutes suffices. Regarding the treatment phase, three linear accelerators are available. They can be operated from 8h00 until 16h30, with a one-hour break. One of the linacs (yellow) can perform 4 treatments per hour, the others (green and blue) 6. As a result, the former has a capacity of 30 slots a day, while the other two linacs can accommodate 45 treatment sessions per day. It can be seen in section 5.3.1 that the capacity currently available in AZ Sint-Lucas is sufficient to schedule all patients on time. Therefore, experiments with a reduction in capacity are also explored in section 5.3. Deliberately setting the capacities at a lower level allows for a better evaluation of the proposed methods.

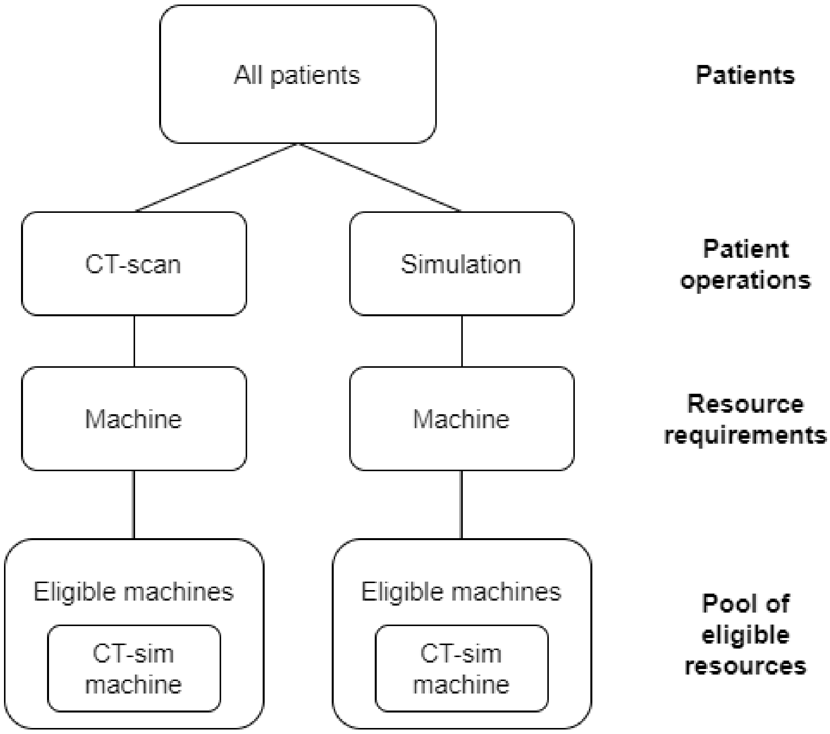


Figure 15: Pre-treatment operations and resources

Upon arrival at the CT/simulation unit, the patient's cancer site is known. In addition, the number of treatment sessions that the patient has to undergo, along with the duration of each session, has been decided. Furthermore, the due date of a patient, i.e. the day by which the patient has to start treatment, can be determined based on the treatment intent, which is also known. Palliative patients have a due date of 5 working days after the release date, whereas non-palliative patients should start treatment within 10 working days. It is decided to give palliative patients a priority of 3 and non-palliatives have a priority equal to 1. In section 5.3, an experiment with different values of these priorities is discussed.

5.1.2 Data collection

To test the performance of the mathematical model and the solution methods, patient data was collected from AZ Sint-Lucas. The dataset contained 1219 treatment requests from patients that started treatment in 2021 and 1260 starting in 2022, including those that arrived at the end of 2020 and 2021 respectively. The latter group arrived at the hospital in one year, but only started treatment in the following year. In order to obtain a coherent subset, only the requests from patients arriving in 2021, respectively 2022 were taken into account which amounts to a total of 1147 and 1212 patients. Moreover, patients treated by brachytherapy or contacttherapy were left out, since they follow a substantially different treatment pathway. Similarly, requests for electrontherapy were not taken into account. It is assumed that the duration of each fraction (z_j) in a treatment request is equal to 1. This matches the current manual approach used at AZ Sint-Lucas. Occasionally, a patient requests more than one treatment simultaneously. In that case, the fractions must be delivered at the same time and hence, multiple treatment requests are combined into one request, but with z_j being multiplied. After these pre-processing steps, 1037 (year 2021) and 1024 (year 2022) different treatment requests remain. The vast majority has a fraction duration of 1 base timeslot. Both years are summarised in table 6.

Duration (z_j) of each fraction (expressed as a number of base timeslots)	**Percentage 2021**	**Percentage 2022**
1	96.24%	93.46%
2	2.80%	5.27%
3	0.96%	1.17%
4	0.00%	0.10%

Table 6: Overview of the fraction durations

The number of fractions per request ranges from 1 to 39, with 20 being the most commonly requested in both years and accounting for 23% and 25% respectively. During 2021, a total of 19407 fractions were requested, averaging 19 fractions per request. In 2022, the average was 18 fractions per request. An overview of the number of fractions per treatment request is given in figure 16.

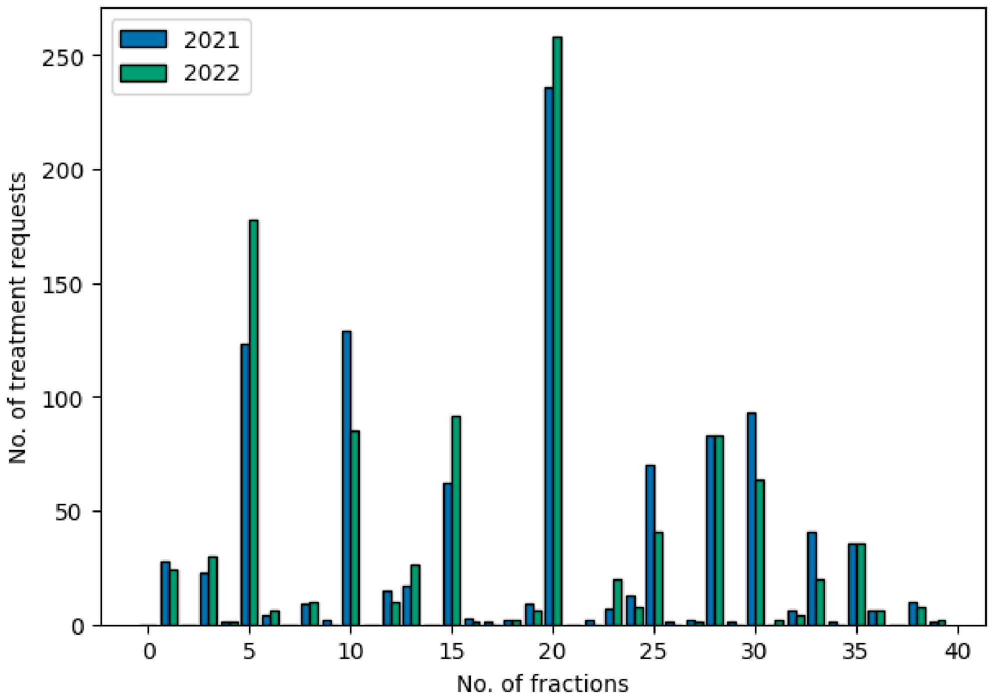
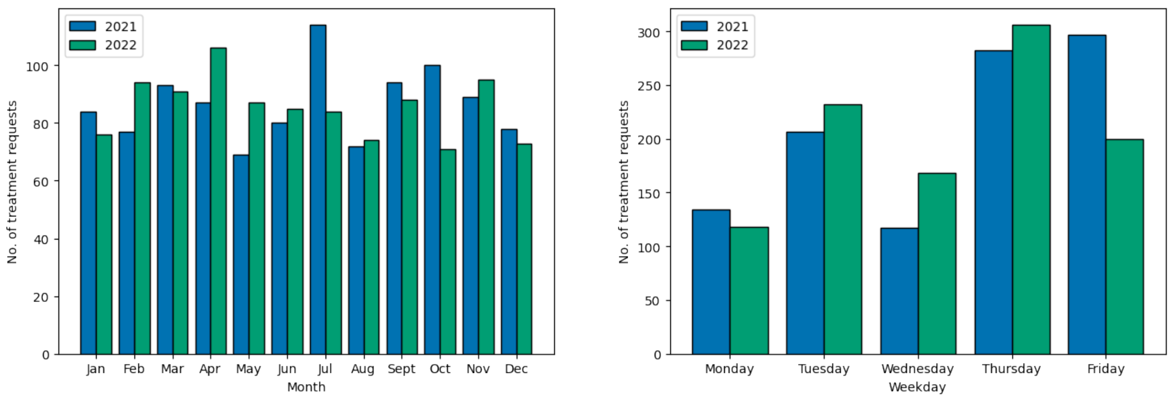


Figure 16: Treatment fractions distribution

In figure 17a, it is shown how the demand is distributed over 12 months for both years. Patients only arrive on weekdays. The distribution of patient arrivals over all weekdays is shown in figure 17b. In both 2021 and 2022, Mondays and Wednesdays appear to be less crowded than other days.



(a) Demand distributed over months

(b) Demand distributed over weekdays

Figure 17: Overview of demand

Palliative patients account for roughly one-fifth (2021: 21%; 2022: 17%) of all treatment requests. These patients must start treatment within 5 working days, following the guidelines stated by AZ Sint-Lucas. Curative patients and patients with the goal of definitive tumor control each have a target waiting time of 10 working days. This distinction based on treatment intent is the only factor deciding the waiting time targets in this dissertation.

Three linear accelerators are available at AZ Sint-Lucas. Furthermore, the most commonly utilised treatment techniques were VMAT and IMRT. In the remainder of this chapter, the solution algorithms developed in chapter 4 are tested on the 2022 data. The data from 2021 is used as historical information in the online stochastic procedure.

5.1.3 Objective function weights

The objective function presented in chapter 3 contains two parts. The first term in the equation minimises the weighted amount of patients that exceed the waiting time targets. This part is assigned weight G_1. The second term minimises the total weighted tardiness of all patients and is given a weight G_2. To determine the weights G_1 and G_2, both terms have to give an output in the same range. Therefore, the second term was normalised. To do so, $\sum_{j \in \mathcal{P}}(w_j \cdot L_j)$ was divided by the weighted sum over all patients of the maximum tardiness of each patient. The maximum tardiness is determined by the patient's due date and the scheduling horizon. Normalising the second term ensures that both terms reach a value in the range $[0, 1]$. To determine the weights, an approach based on Zeleny (1974) is used.

G_1	G_2
1.00	0.00
0.95	0.05
0.90	0.10
...	...
0.10	0.90
0.05	0.95
0.00	1.00

Table 7: Objective function weights

The ILP model is run 21 times. G_1 and G_2 range from 0.00 to 1.00 with steps of 0.05 (see table 7). All other input is equal for each instance. In order to warm up the model, the model was run over 13 consecutive weeks, according to the batch scheduling method.

