Large scale elective surgery scheduling under uncertainty

Fabrício de Reuter Sperandio

A thesis submitted to Faculdade de Engenharia da Universidade do Porto
for the doctoral degree in Industrial Engineering and Management

Supervisors

Professor Bernardo Sobrinho Simões de Almada Lobo
Professor José Luís Cabral Moura Borges

U. PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

2015

Abstract

Since 2009, as a result of the global financial and economic crisis, the health expenditure in Organisation for Economic Co-operation and Development (OECD) countries stopped a long term rising trend and has been stagnating or even falling in many countries. The crisis forced many governments to promote challenging cuts in public expenditure. For instance, Portugal agreed with the European Union, within an economic and financial adjustment programme, to cut 15% on health costs between 2011 and 2013. In this context, many countries promoted reforms in the health sector to increase productivity and efficiency. In addition, in face of the complexity of healthcare management problems, specially due to the strong uncertainty inherent to this type of problems, healthcare decision makers need decision support tools to reduce costs without impacting quality of care. In this context, the field of operations research has an extensive set of techniques that have been applied to healthcare management problems. In particular, due to the high volume of resources assigned to the operating theater (OT), the application of operations resources techniques to OT management problems has been an active research area. Nevertheless, it still presents well known research gaps, among them, the lack of efficient and realistic elective surgery scheduling methods. This thesis proposes a decision support system (DSS) for the elective surgery scheduling problem and four progressive more complex scheduling methods. The DSS tackles the issues of decision support, uncertainty reduction and surgery schedule optimization, through the integration of data mining and optimization techniques. This system was designed based on the needs of surgeons and hospital managers from a large hospital in the north of Portugal. Regarding schedule optimization, the first scheduling method, which is integrated into the DSS and is proposed to automate the process of generating new schedules, consists in a mixed integer programming (MIP) model which uses a discrete representation of time. The second method consists in a new MIP formulation using a continuous representation of time that is able to find better solutions in a reduced amount of time. The third method is composed of a genetic algorithm and a set of local search procedures designed to tackle large scale problems. Finally, the last method consists in a new multi-objective optimization approach based on the integration between simulation and optimization to tackle a stochastic version of the problem with multiple sources of uncertainty. This approach is a proactive way to reduce the impact of uncertainty in the execution of the schedules. The proposed DSS and new scheduling methods tackle an important societal issue and are direct contributions to the scientific community, as they allow for increased productivity and efficiency in the elective surgery scheduling processes.

Resumo

Desde 2009, devido à crise crise económica e financeira mundial, os custos com saúde nos países da Organização para a Cooperação e Desenvolvimento Económico (OCDE) interromperam uma tendência longa de crescimento e estagnaram ou até mesmo caíram em muitos países. A crise forçou muitos governos a promover duras medidas de controlo orçamental. Por exemplo, Portugal acordou com a União Europeia, como parte do programa de ajustamento económico e financeiro, reduzir 15% dos gastos em saúde entre 2011 e 2013. Neste contexto, muitos países foram incentivados a promover reformas no setor da saúde para aumentar a produtividade e a eficiência. Além disso, devido à complexidade dos problemas de gestão hospitalar, em especial por causa da forte incerteza a que estão sujeitos, os gestores necessitam de ferramentas de apoio à decisão que lhes permitam reduzir os custos sem comprometer a qualidade dos cuidados prestados à população. Neste sentido, a área de investigação operacional (IO) contém um amplo conjunto de técnicas que têm sido aplicadas a problemas de gestão no setor da saúde. Em especial, devido ao grande volume de recursos atribuídos ao bloco operatório (BO), a aplicação de técnicas de IO à gestão de BOs tem sido uma área de investigação bastante activa nos últimos anos. Apesar disso, a literatura de IO aplicada à gestão de BOs ainda apresenta algumas lacunas, entre elas, a falta de modelos eficientes e realistas para o escalonamento de cirurgias eletivas. Esta tese propõe um sistema de apoio à decisão (SAD) para o escalonamento de cirurgias electivas nos hospitais portugueses e quatro métodos de escalonamento alternativos, progressivamente mais complexos. O SAD aborda as questões de apoio à decisão, redução da incerteza e optimização do escalonamento cirúrgico através da combinação de técnicas de *data mining* e optimização. Quanto aos métodos de escalonamento, o primeiro método, que está integrado no SAD e é proposto para automatizar o processo de criação de escalas, consiste num modelo de programação inteira mista (PIM ou do inglês MIP) que usa uma representação discreta do tempo. O segundo método consiste numa nova formulação do problema de PIM que usa uma representação contínua do tempo, permitindo assim obter melhores soluções em menor tempo. O terceiro método é composto por um algoritmo genético e um conjunto de procedimentos de melhoria local projectados para abordar problemas de grande escala. Por fim, o quarto método consiste numa nova abordagem multiobjectivo baseada na combinação de simulação e optimização para uma versão estocástica do problema. Esta abordagem é uma abordagem proativa para reduzir o impacto da incerteza no resultado do planeamento. Os métodos propostos nesta tese abordam um tema importante para a sociedade num momento de forte restrição orçamental. Além disso, os novos métodos de escalonamento são contribuições originais para a comunidade científica pois preenchem lacunas específicas da literatura.

Acknowledgments

First of all, I would like to thank my supervisors, Professor Bernardo Almada Lobo and Professor José Luís Borges, for all the support and understanding since my arrival in Portugal, throughout the research projects and scientific papers we worked together, until the delivery of this thesis. Without their close collaboration this work would not be possible. In addition, I would like to thank Professor António Carvalho Brito and Professor José Fernando Gonçalves for their support and for sharing their experience with me. Finally, I would like to thank all Professors of the Industrial Engineering and Management Department (DEGI) of the Faculty of Engineering of the University of Porto (FEUP), which were always very kind and helped me to get a broader understanding of the industrial engineering and management area.

I also would like to thank all the colleagues I had the pleasure to work with in different places at FEUP. In the begining, it was the room 304 in the L building, where I met Carla, Maria, Teresa, Nelson, Danilo and Roberta who became close friends. Next, it came IDMEC and IBM CAS, where I first met Carlos, José Rui and Arnon, and later, António, Marta and Miguel. Next, it came Adjust Consulting and one of the best teams I had the pleasure to work with, composed by: Miguel, Ricardo, Ilda, José Rui, Steeve and Ricardo Brandão. A special thanks to Steeve with whom I could discuss about programming and algorithms. Finally, I would like to thank my colleagues at CEGI — Centre for Industrial Engineering and Management at INESC TEC. In special, Luís, Pedro, Teresa and Gonçalo with whom I could discuss and learn about optimization techniques.

A deep thanks to my family in Brazil, in special, to my mother, who always made everything possible to support me, and to my father, who passed away during the course of this PhD. I am sure he would be glad to see the results of this work. I am also specially grateful to Andreia, who was my safe port, helping to soften the many tough moments along this course. Certainly, without her and the environment we created in our home, it would be harder to finish this course.

Finally, I would like to extend the greetings to all Portuguese I have met during this 5 years, who taught me about culture, history and gastronomy. These are lessons I will take with me. When I left Brazil, I was concerned about my professional life and the career I was leaving in Brazil, but Professor Bernardo warned me about the personal growth I would experience coming to Portugal. Now, in addition to the great professional growth it brought me, I realize he was right, because of the immense bunch of experience I could get with all the people I have met.

Acronyms

BRKGA Biased Random-Key Genetic Algorithm

CP Constraint Programming

COMPASS Convergent Optimization via Most-Promising-Area Stochastic Search

COvS Continuous Optimization via Simulation

DSS Decision Support System

DES Discrete-Event Simulation

DOvS Discrete Optimization via Simulation

EA Evolutionary Algorithm

GA Genetic Algorithm

HIS Hospital Information System

ICU Intensive Care Unit

IP Integer Programming

KPI Key Performance Indicator

LP Linear Programming

MSS Master Surgery Schedule

MIP Mixed Integer Programming

MMBJS Multi-Mode Blocking Job Shop

MOEA Multi-Objective Evolutionary Algorithm

NHS National Healthcare System

NSGA Non-dominated Sorting Genetic Algorithm

OECD Organisation for Economic Co-operation and Development

OR Operating Room

OR/MS Operations Research and Management Science

OT Operating Theater

ORAHS Operational Research Applied to Health Services

OCBA Optimal Budget Computing Allocation

PACU Post-anesthesia Care Unit

RSM Response Surface Method

RKGA Random-Key Genetic Algorithm

SAR Simulation Allocation Rule

SCAP Surgical Case Assignment Problem

SO Simulation Optimization

SIGIC Sistema Integrado de Gestão de Inscritos para Cirurgia

SP Stochastic Programming

WSC Winter Simulation Conference

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Since 2006, after the introduction of the System for Management of Patients Waiting for Surgery ("Sistema Integrado de Gestão de Inscritos para Cirurgia" - SIGIC) program, Portugal has been successfully reducing the number of patients waiting for a surgery, as well as the average waiting time. The last report (ACSS, 2014) shows that by the end of 2013 the number of patients waiting for a surgery was 20.4% lower than in 2006. Also, the waiting times by the end of 2013 were 58.9% lower than in 2006. These results can be partially explained by gains in efficiency, even though they are substantially driven by an expanded capacity. The government hired additional working hours from surgeons working on public hospitals as well as funded surgeries on private hospitals when the waiting time was close to exceed the limits established by the program. As a result, the number of performed surgeries by the end of 2013 achieved a remarkable growth of 57.6% in comparison to 2006.

The Portuguese National Health Service (NHS) has been challenged to deal with an increasing demand for elective surgical services. The system experienced a 42.5% increase in the annual number of new requests for elective surgeries in the period between 2006 and 2013. Such increasing demand is a result of an ageing population, which naturally demands more frequent and intensive healthcare, as well as of better access of the population to surgical treatments, due to an efficient transportation network. On the other hand, the Eurozone financial crisis, which started in late 2008, particularly affected Portugal, imposing tremendous challenges to reduce public expenditure. For instance, the government agreed with the European Union to cut 15% (relative to 2010) on healthcare operational costs in the period between 2011 and 2013 (Ribeiro et al., 2011), which put special programs like SIGIC at serious risk. In order to achieve this target reduction, the amount of extra hours was cut in some hospitals and there was a fear that it could impact quality of care (Escoval et al., 2012). In the end of 2013, the number of patients which exceeded the maximum waiting time before treatment still reported a high value, representing 12.8% of the total number of patients waiting for a surgery (ACSS, 2014). In a scenario of increasing demand for elective surgeries and constrained healthcare budgets there is a clear need for promoting productivity and efficiency in the utilization of resources, allowing hospitals to reduce costs without impacting quality of care.

Operations research has been long helping healthcare institutions to improve efficiency. In the last decade, the application of operations research methods to tackle healthcare problems has been an active research area (Hulshof et al., 2012). In particular, the operating room management area has been attracting large attention since the operating theater (OT) is considered hospitals' largest cost and revenue center. Its effective management impacts several hospital Key Performance Indicators (KPIs), such as: number of patients waiting for a surgery, mean waiting time, average length of stay and case mix index. In the last few years, extensive literature reviews have been dedicated to operating room management problems (Cardoen et al., 2010a; Guerriero and Guido, 2011; May et al., 2011). Such reviews usually classify the problems into three decision levels: strategic, tactical and operational. Among them, the weekly scheduling of elective surgeries at the operational decision level has been the subject of the most part of the studies. This problem consists in assigning a surgery date, an operating room and a starting time to a set of elective patients in the waiting list, thus integrating two sub-problems: advance and allocation scheduling. The first sub-problem consists in selecting the patients from the waiting list and assigning a surgery date for them and the second consists in sequencing the surgeries within each day.

However, the current literature on operations research methods for operating room management presents some well-known research gaps. Two of the main issues concern the high computational cost for solving detailed scheduling problems and the low applicability of results. In order to tackle large size instances researchers are forced to apply simplified models which either lead to low quality or to unrealistic solutions. Both issues can contribute to a reported low implementation rate (Brailsford and Vissers, 2011). One of the main drivers of the high computational cost is the uncertainty inherent to healthcare management problems. In the OT, this uncertainty may come from multiple sources, such as: surgery times, emergency patients, staff no-shows and equipment failures.

Computer simulation is considered one of the most suitable tools to tackle uncertainty in healthcare problems (Guerriero and Guido, 2011). Its modelling flexibility enables analysts to model problems that, due to its overwhelming complexity, could not be modelled using analytical tools. One of its main uses is to perform scenario analysis, a process of comparing the estimated performance under uncertainty of a small number of alternatives, in order to select the best option. However, in order to tackle a large number of alternatives, in which exhaustive search is not feasible, an automatic way of generating different alternatives is required. The combination between combinatorial optimization and computer simulation to automatize the process of finding new alternative solutions to perform scenario analysis gives birth to a new computer simulation field called simulation optimization (Henderson and Nelson, 2006).

1.2 Research Objectives

Given the increasing demand for elective surgeries, constrained healthcare budgets and the complexity of the operational elective surgery scheduling problem, this thesis has the ultimate goal of proposing an advanced decision support system (DSS) for the elective surgery scheduling problem in Portuguese hospitals. In order to accomplish this goal the work is guided towards four main objectives. The workflow of this thesis is illustrated in Figure 1.1.

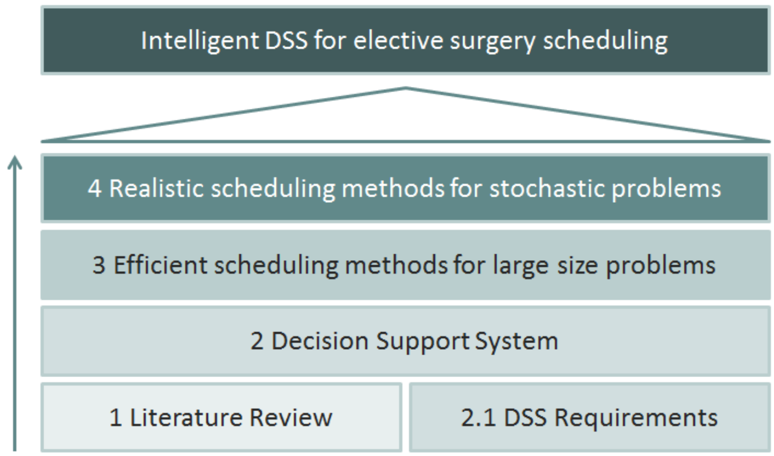


Figure 1.1: Thesis workflow to reach the final objective

The first objective is to conduct a literature review on operating room scheduling problems. The second, based on the literature review and on a case study in a Portuguese hospital, is to propose an intelligent DSS combining data mining and optimization features. The third and forth objectives aim to enhance the scheduling component of the DSS proposing more advanced approaches in deterministic and stochastic settings, respectively. The third objective focuses on providing more efficient scheduling methods to tackle large size instances and the last on scheduling methods under uncertainty. This workflow intends to come up with a DSS well suited for the needs of the Portuguese operating room managers and at the same time including advanced scheduling methods. In the process, this thesis aims to expand the current body of knowledge about elective surgery scheduling methods. The four main objectives and its respective specific objectives are detailed below:

(1) Literature Review: To analyse the literature on operations research for operating room management.

- (1.1) To identify research gaps. This objective has the goal of identifying areas in which this thesis can contribute to the scientific community. These areas encompass new problem settings or situations in which the existing solution approaches leave room for improvement in terms of quality of solutions or computational running time.
- (1.2) To identify possible solution approaches. This objective aims to detect alternative solution approaches for fulfilling the research gaps identified in the previous topic. In particular, it aims to study simulation optimization approaches for tackling operating room scheduling problems under uncertainty, which is a known research gap.

(2) Decision Support System: To propose an intelligent decision support system (DSS) to aid the elective surgery scheduling process in Portuguese public hospitals. The accomplishment of this objective will bring a practical contribution for improving the low implementation rate of operations research projects in the healthcare sector as well as for improving the performance of surgical services. It contributes directly to the society in the sense that it reaches the end user.

(2.1) To analyse the DSS requirements: In order to achieve this objective the waiting list management and scheduling process in Portuguese hospitals must be analysed and the decision support system requirements must be defined. In order to improve usability and applicability of results, the analysis must pay special attention to issues concerning the user interface and integration with hospital information systems.

(2.2) To design and develop an intelligent DSS: This objective concern the actual development of the DSS. In order to achieve this end a working version of the system must be presented.

(2.3) To propose a method for the estimation of surgery durations: The estimation of surgery durations is a key input for the scheduling model and is subject to high variability. In order to achieve this objective a prediction method able to effectively reduce the deviation between the predicted and the actual duration of surgeries must be proposed.

(2.4) To develop an exact scheduling model for automating the construction of operational surgery schedules in the DSS: Together with the prediction of surgery durations this objective aims to provide the intelligence of the system. In order to achieve this goal a scheduling model aligned with the requirements identified in item (2.1), in particular the rules of the Portuguese NHS, must be derived.

(3) Efficient Scheduling Methods: To come up with more efficient scheduling methods for generating operational surgery schedules in deterministic settings. This objective aims to enhance the performance of the scheduling model proposed in (2.4). In order to achieve this goal new scheduling methods must be proposed and compared against the previous model. As the demand for elective surgeries has been increasing the new methods should be designed to tackle large size instances.

(3.1) To propose an exact scheduling model: Exact models are important as they are able to provide a proof of optimality.

(3.2) To propose a heuristic scheduling method: Heuristic solution approaches are important to tackle large size instances.

(3.3) To compare the alternative scheduling methods using instances based on real data. In order to achieve this objective the results of the two alternative scheduling methods must outperform the results of the model proposed in (2.4), and extended in Chapter 3, either in quality of solutions or computational time.

(4) Realistic Scheduling Methods: To propose scheduling methods for generating more realistic (considering multiple sources of uncertainty) operational surgery schedules in stochastic settings. Healthcare management problems are subject to strong uncertainty resulting in large deviations between the initially planned and actually performed activities. In order to achieve this objective the proposed solution approach must generate solutions that mitigate such deviations.

(4.1) To propose a multi-objective optimization method: Since the elective surgery scheduling problem under uncertainty has multiple and conflicting objectives, a multi-objective optimization approach is required to provide the decision maker a set of alternative solutions representing the trade-offs between the conflicting objectives.

(4.2) To propose a simulation model: The simulation model is used for evaluating the performance of the alternative surgery schedules under uncertainty.

(4.3) To propose an integrated simulation optimization approach: The integration between simulation and optimization must be carefully designed because of the high computational cost of combining combinatorial and stochastic problems. In order to achieve this goal the main issues concerning this integration must be identified and a solution must be proposed.

(4.4) To evaluate the performance of the proposed approach using instances based on real data. In order to achieve this objective the results of the proposed multi-objective simulation optimization approach must be compared with the results obtained using fixed planned slacks.

1.3 Thesis Summary

This thesis is organized in four core chapters, each one dedicated to one of the main objectives. Chapters 3, 4 and 5 are written as scientific papers which were submitted to international journals. Each paper has the contribution of a team of researchers. In particular, the work presented in Chapter 3 was developed within a research project funded by the Portuguese Foundation for Science and Technology (FCT).

Chapter 2 contains a literature review to allow the reader to get a general understanding of operating room management problems and possible solution approaches. Regarding the problems, in order to study the influence of different decisions, all sub-problems across the three decision levels (strategic, tactical and operational) are included in the review. Regarding the solution approaches, the review focuses on meta-heuristics, computer simulation and simulation optimization as important methods for tackling large size instances and problems under uncertainty. Each of the subsequent chapters contains its own literature review section focused on the specific topics addressed in each chapter.

Chapter 3 concerns the development of a decision support system for the operating theater and the integration of a scheduling method, involving data mining and optimization techniques. This study covers three main areas: decision support, uncertainty reduction and surgery schedule optimization. The first area includes the information system requirements with a special focus on usability and integration issues. The second area describes the data mining methods used for enhancing the prediction of surgery durations. The third area describes the requirements for a basic scheduling model for automating the generation of elective surgery schedules in the DSS. This scheduling method is afterwards enhanced in Chapters 4 and 5 of this thesis.

Chapter 4 proposes two new scheduling methods for the DSS described in Chapter 3. The first is an exact mixed integer programming (MIP) model using a continuous representation of time and the second is a meta-heuristic based on the biased random key genetic algorithm (BRKGA) framework. In order to evaluate the performance of the proposed approaches, the results of the MIP model using a continuous time representation are compared against the results of the scheduling model used in Chapter 3, which uses a discrete representation of time. Further, the results of the exact model using a continuous representation of time are compared against the results obtained using the heuristic approach.

Chapter 5 proposes an integrated simulation optimization approach for the elective surgery scheduling problem under uncertainty. This chapter describes in detail both simulation and optimization modules of the integrated solution approach as well as it explores the main issues concerning how to allocate the simulation budget (number of simulation replications). The simulation module features a discrete-event simulation model including four sources of uncertainty. The optimization module features a multi-objective evolutionary algorithm based on the non-dominated sorting genetic algorithm (NSGA-II) framework.

Finally, Chapter 6 presents the conclusions, main contributions of the thesis and suggestions for future research.

Chapter 2

Literature Review

The goal of this literature review is twofold. On one hand, it aims to evaluate the current body of knowledge on operating room planning and scheduling in order to identify possible research gaps. On the other hand, it aims to review state of the art simulation optimization approaches that could be applied to operating room management problems, leading to potential research opportunities. In order to do that, this section is organized into four subsections, the first three cover operating room planning and scheduling, computer simulation in healthcare and simulation optimization, while the later presents the research gaps. For the sake of supporting the work developed in Chapter 5, the most popular random search procedures are described in the context of simulation optimization. This literature review is based on selected international journal papers and key conference proceedings, such as annual meetings of the Association of European Operational Research Societies (EURO) Working Group on Operational Research Applied to Health Services (ORAHS) and the Winter Simulation Conference (WSC).

2.1 Operating Room Planning and Scheduling

In the last few years, at least three extensive literature reviews on operating room management problems were published (Cardoen et al., 2010a; Guerriero and Guido, 2011; May et al., 2011). This literature review shares the same structure of these reviews, but highlights papers addressing uncertainty on surgery scheduling and specific characteristics of the Portuguese context. More recently, a comprehensive review of Operations Research and Management Science (OR/MS) methods applied to healthcare problems highlights the relationship between different decisions and emphasizes the need for integrated approaches (Hulshof et al., 2012). However, in order to reduce complexity, authors usually breakdown problems into more manageable sub-problems and develop specific approaches. The operating room management field encompasses a variety of sub-problems, which authors try to organize in different categories and decision levels. For instance, May et al. (2011) organize operating room (OR) management problems into eight categories: capacity planning; process re-engineering; surgical services portfolio; procedure duration estimation; schedule construction; schedule execution; monitoring; and control. However, a common way to organize the literature in this area is to classify studies into three hierarchical decision levels, namely: strategic, tactical and operational. This literature review applies such classification.

2.1.1 Strategic Decision Level

The strategic decision level encompasses the case mix, resource allocation and capacity planning problems. The case mix problem consists in defining the mix and volume of patients treated by each surgical service or specific surgeon over a given planning horizon. The mix of patients is based on classification schemes that cluster patients with similar resource requirements.

Examples of such classification schemes are: the American Diagnosis Related Groups (DRG), Canadian Case Mix Groups (CMG) and Portuguese "Grupos de Diagnósticos Homogéneos" (GDH). Solution approaches depend on the nature of the institutions (public or private) and the funding mechanism in each country. In order to distinguish them, hospitals are classified in two types: profit satisfiers and profit maximizers.

For instance, healthcare in Canada is provided by independent physicians-entrepreneurs working with private non-profit hospitals. In this context, Blake and Carter (2002) developed a goal programming approach to set case mix and volume for physicians, allowing the hospital to achieve the break-even point, while maximizing surgeon's preferences in terms of case mix and desired level of income. López et al. (2008) also used a goal programming model, but for estimating the case mix included in the contract-program that public hospitals subscribe with government in Spain. The model sets case mix and volume for each surgical speciality while minimizing the deviation between target and achieved values of main contract attributes, such as: financing, number of discharges, average length of stay and case-mix index.

On the other hand, Ma and Demeulemeester (2012) assumed hospitals as profit maximizers and proposed an integer linear programming model to find the case mix that maximizes total financial contribution. In general, case mix models are subject to demand (lower and upper bounds on the estimated volume of each patient group) and resource constraints (number of operating rooms and number of beds in each surgery ward).

In Portugal, the funding mechanism of public hospitals is based on a contract-program (Santana, 2003), similar to the one used in Spain. This document encompasses several production lines, such as: inpatient surgery and clinic, outpatient surgery and clinic, emergencies, consultations, and Day-Hospital. The amount of money a hospital receives by its surgical activity is a function of the number of discharges and the complexity of procedures performed by the hospital. Such complexity is defined by the case-mix index, which is the average of the complexity indexes associated to each patient group.

2.1.2 Tactical Decision Level

With the amount of OR time and capacity of resources defined at the upper level, the tactical decision level focuses on strategies to maximize OR utilization. The strategies used at this level can be organized into three categories: block scheduling, open scheduling and modified-block scheduling. The block scheduling system is the most used and requires the construction of a Master Surgery Schedule (MSS). The MSS defines which time blocks are reserved for which surgical speciality, as well as the time blocks available for emergencies, and their respective opening hours. Figure 2.1 shows an example of a Master Surgery Schedule in use at a Portuguese hospital. It is basically a cyclic time-table that defines the operating rooms occupied by each surgical speciality on each day of the week.

ROOMS	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
NC1	NC / NC	NC / NC	NC / NC	NC / NC	LIVRE	NC
NC2	NC / NC	NC / NC	NC / NC	NC / NC	LIVRE	NC
A	OFT / OFT	OFT /OFT	OFT / OFT	OFT / OFT	OFT / OFT	LIVRE
B	ORL / ORL	ORL / ORL	ORL / ORL	ORL/ORL	CMF/ LIVRE	ORL
C	EST / OFT	CAT / C2	CMF / OFT	CAT / OFT	CAT	LIVRE
D	OFT / OFT	OFT / OFT	OFT / OFT	OFT / OFT	OFT / OFT	OFT
A	ORT / ORT	ORT / ORT	ORT / ORT	ORT / ORT	ORT / ORT	ORT
B	ORT DIF / SU	ORT DIF / SU	ORT DIF / SU	ORT DIF / SU	ORT DIF / SU	SU / SU
C	ORT /ORT	ORT / ORT	ORT / ORT	ORT / ORT	ORT / ORT	ORT
D	C3 / C3	C1 / C1	C3 / C3	C3 / C3	C2 / LIVRE	C1
E	C1 / C1	C2 / C1	URO / C1	C1/CV/ SU	CV / SU	SU / SU
F	SU / SU	SU / SU	SU / SU	SU / SU	SU / SU	SU / SU
G	URO / URO	URO / URO	UTCO / UTCO	C1 / C1	C2 / LIVRE	URO
H	CV / CV	CV / CV	C2 / C2	CV / CV	URO / LIVRE	C2
I	URO / C2	URO / Tx DADOR VIVO	URO / URO	CG / LIVRE	URO / LIVRE	LIVRE

Figure 2.1: Example of a Master Surgery Schedule

Throughout the years, mathematical models to generate a MSS have become more complex, in the sense they have started to consider multiple resources and tackle uncertainty more accurately. Blake and Donald (2002) aim to minimize the difference between the target OR time defined by the case mix planning and the actual assigned time in the master surgery schedule. It is important to achieve target OR times in order to preserve case mix and capacity planning efforts carried out on the strategic decision level. Beliën and Demeulemeester (2007) tackle this issue as a demand constraint and develop a number of models aiming to minimize the expected total bed shortage with levelled resulting bed occupancy. van Oostrum et al. (2008) not only intended to level the requirements for subsequent beds (ward and intensive care unit (ICU)), but also to maximize operating room utilization. Both authors emphasize the impact of operating room scheduling on bed capacity and nursing staff requirements. Santibáñez et al. (2007) present an innovative approach for surgical block scheduling in a system of hospitals. The authors developed a flexible model with alternative objective functions. One of them sets levelled bed utilization as a constraint and maximizes the throughput of patients.

The basic resources considered to build a MSS are: the target operating room time each speciality should get (demand) and the availability of operating rooms (capacity). Blake and Donald (2002) consider two types of operating rooms and the respective demand of each speciality. In this case, the different types of operating rooms are used for modelling operating rooms with specific equipment. Then, authors optimize surgery ward beds and ICU beds. Recovery beds are not addressed at the tactical level, unless the model integrates surgery scheduling. Such type of resource is normally addressed in models on the operational level. Santibáñez et al. (2007) constrain their model by the number of surgeons on each speciality, van Oostrum et al. (2008) acknowledge the need of reserving capacity for urgencies and emergencies in the MSS, but do not tackle this issue. Santibáñez et al. (2007) consider emergencies as a separate speciality, which operates after the time designated for elective patients and use a single OR distinct from those used by elective surgeries. An important constraint to take into account is the planning horizon. It closely depends on the degree of flexibility each hospital has. On one hand, most authors agree that the master surgery schedule should change as little as possible from week to week. On the other hand, better results can be achieved by integrated approaches that change the MSS as a function of the elective patient scheduling. For example, Tànfani and Testi (2009) and Marques et al. (2012) developed integrated approaches allowing both the number of blocks and the operating rooms assigned to each surgical speciality to change every week.

The approaches used at the tactical decision level can be distinguished between deterministic and stochastic. Besides the predominance of deterministic models, researchers have been gradually adopting stochastic models. A stochastic model usually evolves from a well-structured deterministic model. Most of the aforementioned papers present a stochastic model preceded by a successful deterministic one. Among the deterministic approaches, the problem is predominantly modelled as an integer programming problem. The deterministic integer programming (IP) model is the base for a stochastic one, as several authors modelled the problem as an IP model with probabilistic constraints. Testi et al. (2007) formulated the problem as a chance constrained stochastic model with probabilistic capacity constraints.

2.1.3 Operational Decision Level

The operational decision level can be organized in off-line (before execution) and on-line (during schedule execution) scheduling. The off-line category can be further distinguished between advance (surgery date) scheduling and allocation (starting time) schedule. In spite of this clear distinction, some studies address both steps in an integrated way. Moreover, inside each category there is a clear distinction between deterministic and stochastic approaches.

2.1.3.1 Off-line

The off-line advance scheduling is the process of fixing a surgery date for a patient. Studies in this category aim to minimize overtime or explore the trade-off between the cost of opening operating rooms and the cost of overbooking operating rooms. Since overtime occurs due to the stochastic nature of the problem, stochastic models are predominant to address this problem, although the approaches are slightly different. Such models are highly dependent on available data, most of them focus on surgery duration and some of them also address emergencies. Despite the characteristics of the problem, to the best of our knowledge, only a few studies have recently addressed the allocation problem using a stochastic approach, Denton et al. (2007); Fei et al. (2008b); Gul et al. (2011); Mancilla and Storer (2011); Lee and Yih (2014). Studies addressing the stochastic elective surgery scheduling problem usually do not include up and downstream resources and focus on the number of operating rooms to open and on the method of assigning surgeries to operating rooms. Lamiri et al. (2007) include patient related constraints, such as a deadline to perform a given surgery.

The second step on the elective planning process is the sequencing of surgeries (defining its starting times). It is a step normally taken immediately after advance scheduling on a two phase approach. Guinet and Chaabane (2003) aim to minimize hospitalization costs, overtime and patient waiting time. Riise and Burke (2011) add a quality of care measure and minimize the waiting time for children in the morning. Jebali et al. (2006) distinguish between undertime and overtime and try to minimize both. Pham and Klinkert (2008) minimize make-span (time to complete all operations), as well as schedule all individual operations as early as possible. It is an unusual objective, as long as there are no cost savings in finishing cases earlier.