The values related to both terms in the objective function are plotted in figure 18. It has to be noted that the combination of $G_1 = 1.00$ and $G_2 = 0.00$ is an outlier and is therefore not included in the scatter plot. The red dot indicates the theoretical minimum, which cannot be achieved in practice. This point is determined by combining the results of using $G_1 = 1.00$ and $G_2 = 0.00$ and $G_1 = 0.00$ and $G_2 = 1.00$. The distance from the theoretical point is subsequently calculated using the euclidean distance formula $D = \sqrt{(X_R - X_1)^2 + (Y_R - Y_1)^2}$ with D being the distance, (X_R, Y_R) the coordinate of the theoretical point and (X_1, Y_1) the coordinate of the point for which to calculate the distance. The green dot in figure 18 has the closest distance to the theoretical minimum and represents two G_1, G_2 combinations: $0.35, 0.65$ and $0.45, 0.55$. In the remainder of this dissertation, it is arbitrarily decided that $G_1 = 0.45$ and $G_2 = 0.55$

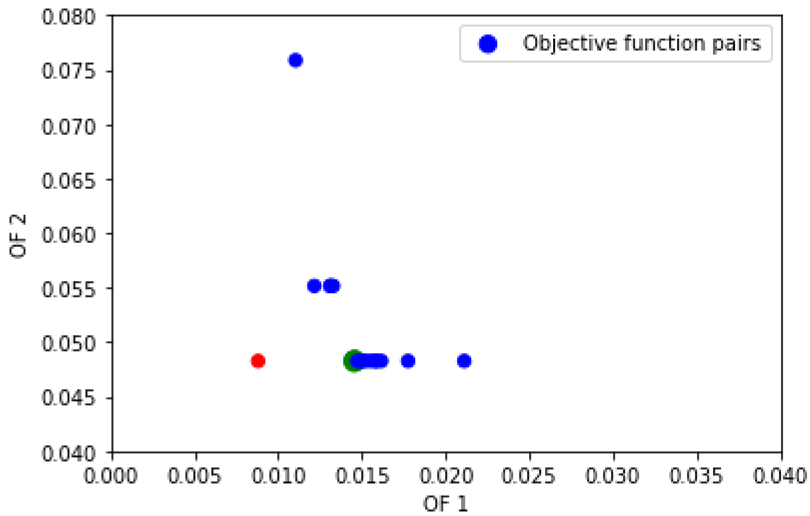


Figure 18: Objective function pairs for different combinations of weights

5.2 Implementation of the solution procedures

In this section, the three solution procedures developed in chapter 4 are implemented with respect to the characteristics of the radiotherapy department at AZ Sint-Lucas.

5.2.1 Tailored batch procedure

In this section, the adapted batch algorithm is presented. Treatment requests are accumulated over 5 working days and this represents the scheduling horizon (A_1). Additionally, the future horizon (A_2) is equal to 15 working days. The last patient from 2022 arrives 260 working days after the start of the year. Since the index of the starting day is equal to zero, the last day receives an index equal to 259. The pseudocode of the procedure is given in algorithm 1.

Algorithm 1 Batch scheduling

procedure MAIN
    for t in [0, 259, step = 5] **do**
        $A_1 \leftarrow$ treatment requests arriving in $[t, t+4]$
        $A_2 \leftarrow$ treatment requests arriving in $[t+5, t+5+14]$
        SOLVEILP(A_1, A_2)
        Update capacities by fixing the schedules from A_1

The next two sections present the online (ASAP) and online stochastic procedure developed in this dissertation. k_1 and v_1 refer to the day and linear accelerator of the first fraction administration, respectively. k_2 and e_2 represent the day and timeslot of the simulation of the isocentre task (i.e. the second pre-treatment task). k_3 and e_3 are the day and timeslot of the CT-scan task (i.e. the first pre-treatment task). These conventions are used in both algorithms.

5.2.2 Tailored ASAP procedure

This section adapts the ASAP algorithm to the characteristics of the scheduling problem at AZ Sint-Lucas. In algorithm 2, a pseudocode representation of the procedure is presented. It makes use of three other functions; *FindFeasTreatmentStart*, *FindFeasSimulationSlot* and *FindFeasCTSlot*. Each of these is briefly explained below and the pseudocode for these functions can be found in the appendix (A.1).

Algorithm 2 ASAP procedure

```
procedure ASAP
    for j in patients do
        (k1, v1) ← FindFeasTreatmentStart(j, a_j + 5)
        (k2, e2) ← FindFeasSimulationSlot(j, k1, 0)
        (k3, e3) ← FindFeasCTSlot(j, e2, k2)

        if k1 =Null, k2 =Null or k3 =Null then        ▷ i.e. if no full schedule found
            (k1, v1) ← FindFeasTreatmentStart(j, a_j + 6)
            (k2, e2) ← FindFeasSimulationSlot(j, k1, 1)
            (k3, e3) ← FindFeasCTSlot(j, e2, k2)

            if k1 =Null, k2 =Null or k3 =Null then    ▷ i.e. if no full schedule found
                control ← True
                counter ← 0
                while control = True do
                    control2 ← True
                    while control2 = True do
                        (k1, v1) ← FindFeasTreatmentStart(j, a_j + 7 * counter)
                        (k2, e2) ← FindFeasSimulationSlot(j, k1, 2)
                        if k2 = Null then
                            counter = counter + 1
                        else
                            control2 =False
                    (k3, e3) ← FindFeasCTSlot(j, e2, k2)
                    if k3 = Null then
                        counter = counter + 1
                    else
                        control =False
                schedule k1, v1, k2, e2, k3, e3
            else
                schedule k1, v1, k2, e2, k3, e3
        else
            schedule k1, v1, k2, e2, k3, e3
```

The goal of function *FindFeasTreatmentStart* is to find the earliest (cf. ASAP) starting date for the treatment of a patient and a certain linac, while ensuring that the linac has enough capacity for the entirety of the treatment duration. It has as input parameters patient j and the earliest day that the function should start searching for an available linear accelerator k_0. The function iterates over the pre-determined horizon. This horizon is set at a level that ensures that all patients can be successfully scheduled. Subsequently, a feasible starting date is searched on any of the linear accelerators. As soon as a feasible date is found, the function is terminated and the first day of treatment k_1 on linac v_1 is returned.

Additionally, *FindFeasSimulationSlot* aims at finding an available slot for the simulation operation, provided that patient j is scheduled to start treatment on day k_1. Parameter r controls for the search space in terms of days. If r is for example equal to 2, the function searches an available slot in days $k_1, k_1 - 1$ and $k_1 - 2$. In the next paragraph, the necessity of this parameter will become apparent. Within a day of the search space, the method finds the last available slot of that day. This may seem counter-intuitive in an ASAP procedure. However, upon closer examination, the underlying rationale for structuring it in this manner becomes apparent. Given that there is a maximum time of 2 days between the simulation appointment and the start of treatment on the one hand and a minimum of 5 days between CT-scan and simulation on the other hand, scheduling the simulation as close as possible to the start of treatment allows for an increase in the search space of function *FindFeasCTSlot*. Figure 19 provides an illustration. Scheduling the simulation appointment on slot 3 of day 12 means that the CT-scan can only be scheduled in slot 1 or 2 of day 7. If, however, the patient is scheduled in slot 8 on day 12, slot 3 to 7 become feasible to plan the CT-scan as well.

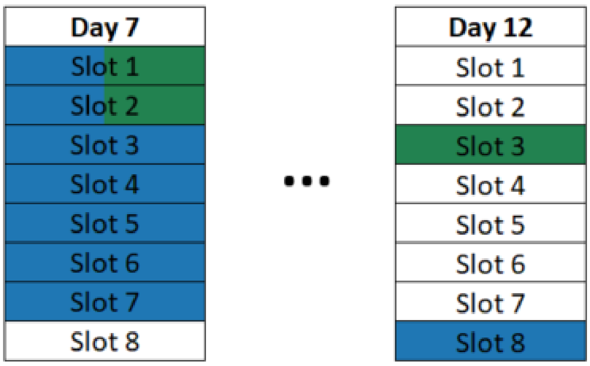


Figure 19: Last available slots

FindFeasCTSlot is similar to *FindFeasSimulationSlot*. The main difference pertains to the search space. Here, all days between the patient's release date a_j and 5 days before the simulation (i.e. $k_2 - 5$) are available for scheduling the CT-scan appointment.

In the main procedure (algorithm 2), each patient receives an appointment date and slot for the CT-scan and simulation tasks. Furthermore, the day and linear accelerator for the first treatment session are returned. First, the procedure tries to schedule the first day of treatment 5 days (or more if no linac is available) after the release date. Given this date, the simulation and CT-scan slot are booked successively. If no complete schedule is found, the same steps are repeated, but now it is ensured that the treatment starts at least 6 days after the release date to give more room for the simulation and CT-scan slot. Again, if no full schedule is found, the methods advances to 7,8,9, ... days after the release date. The reason for doing the scheduling starting from 5 and 6 days after the release date separately is the previously mentioned parameter r. Figure 20 demonstrates how r guarantees that there is sufficient space left in the schedule for the CT-scan slot when scheduling the simulation slot. It is assumed that the release date of the patient is equal to day 0. As soon as a complete schedule is found, the resource capacities are updated and the procedure moves to the next patient.

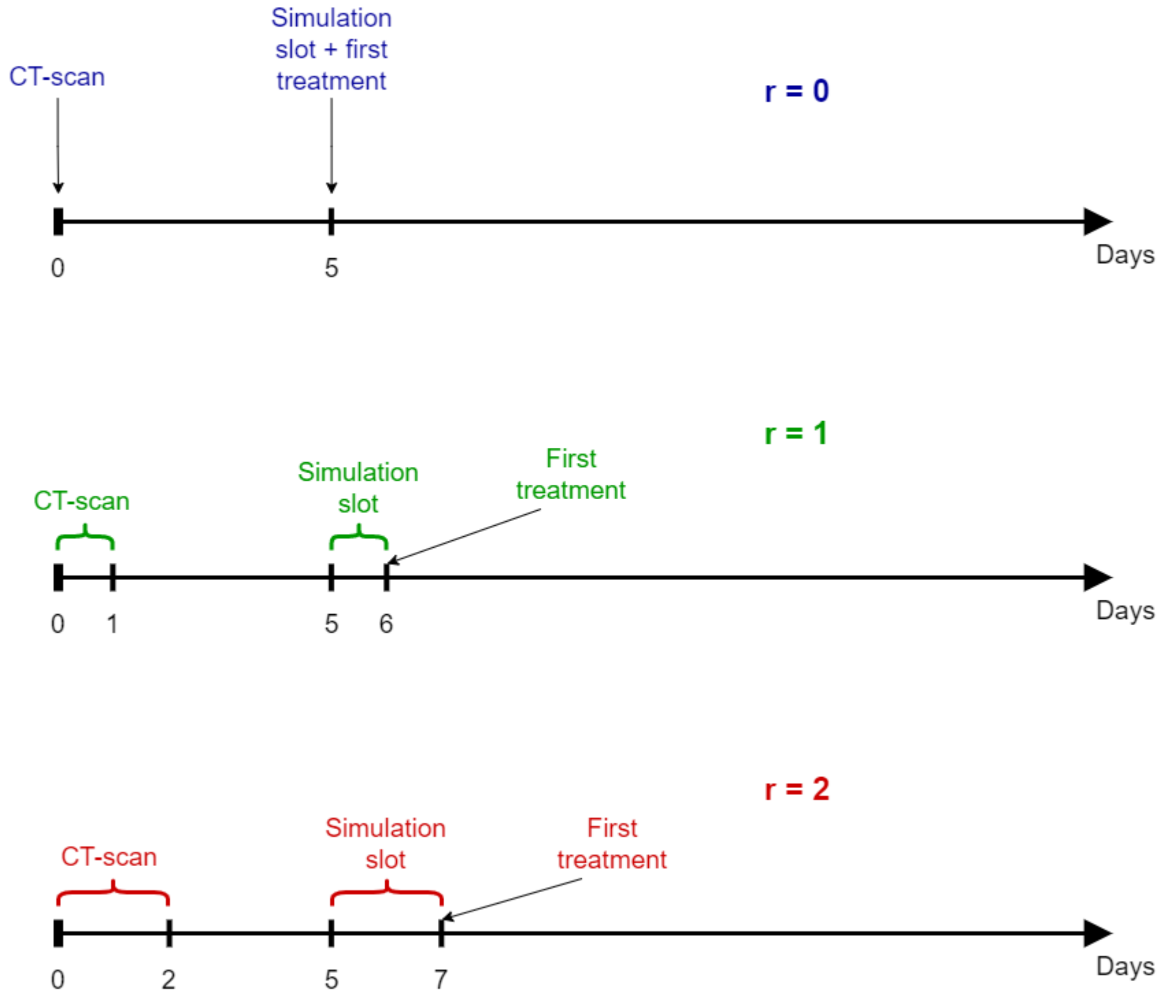


Figure 20: Value of r

5.2.3 Tailored online stochastic procedure

This section elaborates on the case-specific implementation of the stochastic online procedure proposed in section 4.3. Algorithm 3 presents the complete pseudocode representation of the procedure.

Algorithm 3 Online stochastic procedure

1: **procedure** OnlineStochastic
2:     *output* $\leftarrow$ Null
3:     **for** j `in patients` **do**
4:         **for** $i \in [1, 10]$ **do**   ▷ to generate 10 scenarios
5:             $\omega \leftarrow$ GenerateScenario( )
6:             *patientInput* $\leftarrow j$ & ω
7:             *output* = *output* + SolveILP(*patientInput*)
8:         $(k_1, v_1) \leftarrow$ MostFrequentTreatment(*output*)
9:         **if** MostFrequentTreatment gives a tie **then**
10:             **if** j is palliative **then**
11:                 Schedule the soonest of the tied (k_1, v_1)'s
12:             **else**
13:                 Schedule the latest of the tied (k_1, v_1)'s
14:         $(k_2, e_2, k_3, e_3) \leftarrow$ SinglePassHeuristic(k_1, v_1)
15:         **schedule** $(k_1, v_1, k_2, e_2, k_3, e_3)$

On line 2, the *output* object is defined and serves the purpose of storing the output for each of the 10 scenarios. Line 5 assigns a list of future patients and their features (e.g. release date, treatment intent, etc.) to the ω variable. The *GenerateScenario* function uses empirical distributions based on historical treatment requests to construct the scenario. *PatientInput* is assigned a list combined of patient j and future arrivals in scenario ω. In line 7, the ILP solver is used to run the model with *patientInput* as input. The output is stored in the *output* object. To assign the first day of fraction administration and the linear accelerator that will accommodate the entire treatment of patient j, *MostFrequentTreatment* gets the most frequent (k_1, v_1) combination out of the 10 combinations (one for each scenario) that were obtained and stored in *output*. If a tie occurs, a palliative patient is scheduled as soon as possible, whereas curative and definitive tumor control patients are assigned the latest possible day out of the tied combinations. Each of these most frequent (k_1, v_1) combinations is also accompanied by a (k_2, e_2) and (k_3, e_3) that represent the simulation and CT-scan slot respectively. The *SinglePassHeuristic* returns these appointments. In case of a tie on line 9, multiple (k_2, e_2) and (k_3, e_3) combinations may qualify as potential solutions. To resolve this, the single-pass heuristic randomly chooses one of these candidate solutions.

The *GenerateScenario* function in algorithm 3 uses a uniform random number generator followed by the inversion method to sample from the distributions given in section 5.1.2. This approach is based on a lecture given by Maenhout (2022). Scenarios are calculated based on the complete dataset of 2021.

First, the discrete distributions given in section 5.1.2 are converted to their equivalent cumulative probabilities. Subsequently, the identification of the distributions suitable for sampling, demands for the execution of statistical tests. The Kolmogorov-Smirnov (K-S) test, developed by Smirnov (1939), is a non-parametric test that can be performed to test the congruence of two empirical distributions. In a non-parametric test, no assumptions about the form of the underlying population distribution are made (Hoeffding, 1994). However, the K-S test is known to have low power (i.e. low probability of correctly rejecting the null hypothesis of equal distributions) under small sample sizes (Massey Jr, 1951). When doing multiple comparisons during hypothesis testing, it becomes more likely that at least one of these comparisons leads to a false rejection of the null hypothesis. To counter this effect, a correction can be made. A widely used correction is the Bonferroni correction (Bonferroni, 1936). In this method, the original significance level is divided by the number of comparisons. As an example, take 10 samples that are to be tested simultaneously for normality at the 5% significance level. Instead of testing each sample at the 5% significance level, applying the Bonferroni correction requires each sample to be tested individually at the $5\%/10 = 0.5\%$ significance level. Another useful test is the Shapiro-Wilk (S-W) test, which can be used to test the normality of a distribution, also when the sample size is relatively small (Shapiro & Wilk, 1965). It is worth noting that all statistical tests conducted within the remainder of this section adhere to a 5% significance level.

Regarding the monthly patient arrivals, each observation in a monthly sample is the number of patient arrivals on a working day. One month contains around 20 working days. As a consequence, the sample size is relatively small. The Shapiro-Wilk test with Bonferroni correction was therefore performed to examine the possibility of using a parametric test. Based on the results of this test (see table 8), the null hypothesis that every month has a normally distributed distribution is rejected on a 5% significance level. Subsequently, the non-parametric Kolmogorov-Smirnov test with Bonferroni correction is executed. The results of this test indicate that the monthly distributions are not significantly different from each other (appendix A.2).

Month	**Significance level (α)**	**Total # comparisons (n)**	**Corrected α ($= \alpha/n$)**	**p-value**
January	5%	12	0.00417	0.09865
February	5%	12	0.00417	0.00034*
March	5%	12	0.00417	0.02192
April	5%	12	0.00417	0.00449
May	5%	12	0.00417	0.00860
June	5%	12	0.00417	0.03840
July	5%	12	0.00417	0.09155
August	5%	12	0.00417	0.06915
September	5%	12	0.00417	0.17414
October	5%	12	0.00417	0.13836
November	5%	12	0.00417	0.00019*
December	5%	12	0.00417	0.12374

Table 8: S-W test for monthly patient arrivals

The next test to be completed relates to the difference between weekdays. Every weekday has 52 observations over the course of one year. As a result, the Kolmogorov-Smirnov test (with Bonferroni correction) can be used and there is no need to investigate the possibility of using a parametric test. The results are summarised in table 9. In conclusion, a statistically significant difference was found in the distributions of patient arrivals for every weekday.

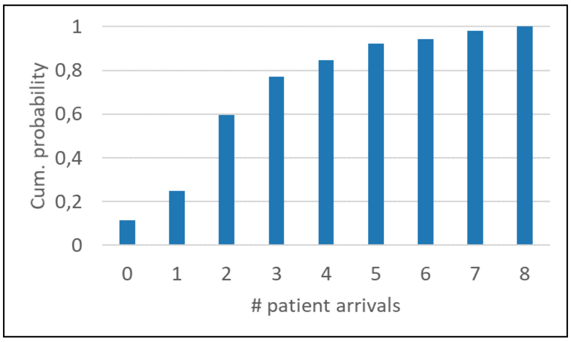
Day 1	**Day 2**	**Significance level (α)**	n	**Corrected α ($= \alpha/n$)**	**p-value**
Monday	Tuesday	5%	10	0.005	0.02592
Monday	Wednesday	5%	10	0.005	0.73897
Monday	Thursday	5%	10	0.005	< 0.00001*
Monday	Friday	5%	10	0.005	< 0.00001*
Tuesday	Wednesday	5%	10	0.005	0.00736
Tuesday	Thursday	5%	10	0.005	0.07734
Tuesday	Friday	5%	10	0.005	0.00736
Wednesday	Thursday	5%	10	0.005	< 0.00001*
Wednesday	Friday	5%	10	0.005	< 0.00001*
Thursday	Friday	5%	10	0.005	0.99830

Table 9: K-S test for patient arrivals on weekdays

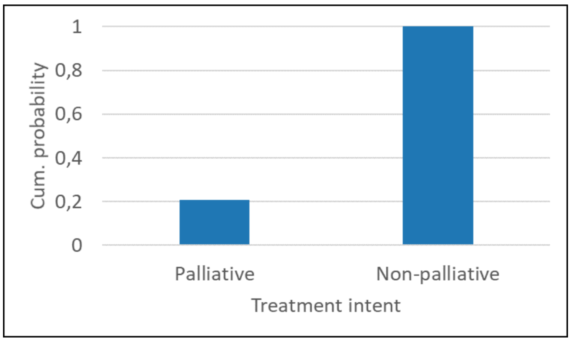
n refers to the number of pairwise comparisons

Upon closely observing the data of 2021, a clear distinction in the treatment duration based on treatment intent (i.e. palliatives vs non-palliatives) seems to be present. Indeed, a Kolmogorov-Smirnov test supports this finding (p-value < 0.05). The same test was performed to test the difference between palliative patients and non-palliative patients in relation to the fraction duration. In that case, the null hypothesis could not be rejected.

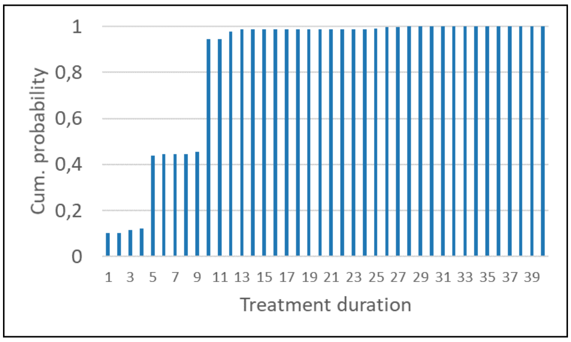
The resulting cumulative distributions can be found in figure 21. With regard to the weekdays, only the distribution for Monday is shown; the distributions for the other weekdays are trivial.