Methods described above try to reduce the complexity of the problem approaching it on two separate phases. However, it can be tackled in an integrated way. For instance, Marques et al. (2012) present an integrated approach aiming to maximize OR utilization, while Conforti et al. (2010) develop a multi-objective model encompassing four objectives: maximization of the utilization of operating rooms, of the number of scheduled patients with high priority and of the number of specialties preferences satisfied; minimization of underutilization of specialties.

The off-line surgery scheduling problem, also known as surgical cases assignment problem, is the main problem addressed in this thesis so that in depth literature reviews are contained within each of the following chapters. Each of the subsequent reviews focus on specific characteristics of the problem, e.g. information systems, problem formulations and stochastic approaches.

2.1.3.2 On-line

The on-line scheduling category deals with the scheduling of add-on cases. Emergency patients cannot be planned in advance and the surgery must start immediately. Many studies report dedicated operating rooms for emergencies. However, such strategy implies additional costs, not only because the staff allocated to the emergency OR, but also because elective surgeries cannot use the OR allocated exclusively to emergencies. Lamiri et al. (2007); Lamiri and Augusto (2008) consider that a random portion of the OR-day capacity is used to serve emergency patients. Pham and Klinkert (2008) model the elective case scheduling problem as an extension of the job shop problem called multi-mode blocking job shop. The authors describe the scheduling of emergency and urgent cases as a job insertion problem.

To the best of our knowledge, it is the only effective model to schedule add-on cases and reschedule previously booked surgeries. Moreover, many authors address this issue reserving additional capacity, but does to provide any model to schedule new surgeries and re-schedule previously booked surgeries. Min and Yih (2010) define the effective capacity for each surgical block, which is calculated by subtracting emergency demand and turnaround time from the planned block capacity.

2.2 Computer Simulation in Healthcare

Computer simulation is one of the most popular operations research tools (Hong and Nelson, 2009). Moreover, it is considered one of the most suitable tools to address healthcare management problems. According to Guerriero and Guido (2011) it represents “the most reliable and efficient tool to handle the complexity and the stochastic aspects” of healthcare problems. Brailsford et al. (2009) highlight “it is the ideal approach for addressing healthcare issues”. However, its effective implementation in healthcare is not as widespread as it is in other areas, such as manufacturing and defence. In order to overcome this problem, a strict collaboration with healthcare practitioners to validate simulation models and gain buy-in and acceptance is essential (Brailsford et al., 2009).

There are different approaches to implement a simulation project, such as: discrete-event simulation (DES), system dynamics (SD), monte carlo simulation and agent-based modelling (ABM). DES and SD are the most popular. Among them, DES is the most widely used simulation approach in healthcare (Brailsford et al., 2009). It is a detailed, stochastic, patient level approach. On the other hand, SD has a whole-system, strategic view. In addition, Monte Carlo simulation and Agent-based modelling has also been used in healthcare. Monte Carlo simulation is the oldest approach and has been applied to stochastic optimization (Goldsman et al., 2010). Agent-based modelling is an emerging and promising approach. In fact, Siebers et al. (2010) claim that DES is dead and Agent-based modelling is the future.

The modelling flexibility of DES is one of its main advantages. Brito and Teixeira (2001) note that analytical procedures are unable to model complex systems, leading to the development of computer simulation to overcome this barrier. In particular, DES models are composed by a network of queues for services in which individual entities flow around. For instance, patients join waiting lists or queues for shared resources, such as: operating rooms, recovery rooms and equipment. Moreover, entities have characteristics that determine their pathway through the network, which are similar to individual patient characteristics, such as: surgical procedure, diseases and surgeon. According to Brailsford et al. (2009) DES "can take account of randomness, variability and uncertainty, as long as enough simulation runs are performed to obtain statistically significant results".

In order to model the uncertainty in processing times and patient arrivals researchers have been using special probability distributions. In particular, lognormal distributions have been used to model surgery times (Strum et al., 2000; Spangler et al., 2004). However, surgery times present a high variability according to characteristics of the surgical procedures, surgical team and group of patients (Li et al., 2009; Eijkemans et al., 2010). In this context, advanced distribution fitting software can make simulation models more valid (Law and McComas, 2011). Such tools are able to accurately determine which probability distribution best represents the data.

2.3 Simulation Optimization

Computer simulation is used for analysing the performance of a given system without using the actual system. In order to perform a simulation, the analyst must create a model of the system. This model is an abstraction of a real system or process and is used to infer the real system's behaviour over time and on different scenarios. Each scenario is represented by a set of input parameters configured by the analyst. The process of testing different configurations of input parameters and observe the system's behaviour is called scenario analysis. Therefore, it is natural to search for the set of parameters that optimizes system's behaviour. However, when the number of possible alternatives is too high and it is not possible to enumerate them all, an optimization procedure is required. Simulation optimization is an area of computer simulation which integrates optimization techniques into simulation analysis.

Fu (1994) presents one of the earliest reviews of simulation optimization techniques. Since then, the area has been gaining popularity motivated by the advances in computer power and memory. Nowadays, simulation optimization has become an active and fast growing research area, offering the most exciting opportunities in the computer simulation field (Hong and Nelson, 2009). Over the last decade, the evolution of simulation optimization approaches has been well documented in regular papers presented at the Winter Simulation Conference (Glover et al., 1999; Fu et al., 2000; Olafsson and Kim, 2002; April et al., 2003; Fu et al., 2005, 2008; Hong and Nelson, 2009; Figueira and Almada-Lobo, 2014). Above all, Fu (2002) represents the most significant study in the area.

There are two main challenges to design simulation optimization approaches: efficiency and statistical validity. Such challenges are inherent to the stochastic nature of the problem and exist because of estimation errors. For instance, deterministic approaches require only one evaluation of the objective function to precisely estimate the performance of a given solution. On the other hand, stochastic approaches require multiple simulation runs, because the output of a single simulation run is random. Therefore, efficiency and statistical validity are directly related with the number of simulation runs performed. Figure 2.2 shows a basic simulation optimization approach instantiated to the surgery scheduling problem. In this example, the optimization box contains a search procedure to generate alternative solutions, while the simulation box assesses the performance of each alternative solution using a simulation model.

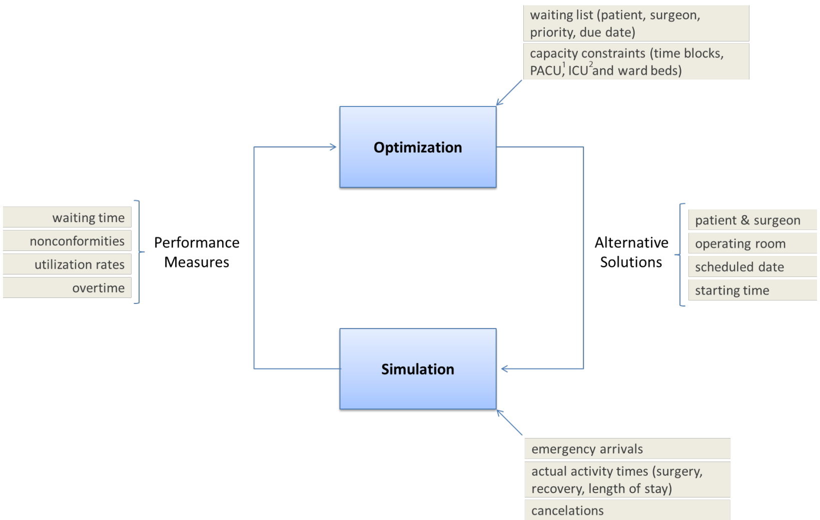


Figure 2.2: Example of a simulation optimization approach

The number of simulation runs performed to estimate the performance of a solution is called simulation cost and the total number of simulation runs available to spend is called simulation budget. As long as the simulation cost increases the statistical validity increases and efficiency decreases. In this context, efficiency concerns how simulation optimization approaches spend the simulation budget in order to improve statistical validity. This is particularly relevant because of the high computational cost of performing additional runs. Moreover, the efficiency of search procedures can also be harmed by estimation errors. Therefore, efficient simulation allocation rules are crucial to find good solutions for practical problems. Simulation opti-

[1] Post-anesthesia Care Unit
[2] Intensive Care Unit

mization approaches are classified into three categories according to the size and structure of the feasible region (Fu et al., 2008). Firstly, if the feasible region is small, problems are classified as ranking and selection problems (R&S). Next, if the feasible region is large and continuous, problems are classified as continuous variable problems (COvS). Finally, if the feasible region is large and parameters are discrete, problems are classified as discrete variable problems (DOvS). It is important to identify the problem within one of these categories because there are specific approaches for each one of them. For instance, surgery scheduling problems are modelled as discrete variable problems. The following paragraphs provide a brief review of the main approaches in each category. Hachicha et al. (2010) give an extensive literature review on simulation optimization methods.

2.3.1 Ranking and Selection Problems

Firstly, ranking and selection problems, or multiple comparison problems, consist in selecting the best design among a small set of alternatives. The most popular approaches to tackle this problem are: ordinal optimization, expected value of information (VIP), optimal computing budget allocation (OCBA) (Chen and Lee, 2010) and indifference zone (IZ) (Hong and Nelson, 2005). OCBA aims to minimize the total simulation budget while achieving a desired optimization level. IZ allocates samples in order to provide a guaranteed lower bound on the probability of correct selection (PCS). VIP describes the evidence of correct selection with Bayesian posterior distributions and allocates samples using decision theory tools to maximize the expected value of information in those samples. For a review of ranking and selection procedures see Kim and Nelson (2007). Branke et al. (2007) conducted an extensive set of experiments to determine the most effective selection procedures. The authors recommend the utilization of OCBA to tackle discrete variable problems, as it showed to be the most efficient approach.

2.3.2 Continuous Variable Problems

Next, continuous variable problems, or continuous optimization via simulation problems (COvS), encompass problems in which feasible region presents a continuous structure. Solution approaches in this category rely on gradient estimators, taking advantage of the structure of the solution space to improve algorithm's efficiency. The most well-known approaches are Stochastic gradient estimation (Fu, 2006), sample path optimization, stochastic approximation and sample average approximation (Kim et al., 2015). Such approaches are based on direct estimation methods, and how the gradient is obtained depends on how much knowledge of the simulation model the algorithm has. Fu (2002) emphasizes that commercial simulation optimization tools see simulation models as a black box and do not take advantage of problems with structure.

Metamodel based approaches also benefit from the gradient, but are obtained in a different way, involving methods such as: linear regression models, quadratic regression models and neural networks. The most well-known approaches in this category are: Response Surface Methodology (RSM) and Kriging-based metamodels. RSM tries to obtain a functional relationship between the input variables and the output objective function. Once a metamodel, also called, surrogate model, is obtained, the search can be carried out with deterministic optimization procedures, since the output is deterministic. The algorithm navigates the solution space using the surrogate model, saving costly simulation runs. Barton and Meckesheimer (2006) present a review of metamodel-based simulation optimization approaches. Li et al. (2008) present a multi-objective simulation optimization approach in which a Kriging-based metamodel is embedded within a multi-objective genetic algorithm. In this work, some of the alternative solutions are evaluated on-line using Kriging metamodels instead of the actual simulation model. Fu (2002) notes that OptQuest, a leading optimization tool for commercial simulation software, uses neural networks for screening out candidates likely to be poor.

2.3.3 Discrete Variable Problems

Finally, discrete variable problems, or discrete optimization via simulation problems, are the most comprehensive category and encompass problems having a discrete search space. This category encompasses most healthcare planning and scheduling problems. Solution approaches in this category cannot explore the structure of the feasible region and are based on random search and metaheuristics adapted from deterministic optimization. Such approaches include, but are not limited to: Evolutionary Algorithms, Simulated Annealing, Tabu Search, Scatter Search, Nested Partitions Method, Cross Entropy Method and COMPASS. They are the most appropriate methods to tackle scheduling problems. However, there is a distinction between them. While the first group comprehends solution approaches adapted from deterministic optimization, the second contains solutions designed specifically to tackle simulation optimization problems. For a complete review of simulation optimization approaches based on random search and metaheuritics see Andradóttir (2006) and Olafsson (2006) respectively.

Boesel et al. (2003) emphasize that methods adapted from deterministic optimization do not offer any performance guarantee and can be inefficient in the presence of high variability. However, these methods are one of the most practical approaches to simulation optimization and integrate most commercial simulation optimization tools. Indeed, such methods must be integrated with statistical ranking and selection frameworks in order to be able to perform well on a stochastic context. Lee et al. (2008) propose the integration of genetic algorithms and OCBA. Chen and Lee (2010) describe in detail the integration of OCBA with different metaheuristics. Olafsson (2006) notes that understanding how to account for simulation noise in metaheuristics may improve their performance in practice.

On the other hand, a group of simulation optimization approaches was specifically designed to simulation optimization problems. For instance, the Convergent Optimization via Most Promising Area Stochastic Search (COMPASS) (Hong and Nelson, 2009; Nelson, 2010) is one of them. This algorithm is designed to be robust to simulation noise and has an integrated simulation allocation rule (SAR) for enhanced efficiency. Also, an extended version of the algorithm to tackle multi-objective problems was proposed Lee et al. (2011). Likewise, the Industrial Strength COMPASS (ISC) is an extended version of the original algorithm that aims to be competitive with the best commercial software and still provides guarantees of convergence (Xu et al., 2010).

2.4 Research Gaps

The current literature review reveals a mismatch between the characteristics of the problems and the features included in most part of the solutions. Problems are described as combinatorial, multi-objective, subject to strong uncertainty and to the availability of multiple resources. However, most part of the solutions is deterministic and considers only a limited number of objectives and constraints. Surgery scheduling approaches contain unrealistic assumptions and focus on specific aspects of the problems in order to reduce complexity. This issue impacts which resources are taken into account and how uncertainty is modeled. Few problems consider, for instance, the characteristics and availability of the surgical team.

Regarding uncertainty, studies focus on elective scheduling, neglecting the impact of non-elective patients, as well as add-on cases, on surgery scheduling. The approaches that address this issue consider only one source of uncertainty. Uncertainty comes from two distinct sources: arrivals and processing times. Patient arrivals are modelled using a single probability distribution function and do not take seasonality and specific demand into account. Stochastic processing times are used to model surgery durations and deterministic values represent the processing times of all other steps along the surgical process. For instance, the cleaning time is considered to be fixed and equal to all surgical specialties. Stochastic optimization approaches are based on Monte Carlo Simulation and do not take the time dimension into account.

The high computational cost prevents researchers to devise more complex models. Efficiency is a main issue when designing a stochastic optimization approach. There is a lack of approaches to tackle this specific issue on the operating room management area. Statistical validity is compromised in order to reduce the computational time. There is a need to bring the latest techniques to reduce computational cost and ensure statistical validity from the simulation optimization theory into the operating room management area.

Regarding the type of decisions, there is a lack of studies focusing on case mix planning on the strategic decision level as well as on-line scheduling on the operational decision level. This gap on how to deal with emergencies and add-on cases can compromise the results of all precedent results. This issue becomes more important if we acknowledge that patients should be scheduled more time in advance. It would be interesting to study the relation between decisions on off-line and on-line phases of the operational decision level.

This thesis approaches some of the gaps identified in the literature review.

In particular, Chapter 4 tackles the high computational cost issue, developing efficient scheduling methods for tackling large scale problems. Chapter 5 tackles the uncertainty issue, presenting a multi-objective approach to surgery scheduling under uncertainty. This approach aims to devise more realistic schedules, reducing the mismatch between the characteristics of the problems and the features included in most part of the solutions. In addition, Chapter 3 tackles the issue of developing a decision support system in order to allow the hospital managers and surgeons to make practical use of the proposed scheduling methods. The hospital information systems available in Portuguese hospitals lack a decision support component. They are used to register data for controlling purposes. They do not transform data on valuable information to support decision making.

Chapter 3

An Intelligent Decision Support System for the Operating Theater: A Case Study

Fabrício Sperandio[1], Carlos Gomes[1], José Borges[1], António Carvalho Brito[1], Bernardo Almada-Lobo[1]

Published in IEEE Transactions on Automation Science and Engineering (Volume: 11, Issue: 1, Pages: 265-273), 2014

Abstract From long to short term planning, decision processes inherent to operating theater organization are often subject of empiricism, leading to far from optimal results. Waiting lists for surgery have always been a societal problem, which governments have been fighting with different management and operational stimulus plans. The current hospital information systems available in Portuguese public hospitals lack a decision support system component that could help achieve better planning solutions. Thus, an intelligent decision support system has been developed, allowing the central-

[1]CEGI – INESC TEC, Faculdade de Engenharia, Universidade do Porto, Portugal

ization and standardization of planning processes, improving the efficiency of the operating theater and tackling the waiting lists for surgery fragile situation. The intelligence of the system derives from data mining and optimization techniques, which enhance surgery duration predictions and operating rooms surgery schedules. Experimental results show significant gains, reducing overtime, undertime, and better resource utilization.

Keywords Operating Theater, Planning, Intelligent Decision Support System

3.1 Introduction

As quality of life improves and societies live longer, health care organizations face significant increases in their demand. It is a vicious loop. Population is ageing due to better health assistance, which is supported by costly technological advances, and aged population requires increased care. These factors increase health care costs and require better management of existing resources. In this context, to maintain good service levels and patient satisfaction, health care organizations are faced with two options: either expand capacity or improve existing resources utilization. The former implies huge capital investments and is therefore a difficult strategic decision. However, improving processes and efficiency entails an organizational development set of actions that can be performed more easily, involving less investment.

The operating theater is often considered the biggest budget consumer and revenue center in a hospital. In addition, its performance has a severe impact on society. Waiting lists for surgery are a critical issue that affect many lives, hence being constantly battled by health care organizations and governments.

In this chapter, motivated by a real world case, we present an intelligent decision support system for operating theater planning and scheduling, and the performance improvement achieved with it. The system was designed for two user profiles, surgeons and hospital managers, providing them a planning framework for tactical and operational problems. The two main functions of the system are: (i) to provide users the means to monitor and to measure the performance of the operating theater; and (ii) to aid users devising better planning alternatives by supporting the task of creating better plans with data mining and optimization techniques.

Our work was integrated within a business process improvement project that took place in a large Portuguese public hospital, allowing the team to gain a fundamental understanding of the surgery scheduling process and the corresponding user needs. The project introduced us to a reality with a heterogeneous way of work across the different specialties and low guideline compliance. These behaviors result from poor organizational monitoring and lack of work flow standardization. To tackle this situation, we have devised a system which helps to standardize the planning processes and to control quality and productivity.

Surgery planning involves taking into account different activities that are to be performed in a very uncertain environment. Such uncertainty leads to frequent deviations between what was planned and what was in fact performed. Several authors defend that the surgery schedule's quality is mainly determined by the accuracy of the surgery duration estimation (Dexter et al., 1999). Thus, in order to reduce deviations, improving duration estimates was sought in this work. Regarding surgery scheduling, we have formulated a mathematical optimization model, which allows finding the optimal allocation of patients to the available operating room shifts.

The novelty introduced in our work concerns the development of a decision support system for the operating theater and the integration of a scheduling method, involving data mining and optimization techniques. The research in this area has been extensive, however, to the best of our knowledge there is no work connecting decision support, uncertainty reduction and surgery schedule optimization.

Following this introduction, we present a literature review of the features and problems addressed in this work. In Section 3.3 we provide a better insight into the operating theater planning problems. The decision support system, the techniques used and its implementation are briefly described in Section 3.4 and final remarks will be given in Section 3.5.

3.2 Literature Review

Operating theater planning problems have been widely covered throughout the literature (Cardoen et al., 2010a; Guerriero and Guido, 2011; Blake and Carter, 1997) and it is a still growing field of research. The most prevalent scientific community in the operating room is the operations research one, which typically studies scheduling problems. However, there is a gap between theory and practice. A Swiss survey has shown that hospital managers are not satisfied with the state of art of scheduling and Hospital Information Systems (HIS) (Sieber and Leibundgut, 2002). Moreover, a recent Portuguese study also criticizes the current HIS used in Portugal, stating that they are functionally and technologically outdated (Gomes et al., 2009).

Operating theater planning is normally divided in three decision levels: (i) operational, (ii) tactical and (iii) strategic. Our work is focused on the first, which corresponds to the periodic (weekly) scheduling of patients to the available operating rooms. The tactical and strategic decision levels concern longer term decision of capacity definition and allocation to the different surgical specialties (operating room timetable).

The operational decision level can be organized in off-line (before execution) and on-line (during execution) scheduling. The off-line category can be further distinguished between advance scheduling, when only the surgery date is defined, and allocation scheduling, when the surgery is sequenced in an operating room. Studies focused on off-line scheduling (Denton et al., 2010; Lamiri et al., 2007; Fei et al., 2008a; Hans et al., 2008), aim to minimize overtime, maximize throughput or explore the trade-off between the cost of opening and overbooking operating rooms. Most of these studies do not include up or downstream resources, dealing just with operating rooms. Lamiri et al. (2007) includes patient related constraints, as the deadline to perform a given surgery; Guinet and Chaabane (2003) aim to minimize hospitalization costs, overtime and patient waiting time; Riise and Burke (2011) add a quality of care measure and minimize the waiting time for children during mornings; Jebali et al. (2006) distinguish between undertime and overtime and try to minimize both. On the other hand, the on-line scheduling category deals with the scheduling of add-on cases (Lamiri et al., 2007; Lamiri and Augusto, 2008; Pham and Klinkert, 2008; Min and Yih, 2010; Persson and Persson, 2010), such as emergency patients, who can not be planned in advance and whose surgery must start as soon as possible. Many studies report dedicated operating rooms for emergencies, however, this strategy implies additional costs, due to staff allocation and maintenance costs, moreover elective patients can not use these operating rooms. Lamiri et al. (2007); Lamiri and Augusto (2008) consider a random portion of the OR-day capacity to serve emergency patients; Pham and Klinkert (2008) model the elective case scheduling problem as an extension of the job shop problem called multi-mode blocking job shop. The authors then describe the scheduling of emergency and urgent cases as a job insertion problem; Min and Yih (2010) define the effective capacity for each surgical block, which is calculated by subtracting emergency demand and turnaround time from the planned block capacity.

The surgical process is characterized by strong uncertainty (Dexter et al., 1999), as different sources of variability emerge from the patient arrival to his postoperative recovery. The surgery duration, including anesthesia and surgical act, is the most studied in the literature. The factor that better defines the duration of a surgery is the combination of surgical procedures (Li et al., 2009). Other significant sources are the main surgeon performing the procedure and his team, anesthesia type, risk class, patient age and gender (Strum et al., 2000). Although those features can explain part of the variability, they also present a major barrier due to the large variety of procedures and the high number of surgeons in a hospital (Macario, 2009).

Researchers have been modeling surgical times targeting different management decisions, but most studies aim to predict surgery duration before it starts (off-line scheduling), others predict the time remaining during surgery execution (on-line scheduling) (Dexter et al., 2009). Finally, another cluster of research focuses on predicting the duration of a series of surgeries, aiming to reduce overtime (Alvarez et al., 2010). However, not every management decision requires an exact point estimate, authors recognize that because the uncertain nature of surgical procedures, it is often better to know its upper and lower bounds than a single estimate (Stepaniak et al., 2009).

With every model and solution method developed, there is a need to bring them into practice and for that (intelligent) decision support systems have the potential to deliver them to the user. The concept of decision support systems can be summarized as information systems designed to support decision making activities. Turban (1982) defines DSSs as interactive, flexible and adaptable information systems proposing possible and better course of actions to the decision maker, they aid decision agents to analyze their options and to find the best alternative among a wide solution space. These systems have long proved to be effective when applied to various domains such as health (Jaspers et al., 2011), where two different applications should be distinguished: (i) management DSS, oriented to organization control; and (ii) clinical DSS. The latter concerns the executional level, where the goal is to mitigate harmful and expensive medical mistakes and help clinical staff to perform their jobs (Jao and Hier, 2010), for example, by providing more accurate diagnoses or safety checklists. These are patient-oriented systems, where the main objective is to improve the clinical work flow, guaranteeing patient care and safety. Intelligent decision support systems move a step further and integrate different techniques (e.g.: decision analysis through data mining) to give these applications an intelligent behavior. Guerlain et al. (2000) identify 7 characteristics of intelligent decision support systems: (i) interactivity; (ii) event and change detection; (iii) representation aiding; (iv) error detection and recovery; (v) information out of data; and (vi) predictive capabilities. This kind of capabilities can be of extreme value to decision agents and provide new decision models to any organization.

3.3 Operating Theater: Portuguese Case Study

According to a Swedish study (Bjornberg et al., 2009), Portugal ranked 21st out of 33 European countries on providing health care services. This result was mainly influenced by the long waiting time for treatments, where Portugal ranked last. On the other hand, on electronic health services Portugal ranked 1st, due to the early, but still in progress, adoption of a national electronic health record (EHR).

In 2004, as an effort to fight the long waiting list for surgery, the Portuguese government introduced a set of policies and guidelines focused on protecting patients' rights and health. The System for Management of Patients Waiting for Surgery (Ministério da Saúde, 2011) introduced a set of waiting time limits according to the patients' priorities. Hospitals are penalized in case patients waiting time limit are exceeded. For example, a high priority patient may only wait for surgery 15 days while a normal patient sees this period extended to 270. To avoid penalizations, existing resources must be used efficiently and to achieve that, surgery schedules must be carefully planned. However, we found that the current hospital information system used in Portuguese public hospitals has limited capabilities to create optimal surgery schedules or even to measure their quality. Decision making processes within the surgery theater are often empiric and the available information systems lack a decision support component, which would help achieving better results. We witnessed surgeons using different methods to devise their planning, such as personal agendas, spreadsheets and online calendars, reducing the level of centralization and integration within the hospital to insignificant levels. Note that it is crucial to share this information internally and with other departments, since operating theater resources are shared among different specialties and people. We reported hundreds of surgical cases being scheduled (inserted into the HIS) after the surgery itself, creating a communication issue between the different departments and the operating theater.

In general, surgeons are not very focused on operational performance and have poor sensibility for optimization, sometimes they are not even aware of how long their patients have been waiting. Even when they are estimating the duration of a surgery they tend not to be very accurate. In fact, improving the accuracy of surgery duration predictions can play a major role in increasing operating theater efficiency. When the duration of a surgery exceeds its prediction (overtime) there is a cascading effect delaying upcoming surgeries, while when the duration is overestimated leading to an early finish (undertime), valuable time is wasted idling, leading to operating room (OR) underutilization. Our analysis has shown that 82.25% of surgeries performed in our case study between 2006 and 2010 had a relative duration deviation of over 10% from their estimation. Table 3.1 summarizes the total sum of undertime and overtime on surgeries performed in that period.

Table 3.1: Summary of overtime and undertime from 2006 to 2010

	Total Time	Number of Surgeries	Average
Undertime	918.066 min	49.029	18,72 min
Overtime	2.092.461 min	33.575	62,32 min

In summary, we have observed in this hospital that there is room for improvement on surgery planning processes and resource management, therefore, benefiting from a decision support system to the operating theater.

3.4 Intelligent Decision Support System

The solution proposed to tackle the long waiting times for surgery is divided into three vectors discussed in this section: (i) decision support system for better information and resource management; (ii) a data mining model to predict surgery durations; and (iii) a weekly elective patient scheduling optimization model. This approach was inspired on the work of Better et al. (2007), who developed a problem solving framework integrating simulation, data mining and optimization techniques.

The decision support system was developed following a user centered approach based on the traditional software engineering life cycle model. The first task of identifying user needs and establishment of the requirements specification was conducted through a series of workshops meant to characterize the operating theater scheduling process and assess where it could be improved. The workshops were not exclusively focused on the decision support system development, but they were essential for understanding and characterizing business processes, as well as to identify the strengths and weaknesses of the current information systems. As a result of these series of workshops, a requirements specification document and a set of low resolution prototypes were produced, which were then presented and validated by key users from the hospital staff. The first trials of the system were initially deployed in two surgical specialties of the hospital as a pilot run.

Having worked closely with a hospital, many features incorporated were requested by surgeons and others were designed to overcome problems detected on the hospital information system currently in use. Another purpose of the decision support system was to integrate data mining and optimization techniques and deliver them to decision agents. The system developed is divided in 3 main modules: (i) resource management; (ii) surgery planning and scheduling; and (iii) performance measurement.

The resource management module is to be used by the operating theater management personnel, grouping features required to define and allocate existing resources (e.g. operating rooms, medical specialties, surgeons and users of the system). The system enables not only the creation of weekly surgery plans, but also the allocation of specialties to operating rooms (master surgery schedule), related to the operating theater tactical decision level.

The surgery planning and scheduling module is the core of the decision support system and makes available a set of features to schedule surgeries. The surgery scheduling interface supports the daily/weekly process of scheduling surgeries and was created to be as functional and easy to use as possible. This agenda shows the operating rooms available for a user's specialty and allows a weekly or daily perspective. The weekly view is an important feature, as it allows the visualization of an entire week operating room plan, which was not available before. To support the surgery scheduling process, we have integrated a data mining model that provides the user an estimation of the surgery duration and an optimization model that gives an optimal scheduling solution according to a given objective function. Figure 3.1 depicts an operating room's agenda and the system's optimization feature.

Here, the user may select different strategies to compute the schedule. Either a dispatching rule that allocates patients on the basis of first in first out, or a mathematical model that optimizes one of the three following objectives: maximization of the number of surgeries, maximization of the OR utilization or minimization of the waiting time.

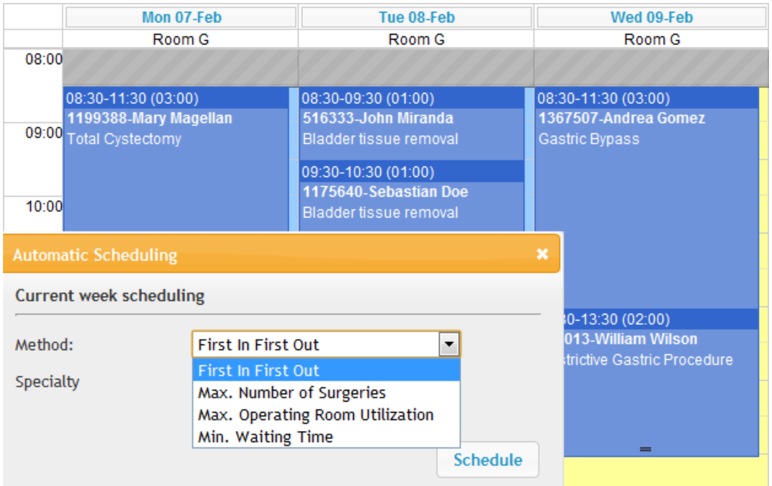


Figure 3.1: Overview of the weekly surgery schedule for a certain operating room. In this particular example, each day corresponds to a different surgical specialty and each block equals a surgery and includes the patient identification, name and procedure

Regarding the patients' waiting list management, two features were specially welcomed by the surgeons: a color scheme that highlights patients according to the time left relatively to the waiting time limit and the possibility of filtering the waiting list by the estimated surgical procedure time duration. The latter gives the means to rapidly identify a surgery adequate to fill a gap in the planning horizon. Details about surgeries and patient information are also easily accessed on the interface. Finally, a non-obtrusive notification system was created providing alerts when operational restrictions are violated. For example, a notification is issued when the expected time duration for the planned surgeries exceeds the limits of the period allocated to the corresponding specialty or when the scheduling violates patients' priorities.

The third module concerns results evaluation through Key Performance Indicators (KPIs), enabling identification of anomalies and opportunities to improve performance. A set of customized charts is provided, such as: operating room/specialty utilization rate over time, the evolution of patient waiting lists over time and the number of penalties due to violation of priorities throughout time. These KPIs are embedded in interactive dashboards that allow an exhaustive benchmark of performance of different surgeons, specialties and the overall operating room.