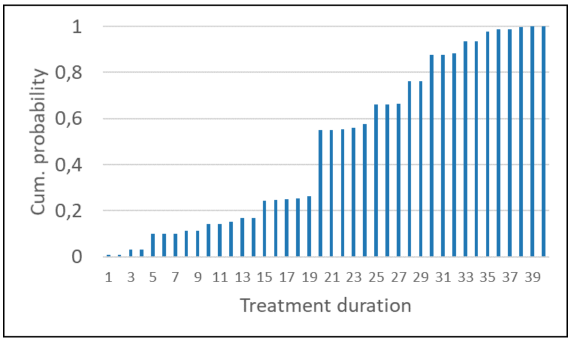
(a) Cum. probability patient arrivals Monday



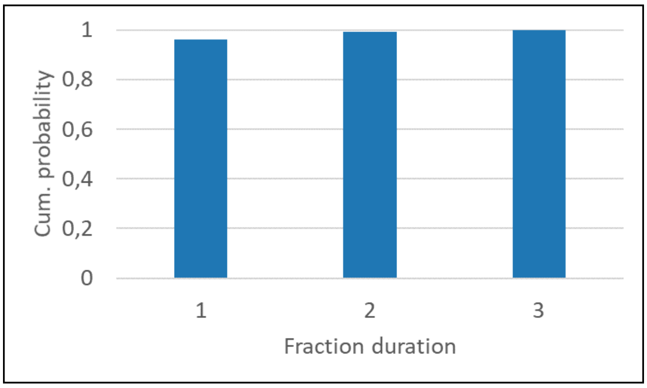
(b) Cum. probability treatment intent



(c) Cum. probability treatment duration - palliative



(d) Cum. probability treatment duration - non-palliative



(e) Cum. probability fraction duration

Figure 21: Cumulative distributions

After defining the cumulative distributions, the *second* step is to generate a uniform random number. In this dissertation, a random number in the range of $[0, 1[$ is generated using the *random()* function from the *random* package in Python, with seed equal to 2. The *third* and final step is to use the inversion method (Maenhout, 2022). To illustrate this method, we take the distribution from figure 21b. If the generated random number lies within the interval $[0; 0.2073288332[$, a palliative patient is generated. If the number is within $[0.2073288332; 1[$, the generated patient is non-palliative. Using the inversion method ensures that the sampling procedure accurately reflects the underlying distributions.

5.3 Results

To examine the performance of the mathematical model and the offline, online and online stochastic algorithms, various experiments have been performed based on the data provided by AZ Sint-Lucas. This section presents the main results and findings of these experiments. The results are obtained with an Intel Core i5-8250U 1.60GHz CPU. The performance of the solution algorithms in terms of due date violations is based on two dimensions. It is important to realise that, although both dimensions relate to the objective function formulated in section 3.2.4, they are calculated without using patient priority weights and normalisation in this section. The first dimension relates to the first part of the objective function, indicating the number of patients that receive the first treatment session after their due date (i.e. patients that are *late*). The second dimension is connected to the second term in the objective function and reflects the total tardiness of patients. The tardiness of each patient is defined as the positive difference in working days between the first linear accelerator appointment and the due date of that patient. To test the statistical significance of the observed differences between the solution methods, the patients are grouped by the week of their release date. This allows for a Wilcoxon signed-rank test to be performed (Wilcoxon, 1945). The null hypothesis of this test states that the median of the differences between two paired samples is equal to zero. The test serves as a non-parametric alternative to the paired samples t-test. Noteworthy, the pairs that have a difference equal to zero are excluded from the test and the resulting sample size is named the reduced sample size, denoted as N in the remainder of this section. The Wilcoxon signed-rank test requires a minimum reduced sample size of 16 pairs (Mundry & Fischer, 1998). Therefore, the test is not performed when the reduced sample size is smaller than 16. Furthermore, all statistical findings are reported on a 5% significance level.

Section 5.3.1 provides a detailed analysis of the offline batch procedure and the online ASAP procedure. In section 5.3.2, the impact of using stochastic information on future patient arrivals is discussed. Finally, section 5.3.3 introduces an additional term to include in the objective function of the ILP model, in response to the findings from section 5.3.1.

5.3.1 Comparison between batch and ASAP procedure

The primary goal of the experiments in this section is to validate the mathematical model presented in section 3.2. To achieve this objective, the batch algorithm is compared to the ASAP algorithm. The latter gives results close to the current manual scheduling approach used at AZ Sint-Lucas. Five experiments are conducted; no capacity reduction, treatment capacity reduction, pre-treatment capacity reduction, a combined reduction of pre-treatment and treatment capacities, and a change in the priority weights. Optimality was not always achieved in the batch procedure; most often the optimality gap was $< 10\%$, but occasionally, the gap could not be decreased to below 50% after one hour of running. In some instances, the optimality gap decreased substantially, even after the 30-minute mark. Therefore, the results for the batch procedure are obtained using a time limit of one hour on the ILP solver and no optimality gap tolerance.

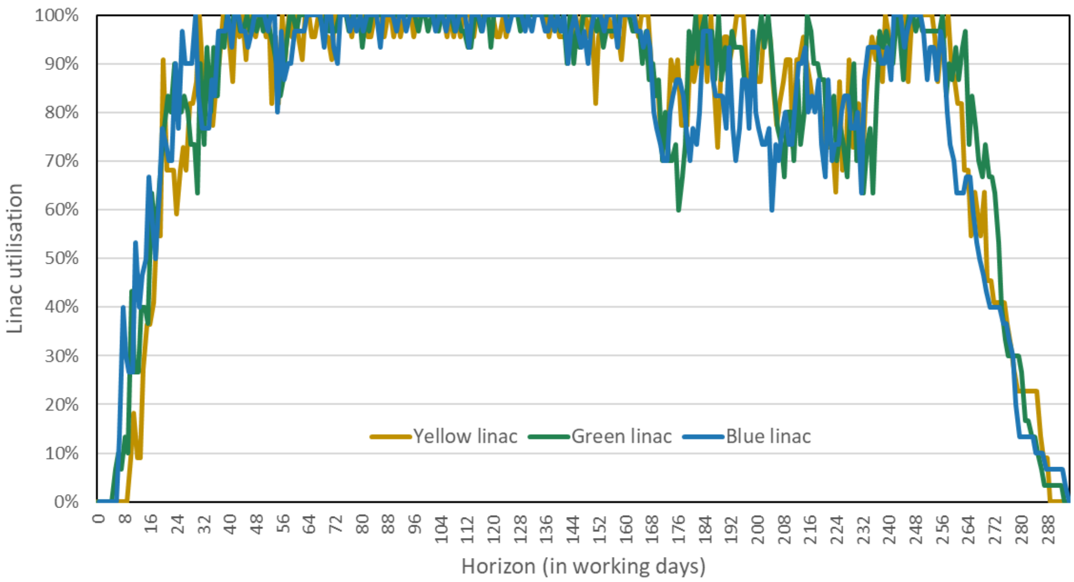
5.3.1.1 No capacity reduction

The first experiment assumes that the capacities are equal to the capacities mentioned in section 5.1.1. More specifically, 18 pre-treatment slots of 30 minutes are available on each day. Regarding the linear accelerators (treatment phase), the yellow linac has 30 slots available, whereas both the green and blue linacs have a capacity of 45 slots per day. Both the batch and ASAP procedure are tested using the complete 2022 dataset. Using the batch procedure resulted in zero due date violations, indicating that enough capacity is available to accommodate all patients. Slightly different results were obtained by using the ASAP method. A total of three palliative patients experienced a one-day delay in meeting their due dates. However, since the manual scheduling practice at AZ Sint-Lucas allows a scheduling nurse to use his or her experience, it is expected that they can adapt the schedule to ensure that those three patients are also scheduled on time.

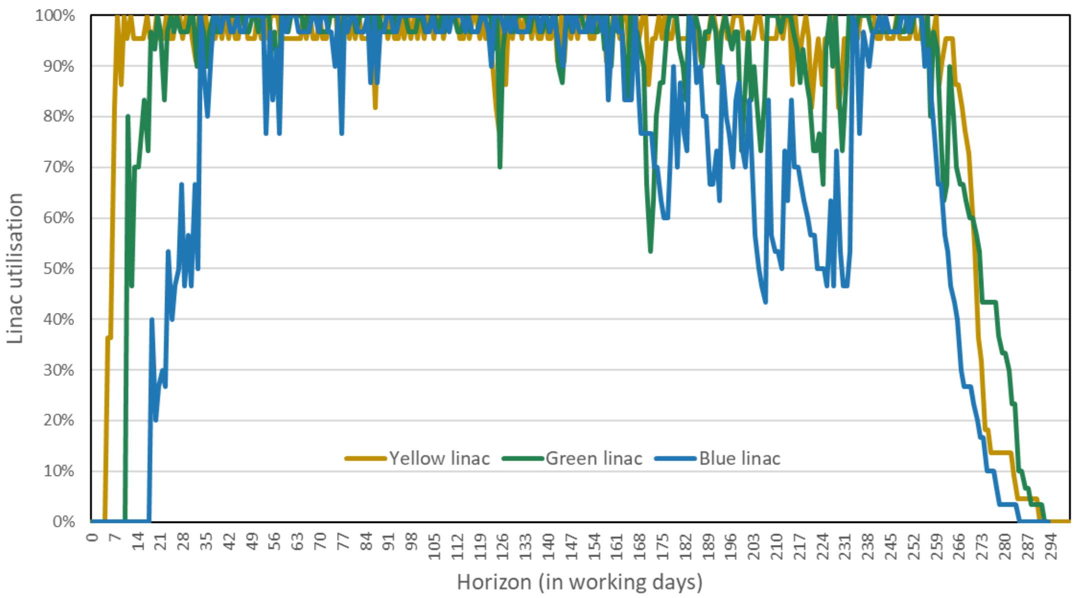
5.3.1.2 Treatment capacity reduction

In the second experiment, the treatment capacities are reduced. The yellow linac receives a capacity reduction of 25% (from 4 to 3 slots per hour). The green and blue linac have their capacities reduced by one-third, i.e. from 6 to 4 slots per hour. The resulting amount of daily slots is 22 for the yellow and 30 for the green and blue linac. The pre-treatment capacities are not subject to a reduction. The batch and ASAP procedure are tested on the complete 2022 dataset.

First, a close examination of the treatment resources is done, based on figure 22. The main difference in the utilisation is the distribution among the linear accelerators. In the ASAP procedure, the timeslots on the linacs are filled in a specific order. Upon searching for a feasible starting day that can accommodate the entire treatment plan of a particular patient, the algorithm iterates over the scheduling horizon. In each iteration, the order in which the linear accelerators are assessed is fixed; yellow, green and subsequently blue. This order is reflected in figure 22b. The utilisation in the batch procedure is more balanced. In addition, the graph appears to be smoother; between days 56 and 150, the sudden decreases in utilisation rate are less pronounced.



(a) Batch



(b) ASAP

Figure 22: Experiment 2: treatment utilisation

Secondly, the performance of the algorithms in relation to the due date violations is discussed. In the batch procedure, a total tardiness of 579 working days is observed. The average over the 1024 patients is equal to 0.56. Furthermore, a total of 43 (4.2%) patients are late. The ASAP procedure resulted in a total tardiness of 662 (14.3% increase) and 219 (21.4%) patients were late. The latter represents an increase of 409.3% compared to the batch procedure. The maximum tardiness is 19 for the batch procedure and 12 for the ASAP procedure. In section 5.3.3, an additional experiment is conducted related to the observed maximum tardiness. Regarding the distinction between palliatives and non-palliatives, the results are pronounced. In the batch procedure, 3 out of the 175 palliative patients are late, resulting in a total tardiness of 24 working days. The ASAP procedure, on the other hand, results in more than half of the palliative patients being late (95 out of 175) and a total tardiness of 373 working days. The batch procedure prioritises the on-time scheduling of palliative patients. Non-palliative patients have an observed total tardiness of 555 working days in the batch procedure, whereas the ASAP procedure has a tardiness of 289 working days. Noteworthy, despite the higher tardiness related to these patients, the batch procedure has a smaller amount of non-palliative patients being late, 40 compared to 124 patients. In section 5.3.1.5, additional experiments are executed to test the influence of the priority weights in the ILP model on the division in tardiness between palliative and non-palliative patients.