According to Guerlain's (2000) framework, this work fits in the intelligent decision support system cluster, as it provides the dimensions discussed on his work and goes further giving a scheduling automation feature. A minimalist overview of the sequential workflow performed by this decision support system is given in Figure 3.2.

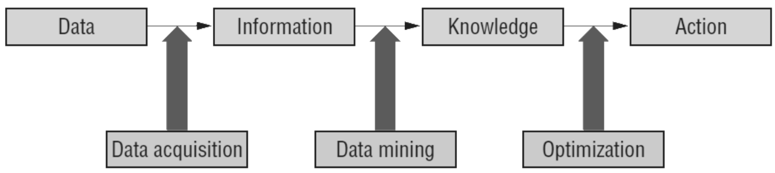


Figure 3.2: Intelligent Decision Support System Process flow

The following subsections will briefly describe the techniques used to provide the intelligent behavior into the system and some of the results achieved. Two appendices were included, where these components are explored further.

3.4.1 Surgery Duration Estimation

While scheduling patients, surgeons have to estimate how long surgeries will take in order to book the operating room in advance, as it is a shared resource. Estimating surgery duration has an important impact on the operating theater schedule and since surgeons' estimates suffer from high deviation (see Table 3.1).

The problem of surgery duration estimation has been widely studied and it has been shown that the operation times can be modeled by lognormal or normal distributions. Several works have reported that the distribution of surgical procedures time can be modeled by a log-normal distribution (Stepaniak et al., 2009). In Eijkemans et al. (2010) the authors conclude that a prediction model aimed at making predictions for individual patients that includes detailed procedure codes and operation, team and patient characteristics may be able to reduce shorter-than-predicted and longer-than-predicted OR times by 12 and 25% respectively. Therefore, the application of data mining techniques seems suitable to address this problem. Data mining concerns the automated discovery of patterns and relationships in data, also known as Knowledge Discovery in Databases (KDD). These techniques work with big and high-dimensional datasets, used to predict future behavior by observing history. Patients and completed surgery databases fit accurately within that description and provide a great source of data to explore (see an example for surgery durations in Figure 3.3).

Experiments were conducted with regression, tree-based and neural network algorithms while using bagging and boosting techniques or not. For our datasets, the best overall performing algorithm was a regression-like model that encompasses two algorithms to predict surgery durations: Bagging and M5 Rules.

Bagging stands for bootstrap aggregating, it is an ensemble meta-algorithm to improve machine learning classification and regression models stability and accuracy, by reducing variance and avoiding over-fitting. This technique generates several versions of the predicting model and uses them to get an aggregated, averaged, predictor. The different versions of the predictor are made by replicating and perturbing the learning set, causing significant changes in the predictors built (Breiman, 1996).

The predictor used, M5 Rules, is based on a decision list built from several M5 model trees (Holmes et al., 1999). During the learning phase, in each iteration a model tree is built and the best leaf (according to some heuristic) is pruned into a rule. Instances covered by this rule are removed from the dataset, so that the process is applied recursively to the remaining instances, terminating when all instances are covered by one or more rules.

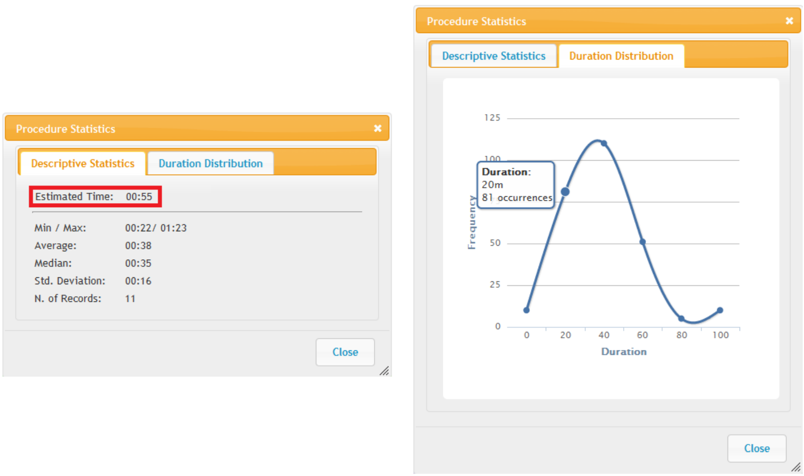


Figure 3.3: Surgery duration estimation (highlighted on the left window), descriptive statistics and the distribution of the durations frequencies for a surgical procedure

Our dataset was built using records from 2006 to 2011 and the last two years of data were separated to validate our results. In Appendix 3.A the structure of the dataset (data types and fields) is presented.

Experiments were conducted with several specialties and herein we report the results of two representative specialties. Experimental results were compared against surgeon duration estimates of two surgical specialties: General Surgery (**GS**) and Vascular Surgery (**VS**). Table 3.2 shows the results obtained in terms of Mean Absolute Error (**MAE**), Mean Absolute Percentage Error (**MAPE**) and Mean Squared Error (**MSE**).

Table 3.2: Comparison between surgeon estimates and data mining results

	Surgeon Estimates			Data Mining		
Specialty	**MAE**	**MAPE**	**MSE**	**MAE**	**MAPE**	**MSE**
GS	70.12	38%	9336.49	47.93	32%	4784.67
VS	32.97	49%	2508.04	24.94	39%	1514.97

Both specialties watch a great improvement on the prediction accuracy. One of the reasons for poor surgeon performance derives from the time period granularity used (multiples of 10/15 minutes) as depicted in Figure 3.4. On the other hand, Figure 3.5 plots the data mining predictions against the real values, reinforcing that surgeon's granularity presents a severe constraint to fine tune schedules. Values below the diagonal line on each figure (optimal predictions) represent surgeries that went overtime, while the others were overestimated leading to operating room under-utilization.

From these results it is clear that there is a strong potential gain by reducing the error in surgery estimation time with our method.

Figures 3.6 and 3.7 present two histograms that enable to compare the distributions of the overestimation and underestimation times for the two selected specialties. The results show that overestimation is more frequent and that underestimation has more extreme values. These extreme values probably correspond to cases in which the surgeon decides to cancel the surgery after the first few minutes due to unexpected factors regarding the patient condition.

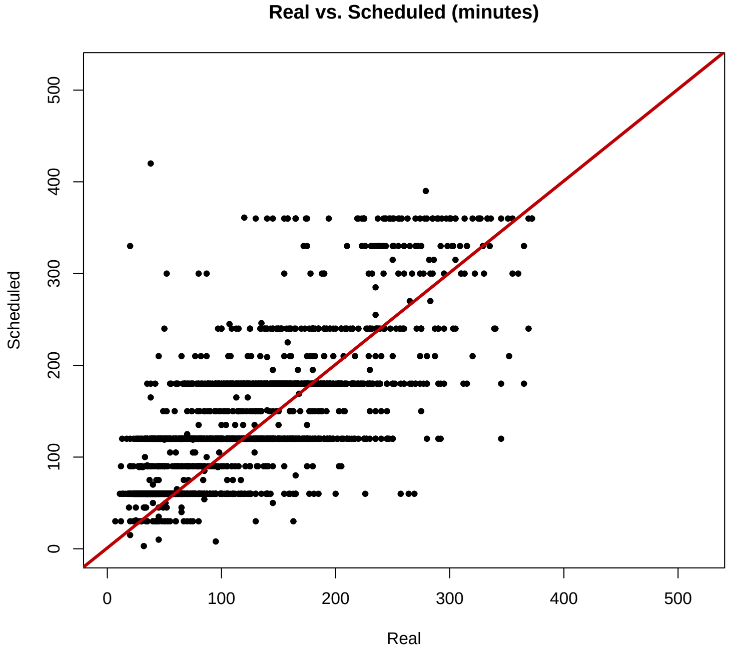


Figure 3.4: Surgeon's scheduled duration vs. real duration

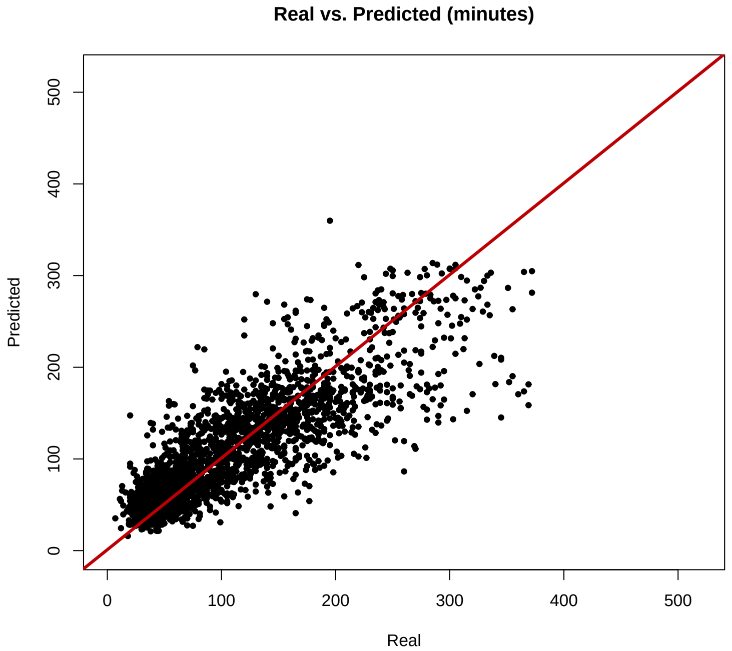


Figure 3.5: Data Mining prediction vs. real duration

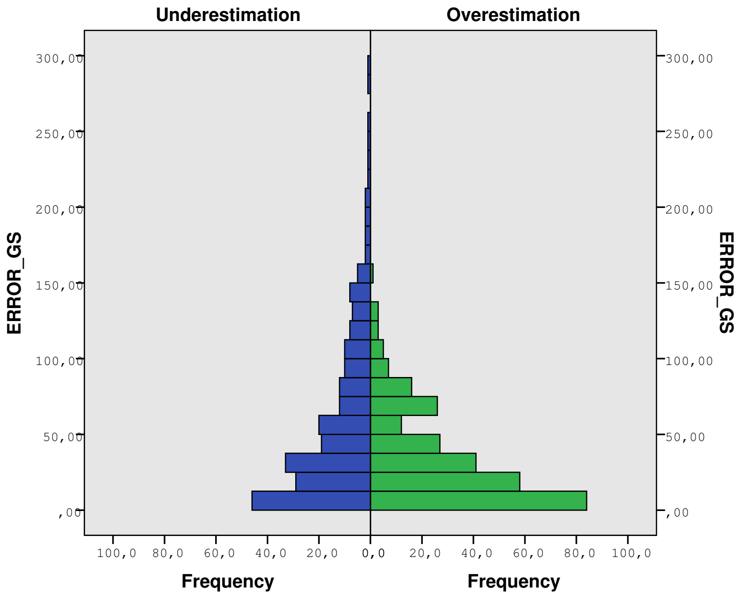


Figure 3.6: Histogram comparing the distribution of underestimation and overestimation times for General Surgery

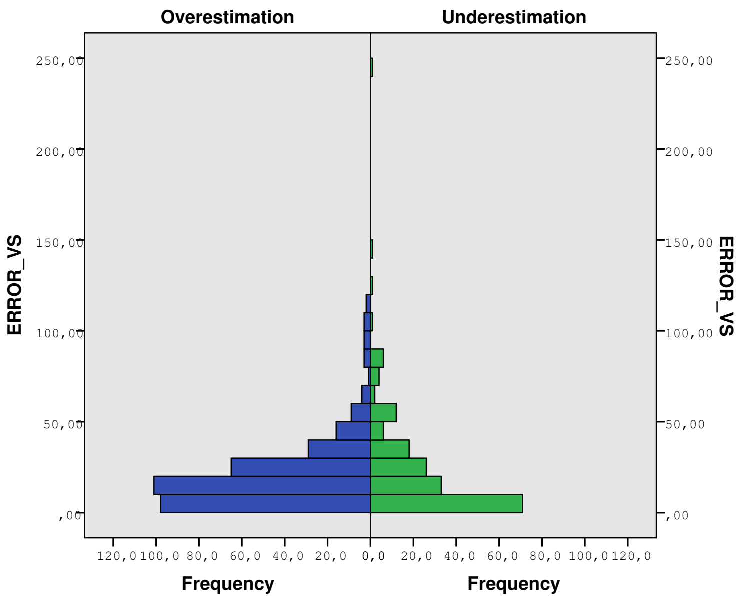


Figure 3.7: Histogram comparing the distribution of underestimation and overestimation times for Vascular Surgery

3.4.2 Optimization Model

Having surgery duration accurate predictions is the first step to devise better schedules. However, there is still the need to find a combination of surgical cases that respect a set of constraints and optimize the surgery plan. We have modeled this problem as an advance scheduling problem, where we take the patients from a surgical specialty and assign them to a certain day of a week and operating room. Our approach is an adaptation of the multiple knapsack problem, where each knapsack corresponds to a morning or afternoon shift on an operating room in a given day. Although we formulate the model as a single-objective problem, three different objective functions are proposed and used in a row: (i) throughput maximization; (ii) utilization maximization; or (iii) waiting time for surgery minimization (days removed from the waiting list).

The model's complete formulation is given in Appendix 3.B and it is solved using CPLEX 12.2. ILOG Concert technology is used to make the bridge between the solver engine and the decision support system. Since the model only deals with elective patients (el—pat) and the problem is addressed in a deterministic fashion (det), according to the framework proposed by Cardoen et al. (2010b) our approach is described as:

(el — pat; date-time; iso — det — single; wait-through-util)

where *iso* means that the operating room is tackled in an isolated way, without taking into consideration downstream and upstream resources.

Computational results for the maximization of one week's number of surgeries are shown in Table 3.3. The first column corresponds to the surgical specialty and columns 2 to 4 correspond to each objective function value. Utilization rate is calculated disregarding clean-up times between surgeries and days removed concern the sum of waiting days for surgery of the scheduled patients, therefore removed from the waiting list. Extended results are given on Appendix 3.B.

Table 3.3: Optimization results - maximizing throughput

Specialty	No. Surgeries	Utilization	Days Removed
GS	43	76.26%	13684
VS	51	70.19%	5598

Comparing these results to reality would lead to a significant disparity towards our results as they are greatly inflated. The operating theater bottleneck lies on the postoperative capacity, the reason to find average operating room utilization rates below 50%. However, we believe the solutions provided by our model can be used as starting points to good planning solutions.

3.5 Conclusions

In summary, this chapter reports the development of a decision support system intended to endorse the process of operating theater planning.

The solution presented is mainly directed to the effective management of the operating theater, where data mining and optimization components are added to allow for more efficient scheduling. To the best of our knowledge, this work is the first to combine the aforementioned techniques to reduce surgery uncertainty and to achieve a better utilization of the existing resources through scheduling optimization within decision support systems.

The results shown, regarding both surgery scheduling and duration estimation, are significantly better than the current reality and can provide the end-user a great advantage when planning, compared to the methods used in the past. Extensions of the optimization model to include other upstream and downstream resources shall be considered in the future, as well as the development of a simulation component to better evaluate generated solutions.

Appendix

3.A Surgery Estimation

The data used was obtained from the patients and completed surgeries databases from the hospital, covering about 90,000 surgeries between 2006 and 2011. From our experience, administrative staff and surgeons are prone to data insertion errors, as we have observed several cases of simple and quick surgeries lasting longer than 12 hours, resulting in the need of data cleansing. We have adopted this procedure for the two specialties for which the results are being reported since such cases were very infrequent (1.9% for GS and 0.3% for VS) and therefore would not have a significant impact in the prediction model. We stress, however, that in case of specialties such as heart or neuron surgery this type of deviation could be natural.

Subsequently, data was divided, where the first 5 years of completed surgeries were used for building the meta-model and the remaining 2 for evaluation. We have adopted a temporal split to mirror the real world scenario in which the model is aimed at predicting the duration of future events from past records.

A mixture of patient, surgeon and procedure information was included into the dataset, resulting in a total of 36 variables. Table 3.4 provides a brief description of these attributes according to their type and meaning.

The results have shown that the variables having stronger influence in the model are the patient gender (#1), the patient's age (#2), the surgery priority (#3), the surgical procedure chosen (#11), the surgeon (#17) and the time estimation given by the surgeon (#35). In the context of the hospital all these variables are available at the time of surgery scheduling, in particular, a surgeon is allocated in advance to a patient and he is asked to provide an estimation of the surgery duration.

Table 3.4: Variables used in the prediction model

#	Type	Description
1	Nominal	Patient Gender
2	Numeric	Patient Age
3	Ordinal	Patient Priority
4	Numeric	Patient Waiting Time for Surgery
5	Nominal	Surgery Specialty Identification
6	Nominal	Surgery Month
7	Nominal	Surgery Weekday
8	Nominal	Surgery Shift
9	Nominal	Patient Diagnosed Disease
10	Numeric	Number of Interventions to be Performed
11	Nominal	Intervention Code 1
12	Nominal	Intervention Code 2
13	Nominal	Intervention Code 3
14	Numeric	Number of Surgeries to Date
15	Numeric	Number of Interventions to Date
16	Binary	If the patient has undergone surgery on other specialties
17	Nominal	Surgeon Identification
18	Nominal	Surgeon Gender
19	Numeric	Number of times the surgeon has dealt with this disease
20	Numeric	Number of times the surgeon has performed the main intervention
21	Binary	If the patient has other diagnosis
22	Binary	If the patient has any circulatory problem
23	Binary	If the patient has diabetes or renal problems
24	Binary	If the diagnosis is recidivist
25	Numeric	Duration of the last similar surgery from this surgeon
26	Numeric	Average Duration of the main procedure of this surgeon
27	Numeric	Standard Deviation of the main procedure of this surgeon
28	Numeric	Average Main Procedure Duration
29	Numeric	standard Deviation Main procedure duration
30	Numeric	Average duration of the surgery act on this combination of procedures
31	Numeric	Median duration of the surgery act on this combination of procedures
32	Numeric	Average total surgery duration
33	Numeric	Median total surgery duration
34	Numeric	Number of records with this combination of interventions
35	Numeric	Scheduled time by the Surgeon
36	Numeric	Surgery Real duration

The WEKA data mining software (Hall et al., 2009) was used for the estimation of surgery durations.

3.B Scheduling Model

The elective patient surgery advanced scheduling problem consists on selecting a sub set of surgeries from the waiting list and assigning them to specific time blocks across the planning week. The time blocks are previously defined and represent a period of time assigned to a surgical specialty on a given OR and day of week. Table 3.5 summarizes the notation used in this elective patient scheduling model.

Note that according to the parameter p_i, a given patient i has a priority to go under surgery proportional to the maximum number of days that he can wait for surgery without the hospital being penalized. p_i may take the value of one, two or three, depending on whether i refers to a normal patient, a high priority patient, or urgent patient.

Table 3.5: Variables used on the elective patient scheduling model

#	Notations
N	Set of Patients
R	Set of Operating Rooms
S	Set of Surgeons
D	Days of the week
T	Parts of the day (Morning or Afternoon)
s_i	Patient i surgeon
d_i	Patient i surgery estimated duration
w_i	Patient i waiting time
p_i	Patient i priority level
A_{rdt}	=1 in case operating room r is available in day d and time t, =0 otherwise
S_{sdt}	=1 in case surgeon s is available in day d and time t, =0 otherwise
ct	Clean up time constant
C	Shift capacity constant

3.B.1 Decision Variables

$$x_{idrt} = \begin{cases} 1 & \text{if surgery } i \text{ starts at period } t \text{ on day } d \text{ in room } r, \\ 0 & \text{otherwise.} \end{cases} \tag{3.1}$$

$$y_{sdrt} = \begin{cases} 1 & \text{if the surgeon } s \text{ is assigned to period } t \text{ on day } d \text{ in room } r, \\ 0 & \text{otherwise.} \end{cases} \tag{3.2}$$

3.B.2 Objective Functions

As mentioned before, there are three objective functions to be optimized. The first concerns the maximization of the number of surgeries scheduled as follows:

$$max\ f_1 = \sum_{i \in N} \sum_{d \in D} \sum_{r \in R} \sum_{t \in T} x_{idrt} \tag{3.3}$$

However, as more surgeries are performed, the utilization decreases since there is the need to prepare and clean up the operating rooms before a surgery, time we consider as waste. Thus, the following expression represents the maximization of the mean utilization of the operating rooms over a week span.

$$max\ f_2 = \frac{\sum_{i \in N} \sum_{d \in D} \sum_{r \in R} \sum_{t \in T} x_{idrt} d_i}{C \sum_{d \in D} \sum_{r \in R} \sum_{t \in T} A_{drt}} \tag{3.4}$$

Lastly, there is also the desire to diminish the patient waiting times for surgery. In order to do that we express the maximization of the waiting time "removed" from the waiting lists. In other words, the summation of the waiting times of scheduled patients. Since high priority patients wait less and have more urgency on being operated, we have weighted the waiting time with the patient's priority level as follows:

$$max\ f_3 = \sum_{i \in N} \sum_{d \in D} \sum_{r \in R} \sum_{t \in T} x_{idrt} w_i 10^{p_i} \tag{3.5}$$

3.B.3 Constraints

The available capacity of the operating rooms in terms of time, must be respected, no overtime is allowed, i.e.:

$$\sum_{i \in N} x_{idrt}(d_i + ct) \leq CA_{drt}, \forall d \in D, r \in R, t \in T \tag{3.6}$$

A patient can only be assigned to a room, day, part of the day, if the room is available for his specialty. Such condition is guaranteed by expression 3.7:

$$x_{idrt} \leq A_{drt}, \forall i \in N, d \in D, r \in R, t \in T \tag{3.7}$$

A surgery can only be scheduled if the surgeon in charge is available at the time.

$$y_{sdrt} \leq S_{sdt}, \forall s \in S, d \in D, r \in R, t \in T \tag{3.8}$$

In each day/shift, which represents a morning or afternoon, a surgeon can be scheduled for at most one OR. In other words, surgeons can not move to different operating rooms in the same working shift.

$$\sum_{r \in R} \sum_{t \in T} y_{sdrt} \leq 1, \forall s \in S, d \in D \tag{3.9}$$

Since we do not allow surgeons to change operating rooms in a morning or afternoon there must be a link between patients and surgeons, so that the latter is also fixed to a shift on an operating room.

$$x_{idrt s_i} \leq y_{sdrt}, \forall i \in N, d \in D, r \in R, t \in T \tag{3.10}$$

The final requirement expresses the domains of the variables:

$$y_{sdrt} \in \{0,1\}, x_{idrt} \in \{0,1\} \tag{3.11}$$

3.B.4 Further Results

Table 3.6: Optimization results - maximizing utilization

Instance	No. Surgeries	Utilization	Days Removed
GS	15	90.81%	4108
VS	23	87.65%	3003

Table 3.7: Optimizations results - minimizing waiting times (maximize days removed considering patient's priority)

Instance	No. Surgeries	Utilization	Days Removed
GS	32	82.46%	18249
VS	45	75.63%	10477

CHAPTER 4

NEW SOLUTION APPROACHES FOR THE SURGICAL CASES ASSIGNMENT PROBLEM: MIXED INTEGER PROGRAMMING VS. BIASED RANDOM-KEY GENETIC ALGORITHM

Fabrício Sperandio[1], José Fernando Gonçalves[2], José Borges[1], Bernardo Almada-Lobo[1]

Technical Report, 2014

Abstract This study addresses the surgical cases assignment problem (SCAP) appearing at large hospitals. The problem consists in generating a weekly surgery schedule assigning operating rooms (ORs), surgery dates and starting times to elective surgeries in the surgical waiting list, hence integrating

[1]CEGI – INESC TEC, Faculdade de Engenharia, Universidade do Porto, Portugal
[2]LIAAD – INESC TEC, Faculdade de Economia, Universidade do Porto, Portugal

advance and allocation scheduling. Admissible schedules are subject to surgeons and ORs capacity constraints as well as patient priority and waiting time rules. Due to long waiting lists and the scarcity of ORs, our aim is to maximize the number of scheduled surgeries as well as the utilization of ORs. Two alternative solution approaches, one exact and one approximate, are proposed and their respective results are compared. The first is based on mixed integer programming (MIP). In this model the problem is formulated as a scheduling problem with block synchronization using a continuous representation of time, which contributes to maximizing the ORs utilization. The MIP model is also compared with a model using a discrete representation of time, which is similar to the model proposed in Chapter 3. The second is a heuristic solution approach based on the biased random-key genetic algorithm (BRKGA). This is a population based approach which uses a vector of random numbers to represent each individual in the population and requires a decoding procedure to translate them into valid surgery schedules. This approach employs an efficient heuristic which keeps track of the resource available times and is able to translate every vector into a high quality solution. The alternative solution methods are compared using instances based on real data from a large hospital. Results show that the proposed MIP model, using a continuous representation of time, outperforms in terms of quality of solutions the model using a discrete representation of time in all instances. In its turn, the BRKGA outperforms the MIP in terms of quality of solutions in the majority of the test instances.

Keywords Surgery scheduling, Mixed integer programming, Genetic algorithm

4.1 Introduction

Healthcare expending continues to rise among OECD countries. In 2012, the overall healthcare expenditure across these countries accounted for 9.3% of GPD on average, higher than the 8.6% accounted before the global financial crisis of 2007–08 (OECD, 2014). Such rise is driven by an increasing demand for healthcare services which in turn is influenced by factors like: new and more expensive technology, ageing population and lifestyle issues (e.g. obesity). In this scenario, healthcare managers face a tough challenge to improve quality and efficiency, while preserving the sustainability of healthcare organizations. This paper is a contribution to the field of operations research and to society, as it promotes the efficient utilization of a core hospital resource, the operating theater (OT).

The OT accounts for up to 40% of hospital revenues and expenses (Gordon et al., 1988; HFM, 2003), making it one of the hospital's most important facilities. Its expenses are driven by a high consumption of human and material resources. Many surgeries are technically complex and require a range of people to be in one place working in harmony, including one or more surgeons, one or more anaesthetists, as well as special theatre nurses, assistants and technicians (Commission, 2003). Mayer et al. (2008) emphasize the importance of optimizing the utilization of such an expensive resource, citing the average cost of running one OT in NHS Scotland (National Health Service for Scotland) facilities to be £ 1.1 million per week. Furthermore, its operation has a direct impact on many other upstream and downstream resources. As a result, the OT is often called the heart of the hospital.

The literature on operations research applied to healthcare includes a high volume of studies tackling operating room (OR) planning and scheduling problems. The largest portion tackles problems in the operational decision level. The surgery scheduling problem in the operational decision level consists in selecting patients from the surgical waiting lists and assigning ORs, surgery dates and starting times to them usually over a one week planning horizon. Due to the complexity of the problem it is often decomposed into two sub-problems: advance and allocation scheduling problems. The advance problem consists in selecting patients from the waiting list and assigning surgery dates while the allocation problem consists in sequencing the surgeries within each day. The majority of the studies tackle each of these problems separately, although there is a trend to adopt integrated approaches. In this last case, in order to reduce the problem complexity, the reported approaches use a discrete representation of time.

This study aims to propose a new modelling approach for the integrated (advance and allocation) surgical cases assignment problem (SCAP) using a continuous representation of time, thus providing a more accurate representation of the problem and a potential higher resource utilization. The modelling challenge is to propose an exact yet efficient mathematical formulation of the problem. The proposed MIP model is inspired by efficient formulations for the travelling salesman problem (TSP), making an analogy between the cities of the TSP and the operating rooms a surgeon works in a given shift. A surgeon is allowed to change between ORs during the same day and working shift. This situation can increase utilization rates since the surgeon's turnover time, the time required for a surgeon to start a new surgery in a different OR, is generally lower than the required cleaning time between two consecutive surgeries in the same OR. In addition, this study aims to propose an original heuristic solution method aiming to find near optimal solutions within a reduced amount of time. The proposed approach is based on the biased random-key genetic algorithm (BRKGA)(Gonçalves and Resende, 2011) framework and on an efficient decoding procedure to translate each individual in the population into a high quality schedule. Finally, it aims to compare the two alternative approaches, a continuous

MIP model and a heuristic based on the BRKGA, analysing the quality of solutions and required computational times.

The performance of the proposed approaches is analysed from three different perspectives. First, the new MIP model is compared with a modification of the model presented in Chapter 3, which uses a discrete representation of time. The modified model is described in the appendix of this chapter. Second, the new MIP model is compared with the heuristic solution approach. Finally, the quality of solutions found by the two proposed approaches along the search progress is compared. All computational experiments were performed over instances generated with real data from a large hospital.

The remainder of this chapter is organized as follows. Section 4.2 reviews existing approaches for the surgery scheduling problem. Section 4.3 describes in detail the particular problem addressed in this paper. Section 4.4 introduces the two proposed approaches: the exact MIP model and the heuristic genetic algorithm. Section 4.5 describes the computational experiments designed to compare both approaches and presents the results. Finally, the last section highlights the main contributions of this paper and indicates some areas for future work.

4.2 Literature Review

4.2.1 Problem Perspective

The management of surgical services entails several complex decision problems. These problems are often classified into three decision levels: strategic, tactical and operational. The strategic level encompasses case mix and capacity planning problems. The first consists in determining the volume and type of surgeries to be performed by each specialty in the long term (1 to 5 years). The second consists in determining the number and capacity of resources dedicated to surgical services as well as their allocation. In the tactical decision level, two main different strategies are used: open scheduling and block scheduling. In the open scheduling strategy ORs are occupied by patients of any specialty. This strategy aims to maximize OR utilization rates. On the other hand, the block scheduling strategy requires to solve a master surgery scheduling (MSS) problem, which consists in determining the ORs reserved for each specialty in each day of the week and working shift. The resulting plan is a weekly timetable implemented in the medium term (6 to 12 months). This is the most used strategy, mainly in large hospitals where the use of a MSS is well established. Regarding the operational problem, in the open strategy it encompasses all specialties together, while in the block strategy it is subdivided among the specialties.

The problems arising in the operational decision level are classified into off-line and on-line scheduling problems. The off-line problem consists in selecting patients from the waiting lists and assigning ORs, surgery dates and starting times over a short term planning horizon (typically 1-week). The on-line problem consists in scheduling daily emergency and high priority cases as well as rescheduling previous elective cases. This study tackles the off-line surgery scheduling problem at the operational decision level, also known as surgical cases assignment problem. The reviewed papers target deterministic versions of the problem only, since, in order to address uncertainty a range of other solution approaches is required, which is out of the scope of this study. For a comprehensive review of surgery planning and scheduling problems see Cardoen et al. (2010a) and Guerriero and Guido (2011). In this literature review, a selected set of papers addressing the SCAP problem was reviewed from both problem perspective and solution perspective and the main characteristics such as decisions, objectives, constraints etc, was reported in Table 4.1 and Table 4.2. The first table focuses on the problem settings while the second on solution approaches.

The SCAP problem can be further decomposed into two sub-problems: advance and allocation scheduling problems. As mentioned in the previous section, the advance problem concerns the surgery dates while the allocation problem concerns the starting times. In most studies the OR assignment is part of the advance problem, however, in Ozkarahan (1995) it is part of the allocation problem with the advance problem consisting only in assigning a surgery date. In addition, some studies combine other decisions, such as: assigning available ORs (Roland et al., 2010; Fei et al., 2010), assigning ORs to specialties (Marques et al., 2012, 2014) and assigning surgeons to patients (Vijayakumar et al., 2013). Most studies tackle the advance and allocation problems separately but there is a growing number of integrated approaches.

The optimization objectives in SCAP problems are either related with resources or patients. Regarding resources, the main objectives are: maximize OR occupancy rates and minimize overtime and makespan. This last objective, along with the objective of minimizing the stay in recovery after closure time, are closely related with minimizing overtime, so that one can infer that even more studies aim to minimize overtime. It is worth noting that this is planned overtime since all studies consider deterministic surgery durations. In this case, there is often a trade-off between opening new ORs, keeping patients waiting and incurring overtime costs. Concerning human resources, Ogulata and Erol (2003) aim to balance the distribution of surgeries among surgeon groups and Meskens et al. (2013) aim to maximize the affinities among members of the surgical team. The patient related objectives include: maximizing the number of scheduled patients, minimizing the patients waiting time and the costs of keeping patients in the hospital waiting to be treated. In addition, Cardoen (2009); Cardoen et al. (2009) focus on particular patient groups, such as: high priority patients, children and patients with long travel distance. In spite of the aforementioned objectives, most approaches combine multiple objectives. This is most often achieved through an aggregated objective function.

With regard to the constraints, they are either related to physical resources, human resources or patients. In the first category, ORs are the main resource followed by post-anaesthesia care unit (PACU) and intensive care unit (ICU). Pham and Klinkert (2008) also consider the preoperative holding unit (PHU). In addition, few studies consider the availability of surgical materials, medical instruments and equipment. Also, Augusto et al. (2010) consider the availability of transporters, since this resource may be a bottleneck, mainly in the beginning of the day. Regarding human resources, the main surgeon in charge is the main resource. Most studies consider the surgeon's availability and a few consider workload and overtime limits. Typically, this surgeon is assigned to the surgical case during the waiting list registration phase. To the best of our knowledge, Vijayakumar et al. (2013) is the only study to assign surgeons to surgical cases during the scheduling phase. In addition, few papers have considered the other members of the surgical team, such as anaesthetists and nurses. Regarding patients, studies have considered constraints in the patient due date and admission date.

4.2.2 Solution Perspective

Regarding continuous and discrete representations of time, this literature review reveals that most studies represent time as discrete intervals. Fifteen minutes is the most common value used for the size of intervals in the computational experiments presented in the papers included in this literature review. Cardoen (2009); Cardoen et al. (2009) use a lower, 5 min intervals, but these studies focus on the allocation problem only (sequencing patients within a day), which is less complex than the integrated problem. The majority of the approaches which use a continuous representation of time tackle the allocation problem alone or decompose the overall problem into

Table 4.1: Summary table of the literature review: modelling Approaches

	Decisions						Objectives															Constraints																	
							Resources									Patients						Physical Resources									Human Resources					Patients			
Reference	A	B	C	D	E	F	A	B	C	D	E	F	G	H	I	A	B	C	D	E	F	A	B	C	D	E	F	G	H	I	A	B	C	D	E	A	B	C	D
Marques et al. (2012)	•	•		•			•										•					•	•								•	•				•			
Marques et al. (2014)	•	•		•			•															•	•								•	•				•			
Roland et al. (2010)		•		•	•				•		•											•	•						•						•		•	•	
Jebali et al. (2006)		•		•			•		•									•				•	•		•	•	•					•	•				•		
Fei et al. (2010)		•			•		•		•		•											•	•	•		•					•						•		
Guinet and Chaabane (2003)		•							•									•				•	•	•			•				•	•					•	•	
Chaabane et al. (2008)		•							•									•				•	•	•			•				•	•					•	•	
Riise and Burke (2011)	•	•		•					•							•						•									•							•	
Cardoen (2009)				•								•	•						•	•	•	•	•			•		•			•								•
Cardoen et al. (2009)				•								•	•						•	•	•	•	•			•		•			•								
Pham and Klinkert (2008)				•				•																	•														
Ogulata and Erol (2003)	•	•					•							•		•						•				•								•		•			
Augusto et al. (2010)				•				•	•													•								•									
Meskens et al. (2013)				•				•							•																	•				•			
Vijayakumar et al. (2013)	•	•		•		•											•										•								•	•			
Zhong et al. (2014)				•				•																															
Fei et al. (2008b)	•	•								•												•	•														•		
Ozkarahan (1995)			•	•			•		•																														
This paper	•	•		•			•	•														•	•	•						•	•	•				•	•		

Decisions

(A) Select patients
(B) Assign OR and date
(C) Assign OR only
(D) Assign starting time
(E) Select open ORs
(F) Assign surgeons to patients

Objectives

Resources

(A) Max OR occupancy rate
(B) Min makespan
(C) Min overtime costs
(D) Min total operating cost
(E) Min costs of opening ORs
(F) Min peak number of bed spaces used in PACU units
(G) Min stay in recovery after the closure time
(H) Balance the distribution of surgeries among surgeons groups
(I) Max affinities among members of the surgical team

Patients

(A) Min patients waiting time
(B) Max number of scheduled surgeries
(C) Min cost of keeping patients in hospital waiting to be treated
(D) Min sum of the starting times of surgeries performed on children
(E) Min sum of the starting times of surgeries performed on prioritized patients
(F) Min number of patients with long travel distance scheduled below a reference time

Constraints

Physical Resources

(A) Operating Rooms (OR)
(B) OR opening hours
(C) OR overtime limits
(D) Intensive care unit (ICU)
(E) Recovery room (PACU)
(F) Equipment
(G) Medical instruments
(H) Pharmaceutical products
(I) Transporters

Human Resources

(A) Surgeons
(B) Surgeon workload limits
(C) Surgeon overtime limits
(D) Surgeon groups
(E) Surgical team

Patients

(A) Patient priority
(B) Patient due date
(C) Patient admission date
(D) Cleaning time for infected patients

two sub-problems, each one approached independently. Zhong et al. (2014) and Ozkarahan (1995) address the integrated problem considering time as a continuous variable but use simple heuristics (longest processing time and shortest processing time) to get an approximate solution. To the best of our knowledge, Pham and Klinkert (2008) are the only to present an exact model using a continuous representation of time. The authors propose an extension of the job shop scheduling problem called multi-mode blocking job shop and conclude that the model can obtain (good) feasible solutions for only small to medium-sized instances. The integrated problem is known to be hard to solve, resulting in long running times, so that an optimal solution appears more as a reference solution important to evaluate the quality of the heuristic.

The high complexity of the SCAP problem makes researchers apply efficient search algorithms, relaxation approaches and search heuristics. Only studies which decompose the overall problem into more manageable subproblems do not rely on approximation algorithms. Among the exact search algorithms with proof of optimality we highlight: branch and bound, column generation, Dantzig–Wolfe decomposition, branch and price and the Hungarian algorithm. In addition, Lagrangian relaxation is used to find an approximate solution. Other solution methods applied to find approximate solutions are: iterated local search, genetic algorithms (GA) and tabu search. Finally, researchers have considered constructive and improvement heuristics.

Random-key genetic algorithms (RKGA) for solving sequencing problems were introduced by Bean (1994). The biased random-key genetic algorithm, proposed by Gonçalves and Resende (2011), is a slight modification of Bean's original method, differing in the way parents are selected for mating and how mating is carried out. Gonçalves et al. (2014b) compared biased and unbiased versions of RKGAs and concluded that the biased variant is faster.

Heuristics based on BRKGAs and RKGAs have already been applied with success on resource constrained project scheduling (Mendes et al., 2009; Gonçalves and Resende, 2011), resource constrained multi-project scheduling (Gonçalves et al., 2008) and job shop scheduling (Gonçalves et al., 2005) problems, which are similar to the SCAP problem. A detailed description of the BRKGA is provided in section 4.4.2.

4.3 Problem Description

The problem consists in assigning a surgery date, an operating room and a starting time to a set of elective patients in the waiting list, thus integrating advance and allocation scheduling. Each surgery of a patient has a pre-assigned surgeon, latest surgery date and estimated surgery duration. The objective is to maximize the number of scheduled surgeries as well as the average OR utilization rate. These are conflicting objectives, as OR cleaning times consume OR capacity and therefore are not considered to contribute to utilization rates. On one hand, when one aims to maximize the number of scheduled surgeries, shorter surgeries are preferred. On the other hand, when one aims to maximize utilization rates, longer surgeries are preferred instead, in order to avoid the setup time involved in cleaning activities. The following items describe the restrictions imposed to admissible surgery schedules.

1. OR cleaning time - Time after each surgery reserved for performing OR cleaning protocol activities, in order to setup the OR for the next event. The next surgery in the same OR can only start after the setup operation (that occurs in between surgeries) is completed;

2. Surgeon turnover time - A surgeon is allowed to have scheduled surgeries in more than one OR in the same shift as long as an offset between

Table 4.2: Summary table of the literature review: solution approaches

	Approach		Planning Horizon		Strategy		Formulation				Solution Approach														Time Modelling		
Reference	A	B	A	B	A	B	A	B	C	D	A	B	C	D	E	F	G	H	I	J	K	L	M	N	A	B	C
Marques et al. (2012)		•		•	•		•		•			•	•												•		15
Marques et al. (2014)		•		•	•			•	•				•												•		15
Roland et al. (2010)		•		•				•		•	•	•													•		15
Jebali et al. (2006)		•		•				•		•			•													•	
Fei et al. (2010)	•			•	•			•		•	•				•			•								•	
Guinet and Chaabane (2003)	•			•				•		•									•						•		60
Chaabane et al. (2008)	•			•		•		•		•									•						•		60
Riise and Burke (2011)		•		•		•		•		•										•					•		15
Cardoen (2009)	•		•			•		•		•			•												•		5
Cardoen et al. (2009)			•			•		•		•					•										•		5
Pham and Klinkert (2008)	•	•	•			•	•		•		•					•										•	
Ogulata and Erol (2003)	•			•		•				•	•														NA	NA	NA
Augusto et al. (2010)		•																		•				•		10	
Meskens et al. (2013)	•		•			•	•																•		•		30
Vijayakumar et al. (2013)		•		•		•	•				•			•											•		15
Zhong et al. (2014)		•		•		•	•								•											•	
Fei et al. (2008b)	•			•	•		•										•							•		•	
Ozkarahan (1995)	•		•			•	•															•				•	
This paper		•		•		•		•		•	•	•														•	

Approach
(A) Decomposed
(B) Integrated

Planning Horizon
(A) 1 day
(B) 1 or 2 week

Strategy
(A) Open scheduling
(B) Block scheduling

Formulation
(A) Single-objective
(B) Multi-objective
(C) Independent
(D) Aggregated

Solution Approach
(A) Linear Programming (LP)
(B) Genetic Algorithms (GA)
(C) LP + Heuristics
(D) Heuristics
(E) Column Generation
(F) Branch and Price
(G) Branch and Bound
(H) Hybrid (GA + Tabu Search)
(I) Hungarian Method
(J) Iterated Search
(K) Lagrangian Relaxation
(L) Goal Programming
(M) Constraint Programming
(N) Dantzig–Wolfe Decomposition

Time Modelling
(A) Discrete
(B) Continuous
(C) Size of Intervals (min)

two consecutive surgeries of the same surgeon is guaranteed. This offset is called turnover time and denotes the required time for a surgeon to change from one OR to another after finishing a surgery;

3. OR time capacity - Each OR has a predefined time capacity on each shift. Naturally, the summation of the scheduled surgeries durations and setup times within each OR and shift must not exceed this predefined capacity. Furthermore, as overtime is not allowed, a surgery must not be scheduled to end after the OR closing time.

4. Surgeon availability and working limits - Each surgeon may be or may be not available to operate in a given shift of a certain day. If a surgeon is not available none of his/her patients must be scheduled for that period. Moreover, surgeons are subject to working limit constraints in terms of the number of working shifts for week. Supposing that these working limits are not guarantee by the surgeon availability itself.

5. Patient priority and waiting time rules - Each patient in the elective surgery waiting list has a predefined priority and a current waiting time. In some countries, there are maximum waiting times established for each priority level. Each surgery in the national health service should respect these times. Moreover, there are maximum scheduling times established for each priority, it means that when a patient reaches the maximum scheduling time he must be scheduled, with the surgery date subject to the maximum waiting time constraint.

4.4 Methodology

4.4.1 Exact Solution Approach: Mixed Integer Programming Model

The first approach is a mixed integer programming (MIP) model which uses a continuous representation of time. The aim is to determine the scheduled patient surgeries in a certain planning horizon, and the respected timings. A surgeon is allowed to perform one or more surgeries within each period of work. A period of work is the time between the start of the first surgery and the end of the last consecutive surgery of surgeon in a shift and OR. Once the surgeons' periods of work are determined by the model, the start and end times of each specific patient are assigned using a simple heuristic. The following paragraphs describe the model in detail. The idea is to define the sequence of surgeons in each OR avoiding potential overlaps between periods of work of the same surgeon in different ORs. The model uses three groups of decision variables: one to decide which patients to schedule in each shift and OR; one to decide on the sequence of surgeons within each shift and OR; and one to determine the start and end times of surgeons in each shift and OR. It is assumed that each patient is waiting only for one surgery and therefore hereafter the terms patient and surgery may be used interchangeably.

We start by introducing the necessary notation:

Sets and indices

I	set of patients (index i)
J	set of working shifts (index j)
K	set of operating rooms (index k)
K_j	set of available ORs in shift j
S	set of surgeons (index s)
I_s	set of patients of surgeon s (index i)
H	set of weeks in the planning horizon (index h)
J_h	set of working days in a given week h (index j)
$I_{maxsched}$	set of patients with maximum scheduling time within the planning horizon
$I_{maxswait}$	set of patients with maximum waiting time within the planning horizon

Parameters

d_i	estimated duration of patient's i surgery
s_i	surgeon in charge of patient's i surgery
max_i	maximum waiting time of patient's i surgery
c_{jk}	available capacity in shift j of OR k
a_{js}	1, if surgeon s is available in shift j; 0, otherwise
day_j	day of shift j
α	weight of the number of scheduled surgeries in the objective function
β	weight of the average OR utilization rate in the objective function
γ	best number of scheduled surgeries
δ	best average OR utilization rate
ct	OR cleaning time
tt	surgeon turnover time
C	total OR capacity
ms	maximum number of shifts per week

Decision variables

$$X_{ijk} = \begin{cases} 1, \text{if patient } i \text{ is scheduled for shift } j \text{ and OR } k \\ 0, \text{otherwise} \end{cases}$$

$$Y_{jkss'} = \begin{cases} 1, \text{if surgeon } s \text{ operates after surgeon } s' \text{ in shift } j \text{ and OR } k \\ 0, \text{otherwise} \end{cases}$$

$$Z_{jks} = \begin{cases} 1, \text{if surgeon } s \text{ is the first to operate in shift } j \text{ and OR } k \\ 0, \text{otherwise} \end{cases}$$

$$W_{jks} = \begin{cases} 1, \text{if surgeon } s \text{ is the last to operate in shift } j \text{ and OR } k \\ 0, \text{otherwise} \end{cases}$$

$$V_{jkk's} = \begin{cases} 1, \text{if surgeon } s \text{ operates in OR } k' \text{ after operated in OR } k \text{ in shift } j \\ 0, \text{otherwise} \end{cases}$$

$$\mu_{jks}^{start} = \text{starting time of surgeon } s \text{ in OR } k \text{ and shift } j$$

$$\mu_{jks}^{end} = \text{end time of surgeon } s \text{ in OR } k \text{ and shift } j$$

Throughout the exposition, i denotes a patient, j denotes a shift, which is a combination of a given day in the planning horizon and a working shift (morning or afternoon), k denotes an operating room and s denotes a surgeon. The weights α and β define the search directions and are normalized, i.e. $\alpha + \beta = 1$.

Regarding the decision variables, the binary variable X pertains to the patients and is used for selecting which patients are scheduled and assigned to the respective shifts and ORs, the binary variables Y, Z, W and V relate to the surgeons and are used for designating the sequence in which the surgeons work in a given shift and OR, and real variables μ^{start} and μ^{end} keep track of the start and end times of each surgeon in a given shift and OR. It is assumed here that each surgeon can work at most one period of work in each shift and OR. It is an assumption made by the authors in favour of the efficiency of the model.

Objective function

Expression (4.1) denotes the objective function which maximizes the number of scheduled surgeries and the average ORs utilization rate. However, as these are competing goals, preferences for objectives are *a priori* declared to form a weighted linear scalarizing function, used to aggregate several objectives into a single one. The value of each objective is normalized based on the maximum values both can take (γ and δ), to prevent the magnitude of each measure to bias the final value of the function, yielding a non-dimensional objective function value.

The objective function value must be minimized as the greater the number of scheduled patients and the average OR utilization rate the lower the function value.

$$min\ F = \alpha \cdot \frac{\gamma - \sum_{i \in I} \sum_{j \in J} \sum_{k \in K} X_{ijk}}{\gamma} + \beta \cdot \frac{\delta - \dfrac{\sum_{i \in I} \sum_{j \in J} \sum_{k \in K} X_{ijk} \cdot d_i}{C}}{\delta} \tag{4.1}$$

Constraints

Constraints are grouped into three categories, related to patients, to surgeons and to time periods. Inequality (4.2) prevents a patient from being scheduled more than once. The surgeries' due dates are defined according to patients' priority and waiting time rules defined in the waiting list manual. Inequality (4.3) expresses the shifts capacity constraint. The summation of all surgery durations and cleaning times in a given shift must be lower than or equal to the capacity of the OR in this shift. These constraints also ensure that each specialty uses only the ORs available to it according to the master surgery schedule. The ORs with a capacity greater than zero are considered available. Note that the model works with only a single specialty.

$$\sum_{j \in J} \sum_{k \in K} X_{ijk} \leq 1, \forall i \in I \tag{4.2}$$

$$\sum_{i \in I} X_{ijk} \cdot (d_i + ct) \leq c_{jk}, \forall j \in J, \forall k \in K \tag{4.3}$$

Expression (4.4) states that surgeries with a maximum scheduling time lower than the planning horizon must be scheduled. Expression (4.5) states that surgeries with a maximum waiting time lower than the planning horizon must be scheduled and inequality (4.6) states that the surgery day must be lower than the maximum waiting time. In the model the maximum scheduling times and maximum waiting times are defined in days relative to the beginning of the planning horizon. The absolute values are defined according to the patient's priority and waiting time rules and can be find in the waiting list manual (Ministério da Saúde, 2011). The aforementioned constraints related specifically to the patients, as the other are related to the surgeons.

$$\sum_{j \in J} \sum_{k \in K} X_{ijk} = 1, \forall i \in I_{maxsched} \tag{4.4}$$

$$\sum_{j \in J} \sum_{k \in K} X_{ijk} = 1, \forall i \in I_{maxwait} \tag{4.5}$$

$$\sum_{j \in J} \sum_{k \in K} X_{ijk} \cdot day_j \leq max_i, \forall i \in I_maxwait \tag{4.6}$$

Thus, the following sets of constraints aims to determine the sequence of surgeons working in each shift and OR. Expressions (4.7) and (4.8) define the surgeons who are the first and the last to operate in each shift and OR, respectively. Inequality (4.9) states that in one shift a surgeon must be the first in at most one OR, otherwise there would be an overlap between surgeries of this particular surgeon as all OR sessions start at the beginning of the shift. Inequality (4.10) ensures that in every shift and OR, a surgeon is either the first one or comes after another surgeon in the sequence.

Similarly, according to expression (4.11), a surgeon is either the last one or precedes another one. These two constraints ensure that a surgeon appears only once in the sequence of each OR and shift, i.e. only one bucket of consecutive work. Expression (4.12) avoids circular references in the sequence of surgeons. Finally, expression (4.13) is the flow equation, which specifies the balance of the inflow and outflow of position for each of the surgeons. It ensures consistency between expressions (4.10) and (4.11) ensuring that there is a link (via variables Y) between all the surgeons in the sequence.

$$\sum_{s \in S} Z_{jks} = 1, \forall j \in J, \forall k \in K \tag{4.7}$$

$$\sum_{s \in S} W_{jks} = 1, \forall j \in J, \forall k \in K \tag{4.8}$$

$$\sum_{k \in K} Z_{jks} \leq 1, \forall j \in J, \forall s \in S \tag{4.9}$$

$$Z_{jks} + \sum_{s' \in S} Y_{jkss'} \leq 1, \forall j \in J, \forall k \in K, \forall s \in S \tag{4.10}$$

$$W_{jks} + \sum_{s' \in S} Y_{jks's} \leq 1, \forall j \in J, \forall k \in K, \forall s \in S \tag{4.11}$$

$$Y_{jkss} = 0, \forall j \in J, \forall k \in K, \forall s \in S \tag{4.12}$$

$$Z_{jks} + \sum_{s' \in S} Y_{jkss'} = W_{jks} + \sum_{s' \in S} Y_{jks's}, \forall j \in J, \forall k \in K, \forall s \in S \tag{4.13}$$

The next set of constraints aims to assign the start and end times of each surgeon in each shift and OR according to the scheduled patients (determined by variable X) and the sequence of surgeons (determined by variables Y,Z,W,V). Expression (4.14) enforces the starting time of the first surgeon to take place at the beginning of the shift. Similarly, constraint (4.15) sets the ending time of each surgeon on each shift and OR, taking into account the starting time, the duration of all the surgeries performed by him and the cleaning times to set up the OR. A surgeon's finishing time is important to check whether the surgeon can afterwards start a surgery in another OR. Moreover, expression (4.16) declares that all end times must be within the capacity (in time units) of the OR in that particular shift.

$$\mu_{jks}^{start} \leq c_{jk} - (c_{jk} \cdot Z_{jks}), \forall j \in J, \forall k \in K, \forall s \in S \quad (4.14)$$

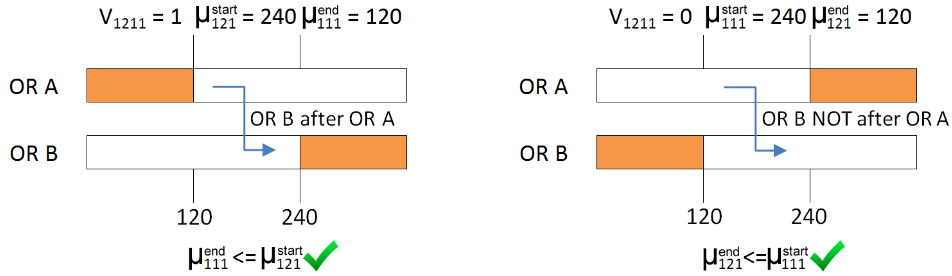
$$\mu_{jks}^{start} + \sum_{i \in I_s} X_{ijk} \cdot d_i + (\max\{1, \sum_{i \in I_s} X_{ijk}\} - 1) \cdot ct \leq \mu_{jks}^{end}, \forall j \in J, \forall k \in K, \forall s \in S \quad (4.15)$$

$$\mu_{jks}^{end} \leq c_{jk}, \forall j \in J, \forall k \in K, \forall s \in S \quad (4.16)$$

Constraint (4.17) aims to eliminate any subtours in the sequence, preventing the occurrence of two or more disconnected groups of surgeons, and guaranteeing that there is a link between all surgeons. The inequality implies that if a surgeon s comes after another surgeon s' then the end time of s' (the previous), denoted by $\mu_{jks'}^{end}$, must be lower than or equal to the start time of s (the next), denoted by μ_{jks}^{start}. In other words, if one surgeon comes after another, then the previous surgeon must end before the beginning of the next.

$$\mu_{jks'}^{end} + (Y_{jkss'} - 1) \cdot (C + ct) \leq \mu_{jks}^{start}, \forall j \in J, \forall k \in K, \forall s \in S, \forall s' \in S \quad (4.17)$$

The timings of each surgeon in different ORs must now be synchronized. The following constraints aim to avoid overlaps between working periods of a given surgeon in different ORs within the same shift. Expression (4.18) states that if a surgeon s works in OR k after having worked in OR k' in a given shift j, denoted by $V_{jkk's}$, then the surgeon's end time in OR k' (the previous) must be lower than the surgeon's start time in OR k (the next), denoted by μ_{jks}^{start}. In contrast, the inverse must also hold according to expression (4.19), if a surgeon s does not work in OR k after having worked in OR k', denoted by $V_{jkk's}$, then the surgeon's start time in OR k', denoted by $\mu_{jk's}^{start}$, must be greater than or equal to the end of the surgeon's working period in OR k, denoted by μ_{jks}^{end}. The two constraints work together, the first validates the situation in which the surgeon works in a given OR after



(a) Surgeon operates in OR A after OR B

(b) Surgeon does not operate in OR A after OR B

Figure 4.1: An illustrative example of the synchronization constraints

another and the second the situation in which he does not work, to determine the working periods in parallel ORs avoiding overlaps.

$$\mu^{end}_{jk's} + tt \leq \mu^{start}_{jks} + c_{jk} \cdot (1 - V_{jkk's}), \forall j \in J, \forall k \in K, \forall k' \in K, \forall s \in S \tag{4.18}$$

$$\mu^{end}_{jks} \leq \mu^{start}_{jk's} + c_{jk} \cdot V_{jkk's}, \forall j \in J, \forall k \in K, \forall k' \in K, \forall s \in S \tag{4.19}$$

Figure 4.1 shows an illustrative example to support the description of the synchronization constraints. Let $J = \{1\}$ be the set of shifts, $K = \{A, B\}$ be the set of ORs and $S = \{1\}$ be the set of surgeons. Figure 4.1(a) shows a case in which surgeon 1 operates in OR B after having operated in OR A, therefore the surgeon's start time in OR B must be greater than the surgeon's end time in OR A. Figure 4.1(b) shows the opposite, when surgeon 1 does not work in OR B after having worked in OR A the surgeon's start time in OR A must be greater than the surgeon's end time in OR B.

The next constraints link the sequence of surgeons to the scheduled surgeries. Expression (4.20) states that if a patient i is scheduled in a shift j and OR k then the surgeon responsible for this operation, denoted by s_i, must be the first to operate or come after another surgeon. In contrast, inequality (4.21) states that if a surgeon s does not have any scheduled patient in shift j and OR k then he must not appear in the sequence. In this expression,

M denotes a big number, greater than the highest possible value for the summation of variables Z, W and Y for this particular surgeon.

$$X_{ijk} \leq \sum_{s' \in S} Y_{jks_is'} + Z_{jks}, \forall i \in I, \forall j \in J, \forall k \in K \quad (4.20)$$

$$Z_{jks} + W_{jks} + \sum_{s' \in S} (Y_{jkss'} + Y_{jks's}) \leq \min\{1, \sum_{i \in I_s} X_{ijk}\} \cdot M, \forall j \in J, \forall k \in K, \forall s \in S \quad (4.21)$$

The output of the model is the set of surgeries scheduled for each shift and OR as well as the sequence of surgeons in each shift and OR and their respective start and end times. In order to determine the starting time of surgeries one must iterate over the sequence of surgeons from the first surgeon in each shift and OR, through each surgeon after him, until the last. Algorithm 1 shows the two functions used in this process. The procedure starts by calling the function generateSchedule which iterates through all shifts and ORs and if the OR is available, iterates through the set of surgeons until it finds the first in the sequence. The next step (line 8), is to print the list of patients of this surgeon, which consists in iterating through the list of patients and for each of the surgeon's patients scheduled for this particular shift and OR, print its starting time. Note that the starting times of patients are relative to the beginning of the surgeon working period, the starting time of the first patient is equal to μ_{jks}^{start} and the starting time of the following is equal to the duration of the previous plus the OR cleaning time. Next (line 16), the procedure verifies if the current surgeon is the last to operate and in this case, it returns to the calling function, otherwise, it finds the next surgeon in the sequence and calls function ListPatients again, recursively (line 21).

Figure 4.2 shows a sample schedule generated by the proposed MIP model. It is a weekly schedule for the Neurosurgery specialty. This schedule has 5 days, 2 shifts each day and 2 ORs. Both ORs are closed on Saturday afternoon.

Algorithm 1: Algorithm for generating a schedule from a solution of the MIP

```
function GenerateSchedule() begin
    for all $j$ in $J$ do
        for all $k$ in $K$ do
            if $c_{jk} > 0$ then
                for all $s$ in $S$ do
                    if $Z_{jks} = 1$ then
                        ListPatients($j$,$k$,$s$)
                        break

function ListPatients($j$, $k$, $s$)
begin
    startTime ← $\mu_{jks}^{start}$
    for all $i$ in $I$ do
        if $s_i = s$ and $X_{ijk} = 1$ then
            PrintPatient($i$,$j$,$k$,startTime)
            startTime=startTime+$d_i$+$ct$
    if $W_{jks'} = 1$ then
        return
    else
        for all $s'$ in $S$ do
            if $Y_{jks's} = 1$ then
                ListPatients($j$,$k$,$s'$)
```

Figure 4.2: Sample Neurosurgery schedule generated by the MIP model

In the picture, each box represents a surgery, the different colors represent different surgeons, and the numbers inside each box mean the surgeon Id (between parenthesis) and start and end time of each surgery. The times are relative to the start of the shift. This schedule has 56 scheduled surgeries and 82.4% of average OR utilization rate taking into account an OR cleaning time of 17 minutes after each surgery and no surgeon turnover time. In this case, the turnover time, which is the required time for a surgeon to switch between ORs, is included in surgery duration. It is worth noting that there are situations in which a surgeon is able to start a new surgery in a different OR before starting the next surgery in the same OR, therefore saving time in which the surgeon would otherwise be idle. These situations are signed with a circle.

4.4.2 Heuristic Solution Approach: Biased Random Key Genetic Algorithm

4.4.2.1 General Genetic Algorithm Description

In this section, as an alternative to the exact approach proposed in the previous section, a heuristic approach based on the biased random-key genetic algorithm (Gonçalves and Resende, 2011) is proposed. Genetic algorithms (Holland, 1975; Goldberg, 1989) are part of a group of nature inspired algorithms based on the concept of natural selection, or survival of the fittest, used to find near-optimal solutions for optimization problems. They are population based algorithms which evolve a set of individuals over a number of generations. Each individual represents a solution for the optimization problem, in our case a surgery schedule. Moreover, each individual has an associated chromosome that encodes a solution. Chromosomes are strings of genes and the value in each gene is an allele. In general, alleles take binary or real values.

In random-key genetic algorithms (RKGAs) (Bean, 1994) each chromosome is encoded as a vector of random-keys and each allele is a random number between 0.0 and 1.0. Figure 4.3 shows a sample RKGA chromosome which indirectly represents a surgery schedule. In this representation, each allele is a random number corresponding to a patient in the waiting list. A decoding procedure, or simply decoder, is required to translate each chromosome into a solution in order to compute the associated performance metrics. In our case, the performance metrics corresponds with the number of scheduled surgeries and the average OR utilization rate. It is worth mentioning that the decoder efficiency plays an important role in the overall algorithm performance as it consumes most of the computational time.

As population based heuristics, GAs evolve populations of solutions through means of recombination and mutation. Recombination consists in selecting two parents from the population and copying sequences of genes from both of them into a new individual, a procedure called crossover. In particular, the proposed GA uses a parametrized uniform crossover (Spears and Jong, 1991). On the other hand, mutation aims to introduce diversity into the population and escape entrapment in local optima. In the case of RKGAs it is achieved by generating completely new individuals, called mutants, and

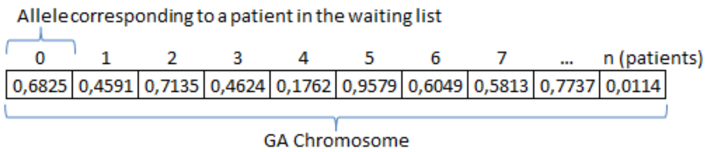


Figure 4.3: Sample chromosome indirectly representing a surgery schedule

introducing them into the populations.

RKGAs use an elitist strategy to evolve populations of solutions over generations. In this strategy, after decoding individuals and computing fitness values (the associated performance metrics), the best individuals are labelled as ELITE. These individuals remain in the population in the next generation as a way to preserve good genes. The biased random-key genetic algorithm differs from standard RKGA in the way individuals are selected for recombination. In a BRKGA, instead of randomly selecting two individuals from the entire population, each new individual is generated by combining one individual from the ELITE part of the population and one from the NON-ELITE part, or from the entire population. This increases the probability of good individuals passing their characteristics to future generations.