The results for the Wilcoxon signed-rank test on the weekly grouped data are presented in table 10. The weekly differences are shown to be statistically significant for palliatives and both categories combined. From the total and weekly average values in the table, we conclude that the batch procedure outperforms the ASAP procedure for palliative patients and for both patient categories together. On the other hand, the choice to prioritise palliative patients in the batch procedure results in more total tardiness for non-palliative patients.

Objective	Batch Total 2022	Batch Weekly avg	Batch N	ASAP Total 2022	ASAP Weekly avg	ASAP N	p-value (a)
Palliative							
#patients late	3	0.06	31	95	1.83	31	< 0.01*
Tardiness	24	0.46	31	373	7.17	31	< 0.01*
Non-palliative							
#patients late	40	0.77	13	124	2.38	13	/
Tardiness	555	10.67	13	289	5.56	13	/
Both categories							
#patients late	43	0.83	31	219	4.21	31	< 0.01*
Tardiness	579	11.13	31	662	12.73	31	0.047*

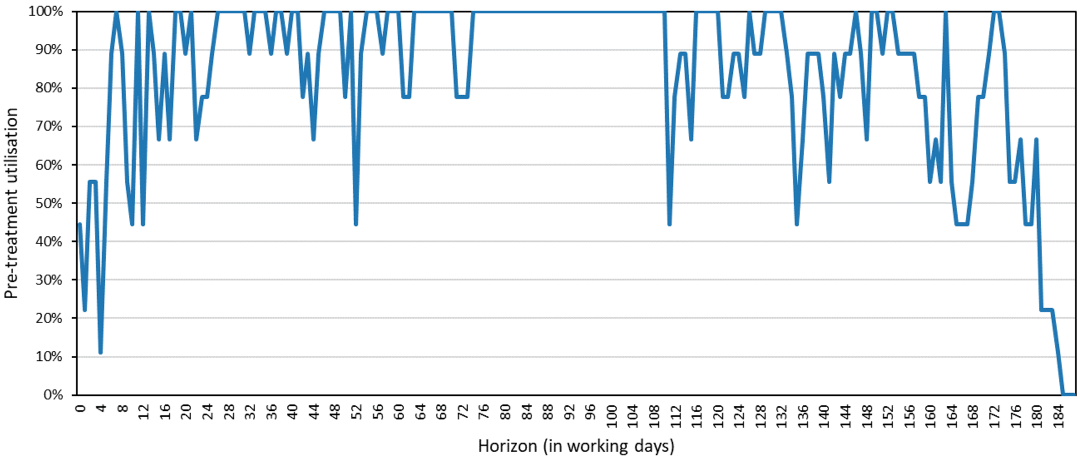
Table 10: Experiment 2: treatment capacity reduction - Patient due date violations

(a) 5% significance level

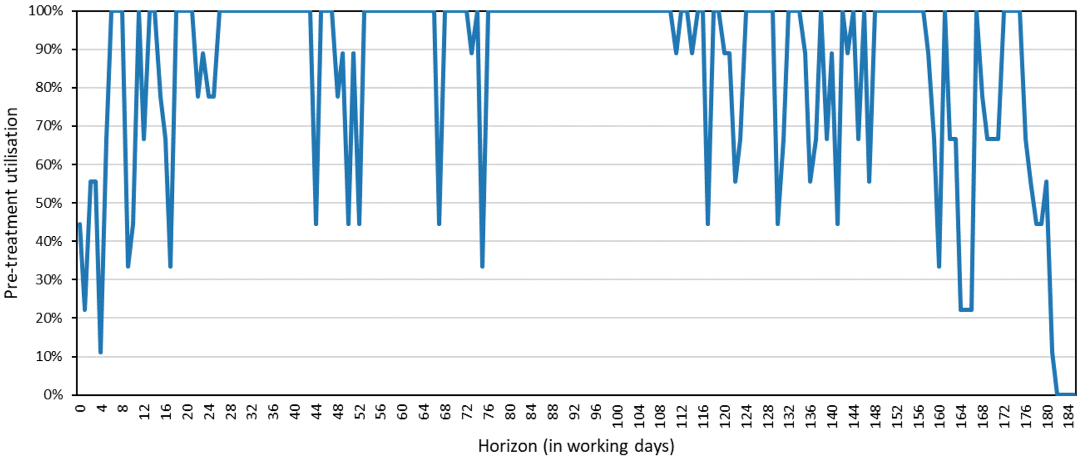
5.3.1.3 Pre-treatment capacity reduction

The third experiment reduces the pre-treatment capacity by 50%, resulting in a daily capacity of 9 timeslots. The treatment capacities, however, are not reduced. To evaluate the batch and ASAP procedures, a subset of the 2022 dataset comprising the first 35 weeks is utilised. By using a shorter timeframe of 35 weeks instead of the full 52 weeks, the running time is reduced without losing the ability to conduct a detailed analysis. Additionally, the statistical significance of the results can still be assessed using the Wilcoxon signed-rank test (for palliatives and both categories combined), since the reduced sample size is larger than 16.

First, a closer look at the pre-treatment utilisation is taken. Figure 23 presents the pre-treatment utilisation for both procedures. It is worth noting that in both procedures, the 50% reduction in capacity results in the pre-treatment resource becoming the bottleneck. Furthermore, visual inspection of the figure does not show any specific trends.



(a) Batch



(b) ASAP

Figure 23: Experiment 3: pre-treatment utilisation

Secondly, the due date violations are evaluated. From the results in table 11, it is clear that the batch procedure is better able to assign timely appointments to patients compared to the ASAP procedure. In addition, the batch procedure was able to schedule all palliative patients on time.

	Batch			ASAP			
Objective	Total	Weekly avg	N	Total	Weekly avg	N	p-value (a)
Palliative							
#patients late	0	0.00	28	68	1.94	28	< 0.01*
Tardiness	0	0.00	28	128	3.66	28	< 0.01*
Non-palliative							
#patients late	3	0.09	2	7	0.20	2	/
Tardiness	26	0.74	3	7	0.20	3	/
Both categories							
#patients late	3	0.09	28	75	2.14	28	< 0.01*
Tardiness	26	0.74	28	135	3.86	28	< 0.01*

Table 11: Experiment 3: pre-treatment capacity reduction - Patient due date violations

(a) 5% significance level

5.3.1.4 Pre-treatment and pre-treatment capacity reduction

In the fourth experiment, the capacity is reduced for both pre-treatment and treatment resources. The pre-treatment capacity is reduced by 50%, similar to the previous experiment. In addition, the treatment capacities receive the same reduction as in the second experiment: -25% for the yellow linear accelerator and -33% for the green and blue linac. the results are obtained using the first 35 weeks of 2022 as input to the algorithms.

The utilisation rates are relatively high for pre-treatment and treatment resources (appendix A.3), indicating that both resources are now impacting patients' due date violations. Finding a good schedule for a patient is more difficult when both the pre-treatment and treatment resources have a low amount of residual capacity.

Next, the performance of both procedures is discussed. The results are presented in table 12. The batch procedure significantly outperforms the ASAP procedure for palliative patients and for the combination of patient categories. Interestingly, the 36.85% difference in weekly tardiness of non-palliative patients is not statistically significant. Although the difference is not statistically significant, it can still be relevant to consider the difference.

Objective	Batch Total	Batch Weekly avg	Batch N	ASAP Total	ASAP Weekly avg	ASAP N	p-value (a)
Palliative							
#patients late	2	0.06	33	100	2.86	33	< 0.01*
Tardiness	22	0.63	33	471	13.46	33	< 0.01*
Non-palliative							
#patients late	50	1.43	18	180	5.14	18	< 0.01*
Tardiness	760	21.71	20	480	13.71	20	0.29
Both categories							
#patients late	52	1.49	33	280	8	33	< 0.01*
Tardiness	782	22.34	33	951	27.17	33	0.018*

Table 12: Experiment 4: pre-treatment and treatment capacity reduction - Patient due date violations

(a) 5% significance level

5.3.1.5 Priority weights

The weights used throughout chapter 5 are 3 for the palliative patients and 1 for the non-palliative patients. The final experiment in this section changes these priority weights. Five weight combinations are assessed, as presented in table 13.

Combination	w_j **palliatives**	w_j **non-palliatives**
a	3	1
b	2	1
c	1	1
d	1	2
e	1	3

Table 13: Experiment 5: priority weights

The results are obtained using the capacity reductions from the second experiment (treatment capacity reduction). The goal of this experiment is to show the influence of the priority weights. Ten weeks are used as input to the model after a warm-up period of 18 weeks. The warm-up period is used to ensure that relevant differences can be obtained. Table 14 summarises the results.

Combination	Palliatives		Non-palliatives	
	#patients late	Tardiness	#patients late	Tardiness
a	2	22	18	230
b	3	44	19	184
c	5	70	13	168
d	14	220	4	51
e	14	225	4	54

Table 14: Experiment 5: priority weights - results

Because of the small sample size, no statements about the significance of the differences can be made. Nevertheless, the results still give a good indication of the impact of changing the priority weights. Combination d and e give more priority to non-palliative patients and this is reflected in the results. Furthermore, assigning more weight to the palliative patients results in a decrease in both the number of palliative patients that are late and the total tardiness of palliative patients.

5.3.2 The impact of including stochastic information

Including stochastic information in the scheduling procedure is expected to reduce the due date violations of patients, compared to the myopic ASAP procedure. To verify this statement, a test design was constructed. First, the batch procedure is used during a predetermined warm-up period. After the warm-up period, a certain amount of patients are scheduled using the online stochastic procedure. For the purpose of comparison, the batch procedure and the ASAP procedure are two alternative techniques that are also used in step 2. Thirdly, the remaining patients are scheduled using the batch procedure.

To determine the warm-up period, the results from the case with treatment capacity reduction, obtained in section 5.3.1.2, are consulted. Figure 24 includes the absolute amount of patients late in every week, for both the batch and ASAP procedure. In addition, the difference between both methods is graphically represented. It can be seen in the figure that the largest difference between both techniques is during week 20. Consequently, the warm-up period is determined to be 18 weeks. This means that the patients to be scheduled in step 2 are those who arrive starting from week 19. It is expected that these patients will influence the scheduling decisions for those patients arriving in week 20 the most.