BRKGAs are based on a generic metaheuristic framework. Figure 4.4 shows an overview of the BRKGA optimization process. Note that this framework makes a clear distinction between the problem dependent and independent parts of the process. The problem independent part includes initialization, selection, recombination and mutation procedures, which are similar among other optimization problems. The problem dependent part encompasses the decoding procedure. This procedure is crucial for the algorithm performance, since it consumes a large portion of the overall computational time. In this paper, we propose a procedure to decode a vector of random-keys into a valid surgery schedule based on lists of available resource time slots. This procedure is able to generate good schedules both in terms of number

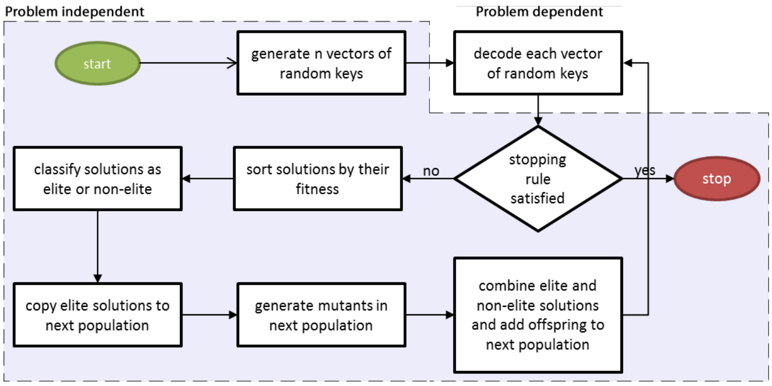
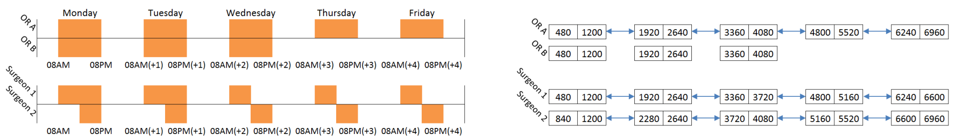


Figure 4.4: Flowchart of the BRKGA framework

of scheduled surgeries as in terms of average OR utilization rate.

Figure 4.5(a) shows a conceptual view of the decoding procedure. The idea is to keep track of the resources availability periods. For instance, consider two ORs (A and B), two surgeons (1 and 2) and a one week planning horizon. In Figure 4.5(a), the highlighted areas denote the available periods of each resource over the planning horizon. It is worth noting that OR B is not available on Thursday and Friday, as well as the two surgeons have distinct available periods. These patterns are directly mapped from the master surgery schedule, which denotes the ORs assigned to each specialty over the week, and from the staff roster, which shows staff working shifts. On the other hand, Figure 4.5(b) shows the same availability periods represented in terms of data structures. The numbers within the cells represent the start and end time of each period in minutes. In this case, lists of time periods (start and end time) represented in minutes from the beginning of the planning horizon until the end. The decoder works using the chromosome of random-keys to determine the scheduling sequence and the lists of available periods to find a time in which both surgeon and OR are free. The following paragraph describes the procedure in detail.

The following steps are used for decoding a vector of random-keys into a



(a) Conceptual view of resource available periods

(b) Lists used to implement the concept of resource available periods

Figure 4.5: Resource available periods from two different perspectives: conceptual and implementation

valid surgery schedule. The sequence of steps is illustrated in Figure 4.6.

1. **Initialize available times** - Creates the data structures to support the procedure. It creates the lists of available periods, as illustrated in Figure 4.5(b), for each OR and surgeon based on the pre-defined OR capacity and surgeon availability;

2. **Sort patients by random-keys** - Sorts the chromosome by ascending order of the random-key in each gene. The resulting sorted vector determines the sequence in which patients are evaluated in the next step, e.g. patients with lower random-keys are evaluated first and patients with higher are evaluated last;

3. **Iterate patients** - Evaluates each patient according to the sequence encoded in the chromosome. If there are no more patients to evaluate, then the procedure ends.

4. **Find surgeon starting time** - Searches the list of available periods of the surgeon responsible for the current surgery and finds the first period that fits the current surgery duration plus cleaning time. If an available period is found, goes to step (5) to search for an OR, otherwise returns to the previous step to evaluate the next patient, because the current patient will not be scheduled due to a lack of time of the responsible surgeon;

5. **Check available OR** - Iterates through the list of available ORs and searches for a time period that fits the current surgery duration plus cleaning time with the surgery starting exactly at the starting time of the surgeon available period defined in the previous step. If an available time is found, the procedure goes to step (5.1) to schedule the surgery, otherwise goes to step (6) to search for an available time in the future;

    5.1. **Schedule surgery** - Updates the output surgery schedule with the current surgery, day, shift, OR and starting time;

    5.2. **Update available time** - Updates the list of available periods of the surgeon responsible for the current patient and the selected OR. In the surgeon case, the surgery duration plus the turnover time must be subtracted from the time period in which the surgery was scheduled, meaning the surgeon is not available at this time. In the case of the OR, it is the surgery duration plus the cleaning time, meaning the selected OR is not available from the beginning of the surgery until the end of the cleaning time.

6. **Find new starting time** - Finds the first OR time period that fits the surgery duration from the current surgeon available time until the end of the planning horizon and returns to step (4) to find a new surgeon available time from this point onwards.

The decoding procedure is able to translate every chromosome into a near feasible solution. The restrictions related to patients' priority and waiting time rules as well as surgeons' workload are not guaranteed. In order to tackle this issue we calculate all the metrics associated with a schedule and penalize the violations in the fitness function. Once we have a schedule as a result of the decoding process we compute the following metrics: number

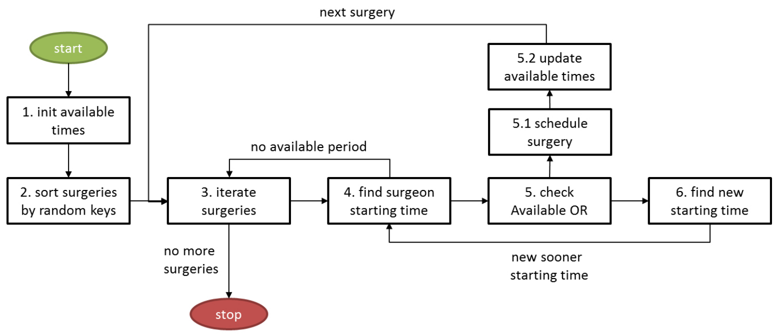


Figure 4.6: Flowchart of the decoding procedure

of scheduled surgeries, average OR occupancy rate, number of violations of surgery due date, number of violations of maximum scheduling date, total deviation from the limit number of working shifts per week. The final objective function is similar to the one used for the models in this chapter, only with the additional terms for the waiting list's violations and surgeon's workload. The decoder is not able to guarantee these problem constraints are respected, therefore, we address them in the objective function. The surgery's due date is defined as a function of the patient's maximum waiting time (time between the day a patient enters the waiting list and the day the surgery is performed) according to the Portuguese legislation and the scheduling date is the maximum time a patient can be in the waiting list without be scheduled for a surgery (time between the day a patient enters the waiting list and the day a surgery date is assigned) (Ministério da Saúde, 2011).

4.4.2.2 Local Search and Chromosome Correction

A local search procedure is performed after the decoding procedure to further enhance the quality of solutions. It uses the decoder supporting data structures to find available time periods in the ORs and to try to switch the surgeries scheduled immediately before and after such available periods by unscheduled surgeries with a larger size. For each available time period, the procedure evaluates all possible movements (changing one surgery for another), ranks them by the benefit (improvement of the objective function) they provide and implement the change that improves the objective function the most. The computational experiments show that these small changes are effective in enhancing the quality of solutions. They enable the algorithm to quickly improve the quality of solutions in particular cases, what would require several generations in the standard evolution process.

After the local search, the chromosome associated with each solution in the population is corrected to represent the actual order in which surgeries are scheduled in the solution. Hence, the local search changes will not be required in the next time the chromosome is evaluated.

4.5 Computational Experiments

4.5.1 Test Instances

Test instances are based on real data provided by a large hospital. There are 10 different surgical specialties (vascular surgery, oral and maxillofacial surgery, neurosurgery, ophthalmology, orthopaedics, urology, otorhinolaryngology, general surgery 1, general surgery 2, general surgery 3) and 2 different sets of 6 instances each for each specialty, summing up 120 instances. These two sets differ in the total number of available ORs. The first contains instances with the same number of ORs in use at the hospital (regular size instances) and the second contains twice this number, simulating a larger size hospital or a capacity expansion at the same hospital (large size instances). Within each set, instances differ by the number of patients and the length of the planning horizon.

Algorithm 2 illustrates the procedure used for generating the test instances.

The sets of parameters used in this procedure are the following: $SP = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, $IS = \{regular, large\}$, $H = \{1, 2\}$, $CM = \{2, 4, 6\}$. In this list, SP represents the set of specialty identifiers, IS the set of instance sizes, H the set of planning horizons (in weeks) and CM the set of capacity multipliers. The capacity multipliers are used for defining the number of patients in each instance, according to the following steps of Algorithm 2. First, the procedure starts by iterating through the set of specialties and getting the time blocks associated with each specialty (a time block denotes a working shift in an OR). This initial set of time blocks corresponds to a standard instance (regular size and 1 week planning horizon). Second, the algorithm iterates through the set of instance types and, in case of large size instances, duplicates the number of time blocks. Third, it iterates through the set of planning horizons and, in case of more than one week, generates new time blocks for the additional weeks. Next, it iterates through the set of capacity multipliers and multiplies each value by the total capacity of the current time blocks, setting an auxiliary variable to represent the maximum value for the sum of surgery durations in this instance. Finally, it iterates through the set of patients in the waiting list, adding them to the current instance and subtracting the auxiliary variable. Line 19 shows that when the value of the auxiliary variable reaches zero, the procedure stops and the current instance is ready.

Table 4.3 and Table 4.4 show the characteristics of the generated regular and large size instances, respectively. In both cases the estimated duration of surgeries is deterministic and based on median values of historical data. The last two columns show the number of patients in each instance whose maximum scheduling time or maximum total waiting time were reached or are within the planning horizon. It is worth mentioning that some of the instances were duplicated due to the lack of patients in the waiting list of each specialty and were excluded from the final analysis, resulting in 96 different instances.

Algorithm 2: Procedure used for generating the testing instances

```
function GenerateInstances(SP,IS,H,CM)
begin
    for sp ∈ SP do /* for each specialty */
        SP_tb ⟵ GetTimeBlocks(sp)
        for is ∈ IS do /* for each set of instance types */
            if is == large then
                SP_tb ⟵ DuplicateTimeBlocks(SP_tb, is)
            for h ∈ H do /* for each planning horizon */
                if h > 1 then
                    SP_tb ⟵ GenerateNewWeeks(SP_tb, h)
                for cm ∈ CM do /* for each capacity multiplier */
                    totalCapacity ⟵ GetTotalCapacity(SP_tb) · cm
                    SP_i ⟵
                    GetPatientsSortedByPriorityAndWaitingTime(sp)
                    for i ∈ SP_i do /* for each patient in the waiting
                    list */
                        if totalCapacity > 0 then
                            AddPatientToInstance(i)
                            totalCapacity ⟵ totalCapacity − (d_i + ct)
                        else
                            break
```

4.5.2 Implementation Details

Both solution methods were coded in C++ and compiled using g++ (GCC) 4.4.7 20120313 (Red Hat 4.4.7-4) with "-O3" and "-fopenmp" compiler options. The exact models use the IBM ILOG CPLEX Optimization Studio V12.4 libraries through the Concert Technology. Further, the GA is based on the application programming interface (API) for the algorithmic framework of biased random-key genetic algorithms, brkgaAPI, presented by Toso and Resende (2014). All computation experiments are performed on machines running Scientific Linux 6 (SL6) distribution and equipped with Intel Xeon Processor E5-2650 CPUs (2 GHz). The number of parallel threads is limited

Table 4.3: Regular size instances: problem instances generated based on real data

Regular Size Instances									
Specialty	Planning Horizon	Capacity Multiplier	No. Patients	No. Time Blocks	No. Surgeons	Avg. Duration	Max. Schedule Date	Max. Surgery Date	Instance Group
1. Vascular surgery									
1.1	1	2	91	9	15	53.74	74	22	1
1.2	1	4	214	9	15	43.52	77	22	1
1.3	1	6	346	9	15	39.22	77	22	3
1.4	2	2	214	18	15	43.52	102	23	1
1.5	2	4	473	18	15	37.79	102	23	3
1.6	2	6	721	18	17	36.95	102	23	3
2. Oral and maxillofacial surgery									
2.1	1	2	32	3	11	47.69	1	0	1
2.2	1	4	75	3	12	38.17	1	0	1
2.3	1	6	115	3	13	36.78	1	0	1
2.4	2	2	75	6	12	38.17	2	0	1
2.5	2	4	165	6	14	33.19	2	0	1
2.6	2	6	179	6	14	33.83	2	0	1
3. Neurosurgery									
3.1	1	2	96	18	10	118.91	39	20	1
3.2	1	4	179	18	15	124.34	39	20	1
3.3	1	6	268	18	16	123.52	39	20	1
3.4	2	2	179	36	15	124.34	46	24	2
3.5	2	4	286	36	16	122.67	46	24	4
4. Ophthalmology									
4.1	1	2	299	24	34	26.59	1	0	3
4.4	2	2	299	48	34	26.59	1	1	4
5. Orthopaedics									
5.1	1	2	143	22	27	90.76	134	133	1
5.2	1	4	282	22	29	92.34	273	272	3
5.3	1	6	416	22	33	94.19	407	380	3
5.4	2	2	282	44	29	92.34	274	272	4
5.5	2	4	558	44	33	93.45	487	393	4
5.6	2	6	861	44	35	90.38	487	393	4
6. Urology									
6.1	1	2	93	12	20	73.37	43	17	1
6.2	1	4	206	12	20	63.73	43	17	1
6.3	1	6	287	12	21	62.86	43	17	3
6.4	2	2	206	24	20	63.73	50	17	1
6.5	2	4	287	24	21	62.86	50	17	3
7. Otolaryngology									
7.1	1	2	87	9	14	56.22	16	1	1
7.2	1	4	170	9	16	57.89	16	1	1
7.3	1	6	253	9	16	58.36	16	1	1
7.4	2	2	170	18	16	57.89	20	2	1
7.5	2	4	335	18	16	58.80	20	2	3
7.6	2	6	448	18	16	58.20	20	2	3
8. General surgery 1									
8.1	1	2	59	9	9	91.17	48	46	1
8.2	1	4	140	9	11	77.19	129	127	1
8.3	1	6	204	9	13	77.10	174	136	1
8.4	2	2	140	18	11	77.19	131	127	1
8.5	2	4	275	18	15	75.91	178	141	1
8.6	2	6	329	18	16	77.34	178	141	3
9. General surgery 2									
9.1	1	2	65	8	8	70.51	31	17	1
9.2	1	4	129	8	8	69.69	31	17	1
9.3	1	6	192	8	8	70.33	31	17	1
9.4	2	2	129	16	8	69.69	41	21	1
9.5	2	4	214	16	8	69.12	41	21	1
10. General surgery 3									
10.1	1	2	64	7	11	66.31	37	19	1
10.2	1	4	123	7	11	63.10	37	19	1
10.3	1	6	163	7	12	63.17	37	19	1
10.4	2	2	123	14	11	63.10	40	24	1
10.5	2	4	163	14	12	63.17	40	24	1

Table 4.4: Large size instances: problem instances generated based on real data

Large Size Instances									
Specialty	Planning Horizon	Capacity Multiplier	Patients	Time Blocks	No. Surgeons	Avg. Duration	Max. Schedule Date	Max. Surgery Date	Instance Group
1. Vascular surgery									
1.7	1	2	205	18	15	44.03	69	15	1
1.8	1	4	457	18	15	37.67	69	15	3
1.9	1	6	691	18	15	37.18	69	15	3
1.10	2	2	457	36	15	37.67	69	15	4
1.11	2	4	767	36	18	36.84	69	15	4
2. Oral and maxillofacial surgery									
2.7	1	2	71	6	12	39.00	1	0	1
2.8	1	4	156	6	14	33.81	1	0	1
2.9	1	6	179	6	14	33.83	1	0	1
2.10	2	2	156	12	14	33.81	1	0	1
2.11	2	4	179	12	14	33.83	1	0	1
3. Neurosurgery									
3.7	1	2	178	36	15	121.92	35	17	2
3.8	1	4	286	36	16	122.67	35	17	4
3.10	2	2	286	72	16	122.67	35	17	4
4. Ophthalmology									
4.7	1	2	299	48	34	26.59	0	0	4
4.10	2	2	299	96	34	26.59	0	0	4
5. Orthopaedics									
5.7	1	2	277	44	29	92.35	268	267	2
5.8	1	4	551	44	33	92.81	453	367	4
5.9	1	6	847	44	35	90.17	453	367	4
5.10	2	2	551	88	33	92.81	453	367	4
5.11	2	4	1153	88	37	87.95	453	367	4
5.12	2	6	1281	88	37	86.59	453	367	4
6. Urology									
6.7	1	2	202	24	20	63.69	34	15	1
6.8	1	4	287	24	21	62.86	34	15	3
6.10	2	2	287	48	21	62.86	34	15	4
7. Otolaryngology									
7.7	1	2	165	18	16	58.13	13	1	1
7.8	1	4	327	18	16	58.73	13	1	3
7.9	1	6	448	18	16	58.20	13	1	3
7.10	2	2	327	36	16	58.73	13	1	4
7.11	2	4	448	36	16	58.20	13	1	4
8. General surgery 1									
8.7	1	2	138	18	11	73.86	126	125	1
8.8	1	4	268	18	15	76.41	168	130	1
8.9	1	6	329	18	16	77.34	168	130	3
8.10	2	2	268	36	15	76.41	168	130	2
8.11	2	4	329	36	16	77.34	168	130	4
9. General surgery 2									
9.7	1	2	126	16	8	70.01	30	13	1
9.8	1	4	214	16	8	69.12	30	13	1
9.10	2	2	214	32	8	69.12	30	13	2
10. General surgery 3									
10.7	1	2	120	14	11	63.32	32	17	1
10.8	1	4	163	14	12	63.17	32	17	1
10.10	2	2	163	28	12	63.17	32	17	2

to 8 and the amount of RAM is limited to 16 GB. This configuration was chosen to represent a standard server available in a hospital by the end of 2014. Additional computational experiments showed that the MIP benefits from more memory. For instance, comparing the MIP limited to 8GB of RAM with the MIP limited to 16GB of RAM, the latter obtained better results in 38% of the instances, with a relative change of 3.9%.

4.5.3 Configuration of Parameters

4.5.3.1 General Parameters

The time limit of each computational experiment is 1 hour. The GA restarts at most 30 times and each evolution runs for 2 min. In addition, the exact models may also stop before the time limit if an optimal solution is found or the memory size limit is reached. Through all the computational experiments the cleaning time is set to 17 min and the surgeon turnover time is set to 0 min. Also, the availability of operating rooms respects the hospital master surgery schedule and surgeons are available at any time. In the discrete model, the time within each shift is discretized in intervals of 15 min, which is the most used value according to the literature review presented in Section 4.2. Finally, the constraints concerning patients priority and waiting time rules as well as surgeons workload are disabled. This configuration makes the problem harder to solve as it expands the feasible region, helping to evidence the differences among the alternative solution methods.

4.5.3.2 Genetic Algorithm Parameters

The BRKGA parameters were defined based on previous studies with the algorithm and on extensive sensitivity analysis. First, sets of values for each parameter were defined based on the recommended values found in previous studies, such as Gonçalves et al. (2014b), Gonçalves et al. (2014a), Toso

Table 4.5: Ranges of each tested GA parameter

Parameter	Tested Sets
Population Size Multiplier	10, 20, 30, 40
Percentage of Elite Population	0.1, 0.15, 0.2, 0.25
Percentage of Mutants	0.15, 0.2, 0.25, 0.3
Probability of Crossover	0.7, 0.75, 0.8, 0.85
No. of Independent Populations	1, 2
No. of Generations until Exchange Best Individuals	50, 100
No. of Generations without improving until Restart	100, 200

Table 4.6: Characteristics of instance groups and best combination of values

Instance Group	Percentage of Instances	No. of Patients	No. of Time Blocks	Best Combination of Parameter Values
1	46%	≤ 283	≤ 27	10, 0.25, 0.15, 0.7, 1, 100, 100
2	9%	≤ 283	> 27	20, 0.1, 0.3, 0.8, 2, 50, 200
3	17%	> 283	≤ 27	40 ,0.15, 0.15, 0.85, 1, 100, 200
4	28%	> 283	> 27	10, 0.2, 0.2, 0.85, 2, 100, 200

and Resende (2014) and Gonçalves and Resende (2013). Table 4.5 shows the pre-defined values for each parameter. Second, every combination of these values was tested on four pilot instances. These instances represent four different groups, based on the number of patients and the number of time blocks. The criteria used for distinguishing the groups were the average number of patients and the average number of time blocks. Table 4.6 shows the characteristics of these four different instance groups alongside with the best combination of parameter values for each group of instances. The last column lists the parameter values in the order they appear in Table 4.5. The best configuration for each instance size was the one that, among all combinations, minimized the objective function value and the running time. It is worth noting that the population size is defined as a multiple of the total number of patients. This approach showed good results in the aforementioned studies enabling the algorithm to adjust according to the instance size. The GA computation time can be adjusted using the number of restarts parameter. In the computational tests, the parameter 30 produced the best results.

4.5.4 Experimental Results

4.5.4.1 Continuous MIP model vs. Discrete IP model

Table 4.7 compares the performance of the proposed MIP model using a continuous representation of time with an IP model using a discrete representation using instances with regular size. The discrete model was able to find an optimal solution for 56% of the instances with an average gap of 0.5% compared to 12% of optimal solutions and 4.5% average gap obtained by the continuous model. However, analysing the quality of solutions, results show that the continuous model produces solutions with a lower objective function value (better) for all the cases. On average, the objective function values of solutions found by the continuous model are 53% lower than the ones found by the discrete model. The last column of Table 4.7 shows the relative change obtained by dividing the difference between the objective function values of the continuous model and the discrete model by the objective function values of the discrete model, used as reference. Even optimal solutions of the discrete model are inferior in quality compared to the solutions found by the continuous model, showing that in fact the discrete model is just an approximation of the real problem. It is worth nothing that the quality of solutions increases with the number of patients in each instance. This was expected because the optimization procedures have more options to find better solutions.

Table 4.8 also compares models with discrete and continuous representations of time but using larger instances with twice the capacity of those in Table 4.8. In these instances the discrete model found an optimal solution in 35% of the cases, with an average gap of 1.8%. In comparison, the continuous model found an optimal solution in 5% of the cases, with an average gap of 11%. This average gap is higher than the 4.5% obtained using regular size instances. Also, the continuous model failed to obtain a feasible solution within the established time limit for 3 instances of Orthopaedics (#5.10, #5.11, #5.12). For these instances the continuous model requires more than 16GB of RAM to find a feasible solution in 1 hour. However, for the other instances, on average, the objective function values of the continuous model are 61% lower than the ones of the discrete model. It means that, compared to the IP model using a discrete representation of time, the proposed MIP model using a continuous representation of time is able to find much better solutions even for large instances.

Table 4.7: Regular Instances: Comparison of mathematical models using discrete and continuous representation of time - Best solutions considering the No. of Scheduled Surgeries and Average OR Occupancy Rate

Regular Size Instances														
	IP Model - discrete time representation						MIP Model - continuous time representation							
Instance	Objective Function Value	No. of Scheduled Surgeries	Avg. OR Occupancy Rate (%)	Status	Gap (%)	Running Time (s)	Objective Function Value	No. of Scheduled Surgeries	Avg. OR Occupancy Rate (%)	Status	Gap (%)	Running Time (s)	Continuous better than discrete?	Relative Change (%)
1. Vascular surgery														
1.1	0.38814	35	64.0	Feasible	0.4	MAX	0.19077	61	64.6	Feasible	3.6	MAX	Yes	-50.8
1.2	0.38533	35	64.5	Feasible	0.4	MAX	0.16672	65	63.5	Feasible	6.7	2528	Yes	-56.7
1.3	0.38304	35	64.9	Feasible	0.3	MAX	0.16443	65	63.9	Feasible	5.4	MAX	Yes	-57.1
1.4	0.38290	57	69.7	Optimal	0.0	258	0.15572	126	63.4	Feasible	4.4	MAX	Yes	-59.3
1.5	0.38107	57	70.0	Optimal	0.0	301	0.15435	126	63.7	Feasible	13.6	1387	Yes	-59.5
1.6	0.37537	57	71.0	Optimal	0.0	3264	0.15279	126	63.9	Feasible	12.8	1571	Yes	-59.3
2. Oral and maxillofacial surgery														
2.1	0.52517	11	61.0	Optimal	0.0	0	0.33127	22	61.9	Optimal	0.0	1	Yes	-36.9
2.2	0.51288	13	56.3	Optimal	0.0	0	0.30268	25	57.0	Optimal	0.0	6	Yes	-41.0
2.3	0.51216	14	52.7	Optimal	0.0	0	0.28425	27	53.5	Optimal	0.0	2	Yes	-44.5
2.4	0.47219	22	61.9	Optimal	0.0	0	0.24813	49	57.9	Optimal	0.0	58	Yes	-47.5
2.5	0.46177	26	57.0	Optimal	0.0	5	0.22513	54	53.6	Feasible	2.0	2559	Yes	-51.2
2.6	0.46070	26	57.2	Optimal	0.0	19	0.22513	54	53.6	Feasible	1.8	1791	Yes	-51.1
3. Neurosurgery														
3.1	0.29099	45	72.3	Feasible	1.4	MAX	0.13730	58	83.5	Feasible	6.5	1805	Yes	-52.8
3.2	0.24645	52	71.0	Feasible	1.4	MAX	0.10848	63	81.9	Feasible	11.6	1340	Yes	-56.0
3.3	0.23277	55	69.4	Optimal	0.0	889	0.09471	65	81.8	Feasible	9.8	1399	Yes	-59.3
3.4	0.25809	88	73.2	Feasible	2.9	MAX	0.09775	114	83.7	Feasible	11.8	MAX	Yes	-62.1
3.5	0.23524	93	73.7	Feasible	5.1	MAX	0.07831	120	82.8	Feasible	18.6	MAX	Yes	-66.7
4. Ophthalmology														
4.1	0.30701	126	51.9	Optimal	0.0	383	0.07795	206	57.5	Feasible	11.4	MAX	Yes	-74.6
4.4	0.10772	262	43.5	Feasible	4.3	MAX	0.00000	299	47.8	Optimal	0.0	551	Yes	-100.0
5. Orthopaedics														
5.1	0.43150	69	69.3	Optimal	0.0	148	0.31872	89	79.5	Feasible	1.5	MAX	Yes	-26.1
5.2	0.42051	87	61.3	Optimal	0.0	184	0.26792	117	73.5	Feasible	1.0	MAX	Yes	-36.3
5.3	0.41547	89	61.2	Optimal	0.0	387	0.24855	128	71.1	Feasible	1.1	MAX	Yes	-40.2
5.4	0.40204	145	67.8	Feasible	0.0	MAX	0.29367	177	78.3	Feasible	5.3	MAX	Yes	-27.0
5.5	0.39059	177	60.3	Optimal	0.0	2185	0.25298	221	72.7	Feasible	9.8	MAX	Yes	-35.2
5.6	0.37826	195	57.2	Feasible	0.3	MAX	0.22763	261	65.4	Feasible	18.4	MAX	Yes	-39.8
6. Urology														
6.1	0.40588	47	61.6	Optimal	0.0	8	0.25025	64	72.7	Feasible	2.0	1393	Yes	-38.3
6.2	0.37313	62	51.7	Optimal	0.0	25	0.18908	81	66.0	Feasible	2.8	MAX	Yes	-49.3
6.3	0.36107	65	50.7	Feasible	0.2	MAX	0.16863	87	63.4	Feasible	0.7	2461	Yes	-53.3

Continued on next page

Table 4.7 – *Continued from previous page*

Regular Size Instances														
	IP Model - discrete time representation						MIP Model - continuous time representation							
Instance	Objective Function Value	No. of Scheduled Surgeries	Avg. OR Occupancy Rate (%)	Status	Gap (%)	Running Time (s)	Objective Function Value	No. of Scheduled Surgeries	Avg. OR Occupancy Rate (%)	Status	Gap (%)	Running Time (s)	Continuous better than discrete?	Relative Change (%)
6.4	0.33289	109	55.6	Feasible	0.0	MAX	0.14872	140	70.4	Feasible	1.7	MAX	Yes	-55.3
6.5	0.31485	117	54.2	Feasible	0.0	MAX	0.12318	152	68.0	Feasible	3.2	MAX	Yes	-60.9
7. Otolaryngology														
7.1	0.39939	33	65.0	Feasible	0.0	MAX	0.25039	47	73.9	Feasible	2.1	705	Yes	-37.3
7.2	0.39892	32	66.3	Optimal	0.0	105	0.23113	52	71.1	Feasible	1.1	846	Yes	-42.1
7.3	0.39837	32	66.4	Optimal	0.0	1965	0.20610	58	68.1	Feasible	1.9	731	Yes	-48.3
7.4	0.33700	72	61.0	Optimal	0.0	4	0.17629	91	74.7	Feasible	5.4	1233	Yes	-47.7
7.5	0.33672	72	61.1	Optimal	0.0	9	0.13880	105	71.0	Feasible	2.1	769	Yes	-58.8
7.6	0.33510	72	61.3	Optimal	0.0	22	0.10984	116	68.0	Feasible	3.9	MAX	Yes	-67.2
8. General surgery 1														
8.1	0.44153	27	68.8	Optimal	0.0	34	0.29224	42	76.4	Feasible	1.2	1552	Yes	-33.8
8.2	0.41806	35	62.2	Optimal	0.0	13	0.20228	62	66.0	Feasible	0.5	MAX	Yes	-51.6
8.3	0.40811	36	62.7	Feasible	0.0	MAX	0.18951	65	64.3	Feasible	1.4	2103	Yes	-53.6
8.4	0.42140	75	55.1	Feasible	0.0	MAX	0.23319	100	72.5	Feasible	0.5	814	Yes	-44.7
8.5	0.39923	70	63.0	Optimal	0.0	533	0.17410	124	66.1	Feasible	1.1	1294	Yes	-56.4
8.6	0.39288	73	62.0	Optimal	0.0	103	0.16420	128	65.0	Feasible	1.4	1394	Yes	-58.2
9. General surgery 2														
9.1	0.37361	27	67.1	Feasible	0.2	MAX	0.19961	40	75.1	Feasible	0.9	2312	Yes	-46.6
9.2	0.34518	32	62.8	Optimal	0.0	2	0.17277	44	72.4	Feasible	0.7	857	Yes	-49.9
9.3	0.33784	33	62.3	Optimal	0.0	8	0.17162	44	72.6	Feasible	6.7	1465	Yes	-49.2
9.4	0.31254	55	67.1	Feasible	0.0	MAX	0.13241	79	75.3	Feasible	3.7	954	Yes	-57.6
9.5	0.28666	66	60.9	Feasible	0.0	MAX	0.09825	88	72.5	Feasible	5.9	1076	Yes	-65.7
10. General surgery 3														
10.1	0.32482	27	59.3	Optimal	0.0	392	0.15395	33	76.1	Feasible	5.1	2707	Yes	-52.6
10.2	0.29064	30	58.7	Feasible	0.3	MAX	0.11723	37	73.8	Optimal	0.0	43	Yes	-59.7
10.3	0.28811	31	57.0	Feasible	0.2	MAX	0.10030	39	72.4	Feasible	1.5	MAX	Yes	-65.2
10.4	0.25573	54	60.1	Feasible	0.2	MAX	0.07230	66	76.4	Feasible	12.0	849	Yes	-71.7
10.5	0.23182	57	60.6	Feasible	0.2	MAX	0.05246	70	75.1	Feasible	5.0	2984	Yes	-77.4