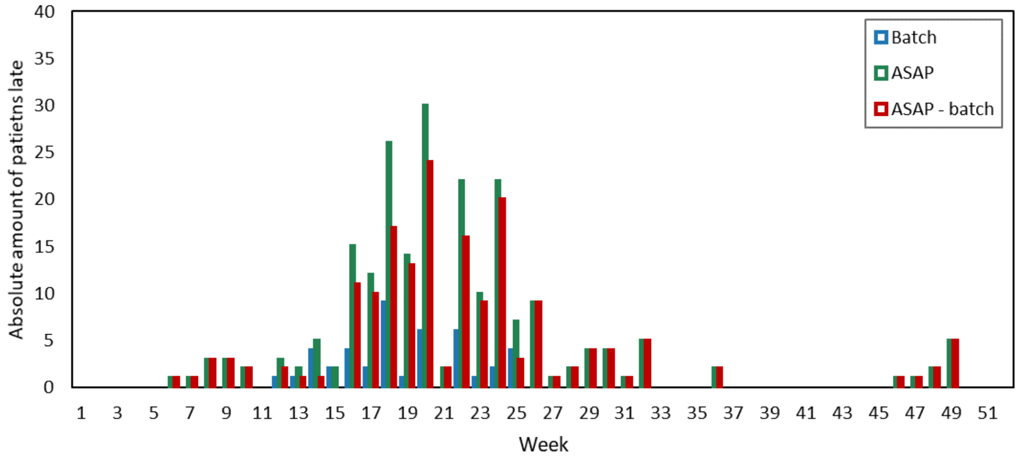


Figure 24: Weekly amount of patients late in batch and ASAP procedure - treatment capacity reduction

The number of patients to be scheduled in step 2 is decided to be equal to 10. In accordance with the findings of Legrain, Fortin, et al. (2015), each of these patients is accompanied with 10 distinct scenarios of future patient arrivals. The resulting run configuration to test the impact of including stochastic information is given in figure 25.

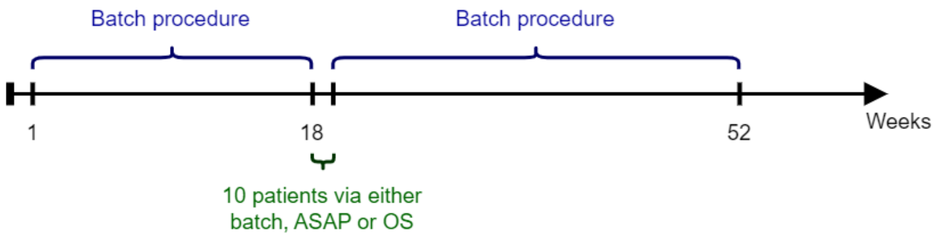


Figure 25: Run configuration to test the influence of including stochastic information

Including 10 patients with 10 scenarios in step 2 results in 100 optimisations. The time limit for each optimisation is set at 60 minutes. In theory, running time could be as high as 100 hours. To reduce this running time, the optimality gap tolerance is decided to be equal to 15%. All patients arriving in 2022 are scheduled. The results for week 19 to 52 are presented in table 15. The header (containing batch, OS and ASAP) refers to the method that is used in the second step.

	Batch		OS		ASAP	
Objective	Total	Weekly avg	Total	Weekly avg	Total	Weekly avg
Palliative						
#patients late	2	0.06	4	0.12	4	0.12
Tardiness	22	0.65	23	0.68	11	0.32
Non-palliative						
#patients late	18	0.53	17	0.50	20	0.59
Tardiness	192	5.65	228	6.71	250	7.35
Both categories						
#patients late	20	0.59	21	0.62	24	0.71
Tardiness	214	6.29	251	7.38	261	7.68

Table 15: Including stochastic information - Patient due date violations (week 19-52)

To analyse the results, the OS method is compared with the ASAP method. Furthermore, the batch procedure is included and acts as a lower bound. In step 2, the differentiating step, only 10 out of 1024 patients are scheduled. As a result, the observed differences between the OS and ASAP techniques are small. Therefore, care has to be taken when formulating statements about the observed differences. In table 15, it can be seen that the OS method tends to perform better (i.e. less total tardiness and fewer patients that are late) in relation to both patient categories combined and in relation to non-palliatives. Surprisingly, the ASAP method resulted in a tardiness of 11 for palliatives, which represents a 50% and 52% decrease when compared to the batch and OS methods, respectively. Closer examination of the results showed that this difference is observed in week 22. A detailed reasoning for this result could not be formulated. To conclude, the experiment conducted in this section suggests an improved scheduling practice when considering stochastic information. However, it is worth noting that the experiment was performed only once. Consequently, further research could include multiple repetitions of the experiment, in order to construct confidence intervals for the related objectives.

5.3.3 An additional term in the objective function

The results in section 5.3.1 have shown that, on average, the batch procedure has a higher tardiness per patient that is late. For example, in experiment 2, 43 patients had a total tardiness of 579 in the batch procedure (13.47 per late patient), whereas 219 patients in the ASAP procedure had a total tardiness of 662 (3.02 per late patient). Furthermore, the maximum lateness was equal to 19 in the batch procedure, compared to 12 in the ASAP procedure. Therefore, in this section, the possibility of reducing the average tardiness per late patient and the maximum tardiness by including an extra term in the objective is examined. The third term to include in the objective function of the ILP model presented in section 3.2 is based on Castro and Petrovic (2012). This term represents the (unweighted) maximum tardiness over all patients in the horizon. In addition, the following elements are added to the model:

- a variable: ML with $ML \geq 0$
- a constraint:

$$L_j \leq ML \qquad \forall P_j \in \mathcal{P} \tag{14}$$

The variable ML represents the maximum tardiness over all patients. The constraint is added to ensure that the variable ML is larger than or equal to the tardiness of every patient. No upper bound is imposed since the variable is minimised in the objective. In order to normalise, the additional term in the objective function is divided by the maximum possible lateness, given the horizon of the optimisation instance. Consequently, the adapted objective function is (d_* represents the earliest due date of all the patients in the horizon):

$$\begin{aligned}
&minimise \quad G_1 \cdot (\sum_{j \in \mathcal{P}} w_j \cdot U_j) / \sum_{j \in \mathcal{P}} w_j \\
&+ G_2 \cdot \sum_{j \in \mathcal{P}} (w_j \cdot L_j) / \sum_{j \in \mathcal{P}} ((max(\mathcal{K}) - d_j) \cdot w_j) \\
&+ G_3 \cdot (ML / (max(\mathcal{K}) - d_*))
\end{aligned}$$

To reflect on the influence of the adapted objective function, a small experiment with the same capacity reductions as in the second experiment is executed. The warm-up period is 18 weeks and the original batch procedure is used during this period. Subsequently, 1 month of patient arrivals is scheduled. Scenarios a, b, c, d and e all use the same input, but they differ in the scheduling of the one month of patient arrivals. Scenarios a, b, c and d use the ILP model with the new objective function and the weights are arbitrarily determined to be $G_1 = 0.45$, $G_2 = 0.55$ and $G_3 = 0$ for scenario a (equal to using the original OF), $G_1 = 0.3$, $G_2 = 0.3$ and $G_3 = 0.4$ for scenario b, $G_1 = 0.2$, $G_2 = 0.2$ and $G_3 = 0.6$ in c and $G_1 = 0$, $G_2 = 0$ and $G_3 = 1$ in scenario d. The final scenario uses the ASAP method to schedule 1 month of patients, instead of the ILP model. The results can be found in table 16. The 'ratio' row presents the ratio of total tardiness divided by the number of patients that are late. It is also worth noting that these results are related to both palliative and non-palliative patients combined.

Objective	Old OF (a)	New OF (b)	New OF (c)	New OF (d)	ASAP (e)
Max	18	11	10	5	12
# patients late	11	19	19	48	53
Tardiness	122	146	126	163	198
Ratio	11.09	7.68	6.63	3.40	3.74

Table 16: Adaptation of the objective function - Patient due date violations

(a) $G_1 = 0.45$, $G_2 = 0.55$ and $G_3 = 0$ - equal to using the original OF

(b) $G_1 = 0.30$, $G_2 = 0.30$ and $G_3 = 0.40$

(c) $G_1 = 0.20$, $G_2 = 0.20$ and $G_3 = 0.60$

(d) $G_1 = 0$, $G_2 = 0$ and $G_3 = 1$

The results indicate that the inclusion of the third objective can be a very meaningful addition. By only using the maximum tardiness as an objective (scenario d), the ILP solver performs better than the ASAP procedure on all aspects considered in the table. It is worth noting that optimality was not achieved, except for scenario d. This could explain why scenario c outperforms scenario b in terms of total tardiness. Furthermore, the results obtained in this experiment show that the decision makers can direct their focus to the objective that is perceived as most important, by adjusting the weights of the objective function.

6 Conclusion

The first aim of this master's dissertation was to formulate a model for radiotherapy scheduling that includes aspects from both the pre-treatment and treatment phase and that can be easily tailored to specific radiotherapy departments. The second goal was to include stochastic information in an online scheduling paradigm and to examine its impact. Relating to the first goal, an ILP model is proposed and the performance of the model is tested. According to Lievens et al. (2020), resource shortage is one of the many possible reasons why some patients fail to receive appropriate radiotherapy treatment. Therefore, several experiments with a reduced capacity were performed. In general, the results were positive and the performance related to the objectives included in the model is better compared to the ASAP procedure. Furthermore, the model enables specific hospitals to adjust the objective function weights to reflect the objective that they perceive as most important. In relation to the second goal, the proposed online stochastic method is compared with an offline procedure (batch) that acts as a lower bound and an online myopic procedure (ASAP) that serves as an upper bound. The results of including stochastic information gave a strong indication of a performance increase. Despite these promising results, some limitations and future research directions are identified.

The proposed mathematical model contains several concepts that are important in radiotherapy scheduling. Nevertheless, some additional concepts were not considered. First, it was assumed that the duration of the fractions in the treatment phase is a stable multiple of a base time period. While this is a common assumption to make, varying treatment times are also considered in the literature (e.g. Conforti et al., 2010). Secondly, it is possible that patients have a preference for the starting and/or end times of the pre-treatment tasks. This is not explicitly included in the model and future research could incorporate patient preferences. Thirdly, the model does not include the possibility to schedule operations using overtime capacity. When overtime is included, the boundary between regular capacity and overtime capacity is very narrow and a fine-grained analysis is required. Therefore, overtime was not included, but we recognise that this is an interesting avenue for future research. A fourth consideration is the inclusion of maximum tardiness in the objective. Small experiments were conducted and showed the potential of including the extra term in the objective function. Therefore, we advise future researchers to include this extra objective and to test its inclusion more extensively.

Furthermore, the conducted experiments in chapter 5 are performed only once, due to time constraints. The resulting inability to formulate statements regarding the statistical significance of the results is a substantial drawback. To statistically validate the findings of this dissertation, the experiments have to be executed several times, in order to construct confidence intervals. In addition, future research could include data instances based on other radiotherapy departments.

Finally, the proposed online stochastic algorithm is inefficient in terms of the required computational time. Although using the ILP solver to optimise each scenario was able to indicate the positive influence of using stochastic information, this method is of little relevance in practice because of the long running times. Consequently, future research is needed towards time-efficient (meta-)heuristic alternatives to the ILP method as part of the online stochastic algorithm.

References

Abdel-Basset, M., Abdel-Fatah, L., & Sangaiah, A. K. (2018). Chapter 10 - metaheuristic algorithms: A comprehensive review. In A. K. Sangaiah, M. Sheng, & Z. Zhang (Eds.), *Computational intelligence for multimedia big data on the cloud with engineering applications* (pp. 185–231). Academic Press.

Akartunalı, K., Mak-Hau, V., & Tran, T. (2015). A unified mixed-integer programming model for simultaneous fluence weight and aperture optimization in vmat, tomotherapy, and cyberknife. *Computers & Operations Research*, *56*, 134–150.

Awasthi, P., & Sandholm, T. (2009). Online stochastic optimization in the large: Application to kidney exchange. *IJCAI*, *9*, 405–411.

Bedford, J. L., & Warrington, A. P. (2009). Commissioning of volumetric modulated arc therapy (vmat). *International Journal of Radiation Oncology* Biology* Physics*, *73*(2), 537–545.

Bent, R., & Van Hentenryck, P. (2005). Online stochastic optimization without distributions. *ICAPS*, *5*, 171–180.

Bertsimas, D., Cacchiani, V., Craft, D., & Nohadani, O. (2013). A hybrid approach to beam angle optimization in intensity-modulated radiation therapy. *Computers & Operations Research*, *40*(9), 2187–2197.

Bonferroni, C. (1936). Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commericiali di Firenze*, *8*, 3–62.

Bortfeld, T. (2006). Imrt: A review and preview. *Physics in Medicine & Biology*, *51*(13), R363.

Brahme, A. (1984). Dosimetric precision requirements in radiation therapy. *Acta Radiologica: Oncology*, *23*(5), 379–391.

Castro, E., & Petrovic, S. (2011). A genetic algorithm for radiotherapy pre-treatment scheduling. *Applications of Evolutionary Computation: EvoApplications 2011: EvoCOMNET, EvoFIN, EvoHOT, EvoMUSART, EvoSTIM, and EvoTRANSLOG, Torino, Italy, April 27-29, 2011, Proceedings, Part II*, 454–463.

Castro, E., & Petrovic, S. (2012). Combined mathematical programming and heuristics for a radiotherapy pre-treatment scheduling problem. *Journal of Scheduling*, *15*(3), 333–346.

Chang, W., Lo, S. M., Chen, T.-L., Chen, J. C., & Wu, H.-N. (2020). Utilizing online stochastic optimization on scheduling of intensity-modulate radiotherapy therapy (imrt). *Journal of Biomedical Informatics*, *108*, 103499.

Chen, Z., King, W., Pearcey, R., Kerba, M., & Mackillop, W. J. (2008). The relationship between waiting time for radiotherapy and clinical outcomes: A systematic review of the literature. *Radiotherapy and Oncology*, *87*(1), 3–16.

Conforti, D., Guerriero, F., & Guido, R. (2008). Optimization models for radiotherapy patient scheduling. *4or*, *6*(3), 263–278.

Conforti, D., Guerriero, F., & Guido, R. (2010). Non-block scheduling with priority for radiotherapy treatments. *European Journal of Operational Research*, *201*(1), 289–296.

Conforti, D., Guerriero, F., Guido, R., & Veltri, M. (2011). An optimal decision-making approach for the management of radiotherapy patients. *OR Spectrum*, *33*(1), 123–148.

Diamant, A., Milner, J., & Quereshy, F. (2018). Dynamic patient scheduling for multi-appointment health care programs. *Production and Operations Management*, *27*(1), 58–79.

Dursun, P., Taşkın, Z. C., & Altınel, İ. K. (2016). Mathematical models for optimal volumetric modulated arc therapy (vmat) treatment planning. *Procedia Computer Science*, *100*, 644–651.

Dursun, P., Taşkın, Z. C., & Altınel, İ. K. (2019). Using branch-and-price to determine optimal treatment plans for volumetric modulated arc therapy (vmat). *Computers & Operations Research*, *110*, 1–17.

El-Omari, N. (2021). Sea lion optimization algorithm for solving the maximum flow problem.

General information about radiotherapy. (n.d.). Retrieved June 2, 2023, from https://siope.eu/General-information-about-radiotherapy

Gocgun, Y. (2018). Simulation-based approximate policy iteration for dynamic patient scheduling for radiation therapy. *Health care management science*, *21*(3), 317–325.

Harper, R. (2016). *Practical foundations for programming languages*. Cambridge University Press.

Hoeffding, W. (1994). A non-parametric test of independence. *The Collected Works of Wassily Hoeffding*, 214–226.

Jensen, A. R., Nellemann, H. M., & Overgaard, J. (2007). Tumor progression in waiting time for radiotherapy in head and neck cancer. *Radiotherapy and oncology*, *84*(1), 5–10.

Kapamara, T., & Petrovic, D. (2009). A heuristics and steepest hill climbing method to scheduling radiotherapy patients. *Proceedings of the 35th International Conference on Operational Research Applied to Health Services, Catholic University of Leuven, Belgium.*

Kapamara, T., Sheibani, K., Haas, O. C., Reeves, C. R., & Petrovic, D. (2006). A review of scheduling problems in radiotherapy. *Proceedings of the Eighteenth International Conference on Systems Engineering (ICSE2006), Coventry University, UK*, 201–207.

Kotha, V., Rewari, A., Lakhiani, C., Zolper, E., Orio, P., Stimac, G., Chadab, T., Fan, K., Evans, K., & Song, D. (2021). Radiation oncology applications in plastic and reconstructive surgery: A nonsystematic review of concepts and principles. *Plastic Reconstructive Surgery*, *147*, 314e–324e.

Legrain, A., Fortin, M.-A., Lahrichi, N., & Rousseau, L.-M. (2015). Online stochastic optimization of radiotherapy patient scheduling. *Health care management science*, *18*(2), 110–123.

Legrain, A., Widmer, M., Fortin, M.-A., Lahrichi, N., & Rousseau, L.-M. (2015). Stochastic optimization of the scheduling of a radiotherapy center. *Journal of Physics: Conference Series*, *616*(1), 012008.

Liang, H., Xiang, Y.-Q., Lv, X., Xie, C.-Q., Cao, S.-M., Wang, L., Qian, C.-N., Yang, J., Ye, Y.-F., Gan, F., et al. (2017). Survival impact of waiting time for radical radiotherapy in nasopharyngeal carcinoma: A large institution-based cohort study from an endemic area. *European journal of cancer*, *73*, 48–60.

Lievens, Y. (2017). Access to innovative radiotherapy: How to make it happen from an economic perspective? *Acta Oncologica*, *56*(11), 1353–1358.

Lievens, Y., Borras, J. M., & Grau, C. (2020). Provision and use of radiotherapy in europe. *Molecular Oncology*, *14*(7), 1461–1469.

Mackillop, W. J. (2007). Killing time: The consequences of delays in radiotherapy. *Radiotherapy and Oncology*, *84*(1), 1–4.

Maenhout, B. (2020). Introduction to or/ms modelling.

Maenhout, B. (2022). Random number generation.

Massey Jr, F. J. (1951). The kolmogorov-smirnov test for goodness of fit. *Journal of the American statistical Association*, *46*(253), 68–78.

Mundry, R., & Fischer, J. (1998). Use of statistical programs for nonparametric tests of small samples often leads to incorrect pvalues: Examples from animal behaviour. *Animal behaviour*, *56*(1), 256–259.

Peng, F., Jia, X., Gu, X., Epelman, M. A., Romeijn, H. E., & Jiang, S. B. (2012). A new column-generation-based algorithm for vmat treatment plan optimization. *Physics in Medicine & Biology*, *57*(14), 4569.

Pérez, E., Ntaimo, L., Malavé, C. O., Bailey, C., & McCormack, P. (2013). Stochastic online appointment scheduling of multi-step sequential procedures in nuclear medicine. *Health care management science*, *16*(4), 281–299.

Petrovic, D., Morshed, M., & Petrovic, S. (2009). Genetic algorithm based scheduling of radiotherapy treatments for cancer patients. *Conference on Artificial Intelligence in Medicine in Europe*, 101–105.

Petrovic, D., Morshed, M., & Petrovic, S. (2011). Multi-objective genetic algorithms for scheduling of radiotherapy treatments for categorised cancer patients. *Expert Systems with Applications*, *38*(6), 6994–7002.

Petrovic, S., Leung, W., Song, X., & Sundar, S. (2006). Algorithms for radiotherapy treatment booking. *25th Workshop of the UK planning and scheduling special interest group*, 105–112.

Pham, T.-S., Legrain, A., De Causmaecker, P., & Rousseau, L.-M. (2021). A prediction-based approach for online dynamic radiotherapy scheduling: A case study in radiotherapy treatment. *arXiv preprint arXiv:2112.08549*.

Pudsey, L., Cutajar, D., Wallace, A., Saba, A., Schmidt, L., Bece, A., Clark, C., Yamada, Y., Biasi, G., Rosenfeld, A., & Poder, J. (2021). The use of collimator angle optimization and jaw tracking for vmat-based single-isocenter multiple-target stereotactic radiosurgery for up to six targets in the varian eclipse treatment planning system. *Journal of Applied Clinical Medical Physics*, *22*.

Riff, M.-C., Cares, J. P., & Neveu, B. (2016). Rason: A new approach to the scheduling radiotherapy problem that considers the current waiting times. *Expert Systems with Applications*, *64*, 287–295.

Salari, E., & Romeijn, H. E. (2012). Quantifying the trade-off between imrt treatment plan quality and delivery efficiency using direct aperture optimization. *INFORMS Journal on Computing*, *24*(4), 518–533.

Saure, A., Patrick, J., Tyldesley, S., & Puterman, M. L. (2012). Dynamic multi-appointment patient scheduling for radiation therapy. *European Journal of Operational Research*, *223*(2), 573–584.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3-4), 591–611.

Smirnov, N. V. (1939). On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Bull. Math. Univ. Moscou*, *2*(2), 3–14.

Van Hentenryck, P., Bent, R., & Upfal, E. (2010). Online stochastic optimization under time constraints. *Annals of Operations Research*, *177*(1), 151–183.

Verweij, B., Ahmed, S., Kleywegt, A. J., Nemhauser, G., & Shapiro, A. (2003). The sample average approximation method applied to stochastic routing problems: A computational study. *Computational optimization and applications*, *24*, 289–333.

Vogl, P., Braune, R., & Doerner, K. F. (2019). Scheduling recurring radiotherapy appointments in an ion beam facility. *Journal of Scheduling*, *22*(2), 137–154.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, *1*(6), 80–83.

Zeleny, M. (1974). A concept of compromise solutions and the method of the displaced ideal. *Computers Operations Research*, *1*(3), 479–496.