Table 4.8: Large Instances: Comparison of mathematical models using discrete and continuous representation of time - Best solutions considering the No. of Scheduled Surgeries and Average OR Occupancy Rate

Large Size Instances														
	IP Model - discrete time representation						MIP Model - continuous time representation							
Instance	Objective Function Value	No. of Scheduled Surgeries	Avg. OR Occupancy Rate (%)	Status	Gap (%)	Running Time (s)	Objective Function Value	No. of Scheduled Surgeries	Avg. OR Occupancy Rate (%)	Status	Gap (%)	Running Time (s)	Continuous better than discrete?	Relative Change (%)
1. Vascular surgery														
1.7	0.38290	57	69.7	Optimal	0.0	51	0.15572	126	63.4	Feasible	5.2	MAX	Yes	-59.3
1.8	0.38107	57	70.0	Optimal	0.0	3521	0.15279	126	63.9	Feasible	12.7	2072	Yes	-59.9
1.9	0.37537	57	71.0	Feasible	0.0	MAX	0.15435	126	63.7	Feasible	13.6	1068	Yes	-58.9
1.10	0.36373	138	56.1	Feasible	1.7	MAX	0.12122	241	62.7	Feasible	30.9	MAX	Yes	-66.7
1.11	0.34830	126	62.0	Feasible	0.3	MAX	0.10740	244	63.9	Feasible	29.0	MAX	Yes	-69.2
2. Oral and maxillofacial surgery														
2.7	0.47219	22	61.9	Optimal	0.0	0	0.24813	49	57.9	Feasible	0.1	MAX	Yes	-47.5
2.8	0.46177	26	57.0	Optimal	0.0	7	0.22513	54	53.6	Feasible	1.8	MAX	Yes	-51.2
2.9	0.46070	26	57.2	Optimal	0.0	9	0.22513	54	53.6	Feasible	1.9	2141	Yes	-51.1
2.10	0.41248	44	61.7	Feasible	0.2	MAX	0.17225	104	55.3	Feasible	0.4	MAX	Yes	-58.2
2.11	0.41154	44	61.9	Optimal	0.0	23	0.16837	106	54.4	Feasible	1.4	MAX	Yes	-59.1
3. Neurosurgery														
3.7	0.26299	86	73.8	Feasible	5.1	MAX	0.10144	113	83.7	Feasible	15.0	MAX	Yes	-61.4
3.8	0.23953	92	73.6	Feasible	7.1	MAX	0.08393	118	83.2	Feasible	24.1	MAX	Yes	-65.0
3.10	0.24075	163	71.3	Feasible	3.0	MAX	0.10989	202	79.1	Feasible	67.7	MAX	Yes	-54.4
4. Ophthalmology														
4.7	0.10472	263	43.6	Feasible	1.5	MAX	0.00000	299	47.8	Optimal	0.0	259	Yes	-100.0
4.10	0.00000	299	23.9	Optimal	0.0	83	0.00000	299	23.9	Optimal	0.0	204	No	-100.0
5. Orthopaedics														
5.7	0.40208	145	67.7	Feasible	0.0	MAX	0.30693	174	76.7	Feasible	9.4	MAX	Yes	-23.7
5.8	0.39290	177	59.8	Feasible	0.6	MAX	0.24725	222	73.4	Feasible	7.7	MAX	Yes	-37.1
5.9	0.37768	196	57.0	Feasible	0.2	MAX	0.20615	262	69.1	Feasible	9.9	MAX	Yes	-45.4
5.10	0.40409	227	68.1	Feasible	14.8	MAX	0.28901	323	71.9	Feasible	27.4	MAX	Yes	-28.5
5.11	0.40409	227	68.1	Feasible	21.7	MAX	1.00000	nfs	nfs	nfs	nfs	MAX	No	147.5
5.12	0.40409	227	68.1	Feasible	23.2	3547	1.00000	nfs	nfs	nfs	nfs	MAX	No	147.5
6. Urology														
6.7	0.33289	109	55.6	Feasible	0.0	MAX	0.15253	139	70.3	Feasible	4.1	MAX	Yes	-54.2
6.8	0.31485	117	54.2	Feasible	0.0	MAX	0.12318	152	68.0	Feasible	3.2	MAX	Yes	-60.9
6.10	0.24980	201	57.5	Feasible	1.6	MAX	0.09049	237	72.0	Feasible	34.0	MAX	Yes	-63.8
7. Otolaryngology														
7.7	0.33700	72	61.0	Optimal	0.0	5	0.17809	91	74.4	Feasible	6.4	851	Yes	-47.2
7.8	0.33672	72	61.1	Optimal	0.0	12	0.14022	105	70.7	Feasible	3.1	1299	Yes	-58.4
7.9	0.33510	72	61.3	Optimal	0.0	22	0.10908	116	68.1	Feasible	3.3	MAX	Yes	-67.4

Continued on next page

Table 4.8 – *Continued from previous page*

Large Size Instances														
	IP Model - discrete time representation						MIP Model - continuous time representation							
Instance	Objective Function Value	No. of Scheduled Surgeries	Avg. OR Occupancy Rate (%)	Status	Gap (%)	Running Time (s)	Objective Function Value	No. of Scheduled Surgeries	Avg. OR Occupancy Rate (%)	Status	Gap (%)	Running Time (s)	Continuous better than discrete?	Relative Change (%)
7.10	0.27136	150	58.8	Feasible	0.0	MAX	0.08965	186	74.0	Feasible	8.6	MAX	Yes	-67.0
7.11	0.26852	157	56.4	Feasible	0.0	MAX	0.07446	196	72.4	Feasible	15.7	MAX	Yes	-72.3
8. General surgery 1														
8.7	0.42140	75	55.1	Feasible	0.0	MAX	0.23445	100	72.3	Feasible	1.0	1554	Yes	-44.4
8.8	0.39923	70	63.0	Optimal	0.0	132	0.17535	124	65.8	Feasible	1.8	1953	Yes	-56.1
8.9	0.39288	73	62.0	Optimal	0.0	247	0.16412	128	65.0	Feasible	1.4	1869	Yes	-58.2
8.10	0.34611	152	54.6	Feasible	0.0	MAX	0.14376	195	73.2	Feasible	2.4	MAX	Yes	-58.5
8.11	0.33959	156	54.1	Optimal	0.0	3524	0.11882	211	70.9	Feasible	2.8	MAX	Yes	-65.0
9. General surgery 2														
9.7	0.31254	55	67.1	Feasible	0.1	MAX	0.13241	79	75.3	Feasible	3.7	778	Yes	-57.6
9.8	0.28666	66	60.9	Feasible	0.0	MAX	0.09856	88	72.5	Feasible	6.2	861	Yes	-65.6
9.10	0.25254	115	64.3	Optimal	0.0	2475	0.06204	153	75.9	Feasible	8.1	1050	Yes	-75.4
10. General surgery 3														
10.7	0.25573	54	60.1	Feasible	0.1	MAX	0.07268	66	76.4	Feasible	12.5	1125	Yes	-71.6
10.8	0.23182	57	60.6	Feasible	0.1	MAX	0.05799	69	75.3	Feasible	14.1	1165	Yes	-75.0
10.10	0.23188	101	61.5	Feasible	1.3	MAX	0.03823	127	76.7	Feasible	32.6	3168	Yes	-83.5

nfs = no feasible solution until the time limit

4.5.4.2 Continuous MIP model vs. BRKGA heuristic

Table 4.9 compares the results of the MIP model using a continuous representation of time with the results obtained using the BRKGA presented in Section 4.4.2 for the regular size instances. The continuous model's results are repeated in this table to make the comparison easier. The GA found a solution with lower objective function value under the specified stopping criteria in 62% of the instances with 29% of the continuous MIP model. However, the differences in quality of solutions between the two proposed approaches are very small. In the instances in which the GA is better, the relative improvement was only 1.1% against 2.6% for the instances in which the continuous model was better. The GA has a better performance on instances that require more memory, such as Orthopaedics #5.6. On the other hand, the highest difference in favour of the exact model is in instances Neurosurgery #3.5 and General surgery 3 #9.5. These instances are characterized by a high number of parallel ORs, showing that the exact model deals with this issue better than the BRKGA. The BRKGA decoder is able to prevent overlaps but may leave some idle time between the surgeries, what is difficult to improve only through crossover and mutation. A local search procedure is required to eliminate the idle time.

The GA computation time can be adjusted using the number of restarts parameter. In the computational tests, a value of 30 was used for the number of restarts for producing the best results under the specified time limit. The last column in Table 4.9 shows that many times the algorithm did not improve after the first restarts. In this situation one can reduce the number of restarts and save computational time.

Table 4.10 compares the proposed continuous model with the proposed BRKGA heuristic using large size instances. In this instance set, each approach found 45% of solutions with lower objective function values. However, among the solutions in which the GA obtained a better value, the average improvement was 13% compared to 6% of the exact model. The heuristic obtained better values in instances that the model requires more memory, such as the largest instances of Orthopaedics in which the model did not obtain any feasible solution. In its turn, the GA lost more comparisons among medium size instances, like the ones of General Surgery. These instances are characterized by a relative low number of surgeons and large surgery durations, which increases the chance of occurring overlaps. In this case, the GA would benefit from a local search procedure to make small improvements in the quality of solutions that are difficult to promote with the GA alone.

Table 4.9: Regular instances: MIP Model vs. GA Heuristic - Best solutions considering No. of Scheduled Surgeries and Average OR Occupancy Rate

Regular Size Instances													
	MIP Model - continuous time				Genetic Algorithm							Comparison	
Instance	Objective Function Value	No. of Scheduled Patients	Avg. OR Occupancy Rate (%)	Running Time (s)	Objective Function Value	No. of Scheduled Patients	Avg. OR Occupancy Rate (%)	No. of Restarts	No. of Improvements	Last Improvement	Running Time (s)	GA better than or equal to MIP?	Relative Change (%)
1. Vascular surgery													
1.1	0.19077	61	64.6	MAX	0.19042	61	64.6	12	1	3	MAX	Yes	-0.2
1.2	0.16672	65	63.5	2528	0.16637	65	63.5	12	1	3	MAX	Yes	-0.2
1.3	0.16443	65	63.9	MAX	0.16421	65	63.9	12	1	3	MAX	Yes	-0.1
1.4	0.15572	126	63.4	MAX	0.15710	126	63.1	12	1	3	MAX	No	0.9
1.5	0.15435	126	63.7	1387	0.15591	126	63.3	12	1	3	MAX	No	1.0
1.6	0.15279	126	63.9	1571	0.15435	126	63.6	12	2	11	MAX	No	1.0
2. Oral and maxillofacial surgery													
2.1	0.33127	22	61.9	1	0.33111	22	61.9	12	1	3	MAX	Yes	0.0
2.2	0.30268	25	57.0	6	0.30262	25	57.0	12	1	3	MAX	Yes	0.0
2.3	0.28425	27	53.5	2	0.28419	27	53.5	12	1	3	MAX	Yes	0.0
2.4	0.24813	49	57.9	58	0.24810	49	57.9	12	1	3	MAX	Yes	0.0
2.5	0.22513	54	53.6	2559	0.22513	54	53.6	12	1	3	MAX	Yes	0.0
2.6	0.22513	54	53.6	1791	0.22513	54	53.6	12	1	3	MAX	Yes	0.0
3. Neurosurgery													
3.1	0.13730	58	83.5	1805	0.13760	58	83.4	12	3	7	MAX	No	0.2
3.2	0.10848	63	81.9	1340	0.11417	62	82.2	12	3	7	MAX	No	5.2
3.3	0.09471	65	81.8	1399	0.10046	64	82.1	12	3	12	MAX	No	6.1
3.4	0.09775	114	83.7	MAX	0.09570	115	83.3	12	4	6	MAX	Yes	-2.1
3.5	0.07831	120	82.8	MAX	0.08464	118	83.1	12	3	12	MAX	No	8.1
4. Ophthalmology													
4.1	0.07795	206	57.5	MAX	0.07781	210	56.3	12	2	4	MAX	Yes	-0.2
4.4	0.00000	299	47.8	551	0.00000	299	47.8	12	1	3	MAX	Yes	0.0
5. Orthopaedics													
5.1	0.31872	89	79.5	MAX	0.32105	89	79.0	12	3	10	MAX	No	0.7
5.2	0.26792	117	73.5	MAX	0.27023	117	73.1	12	2	9	MAX	No	0.9
5.3	0.24855	128	71.1	MAX	0.24996	128	70.8	12	3	8	MAX	No	0.6
5.4	0.29367	177	78.3	MAX	0.29913	177	77.3	12	1	3	MAX	No	1.9
5.5	0.25298	221	72.7	MAX	0.24983	225	72.0	12	3	9	MAX	Yes	-1.2
5.6	0.22763	261	65.4	MAX	0.20960	267	66.9	12	3	11	MAX	Yes	-7.9
6. Urology													
6.1	0.25025	64	72.7	1393	0.24798	64	73.1	12	2	8	MAX	Yes	-0.9
6.2	0.18908	81	66.0	MAX	0.18574	82	65.6	12	1	3	MAX	Yes	-1.8
6.3	0.16863	87	63.4	2461	0.16837	87	63.5	12	1	3	MAX	Yes	-0.2
6.4	0.14872	140	70.4	MAX	0.15258	139	70.3	12	1	3	MAX	No	2.6
6.5	0.12318	152	68.0	MAX	0.12251	153	67.6	12	4	7	MAX	Yes	-0.5
7. Otolaryngology													
7.1	0.25039	47	73.9	705	0.24946	47	74.1	12	3	7	MAX	Yes	-0.4
7.2	0.23113	52	71.1	846	0.23032	52	71.2	12	3	9	MAX	Yes	-0.4
7.3	0.20610	58	68.1	731	0.20608	58	68.1	12	2	4	MAX	Yes	0.0

Continued on next page

Table 4.9 – *Continued from previous page*

Regular Size Instances														
	MIP Model - continuous time				Genetic Algorithm								Comparison	
Instance	Objective Function Value	No. of Scheduled Patients	Avg. OR Occupancy Rate (%)	Running Time (s)	Objective Function Value	No. of Scheduled Patients	Avg. OR Occupancy Rate (%)	No. of Restarts	No. of Improvements	Last Improvement	Running Time (s)	GA better than or equal to MIP?	Relative Change (%)
7.4	0.17629	91	74.7	1233	0.17605	91	74.7	12	2	7	MAX	Yes	-0.1
7.5	0.13880	105	71.0	769	0.13830	105	71.1	12	3	10	MAX	Yes	-0.4
7.6	0.10984	116	68.0	MAX	0.10718	117	67.7	12	3	12	MAX	Yes	-2.4
8. General surgery 1													
8.1	0.29224	42	76.4	1552	0.28979	42	76.9	12	1	3	MAX	Yes	-0.8
8.2	0.20228	62	66.0	MAX	0.20188	62	66.1	12	2	4	MAX	Yes	-0.2
8.3	0.18951	65	64.3	2103	0.19267	64	65.1	12	2	5	MAX	No	1.7
8.4	0.23319	100	72.5	814	0.23241	100	72.7	12	4	11	MAX	Yes	-0.3
8.5	0.17410	124	66.1	1294	0.17440	124	66.0	12	3	10	MAX	No	0.2
8.6	0.16420	128	65.0	1394	0.16434	128	65.0	12	6	12	MAX	No	0.1
9. General surgery 2													
9.1	0.19961	40	75.1	2312	0.19977	40	75.0	12	3	6	MAX	No	0.1
9.2	0.17277	44	72.4	857	0.17911	43	73.1	12	3	12	MAX	No	3.7
9.3	0.17162	44	72.6	1465	0.17162	44	72.6	12	2	11	MAX	Yes	0.0
9.4	0.13241	79	75.3	954	0.13249	79	75.3	12	1	3	MAX	No	0.1
9.5	0.09825	88	72.5	1076	0.09525	89	72.1	12	1	3	MAX	Yes	-3.0
10. General surgery 3													
10.1	0.15395	33	76.1	2707	0.15374	33	76.1	12	1	3	MAX	Yes	-0.1
10.2	0.11723	37	73.8	43	0.11744	37	73.8	12	2	5	MAX	No	0.2
10.3	0.10030	39	72.4	MAX	0.10897	38	73.0	12	1	3	MAX	No	8.6
10.4	0.07230	66	76.4	849	0.06981	67	75.7	12	3	12	MAX	Yes	-3.4
10.5	0.05246	70	75.1	2984	0.05768	69	75.4	12	1	3	MAX	No	9.9

Table 4.10: Large instances: MIP Model vs. GA Heuristic - Best solutions considering No. of Scheduled Surgeries and Average OR Occupancy Rate

Large Size Instances														
	MIP Model - continuous time				Genetic Algorithm							Comparison	
Instance	Objective Function Value	No. of Scheduled Patients	Avg. OR Occupancy Rate (%)	Running Time (s)	Objective Function Value	No. of Scheduled Patients	Avg. OR Occupancy Rate (%)	No. of Restarts	No. of Improvements	Last Improvement	Running Time (s)	GA better than or equal to MIP?	Relative Change (%)
1. Vascular surgery													
1.7	0.15572	126	63.4	MAX	0.15710	126	63.1	12	4	7	MAX	No	0.9
1.8	0.15279	126	63.9	2072	0.15591	126	63.3	12	1	3	MAX	No	2.0
1.9	0.15435	126	63.7	1068	0.15591	126	63.3	12	1	3	MAX	No	1.0
1.10	0.12122	241	62.7	MAX	0.11734	243	62.6	12	3	10	MAX	Yes	-3.2
1.11	0.10740	244	63.9	MAX	0.12169	241	62.6	12	3	10	MAX	No	13.3
2. Oral and maxillofacial surgery													
2.7	0.24813	49	57.9	MAX	0.24810	49	57.9	12	1	3	MAX	Yes	0.0
2.8	0.22513	54	53.6	MAX	0.22513	54	53.6	12	1	3	MAX	No	0.0
2.9	0.22513	54	53.6	2141	0.22513	54	53.6	12	1	3	MAX	No	0.0
2.10	0.17225	104	55.3	MAX	0.17208	104	55.3	12	2	6	MAX	Yes	-0.1
2.11	0.16837	106	54.4	MAX	0.16866	106	54.3	12	3	7	MAX	No	0.2
3. Neurosurgery													
3.7	0.10144	113	83.7	MAX	0.10352	114	82.5	12	3	5	MAX	No	2.0
3.8	0.08393	118	83.2	MAX	0.08709	118	82.6	12	4	7	MAX	No	3.8
3.10	0.10989	202	79.1	MAX	0.08713	209	80.3	12	2	4	MAX	Yes	-20.7
4. Ophthalmology													
4.7	0.00000	299	47.8	259	0.00000	299	47.8	12	1	3	MAX	No	0.0
4.10	0.00000	299	23.9	204	0.00000	299	23.9	12	1	3	MAX	No	0.0
5. Orthopaedics													
5.7	0.30693	174	76.7	MAX	0.30394	177	76.3	12	3	6	MAX	Yes	-1.0
5.8	0.24725	222	73.4	MAX	0.25842	223	71.0	12	3	8	MAX	No	4.5
5.9	0.20615	262	69.1	MAX	0.21954	263	66.2	12	2	4	MAX	No	6.5
5.10	0.28901	323	71.9	MAX	0.24426	345	76.1	12	2	9	MAX	Yes	-15.5
5.11	1.00000	nfs	nfs	MAX	0.16479	467	68.8	12	4	6	MAX	Yes	-83.5
5.12	1.00000	nfs	nfs	MAX	0.14881	488	68.0	12	2	6	MAX	Yes	-85.1
6. Urology													
6.7	0.15253	139	70.3	MAX	0.15302	139	70.2	12	3	10	MAX	No	0.3
6.8	0.12318	152	68.0	MAX	0.12565	152	67.6	12	4	11	MAX	No	2.0
6.10	0.09049	237	72.0	MAX	0.08705	240	71.5	12	3	10	MAX	Yes	-3.8
7. Otolaryngology													
7.7	0.17809	91	74.4	851	0.17634	91	74.7	12	2	6	MAX	Yes	-1.0
7.8	0.14022	105	70.7	1299	0.13894	105	70.9	12	2	4	MAX	Yes	-0.9
7.9	0.10908	116	68.1	MAX	0.11038	116	67.9	12	3	9	MAX	No	1.2

Continued on next page

Table 4.10 – *Continued from previous page*

Large Size Instances													
	MIP Model - continuous time				Genetic Algorithm							Comparison	
Instance	Objective Function Value	No. of Scheduled Patients	Avg. OR Occupancy Rate (%)	Running Time (s)	Objective Function Value	No. of Scheduled Patients	Avg. OR Occupancy Rate (%)	No. of Restarts	No. of Improvements	Last Improvement	Running Time (s)	GA better than or equal to MIP?	Relative Change (%)
7.10	0.08965	186	74.0	MAX	0.10113	185	72.5	12	3	5	MAX	No	12.8
7.11	0.07446	196	72.4	MAX	0.07438	198	71.6	12	4	10	MAX	Yes	-0.1
8. General surgery 1													
8.7	0.23445	100	72.3	1554	0.23539	100	72.1	12	3	8	MAX	No	0.4
8.8	0.17535	124	65.8	1953	0.17901	123	65.8	12	4	8	MAX	No	2.1
8.9	0.16412	128	65.0	1869	0.16976	127	64.6	12	4	11	MAX	No	3.4
8.10	0.14376	195	73.2	MAX	0.16538	192	70.5	12	7	12	MAX	No	15.0
8.11	0.11882	211	70.9	MAX	0.14216	207	68.3	12	5	12	MAX	No	19.6
9. General surgery 2													
9.7	0.13241	79	75.3	778	0.13327	79	75.2	12	4	9	MAX	No	0.7
9.8	0.09856	88	72.5	861	0.10011	88	72.2	12	2	11	MAX	No	1.6
9.10	0.06204	153	75.9	1050	0.06893	152	75.2	12	6	11	MAX	No	11.1
10. General surgery 3													
10.7	0.07268	66	76.4	1125	0.07082	67	75.5	12	3	7	MAX	Yes	-2.6
10.8	0.05799	69	75.3	1165	0.05901	69	75.1	12	2	9	MAX	No	1.8
10.10	0.03823	127	76.7	3168	0.05932	124	75.0	12	3	9	MAX	No	55.2

nfs = no feasible solution until the time limit

Table 4.11: Summary of the computational experiments

Regular size instances

	Discrete exact model	Continuous exact model	Continuous exact model	BRKGA heuristic
Percentage of better solutions	0%	100%	29%	62%
Avg. relative difference in better instances	-	53%	2.6%	1.1%

Large size instances

	Discrete exact model	Continuous exact model	Continuous exact model	BRKGA heuristic
Percentage of better solutions	0%	100%	45%	45%
Avg. relative difference in better instances	-	61%	13%	6%

Table 4.12: Percentage of better solutions by solution method in each instance group

Instance Group	Continuous exact model (%)	BRKGA heuristic (%)
1	22	65
2	50	50
3	59	41
4	45	40

Table 4.11 summarizes the results of the computational experiments based on the percentage of better solutions that each alternative approach obtained on each comparison and on the relative change between the objective function values. The continuous model shows to be clearly better than the discrete model as it found better solutions for all the instances. These solutions are substantially better, 53% in regular size instances and 61% on large size instances. In its turn, the heuristic found better solutions than the continuous model for 62% of the regular size instances and 45% of the large size instances. Surprisingly, the heuristic was able to find a higher proportion of better solutions among smaller size instances. Table 4.12 shows the proportion of better solutions obtained by the exact model and the heuristic in each group of instances. The proportions are balanced, except the better result of the heuristic in smaller instances.

4.6 Discussion and Future Work

This chapter proposed two alternative solution methods for the integrated SCAP: (1) an exact and (2) a heuristic. In the first case, our contribution is a new formulation using a continuous representation of time. Compared to a model using a discrete representation of time, which is an adaptation of the model presented in Chapter 3, this new formulation found better solutions for all instances. In comparison with the heuristic procedure proposed in this paper it found better quality solutions for instances with a high number of parallel ORs. It shows that the proposed exact formulation is very effective in synchronizing the utilization of parallel resources. The downside of this exact formulation is the required amount of memory. In the second case, our contribution is an approximation method based on the biased random-key genetic algorithm featuring an original decoding procedure as well as additional local search procedures. The heuristic was able to obtain better solutions than the continuous model in 62% of the regular size instances and 45% of the large size instances. Surprisingly, the results are better in small size instances. The reason is that the GA is not able to make certain small changes to enhance the quality of solutions. The implemented local search procedures helped to improve most part of the solutions but have a limited number of movements. Currently, these movements are able to increase the utilization of ORs but not the number of scheduled surgeries.

In future work, the authors intend to enhance the performance of the heuristic with the addition of new local search procedures. The implemented local search procedures provide good results, enabling the GA to find better quality solutions in almost all instances. However, we implement only two simple movements that change one surgery by another. New movements should be implemented to change one scheduled surgery for multiple unscheduled ones. Also, the problem of finding the best combination of parameters should be addressed to allow the GA to have a more uniform performance across different instances. Furthermore, in what concerns different problem settings, we intend to evaluate the performance of the proposed approaches in a rolling horizon framework. In this case, additional constraints are required to minimize the rescheduling of previously scheduled patients as well as to minimize situations in which the sequence of the waiting list, determined by priority and waiting time rules, is broken. This framework would allow to compare the performance of alternative objective functions in the long term, to better understand the impact of prioritizing the number of scheduled patients or the average OR utilization rate.

Acknowledgments

This work has been partially supported by projects PTDC/EGE-GES/117692/2010 and NORTE-07-0124-FEDER-000057 funded by the North Portugal Regional Operational Programme (ON.2 – O Novo Norte), under the National Strategic Reference Framework (NSRF), through the European Regional Development Fund (ERDF) and the Programme COMPETE, and by national funds, through the Portuguese funding agency, Fundacão para a Ciência e a Tecnologia (FCT).

Appendix

4.A Exact Model Using a Discrete Representation of Time

This section briefly describes the discrete model compared with the proposed continuous model. The discrete model is described in detail in Chapter 3. Similar models are also presented in Marques et al. (2012) and Guinet and Chaabane (2003).

4.A.1 Sets and Indices

The sets and indices are equal to the ones presented in Section 4.4.1, except for the introduction of a new set L to denote the intervals in which surgeries are allowed to start in each working shift. These discrete intervals are a result of the discretization of time and their size usually ranges from 10 min to 1 hour, with the most used value being 15 min. Among the parameters the only new entry is parameter n to denote the number of intervals in each shift. This value is determined by dividing the capacity of ORs by the selected size of interval, e.g. $360/15 = 24$.

I	set of patients (index i)
J	set of working shifts in the planning horizon (index j)
K	set of operating rooms (index k)
K_j	set of available ORs in shifts j
S	set of surgeons (index s)
I_s	set of patients of surgeon s (index i)
H	set of weeks in the planning horizon (index h)
J_h	set of days in a given week h (index j)
L	set of intervals in each shift j (index l)
$I_{maxsched}$	set of patients with maximum scheduling time within the planning horizon
$I_{maxswait}$	set of patients with maximum waiting time within the planning horizon

4.A.2 Parameters

d_i	estimated duration in minutes of patient's i surgery
s_i	surgeon in charge of patient's i surgery
max_i	maximum waiting time of patient's i surgery
c_{jk}	available capacity in shift j of OR k
a_{js}	availability in shift j of surgeon s
day_j	day of shift j
α	weight of the number of scheduled surgeries in the objective function
β	weight of the average OR utilization rate in the objective function
γ	best number of scheduled surgeries
δ	best average OR utilization rate
ct	OR cleaning time
tt	surgeon turnover time
C	total OR capacity
ms	maximum number of shifts per week
n	number of intervals per shift

The discrete model has only one decision variable to represent the scheduled patients. Variable X_{ijkl} represents all at once the selected patient, day, shift, OR and starting time. The objective function (4.23) is very similar to the objective function of the continuous model. The only difference is that it has one more cycle, through the set L, to determine the scheduled patients. For a detailed description of the objective function used in the continuous model see Section 4.4.1.

4.A.3 Decision Variables

$$X_{ijkl} = \begin{cases} 1, \text{if patient } i \text{ is scheduled for shift } j, \text{ OR } k \text{ and period } l \\ 0, \text{otherwise} \end{cases}$$

4.A.4 Objective Function

$$min\ F = \alpha \cdot \frac{\gamma - \sum_{i\in I}\sum_{j\in J}\sum_{k\in K}\sum_{l\in L} X_{ijkl}}{\gamma} + \beta \cdot \frac{\delta - \dfrac{\sum_{i\in I}\sum_{j\in J}\sum_{k\in K}\sum_{l\in L} X_{ijkl}\cdot d_i}{C}}{\delta} \tag{4.22}$$

4.A.5 Constraints

The first set of constraints provide the basic structure of the model. Inequality (4.23) prevents a patient from being scheduled more than once, expression (4.24) restricts the scheduling of patients to the capacity of available shifts and ORs, and constraint (4.25) prevents surgeries from having a scheduled end time greater than the surgical suite closing time.

$$\sum_{j\in J}\sum_{k\in K}\sum_{l\in L} X_{ijkl} \leq 1, \forall i \in I \tag{4.23}$$

$$\sum_{l\in L}\sum_{i\in I} X_{ijkl}\cdot(d_i + ct) \leq c_{jk}, \forall j \in J, \forall k \in K \tag{4.24}$$

$$\sum_{l\in L | l+d_i+ct\leq n}\ \sum_{i\in I} X_{ijkl} \leq c_{jk}, \forall j \in J, \forall k \in K \tag{4.25}$$

(4.26)

Expression (4.27) states that surgeries with a maximum scheduling time lower than the planning horizon must be scheduled. Expression (4.28) states that surgeries with a maximum waiting time lower than the planning horizon must be scheduled and inequality (4.29) states that the surgery day must be lower than the maximum waiting time. These constraints are equal to the ones used in the continuous model and are designed to respect patients' priority and waiting time rules.

$$\sum_{j\in J}\sum_{k\in K} X_{ijk} = 1, \forall i \in I_{maxsched} \tag{4.27}$$

$$\sum_{j\in J}\sum_{k\in K} X_{ijk} = 1, \forall i \in I_{maxwait} \tag{4.28}$$

$$\sum_{j\in J}\sum_{k\in K} X_{ijk} \cdot day_j \leq max_i, \forall i \in I_maxwait \tag{4.29}$$

The next group prevents overlap of patients in the same room and the overlap of patients from the same surgeon in different rooms. It is worth mentioning that these constraints are not required in the proposed continuous model. Thus, it is one of the main differences between the models. Constraint (4.30) prevents the overlap of surgeries in the same shift and OR also ensuring the cleaning time after each surgery, while constraint (4.31) avoids the overlap of patients of the same surgeon in different ORs in the same shift observing surgeons' turnover time.

$$\sum_{i\in I}\sum_{j'\in J|j'\geq j-d_i+1-ct \text{ and } j'\leq j} X_{ij'kl} \leq 1, \forall j \in J, \forall k \in K, \forall l \in L \tag{4.30}$$

$$\sum_{i\in I_s}\sum_{k\in K}\sum_{l'\in L|l'\geq 0 \text{ and } l'\geq l-di+tt+1 \text{ and } l'\leq l \text{ and } l'<n} X_{ijkl'} \leq 1, \forall s \in S, \forall j \in J \tag{4.31}$$

Finally, the last set concerns surgeon availability and workload. Constraint (4.32) restricts the scheduling of patients for a given surgeon to his/her availability and constraint (4.33) constrains the surgeon's workload in terms of number of working shifts per week.

$$\min\{1, \sum_{k\in K}\sum_{l\in L}\sum_{i\in I_s} X_{ijkl}\} \leq a_{js}, \forall j \in J, \forall s \in S \tag{4.32}$$

$$\sum_{j\in J_h} \min\{1, \sum_{k\in K}\sum_{l\in L}\sum_{i\in I_s} X_{ijkl}\} \leq ms, \forall s \in S, \forall h \in H \tag{4.33}$$

Chapter 5

Multi-objective Simulation Optimization for Surgery Scheduling under Uncertainty

Fabrício Sperandio[1], **José Borges**[1], **Bernardo Almada-Lobo**[1]

Technical Report, 2014

Abstract Surgical management processes are subject to high variability resulting in significant deviations between intended and actual performance of surgical plans. For instance, when surgeries take longer than predicted or emergency patients arrive, it often results in overtime and possible cancellation of surgeries. In order to control such effects, the variability in surgical processes should be embedded into scheduling models. This paper proposes a Simulation Optimization (SO) approach to the stochastic surgery scheduling problem. It integrates a multi-objective evolutionary algorithm (MOEA) to search for alternative surgery schedules with a discrete-event simulation (DES) model to estimate the schedule's performance un-

[1]CEGI – INESC TEC, Faculdade de Engenharia, Universidade do Porto, Portugal

der uncertainty. This multi-objective approach offers operating room (OR) managers a set of schedules to choose from instead of only one as in most stochastic approaches found in the literature. The aim is to devise schedules maximizing the number of performed surgeries and average occupancy rate as well as minimizing the number of cancellations and total overtime minutes. The schedule's performance is estimated using a DES model featuring four stochastic variables: surgery duration, emergencies, cancellations and delays/advances starting the first surgery in each shift. The proposed approach is compared with a standard deterministic MOEA based on fixed planned slacks. Moreover, the performance of each alternative configuration is evaluated using a comprehensive methodology for performance assessment of multi-objective stochastic optimizers. Experimental results show that SO outperforms planned slacks in all tested instances. Therefore, generating more realistic surgery schedules and offering decision makers more choices to choose from.