A Appendix

A.1 Functions in the ASAP procedure

Algorithm 4 Find feasible treatment start

```
function FindFeasTreatmentStart(j, k_0)
    found ← False
    k_1 ← Null
    v_1 ← Null
    while found ≠ True do
        for k ∈ horizon do
            for v ∈ linacs do
                if v can accommodate entire treatment of j then
                    found ← True
                    k_1 ← k
                    v_1 ← v
                    break all loops
    return k_1, v_1                ▷ Returns the day of first treatment and the linac
```

Algorithm 5 Find feasible simulation slot

```
function FindFeasSimulationSlot(j, k_1, r)
    found ← False
    k_2 ← Null
    e_2 ← Null
    while found ≠ True do
        for k ∈ [k_1, k_1 − r] do
            if One or more slots available on day k then
                found ← True
                k_2 ← k
                e_2 ← last available slot on day k
                break all loops
    return k_2, e_2        ▷ Returns the day and slot of the simulation appointment
```

Algorithm 6 Find feasible CT-scan slot

```
function FindFeasCTSlot(j, k_2, e_2)
    found ← False
    k_3 ← Null
    e_3 ← Null
    while found ≠ True do
        for k ∈ [k_2 − 5, a_j] do
            if One or more slots available on day k then
                found ← True
                k_3 ← k
                e_3 ← last available slot on day k
                break all loops
    return k_3, e_3        ▷ Returns the day and slot of the ct-scan appointment
```

A.2 Kolmogorov-Smirnov test for monthly patient arrivals

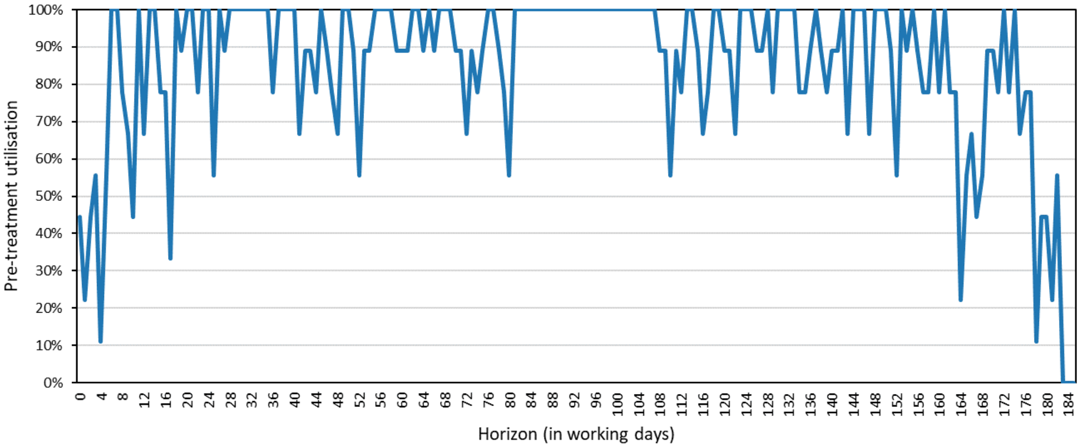
Month 1	**Month 2**	**Significance level (α)**	n	**Corrected α ($=\alpha/n$)**	**p-value**
January	February	5%	66	0.00076	0.62534
January	March	5%	66	0.00076	0.94867
January	April	5%	66	0.00076	0.53517
January	May	5%	66	0.00076	0.45041
January	June	5%	66	0.00076	0.32836
January	July	5%	66	0.00076	0.06809
January	August	5%	66	0.00076	0.97747
January	September	5%	66	0.00076	0.99943
January	October	5%	66	0.00076	0.96156
January	November	5%	66	0.00076	0.62534
January	December	5%	66	0.00076	0.80806
February	March	5%	66	0.00076	0.66970
February	April	5%	66	0.00076	0.83197
February	May	5%	66	0.00076	0.74666
February	June	5%	66	0.00076	0.73521
February	July	5%	66	0.00076	0.00975
February	August	5%	66	0.00076	0.76126
February	September	5%	66	0.00076	0.48536
February	October	5%	66	0.00076	0.20189
February	November	5%	66	0.00076	0.83197
February	December	5%	66	0.00076	0.58943
March	April	5%	66	0.00076	0.81990
March	May	5%	66	0.00076	0.27949
March	June	5%	66	0.00076	0.21837
March	July	5%	66	0.00076	0.06909
March	August	5%	66	0.00076	0.60980
March	September	5%	66	0.00076	0.99006
March	October	5%	66	0.00076	0.61687
March	November	5%	66	0.00076	0.66970
March	December	5%	66	0.00076	0.98681
April	May	5%	66	0.00076	0.56157

April	June	5%	66	0.00076	0.40871
April	July	5%	66	0.00076	0.05844
April	August	5%	66	0.00076	0.44164
April	September	5%	66	0.00076	0.54577
April	October	5%	66	0.00076	0.87813
April	November	5%	66	0.00076	0.83197
April	December	5%	66	0.00076	0.81035
May	June	5%	66	0.00076	0.82910
May	July	5%	66	0.00076	0.04446
May	August	5%	66	0.00076	0.99433
May	September	5%	66	0.00076	0.91784
May	October	5%	66	0.00076	0.11557
May	November	5%	66	0.00076	0.94992
May	December	5%	66	0.00076	0.45041
June	July	5%	66	0.00076	0.04386
June	August	5%	66	0.00076	0.83889
June	September	5%	66	0.00076	0.87168
June	October	5%	66	0.00076	0.06716
June	November	5%	66	0.00076	0.98190
June	December	5%	66	0.00076	0.32836
July	August	5%	66	0.00076	0.06371
July	September	5%	66	0.00076	0.07260
July	October	5%	66	0.00076	0.31495
July	November	5%	66	0.00076	0.03726
July	December	5%	66	0.00076	0.15317
August	September	5%	66	0.00076	0.99989
August	October	5%	66	0.00076	0.55122
August	November	5%	66	0.00076	0.91466
August	December	5%	66	0.00076	0.93487
September	October	5%	66	0.00076	0.67875
September	November	5%	66	0.00076	0.69719
September	December	5%	66	0.00076	0.98931
October	November	5%	66	0.00076	0.32020
October	December	5%	66	0.00076	0.82261
November	December	5%	66	0.00076	0.88001

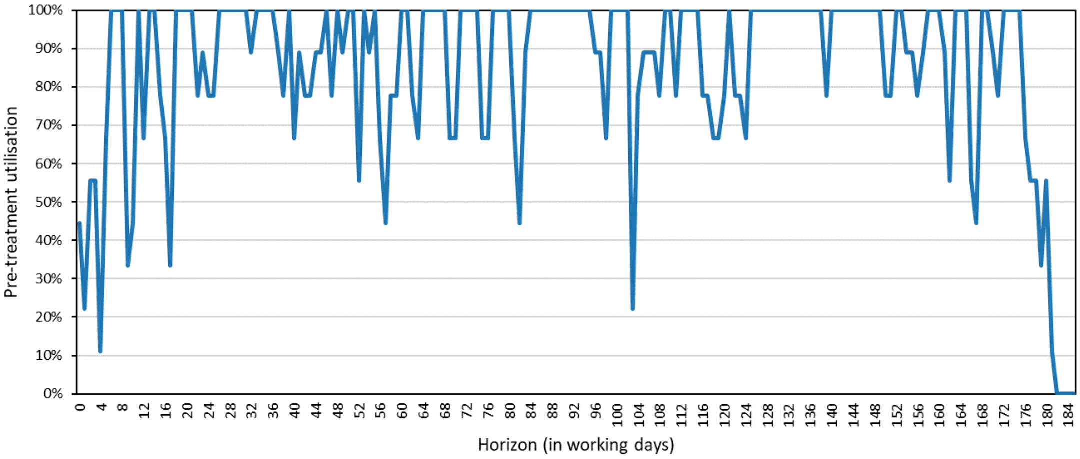
Table 17: K-S test for monthly patient arrivals

n refers to the number of pairwise comparisons

A.3 Experiment 4: resource capacities

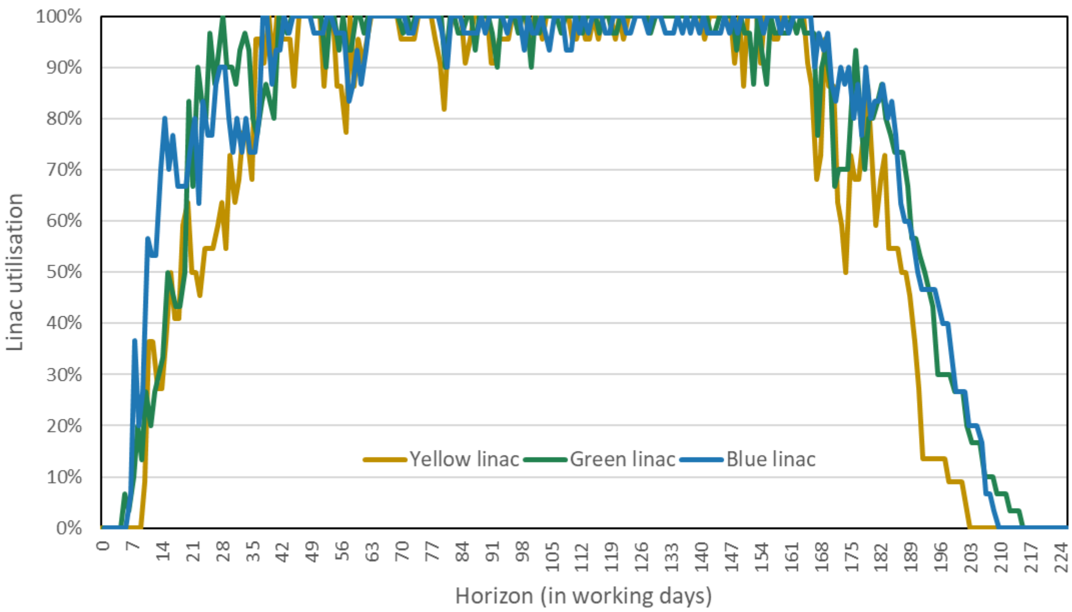


(a) Batch

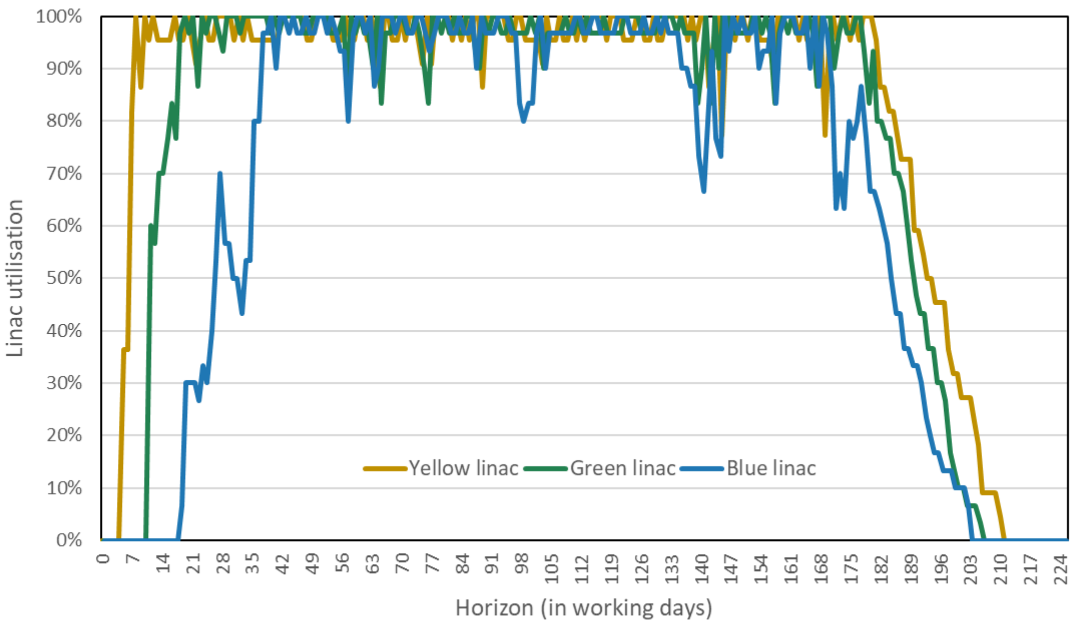


(b) ASAP

Figure 26: Experiment 4: pre-treatment utilisation



(a) Batch



(b) ASAP

Figure 27: Experiment 4: treatment utilisation