Keywords Multi-objective, Simulation Optimization, Operating Room, Scheduling, Uncertainty, Stochasticity

5.1 Introduction

Nowadays, healthcare managers are facing great challenges to preserve quality of care under a budget constrained scenario. On one hand, a set of structural forces, such as an ageing population and the introduction of new technologies, is driving a natural rise on healthcare costs. On the other hand, the recent financial crisis is forcing abrupt and extensive cost containments (de la Maisonneuve and Martins, 2013). For instance, the average healthcare expenditure among Organisation for Economic Co-operation and Development (OECD) countries was rising steadily until 2010, when it fell sharply, and until now did not recover its historical growth rates (OECD, 2013). In this context, healthcare managers need intelligent decision support tools to help them reduce costs without impacting quality of care.

Public administrations have been experiencing successive cuts on budgets. For instance, the Portuguese government has agreed with the European Commission to cut 30% on healthcare expenses on the period comprised between 2011 and 2013 (Ribeiro et al., 2011) and the budget for 2014 was 200 million Euro shorter than 2013. In this context, hospital care and specially its surgical activity represent major opportunities for cost reduction. Hospitals account for the largest share of national healthcare expenses. Likewise, the OR represents the major source of revenue as well as the largest cost center within a hospital. It is considered a core and expensive resource which influences many other pre and post-operation processes.

Management of surgical services encompasses a number of complex decision problems, such as: capacity planning, case mix planning, resource allocation, surgery scheduling and staff scheduling problems. These problems share three main characteristics which contribute to increase its complexity: a large number of alternatives, multiple stakeholders with sometimes conflicting objectives and high uncertainty. The first two characteristics have been subject to an extensive number of studies in the field of operations research applied to healthcare. However, the last characteristic has received considerably less attention. For instance, Guerriero and Guido (2011) concluded that the majority of published papers assume that processing times and recovery times are known in advance. In addition, Cardoen et al. (2010a) highlighted that only limited research has been applied to non-elective patient scheduling. Such class of patients encompasses emergencies and high-priority cases whose arrival is highly uncertain. In this context, how to deal with uncertainty in OR management problems still represents an open challenge.

Uncertainty is an intrinsic characteristic of OR planning and scheduling problems related to the human nature of the activities performed. According to May et al. (2011) surgery scheduling is a challenging task because "every detailed plan is almost certainly to deviate significantly from what actually transpires in the course of a surgical day". Nevertheless, taking uncertainty into account requires more complex models and respectively higher computational costs. This explains the trend of researchers to focus on deterministic approaches (Cardoen et al., 2010a). However, it also results in unrealistic plans with low performance in practice compromising the acceptance of optimization tools among doctors and hospital managers. For instance, uncertainty in the actual surgery duration impacts OR occupancy rates and patient waiting times. More specifically, if a surgery is shorter than predicted, resources may not be ready to start the next one and OR becomes idle resulting in low occupancy rates. On the other hand, if a surgery takes longer than predicted, subsequent surgeries have to be postponed resulting in patient waiting time, human resource's (HR) overtime and ultimately in cancelled surgeries.

Computer Simulation is considered the most suitable method to address OR management problems under uncertainty (Guerriero and Guido, 2011). It allows analysts to build more detailed models including relevant aspects of the problem that are harder (or even impossible) to model with other approaches. Furthermore, Simulation Optimization (SO) offers an extensive set of methods for optimizing simulation models as well as for reducing the required computational time. The growth of SO literature allied to a low number of applications to OR planning and scheduling problems configures a research opportunity. Solution approaches designed specifically for SO problems are able to reduce the required computational cost exploring statistical information of simulation samples.

This study proposes a multi-objective simulation optimization approach to the surgery scheduling problem under uncertainty. This approach encompasses an optimization component and a simulation component. The former features a multi-objective evolutionary algorithm (MOEA) to find surgery schedules which maximize the number of performed surgeries and the average OR occupancy rate as well as minimize the number of cancelled surgeries and total overtime minutes. The latter features a Discrete-Event Simulation model including four sources of uncertainty: surgery duration, emergencies, cancellations and delays/advances starting the first surgery in each shift.

The contribution of this paper is three-fold. To the best of our knowledge, it is the first multi-objective optimization approach to tackle the general stochastic surgery scheduling problem. In this solution approach the OR manager is provided a set of surgery schedules to choose from, illustrating the trade-off between conflicting objectives. Moreover, it is the first approach to take into account four important sources of uncertainty arising in a large Portuguese hospital and to model surgery duration considering its main determinant attributes. Finally, it tackles scheduling and sequencing decisions at once, allowing surgeons to change between ORs within the same shift, which is a common assumption in the context of this study and allows to improve OR occupancy rates.

Computational experiments are performed on instances built with real data from a large Portuguese hospital. First, a deterministic version of the algorithm is tested with alternative planned slacks. Second, the proposed simulation optimization approach is evaluated with an alternative number of replications. Finally, a comparison between the best configurations of each approach is performed. The evaluations and comparisons are based on a comprehensive methodology for performance assessment of multi-objective optimizers including a combination of quality indicators and suitable statistical tests to assess the statistical significance of the results.

The remainder of this paper is organized as follows: literature review, problem description, solution approach, computational experiments, discussion and future work. The first section reviews stochastic approaches to the surgery scheduling problem. The solution approach section is split into two subsections describing in detail the two components of the integrated solution: optimization and simulation. The computational experiments section describes the experiments performed, the methodology applied to evaluate them and their respective results. Finally, the last section summarizes the study highlighting the strengths and weaknesses of the proposed approach and pointing out areas for future work.

5.2 Literature Review

The management of surgical services encompasses a set of complex planning and scheduling problems. In order to reduce such complexity researchers classify problems into three decision levels: strategic, tactical and operational. In the strategic decision level, the case mix planning problem consists in determining the number and type of surgeries to be performed by each surgical specialty in the long term. In the tactical level, the master surgery scheduling problem consists in building a weekly time-table determining the operating rooms (ORs) assigned to each specialty in each day of week. Finally, in the operational level, the surgery scheduling problem consists in selecting a sub-set of patients from the elective surgery waiting list and determining a surgery date, OR and starting time for them. This review focuses on stochastic approaches for the operational problem. For a complete review on surgical management problems see Cardoen et al. (2010a), Guerriero and Guido (2011) and May et al. (2011). Table 5.1 summarizes the main characteristics of all papers that, to the best of out knowledge, address the surgery scheduling problem under uncertainty. Papers are sorted by sub-problem in order to group similar characteristics. A paper may only partially show a characteristic, which in these cases is explained in the following paragraphs.

The surgery scheduling problem at the operational decision level can be decomposed into two sub-problems: advance and allocation scheduling problems. According to Table 5.1 these problems have been addressed separately. The first columns show that the majority of the studies addressed the advance problem alone and only one has the integrated problem of advance and allocation scheduling. In general, the advance scheduling problem consists in selecting a sub-set of patients from the waiting list and assigning them to a specific OR and day over a weekly planning horizon. However, there are small variations of this sub-problem. For instance, Lamiri and Augusto (2008); Lamiri et al. (2009) focus only on determining the set of elective patients to be operated in each day, leaving the assignment of a specific OR to a later stage. Moreover, in Hans et al. (2008) the set of patients to be scheduled in a given week is pre-defined and no patient is postponed for the next planning period. In its turn, the allocation scheduling problem consists in sequencing the surgeries in each OR-day. Studies addressing this problem usually consider multiple ORs. In contrast, Denton et al. (2007) and Mancilla and Storer (2011) consider only a single OR. We propose to integrate advance and allocation scheduling problems as well as consider multiple ORs. Addressing both problems simultaneously leads to better solutions to the overall problem as, assuming a surgeon is allowed to change ORs during the same working shift, often the best solution to the allocation problem requires changing the advance scheduling solution.

The main objective addressed in stochastic versions of the surgery scheduling problem is to reduce the risk of overtime. Table 5.1 shows that 11 studies take this objective explicitly into account. In contrast, Shylo et al. (2012) and Addis et al. (2014) do not consider it in the objective function. However, these studies rely on robust optimization, which guarantees acceptable levels

Table 5.1: Summary table of the literature review

	Problems		Objectives					Resources				Constraints			Uncertainty				Solution Approaches			
Reference	A	B	A	B	C	D	E	A	B	C	D	A	B	C	A	B	C	D	A	B	C	D
Addis et al. (2014)	•		•					•	•			•	◦		•				•			
Bruni et al. (2014)	•		◦			◦		•	•			•	◦		•	•			•			
Hans et al. (2008)	◦			•		•		•	•		◦	•			•					•		
Lamiri and Augusto (2008)	•		•			•		◦	•			•	•			•			•			
Lamiri et al. (2009)	•		•			•		◦	•			◦	◦			•			•	•		
Min and Yih (2010)	•					•		•	•			•			•	•			•			
Rachuba and Werners (2014)	•		•	•		•		•	•			•			•	•			•			
Shylo et al. (2012)	•			•					•			•			◦				•			
Batun et al. (2010)		•		•		•			•	•	◦	•			•				•			
Denton et al. (2007)		•		•		•	◦					•			•				•	•		
Gul et al. (2011)		•				•	•		•		◦	•			•					•	•	•
Lee and Yih (2014)		•		•		•	•		•		◦	•			•				•			
Mancilla and Storer (2011)		•		•		•	•					•			•				•	•		
Dexter and Macario (1999)	•	•		•				◦	•	•		•	◦		•					•	•	
# papers (out of 14)	9	6	5	8	0	11	4	8	12	2	4	14	5	0	12	5	0	0	11	6	2	1
This paper	•	•	•	•	•	•		•		•	•		•	•	•	•		•	•	•		

Problems
(A) Advance scheduling
(B) Allocation scheduling

Objectives
(A) Max. number of performed surgeries
(B) Max. OR occupancy rate
(C) Min. number of cancelled surgeries
(D) Min. total minutes of overtime
(E) Min. patient and/or OR team waiting in the process flow

Resources
(A) Time blocks (OR-days)
(B) Multiple ORs
(C) Multiple surgeons
(D) Pre/post-surgical resources, surgical staff

Constraints
(A) Resources capacity
(B) Patient priority and waiting time
(C) Surgeon availability and working limits

Sources of uncertainty
(A) Surgery duration
(B) Emergencies
(C) Cancelled surgeries
(D) Delays/advances starting the first surgery in each shift

Solution Approaches
(A) Stochastic programming, robust optimization or chance constraints
(B) Heuristics or metaheuristics
(C) Discrete-Event Simulation
(D) Multi-objective optimization

of overtime. Other objectives are closely related to the specific sub-problem being addressed. Studies addressing the advance scheduling problem focus on minimizing patient-related costs. These costs are associated with patient waiting time in the waiting list, urgency and tardiness (maximum waiting time). In other words, such approaches aim to maximize the number of patients scheduled and establish an order among them. In addition, studies in this category aim to maximize OR occupancy rates. Besides reducing overtime, studies addressing specifically the allocation problem focus on reducing waiting time in the process flow, synchronizing the utilization of resources. Nevertheless, the perspective may be different, since a set of studies focus on the patient (Denton et al., 2007; Gul et al., 2011) and another on clinical resources (Batun et al., 2010; Lee and Yih, 2014; Mancilla and Storer, 2011). We propose to take four objectives into account: (1) maximize the number of performed surgeries, (2) maximize average OR occupancy rates, (3) minimize the number of cancelled surgeries and (4) minimize total minutes of overtime. The third objective was not explicitly addressed by any of the reviewed papers. It is often considered a result of excessive overtime. We explicitly consider it an objective because "lack of OR time" is a common reason for cancelling surgeries in the hospital under analysis and must be controlled. Cancelled surgeries reduce patient quality of service, increase hospital costs and impact subsequent elective schedules.

In studies addressing the advance scheduling problem only, time blocks are the main resources. In general, in these studies, a time block consists in a combination of OR and day. In contrast, Dexter and Macario (1999) consider only surgeon block time, a combination between surgeon and day, and Lamiri and Augusto (2008); Lamiri et al. (2009) consider only days of the planning horizon. Also, Rachuba and Werners (2014) are the only to consider two surgical blocks per room each day, i.e. morning and afternoon. On the other hand, in studies addressing the allocation scheduling problem only, the bottleneck are ORs. In addition, Batun et al. (2010) consider the intensive care unit (ICU), Lee and Yih (2014) address the post-anaesthesia care unit (PACU) and Gul et al. (2011) look at pre/post-surgical resources (waiting area and intake/recovery rooms). Hans et al. (2008) are the only to consider additional OR personnel. Finally, regarding resources, only two of the reviewed papers take surgeons explicitly into account (Dexter and Macario, 1999; Batun et al., 2010). Papers that do not consider it require general assumptions about the surgeon workload and availability. Often, surgeons are pre-assigned to specific time blocks on a previous stage and do not change rooms in the same day. We propose to consider surgeons explicitly which allows a surgeon to work in more than one OR in the same working shift. It helps to increase OR occupancy rates since surgeons are available to start another surgery in a different OR without waiting for the cleaning of the previous OR, as well as promote the productivity of the surgeon.

Naturally, all reviewed papers include resource capacity constraints. In contrast, just a few include additional business logic constraints. Exceptions are OR cleaning times and surgeon turnover times. Patient urgency and waiting time limits are often addressed using penalties in the objective function. Table 5.1 indicates the papers which address this issue in the objective function. In addition, Rachuba and Werners (2014) consider the first feasible day for a surgery and limit the maximum amount of overtime. The first feasible day derives from restrictions in the clinical pathway, which may include pre-surgical analysis. Beyond constraints in the number and availability of resources, we propose to limit surgeon daily and weekly workload as well as consider a surgery due date which is determined by the patient urgency and waiting time. These constraints are derived from the Portuguese legislation.

The variability in surgery durations is the main source of uncertainty taken into account in the literature. Moreover, few papers take into account the

OR-time occupied by emergencies and Min and Yih (2010) consider the length of stay in the intensive case unit. It is worth mentioning that the approaches to model the behaviour of the stochastic variable representing the uncertain surgery durations differs broadly. They vary from fitting probability distributions to historical data (Min and Yih, 2010) and sampling directly from historical data (Denton et al., 2007) to using uniform probability distributions with fixed parameters (Lamiri and Augusto, 2008; Lamiri et al., 2009). In addition, Batun et al. (2010) decompose the surgery duration in pre-incision, incision and post-incision and Shylo et al. (2012) consider the distribution of the sum of durations only. Regarding how historical data is grouped to be analysed, most approaches group it by surgical department. For instance, Min and Yih (2010) highlight that in practice the surgery duration depends on the surgery type, the surgeon and the patient. However, the study assumes all surgeries in the same surgical department follow identical probability distributions, usually a log-normal one. In contrast, Hans et al. (2008) cluster surgeries into 4 to 8 categories within each surgical department sharing the same mean and standard deviation. We propose to take into account 4 sources of uncertainty and to model the behaviour of surgery durations using its main predictive factors.

Concerning the solution approaches most papers rely on Stochastic Programming. In particular, the formulation of two-stage problems and its resolution by Monte Carlo sampling and the Sample Average Approximation method. In order to reduce the computational cost researchers have been applying decomposition approaches such as Bender's decomposition (Mancilla and Storer, 2011) and the L-Shaped method (Batun et al., 2010). In addition, solution approaches based on constructive and improvement heuristics as well as meta-heuristics have been applied to solve real size instances. Heuristic approaches usually explore statistical information on the variability of surgery durations based on historical data. Finally, Shylo et al. (2012) propose a chance-constrained model to ensure acceptable levels of overtime and Rachuba and Werners (2014) apply fuzzy sets to merge the interests of different stakeholders. We propose an approach based on simulation optimization combining a multi-objective evolutionary algorithm and discrete-event simulation (DES).

Simulation optimization is an active research area within the Computer Simulation field and healthcare is one of the main application areas of Computer Simulation. Guerriero and Guido (2011) highlight that due to its modelling flexibility, Computer Simulation is the most reliable and efficient tool to address the complexity and stochasticity that arises in healthcare management problems. It has been successfully applied to perform scenario (what-if) analysis. For instance, Azari-Rad et al. (2014) propose a DES model for perioperative process improvement and Konrad et al. (2013) a DES model targeting the emergency department (ED). However, this process considers only a limited number of alternatives. When the number of alternatives is high some sort of optimization procedure is required to search for the best ones. The integration between Computer Simulation and optimization tools have been given multiple names, e.g. Simulation-based Optimization (Shapiro, 1996), Optimization via Simulation (Fu, 1994), Simulation Optimization (Fu, 2002). In this paper we use the latter definition.

In simulation optimization, the optimization role is to search for alternative solutions to the underlying optimization problem and the simulation role is to evaluate its performance under uncertainty. In the last decades, several authors published literature reviews about simulation optimization (Fu, 1994; Glover et al., 1999; Fu, 2002; April et al., 2003; Fu et al., 2005, 2008; Hong and Nelson, 2009; Figueira and Almada-Lobo, 2014), most of them in the proceedings of the Winter Simulation Conference. For instance, Fu (2002) presented an extensive literature review on the topic describing the main solution approaches and discussing efficiency issues. The author highlights that different from deterministic optimization, in simulation optimization the estimation cost is higher than the search cost - and discusses the integration of statistical procedures to deal with the stochastic nature of the problem. The referred estimation cost is determined by the number of simulation replications performed to estimate the performance of each alternative solution.

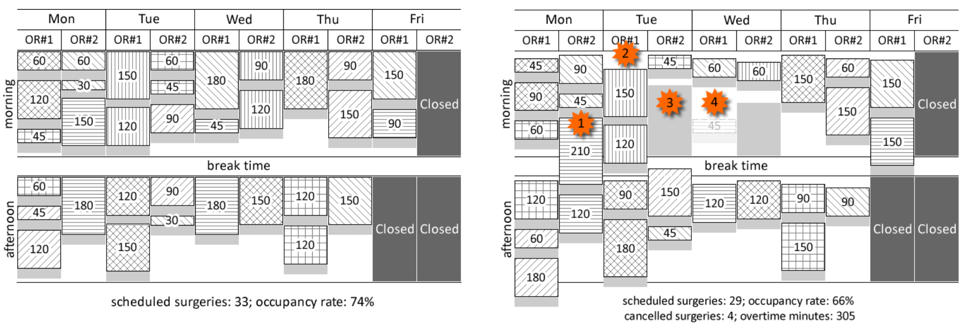
5.3 Problem Description

This paper focuses on the stochastic surgery scheduling problem at the operational decision level. The problem consists in selecting a sub-set of patients from the elective surgery waiting list and assign a surgery date, operating room and starting time for them. Thus, it integrates simultaneously advance and allocation scheduling problems. In addition, there is a problem of assigning sufficient planned slack to each working shift to deal with unveiled uncertainty.

In Portugal, the Integrated Management System of Registered Patients for Surgery ("Sistema Integrado de Gestão de Inscritos para Cirurgia" - SIGIC) (Ministério da Saúde, 2011) program was introduced in 2004 to tackle excessive waiting times. Once the need for an elective surgery is identified, patients are added to the waiting list in their main hospitals and wait for their surgeries to be scheduled. If the surgery is not scheduled within 75% of the maximum waiting time according to each priority level the patient is allowed to perform the surgery in another hospital, either in public or private networks, and his origin hospital is responsible for paying the treatment. Marques et al. (2012) present a deterministic approach for surgery scheduling in Portuguese hospitals.

Surgery schedules are built for each surgical department on a weekly basis. Every Thursday, the head of each surgical department is responsible for launching the schedule for the following week. The schedule is elaborated manually, a task that consumes time that could be applied to perform clinical activities or perform what-if-analysis to different plans. In fact, a decision support system to support scheduling activities in Portuguese hospitals was proposed in chapter 3. In order to illustrate the scheduling problem, Figure 5.1(a) shows a valid weekly schedule for a hypothetical surgical department.

The example schedule in Figure 5.1(a) shows a standard working week with 2 ORs in each day, designated by OR#1 and OR#2. The ORs operate in two working shifts (morning and afternoon), with a time break between them. Note that some ORs may be closed in specific shifts and days of week. Hereafter, an open OR in a given shift and day of week is defined as a time block. Moreover, in this example, scheduled surgeries are represented as boxes inside each time block and numbers inside each box represent respective surgery durations. The different graphic patterns indicate different surgeons, the required cleaning time after each surgery is represented in light gray and the empty space at the end of each time block represents the planned slack (idle time).



(a) Weekly surgery schedule: planned (b) Weekly surgery schedule: executed

Figure 5.1: Impact of uncertainty in the weekly surgery schedule

Figure 5.1(b) shows the example schedule after its execution. In this example the number inside each box represents actual surgery durations. It is worth of note the impact of each source of uncertainty. The sign "1" indicates a surgery which took much longer than predicted resulting in overtime in this shift. Sign "2" shows a delay on the starting time of the first surgery in the morning resulting in OR underutilization. Sign "3" indicates OR time being occupied by unexpected emergencies resulting in the cancellation of previously scheduled elective surgeries. Sign "4" indicates OR underutilization as a consequence of cancelled elective surgeries. Since schedule performance measures are affected by uncertainty our aim is to optimize the estimated performance measures of the execution of the plan.

In summary, our goal is to optimize the following four objectives: (1) maximize the number of surgeries performed; (2) maximize the average OR occupancy rate; (3) minimize the number of surgeries cancelled; (4) minimize the total overtime. Feasible surgery schedules are subject to the following six families of constraints: (1) the duration of the surgeries (plus cleaning times) within each time block must not exceed the time block's length; (2) a surgery must not be scheduled to end after OR closing time; (3) patient priority and waiting time rules imposed by the Portuguese legislation must not be violated; (4) a surgeon must not be scheduled to work for more than a certain number of hours a day and a certain number of hours a week; (5) lower bound on the time between consecutive surgeries in the same OR; (6) lower bound on the time between consecutive surgeries of the same surgeon in different ORs.

Finally, we assume that other human and material resources do not compromise the implementation of the proposed plans. For instance, the operating rooms work with fixed nursing teams and the capacity of the post-anaesthesia care unit and surgery wards are not a bottleneck.

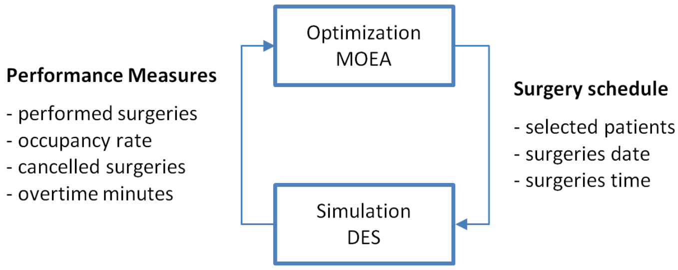


Figure 5.2: The Simulation Optimization loop

5.4 Solution Approach

The proposed solution approach encompasses optimization and simulation modules. The optimization module aims to search for solutions to the problem, while the simulation module assesses the performance of each alternative solution under uncertainty. The former module features a multi-objective evolutionary algorithm and the latter incorporates a discrete-event simulation model. The integration between the two modules takes place in the fitness evaluation function of the MOEA. In this step the simulation model runs a pre-determined number of replications and the average performance measures are calculated. Thus, simulation average performance metrics become the optimization objectives. Figure 5.2 illustrates the integration between simulation and optimization.

5.4.1 Multi-Objective Evolutionary Algorithm

The optimization module implements a multi-objective evolutionary algorithm. The actual algorithm is a customized version of the NSGA-II (Deb et al., 2002) algorithm for multi-objective optimization. The NSGA-II implements the concept of crowding distance, which is a measure of how close an individual/solution is from its neighbours. Large average crowding distance will result in better diversity in the population. The algorithm is a genetic algorithm (GA) which evolves a population of solutions towards the set of optimal Pareto solutions. This set comprises alternative surgery schedules representing trade-offs between conflicting objectives. The following paragraphs describe the encoding scheme and genetic operations. The fitness evaluation is performed invoking the DES model.

The encoding scheme and respective decoding procedure are key determinants of MOEA performance. We propose an encoding scheme based on a vector of real variables and a two-phase decoding procedure to translate each GA chromosome into a feasible solution. This encoding scheme is based on the biased random-key genetic algorithm (BRKGA) proposed by Gonçalves and Resende (2011). Preliminary results show that this approach outperforms approaches based on encoding schemes using binary variables, like the one proposed by Conforti et al. (2010). In fact, GAs were originally designed for unconstrained optimization problems, therefore they are more efficient searching in the feasible solution space only.

Figure 5.3 illustrates an example GA chromosome representation using real variables. For simplification purposes, in this figure as well as in Figure 5.4, only the first 5 surgeries and the last one appear. Each individual in the population represents a valid surgery scheduled and is associated to one of these chromosomes. Each real variable is assigned a random number, known as random key, ranging from 0 to 1. Furthermore, each chromosome is split in two parts. The first part determines the sequence in which surgeries are scheduled inside each time block, while the second part determines the planned slack assigned to each time block. Each random number in the first part of the chromosome corresponds to one surgery in the waiting list. Also, each random number in the second part corresponds to one of the available time blocks. A special decoding procedure translates each chromosome into an admissible surgery schedule.

Figure 5.4 illustrates the decoding procedure. First, Figure 5.4(a) shows an

	assign scheduled surgeries							assign planned slacks					
identifiers	1	2	3	4	5	...	n	1	2	3	4	...	m
random numbers	0,0904	0,4173	0,4591	0,2173	0,1930		0,6105	0,9158	0,8569	0,7261	0,9563		0,5371

Figure 5.3: An example GA chromosome representation based on real variables

example unordered input set of surgeries with associated surgeon, expected duration and random number. The associated surgeon and expected surgery duration are inputs of the problem, while the random numbers are assigned every time a chromosome is created. In Portugal, the main surgeon in charge is associated to the respective elective surgery at the moment of the waiting list registration. Next, Figure 5.4(b) illustrates an important step of the decoding procedure which consists in sorting the set of surgeries by ascending order of random numbers: the order in which surgeries are scheduled is determined. Next, Figure 5.4(c) shows the associated time blocks and starting times assigned by the decoding procedure and Figure 5.4(d) illustrates the resulting surgery schedule. For simplicity, the example schedule highlights only the first five surgeries. Moreover, it shows a solid line at the end of each open time block. This line represents the maximum end time for scheduled surgeries, resulting from multiplying the random number associated with each time block in the chromosome by 60 (minutes) - in practice the necessary slack per shift is not given more than one hour. The space between this line and the end of the time block is the planned slack. In summary, the decoding procedure consists in going through the set of surgeries in ascending order of random numbers and schedule each surgery in the next time block it fits (considering the planned slack). Time blocks are sorted in ascending order of day of week, operating room and working shift. The following paragraph describes the procedure in detail.

The algorithm to decode a chromosome into a feasible surgery schedule is composed of two phases. The first phase generates schedules meeting all

surgery	surgeon	expected duration	random number
1	3	120	0,0904
2	4	180	0,4173
3	5	180	0,4591
4	2	60	0,2173
5	1	120	0,1930
...	...	...	...
n	3	150	0,6105

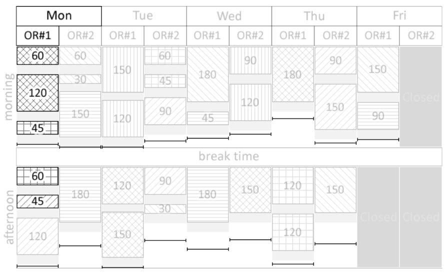
(a) Input set of surgeries with associated random numbers

surgery	surgeon	expected duration	random number
2	1	60	0,0038
4	1	120	0,0099
40	2	45	0,0103
21	2	60	0,0265
3	3	45	0,0406
...	...	...	...
8	2	60	0,9654

(b) Input set ordered by ascending order of random numbers

surgery	surgeon	expected duration	random number	schedule	time block	starting time
2	1	60	0,0038	TRUE	1	0
4	1	120	0,0099	TRUE	1	90
40	2	45	0,0103	TRUE	1	240
21	2	60	0,0265	TRUE	2	0
3	3	45	0,0406	TRUE	2	90
...	...	...	...	...	...	...
8	2	60	0,9654	FALSE	-	-

(c) Time block and starting timed assigned according to the decoding procedure



(d) Resulting example surgery schedule

Figure 5.4: An illustrative example of the decoding procedure

the requirements but the patient priority and waiting time rules, which are tackled in the second phase. Algorithm 3 illustrates the complete decoding procedure, which is described in high-level in the following lines and with more detail about each function in the following paragraph. The first phase encompasses lines 4-15 and the second lines 16-19. It starts by iterating through the set of time blocks (line 5) and through the set of surgeries (line 7). If a surgery meets all the requirements to be scheduled in the current time block (line 8), the schedule is confirmed (lines 9 to 13). Otherwise, the inner loop breaks and another time block is evaluated (line 15). When all the time blocks are evaluated the first phase is completed. Next, the second phase consists in iterating through the set of surgeries and checking if each surgery scheduled meets the respective patient priority and waiting time rules (line 17). In case they do not meet, a new random number is generated (line 18) which forces the surgery to be scheduled until its due date. In case the schedule does not meet the waiting time rules, the procedure repeats the main loop and with the new random numbers the surgeries should be assigned to feasible dates. On the other hand, if the solution meets all the requirements, the procedure completes and returns a solution, otherwise the main loop is repeated.

Algorithm 3: Procedure for decoding a chromosome encoded with random keys into a feasible surgery schedule

```
Data: GA chromosome
Result: Feasible surgery schedule
begin
    solution ⟵ getInitialSolution(chromosome)
    repeat
        currentIndex ⟵ 0
        for i ← 0 to nTimeBlocks do
            startTime ⟵ 0
            for j ← currentIndex to nSurgeries do
                if timeBlockCapacity(i,j,solution) and
                surgeonWorkload(i,j,solution) and
                surgeonAvailability(i,j,solution) then
                    solution[j].timeBlock ⟵ i
                    solution[j].scheduled ⟵ true
                    solution[j].startTime ⟵ startTime
                    currentIndex ⟵ currentIndex + 1
                    startTime ⟵
                    startTime + solution[i].duration + cleaningTime
                else
                    break
        for i ← 0 to nSurgeries do
            if not priorityAndWaitingTime(i, solution) then
                solution[i].randomNumber ⟵
                newRandomNumber(i, solution)
    until isFeasible(solution)
    return solution
```

The getInitialSolution procedure gets the chromosome as an array of random numbers and returns an initial solution. First, the procedure inserts one surgery object in the solution array for each surgery in the chromosome. Next, it sorts the solution array by the random numbers assigned to each surgery. Initially, each surgery in the solution has the schedule property set to false. The timeblockCapacity procedure checks if the current surgery does not exceed the capacity of the time block. The surgeonWorkload procedure checks if the current surgery does not violate surgeons' daily and weekly workloads. The surgeonAvailability procedure checks if the same surgeon is not scheduled to be working on another OR at the same time. If that is the case, then the procedure delays the start of the current surgery until the end of the previous one. The priorityAndWaitingTime procedure checks if patient's maximum schedule date is met. In other words, for instance, if a patient must be scheduled until Tuesday and the procedure schedules to Friday, or not schedule at all, it breaks the rule. In these cases, the newRandomNumber procedure samples a new random number for these patients. The new random number is sampled from 0 to the maximum number of surgeries scheduled in the latest to avoid breaking the rule.

The crossover and mutation operators introduce diversity into the populations. Its impact is controlled by crossover and mutation rates parameters. The proposed GA uses the simulated binary crossover (SBX) and polynomial mutation operators. These operators were proposed by Deb et al. (2002) for real-coded GAs.

5.4.2 Discrete-Event Simulation

The simulation module of the integrated solution approach implements a stochastic Discrete-Event Simulation model. The model structure and behaviour are based on the Adevs (Nutaro, 2010) framework for fast discrete event simulation. This framework was selected based on its performance, flexibility and scalability. Indeed, performance is a key requirement of any simulation optimization approach. Flexibility and scalability are also key requirements to build complex OR models with different resources and complex relationships among them. The stochasticity is modelled with 4 random variables whose behaviour is based on historical data from a large Portuguese hospital.

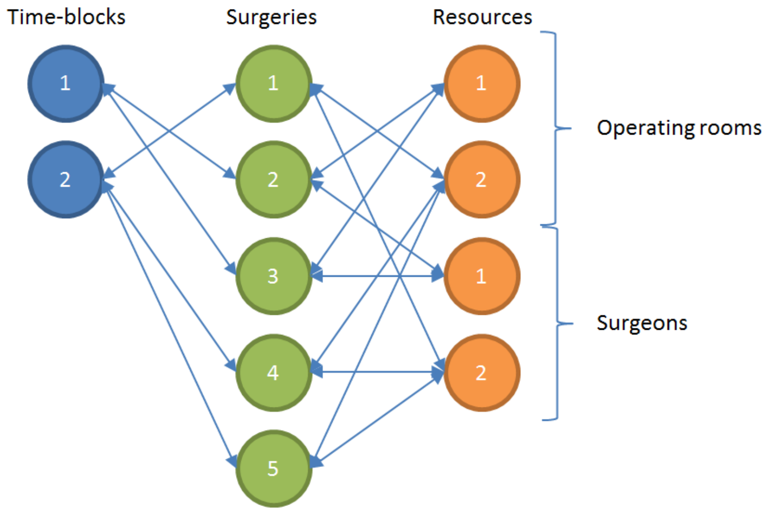


Figure 5.5: A simple network model showing the three types of components and the connections among them

The simulation model consists in a network of atomic models connected through input and output ports enabling the exchange of messages among them. Each atomic model implements the behaviour of a specific component of the system. In this case, the proposed implementation uses three types of components: Surgery, Resource and Time block. Figure 5.5 illustrates the network with arrows representing the connection between the components. It should be highlighted that the arrows are double-sided meaning that each component sends output messages and receives input messages from the components connected to it.

In our case, components of type Resource are used to model the behaviour of surgeons and operating rooms, but can be extended to model other resources. Moreover, each atomic model is a state machine, characterized by a set of states and state transition functions. There are two types of state transition functions: internal and external. Internal functions are called when an internal event occurs, for instance, the end of a surgery. External functions are called when a component receives a message from another component, for instance, at the end of a surgery the Surgery component

Surgery	Resource	Time block
1. Pre-operative	Queue	Queue
2. Ready	1. Free	
3. Requesting	2. Busy	
4. Waiting		
5. Working		
6. Cleaning		
7. Releasing		
8. Post-operative		

Table 5.2: Set of sequential states of each atomic model

sends a message to the Resource components to release them.

Table 5.2 lists the set of sequential states in each component. These states change with the exchange of messages between the components. Initially, the time block component has one surgery in the queue, the surgery is on state 'pre-operative', and both resources are on state 'free'. Next, the time block component sends a message to the surgery component, which makes the surgery change state to 'requesting' and send messages to the two required resources (surgeon and operating room). Both resources receive the messages, change state to 'busy' and send a message back to the surgery. The surgery receives both messages, change state to 'working' and schedules the next event to the end of the simulated surgery. When the scheduled event is triggered, the surgery changes state to 'releasing' and sends messages to the resources used to release them. Both resources receive the messages, change state to 'free' and send a message back to the surgery. The surgery receives the messages, changes state to 'post-operative' and sends a message to the time block requesting another surgery.

Note that each surgery starts as soon as the required resources are available (operating room and surgeon), assuming the other resources are ready. If surgeries were started only after the scheduled time, the amount of overtime and cancelled surgeries would be presumably much higher.

The stochastic behaviour of the simulation is modelled by 4 stochastic variables: (1) duration of surgical procedures; (2) cancelled surgeries; (3) total

Source of uncertainty	Attributes
Surgery duration	specialty, combination of surgical procedures, main surgeon in charge
Emergencies	specialty, operating room, day of week, working shift
Cancelled surgeries	specialty
Delays/Advances	specialty, working shift

Table 5.3: Attributes used for modelling the behaviour of stochastic variables

time (in minutes) occupied by emergencies; (4) delays/advances (in minutes) on the start of the first surgery in each shift. Table 5.3 lists the attributes used to query the database for historical data and model the behaviour of each stochastic variable. The aim is to reduce variability, which benefits the simulation optimization approach, and to create a more realistic simulation model.

Simulation model performance measures are computed at the end of each simulation based on the analysis of the simulated ending times of each surgery. The computed performance measures become the fitness values of each solution of the MOEA. The number of performed surgeries is the number of scheduled surgeries with simulated starting times before the surgical suite closing time. The number of cancelled surgeries represents the number of scheduled surgeries with simulated starting times after the surgical suite closing time (we assume that these surgeries are cancelled by lack of OR time, which is common practice in the hospital under analysis). The estimated occupancy rate is the sum of the simulated durations of each performed surgery over the total time block's capacity. Note that it does not include turnover times. The total overtime is the difference between surgery's simulated ending time for each surgery that ends after surgical suite's closing time and the surgical suite's closing time.

5.5 Computational Experiments

5.5.1 Types of Experiments

This section describes the computational experiments, testing instances and performance assessment methodology used for evaluating the proposed simulation optimization approach. In summary, three sets of results are analysed: the results of two computational experiments with different versions of the MOEA and the comparison between them. The first experiment consists in running a deterministic version of the MOEA with alternative planned slacks. The second experiment consists in running the proposed simulation optimization approach with alternative number of replications. Finally, the comparison is made between the best configuration of each experiment.

The first experiment consists in running a deterministic version of the MOEA. It is similar to standard deterministic approaches found in the literature. Also, it shares the same encoding scheme and genetic operators with the simulation optimization version, but aims at maximizing only two objectives: the number of scheduled surgeries and the occupancy rate. During the search, objective values are computed by an analytical function. The other two objectives can only be estimated by means of simulation. The search runs for 1 minute and then each solution is simulated 1000 times to estimate the 4 performance measures with high confidence. At this point, the objective values associated to each solution are the samples' averages. Moreover, three different planned slack configurations are tested: 0, 10% and 20%. In the first configuration no planned slack is used. In the other two a percentage of the time block's total length is left empty (in the end) to prevent overtime and cancelled surgeries in case of unexpected events.

The second experiment consists in running the proposed simulation optimization approach with a varying number of simulation replications, in order to estimate the performance measures for each solution. Six alternative configurations are evaluated, each with the following number of replications: 5, 25, 50, 75, 100 and 150. This experiment aims to evaluate the impact of the number of replications in the algorithm's performance and to determine the configuration which provides the best performance under a fixed time limit. All experiments run for 1 min and final non-dominated solutions are simulated 1000 times.

After the two computational experiments, the best configurations of the deterministic and the simulation optimization approaches are compared. It enables us to determine the benefits of the proposed simulation optimization approach over a standard deterministic approach. Both configurations run for the same fixed amount of time (1 min). Thus, simulation optimization should be much more efficient since its computational cost is higher and the number generations performed is several times smaller.

Algorithm 4: Sequence of steps performed to assess the performance of each alternative configuration

```
begin
    for all specialties do
        for all configurations do
            for i ← 0 to 30 do
                // runs the MOEA for 1 min
                runExperiment(1)
                // simulates final non-dominated solutions 1000
                   times
                simulateFinalParetoSet(1000)
            normalizeObjectives()
            computeEmpiricalAttainmentFunctions()
            findReferencePoint()
            computeHypervolumeIndicator()
            findReferenceSet()
            computeEpsilonIndicator()
            computeRIndicator()
            performKruskalWallisStatisticalTest()
```

Table 5.4: Characteristics of the testing instances

Surgical specialty	# Patients	# Surgeons	# Procedures	# Time blocks	Surgery duration Avg. Length	Std. Dev.
Vascular surgery	115	13	4	9	60	29
Oral and maxillofacial surgery	58	10	18	3	44	26
Neurosurgery	66	9	30	18	188	95
Ophthalmology	499	34	61	24	41	23
Orthopaedics	126	21	57	22	97	56
Urology	109	16	45	12	89	51
Otolaryngology	80	16	47	9	82	29
General surgery 1	49	9	27	9	172	101
General surgery 2	51	6	26	8	114	42
General surgery 3	48	7	10	7	117	50

5.5.2 Testing Instances

The computational experiments are performed over a set of 10 testing instances built with real data from a large Portuguese hospital. Each instance concerns a different surgical specialty and represents different testing settings in terms of the size of the problem and the degree of uncertainty. Table 5.4 describes the characteristics of the testing instances. Ophthalmology and Vascular surgery are the most demanding instances.

The procedure to generate the instances consisted in consulting the surgical waiting list on a given date and selecting patients from higher to lower priority and waiting time until the sum of the expected surgery durations reaches twice the capacity of the time blocks. Indeed, this is the procedure suggested in the surgical waiting list's manual (Ministério da Saúde, 2011).

5.5.3 Performance Assessment Methodology

The performance assessment methodology applied in the evaluation of results relies on the literature about performance assessment of stochastic multi-objective optimizers, mainly on the studies presented by Knowles et al. (2006) and Zitzler et al. (2003). The methodology uses the dominance ranking approach, a combination of quality indicators, empirical attainment functions and the respective statistical testing procedures to assess the statistical significance of the results. The results of the different approaches are evaluated in the following order: dominance ranking, quality indicators and empirical attainment functions. Algorithm 4 shows the steps performed to compute the performance measures for all specialties and alternative configurations. Note that, for each configuration, the RunExperiment function is called 30 times. This function runs the MOEA with each configuration's parameters for 1 minute. Next, to estimate performance measures with high confidence, the solutions in the final Pareto approximation set are simulated 1000 times.

The dominance ranking approach consists in performing a non-dominated sorting on the combined set of all approximation sets generated by one or more alternative configurations being compared. Next, a statistical rank test is applied to pairs of configurations to determine whether the ranks associated to one of them are significantly lower than the ranks associated to the other. This approach is able to determine the best configuration in case one configuration is significantly better than another. However, if the results of the statistical test are inconclusive, the remaining approaches are applied.

Quality indicators are used to characterize further the differences between the approximation sets. There is a variety of quality indicators, some of them are compliant with the concept of Pareto dominance and some are not. In this study we use only indicators in the former group, since these indicators are designed to assess how close a Pareto front approximation is from the Pareto optimal front. Moreover, different indicators are more sensible to different features of the approximation sets, for instance: distance, diversity, spread or cardinality. Therefore it is recommended to use a combination of them to yield more sound results. We use three different quality indicators: the Hypervolume indicator, the Epsilon indicator and the R2 indicator.

The Hypervolume indicator considers the volume of the objective space dominated by an approximation set. In other words, it measures the size of the space covered by an approximation set. The Epsilon indicator gives the factor by which an approximation set is worse than other in all objectives. In a single-objective case it refers to the ratio between the two objective values represented by the two approximation sets. Intuitively, it represents how much an approximation set A needs to translate/scale so that it covers the reference set. Finally, the R2 indicator measures the difference in the mean distance of an approximation set A to a reference set R, from an ideal point.

Empirical Attainment Functions (EAF) are used for visualizing the outcomes of multiple runs of a given configuration. Due to the stochastic behaviour of the algorithm, different runs of the same configuration can yield different results. Therefore, a plot illustrating the solutions generated by a given configuration, or comparing the solutions generated by two alternative configurations should take stochasticity into account. To compute the EAF from non-dominated sets of 4 objective vectors, the algorithm developed by Guerreiro (2011) is applied. Also, in order to plot 4 objectives, the parallel coordinates plot is applied.

The Kruskal-Wallis test is used to assess the statistical significance of the results. It compares sample indicator values of two alternative configurations under the hypothesis that there is no statistical significance between them. Considering a 95% significance level, if the test statistics (p-value) is lower than 0.05 we reject the null hypothesis and accept the alternative hypothesis that the first configuration is better than the second. To analyse the results a set of matrices of configurations is used.

Figure 5.6 illustrates the assessment procedure showing the results of different configurations of the deterministic approach applied to the Ophthalmology specialty. In this example, configurations 1, 2 and 3 represent 0, 10%

	Dominance Ranking			Quality Indicators											
				Epsilon			Hypervolume			R2			Summary		
	1	2	3	**1**	2	3	1	2	3	1	2	3	1	2	3
1		0,83	0,84		1,00	0,80		1,00	0,22		0,99	0,20		no	no
2	0,17		0,71	**0,00**		**0,00**	**0,00**		**0,00**	**0,01**		**0,00**	**yes**		**yes**
3	0,16	0,29		0,20	1,00		0,78	1,00		0,80	1,00		no	no	

Figure 5.6: An illustrative example of the analysis of the performance assessment measures

and 20% planned slack respectively. Matrices should be read from rows to columns. For instance, the highlighted cell shows the result of comparing configuration 2 with configuration 1 in the Epsilon indicator. In this case, as the result of the statistical test (p-value) is lower than the 0.05, considering this indicator and the predefined significance level, configuration 2 is better than configuration 1. The results are inconclusive in the dominance ranking approach and consistent in all quality indicators. The summary matrix shows "yes" if the results of the quality indicators are all lower than 0.05 and "no" otherwise. Based on this, the question "Is configuration 2 better than configuration 1?", can be answered positively in this case.

Both approaches, planned slacks and simulation optimization, were implemented in C++ and the computational experiments were carried out on an Intel Xeon 3.00 GHz processor running version 6 of the Scientific Linux operating system. Each experiment was limited to use a maximum of 10 processor cores.

5.5.4 Experimental Results

5.5.4.1 Deterministic Approach

Table 5.5 shows a summary of the statistical tests applied to quality indicators across all specialties. It depicts the total number of "yes" in summary tables like the one showed in the previous examples. The results show that using some planned slack is better than using no planned slack. On one hand, comparing configuration 1 with the other two, none of the statistical

	0	10%	20%
0	-	0	0
10%	8	-	2
20%	7	1	-

Table 5.5: Summary of the comparison of different configurations of the deterministic approach

tests indicates that configuration 1 is better. On the other hand, comparing 2 with 1, in 8 out of 10 instances the tests indicate that 2 is better. Furthermore, comparing 3 with 1, in 7 out of 10 instances 3 is better. Some planned slack clearly helps to reduce the impact of uncertainty, decreasing the number of cancelled surgeries and minutes of overtime.

Regarding the amount of planned slack the results are not clear. Configuration 2 (10%) is better than 3 (20%) in 2 instances and 3 is better than 2 in only 1 instance. The benefits of using a fixed planned slack start to decrease as the amount of planned slack increases. In fact, as the empty space inside each time block increases the number of scheduled surgeries as well as the surgical suite occupancy rate decreases. The results seem to indicate that the planned slack should not be fixed, but instead it should be adaptive and take into account the uncertainty intrinsic to each instance.

5.5.4.2 Stochastic Simulation Optimization Approach

In the simulation optimization approach the different configurations represent alternative number of replications applied to estimate performance measures for each solution during the optimization process. Table 5.6 summarizes the results of the statistical tests performed over the quality indicators data. Clearly, 5 replications are not enough to estimate performance measures accurately. The first row shows that 5 replications are a better option only in 2 instances, representing 4% of the total comparisons. Also, in 80% of the total comparisons, it is beneficial to use a number of replications higher than 5. A low number of replications enables the algorithm to

	5	25	50	75	100	150	better
5	-	0	0	0	0	2	4%
25	10	-	2	3	4	4	46%
50	9	4	-	2	4	6	50%
75	8	3	0	-	4	6	42%
100	7	1	0	0	-	3	22%
150	6	1	1	0	0	-	16%
worse	80%	18%	6%	10%	24%	42%	

Table 5.6: Summary of the comparison between different configurations of the simulation optimization approach

perform a high number of generations under a fixed time limit. However, the lack of precision in the estimates is unable to guide the algorithm to find better solutions, resulting in poor quality solutions.

A too high number of replications is also not an appropriate choice. The last row shows that using 150 replications is a better choice in 16% of the cases, but most of them are comparisons with 5 replications (in 2 comparisons only). The results also show that in 42% of the cases a number of replications lower than 150 is better. As the number of replications increases, the accuracy also increases. However, the number of generations performed under a fixed time limit decreases “exponentially”. Therefore, a number of replications too high does not pay off as the algorithm does not run enough generations to find good solutions.

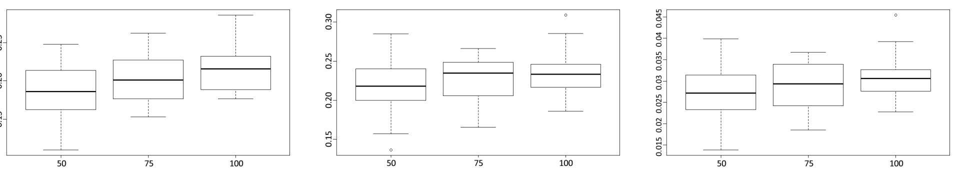
When the number of replications is between 5 and 150 the aggregated results are not so clear and can be misleading. The results indicate that 50 replications are better than other options in 50% of the comparisons. However, it clearly depends on the characteristics of the instance, such as the problem size and degree of uncertainty. For instance, Table 5.7 marks with “X” the best configurations for each surgical specialty. In Urology and General surgery 1, 25 replications is clearly a better option. In contrast, other instances show a few ties. On average, at a 95% significance level, the approach based on quality indicators produced 2.8 ties. In order to further characterize the influence of the number of replications in the algorithm

	5	25	50	75	100	150
Vascular surgery			X	X	X	
Oral and maxillofacial surgery			X	X	X	X
Neurosurgery		X	X	X		
Ophthalmology		X	X	X	X	X
Orthopaedics		X	X			
Urology		X				
Otolaryngology			X	X		
General surgery 1		X				
General surgery 2				X	X	X
General surgery 3			X	X	X	X

Table 5.7: Configurations in which it is not able to determine a better configuration

performance it is necessary to analyse indicators data.

Figure 5.7 shows 3 box plots comparing indicator data in 3 alternative number of replications for the Vascular Surgery instance. It is not able to determine which one of them is a better option with the Kruskal–Wallis test at a 95% confidence level. It should be taken into account that in the indicator values a lower value is better. The plots are consistent among the indicators. In general, a lower number of replications, in this case 50, is able to generate lower median indicator's values. However, the variability is higher than in the other configurations. Thus, in order to get more predictable results, a higher number of replications is a better configuration.



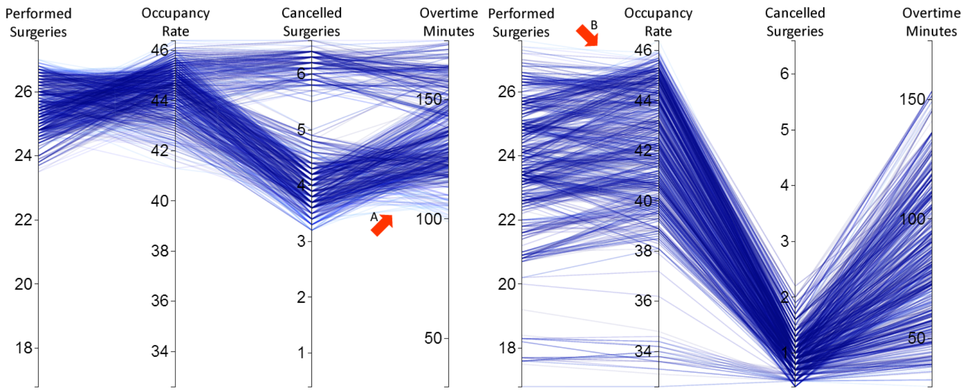
(a) Epsilon indicator (b) Hypervolume indicator (c) R indicator

Figure 5.7: Descriptive statistics for the Vascular surgery's indicators data

5.5.4.3 Comparison between the approaches

In order to identify differences in performance between the approaches as well as to visualize the outcomes of multiple runs, Empirical Attainment

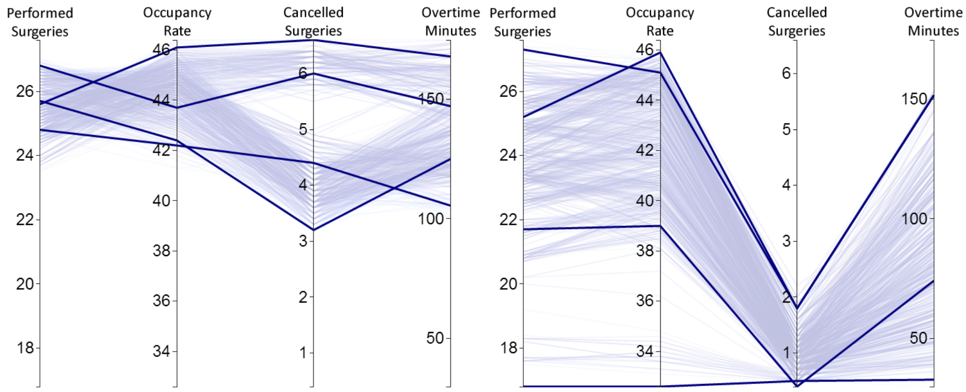
Functions (EAFs) are applied. EAFs are able to determine attainment levels (or super-levels) which are regions in the objective space with associated probabilities of a single run of the MOEA generate a solution within it. Figure 5.8 shows the points delimiting 3 of those regions with respective probabilities of 1/3, 2/3 and 3/3. In this example, points delimiting regions with lower probability appear lighter and with higher probability darker. The lighter lines are better noticed on the edges of the axis (see signs “A” and “B”), close to darker lines, meaning the algorithm has a high probability to achieve good solutions. The data correspond to the deterministic approach with 10% planned slack and the simulation optimization approach with 75 replications applied to Vascular Surgery.



(a) Deterministic approach (b) Simulation optimization approach

Figure 5.8: Points delimiting the attainment surfaces of multiple runs of the MOEA for the Vascular Surgery department

Figure 5.8(a) shows that the deterministic approach generates solutions with an excessive number of cancelled surgeries and overtime minutes. On the other hand, Figure 5.8(b) shows that the simulation optimization approach is able to generate solutions with a lower number of cancelled surgeries and overtime minutes. Moreover, such solutions have high number of performed surgeries and occupancy rate as well. Figure 5.9 highlights the best points in each objective generated by each approach. Indeed, the simulation optimization approach yields the best (lowest) values regarding the number of cancelled surgeries and overtime minutes.

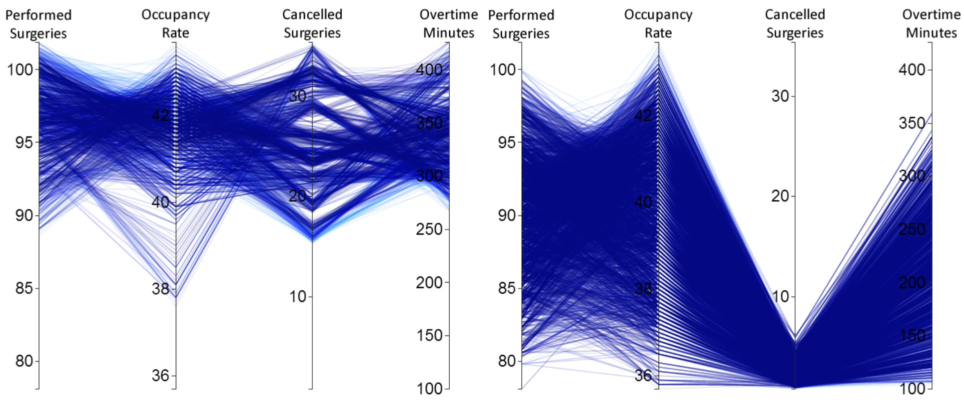


(a) Deterministic approach (b) Simulation optimization approach

Figure 5.9: Best solutions in each objective for the Vascular Surgery department

Similar results were obtained to other instances. For instance, Figure 5.10 shows the points delimiting attainment surfaces for the Ophthalmology department. It is a quite demanding instance type due to the short duration of the surgeries and high variance. The deterministic approach with a 10% planned slack generates a high number of cancelled surgeries as well as a high amount of overtime. In contrast, the simulation optimization approach generated solutions with a reduced number of cancelled surgeries and overtime minutes, while keeping high numbers of performed surgeries and occupancy rates.

Figure 5.11 shows a set of matrices comparing the best configurations of the deterministic MOEA with the best configurations of the simulation optimization approach based on the results of the Kruskal-Wallis statistical testing procedure in three different quality indicators. These results indicate that the simulation optimization approach is better than the deterministic MOEA in the majority of the tested instances (80%). The proposed simulation optimization approach did not reveal to be better for two instances



(a) Deterministic approach

(b) Simulation optimization approach

Figure 5.10: Points delimiting the attainment surfaces of multiple runs of the MOEA for the Ophthalmology department

only: Orthopedics and General surgery 3. Preliminary results show that these instances present better results with a different set of parameters than the one used in the previous analysis (running time, number of replications and maximum planned slack). For instance, Orthopedics indicates to require more computational time due to the size of the instance and its degree of uncertainty. The best simulation optimization configuration outperforms the best fixed planned slack when the algorithm runs for 5 minutes. In addition, General surgery 3 shows to require a higher maximum planned slack as well as a higher number of replications. In this particular instance, the SO approach improves as the number of replications increases and outperforms the deterministic approach when the number of replications reaches 250 and the maximum planned slack is equal to 20% of the time block size (the highest value in the deterministic approach).

5.6 Discussion and Future Work

This paper presents a multi-objective simulation optimization approach for the surgery scheduling problem under uncertainty. The aim is to generate

Vascular surgery

	Epsilon Indicator			Hypervolume Indicator			R Indicator			Overall result Is line i better than column j?		
	0	10%	20%	0	10%	20%	0	10%	20%	0	10%	20%
5	2,85037E-10	4,06438E-07	1,08103E-07	2,00331E-10	2,81651E-07	4,14325E-07	6,26118E-10	6,99654E-07	6,80291E-07			
25	2,87107E-35	3,06179E-29	2,12041E-30	2,35374E-42	6,55596E-36	1,53899E-35	7,42892E-41	1,58785E-34	1,49126E-34			
50	7,64016E-32	4,36068E-26	3,43359E-27	5,57859E-52	5,43829E-45	1,38118E-44	8,70102E-43	2,46743E-36	2,31427E-36	Yes	Yes	Yes
75	5,78677E-37	8,24502E-31	5,37967E-32	2,01351E-63	6,40125E-56	1,75208E-55	1,97534E-51	1,7208E-44	1,60552E-44	Yes	Yes	Yes
100	1,07983E-48	8,51541E-42	3,90264E-43	3,76866E-79	4,05434E-71	1,19966E-70	6,47948E-66	2,43792E-58	2,25895E-58	Yes	Yes	Yes
150	1,50881E-52	1,90253E-45	7,91066E-47	1,4696E-95	3,88369E-87	1,21808E-86	9,20613E-80	9,74107E-72	8,98155E-72			

Oral and maxillofacial surgery

	Epsilon Indicator			Hypervolume Indicator			R Indicator			Overall result		
	0	10%	20%	0	10%	20%	0	10%	20%	0	10%	20%
5	5,50538E-07	7,07355E-05	0,000052035	7,62589E-10	9,08872E-07	3,60961E-08	1,40949E-09	1,60639E-06	3,52036E-08			
25	9,84323E-25	1,97212E-20	1,01089E-20	2,78445E-41	7,97755E-35	7,00674E-38	7,76867E-33	6,30064E-27	3,13191E-30			
50	1,39552E-31	9,33206E-27	4,38364E-27	4,32615E-56	8,24255E-49	2,83414E-52	3,67554E-49	3,74768E-42	4,27752E-46	Yes	Yes	Yes
75	2,1244E-52	1,89848E-46	7,38024E-47	4,05537E-52	4,93401E-45	2,11922E-48	9,29739E-52	1,30755E-44	1,23729E-48	Yes	Yes	Yes
100	4,31946E-69	1,49976E-62	5,27486E-63	1,03497E-55	1,89246E-48	6,6413E-52	1,29928E-63	6,42864E-56	2,91728E-60	Yes	Yes	Yes
150	1,05092E-75	5,53824E-69	1,88845E-69	6,44981E-51	6,79803E-44	3,13508E-47	8,62171E-57	2,15145E-49	1,45678E-53	Yes	Yes	Yes

Neurosurgery

	Epsilon Indicator			Hypervolume Indicator			R Indicator			Overall result		
	0	10%	20%	0	10%	20%	0	10%	20%	0	10%	20%
5	6,70594E-34	9,69755E-13	5,83607E-12	1,17076E-21	0,00077831	0,000280402	1,55811E-26	2,00215E-05	4,60442E-06			
25	5,67185E-77	2,55674E-45	6,37472E-44	1,0678E-77	5,20602E-40	1,73838E-41	5,49485E-88	3,72823E-47	5,78369E-49		Yes	Yes
50	1,71402E-61	1,20125E-32	2,01242E-31	2,11363E-73	1,35286E-36	5,07396E-38	3,2077E-82	1,97353E-42	3,58421E-44		Yes	Yes
75	7,16338E-55	1,67494E-27	2,28972E-26	3,27699E-64	1,53938E-29	7,60855E-31	7,36218E-72	3,04855E-34	7,63106E-36		Yes	Yes
100	1,08641E-41	6,69634E-18	5,66863E-17	1,03411E-56	4,55663E-24	2,90994E-25	8,35757E-61	5,32176E-26	1,99371E-27			
150	2,09117E-32	8,34166E-12	4,67397E-11	2,73938E-43	4,50739E-15	4,95058E-16	1,27745E-44	4,0378E-15	3,1387E-16			

Ophthalmology

	Epsilon Indicator			Hypervolume Indicator			R Indicator			Overall result		
	0	10%	20%	0	10%	20%	0	10%	20%	0	10%	20%
5	0,0980582	0,998004	0,329606	0,0421761	0,821366	0,00652258	0,0901484	0,840149	0,0146829			
25	7,84054E-10	0,028012	9,74373E-08	3,73796E-68	1,37678E-52	7,75747E-73	4,31163E-47	4,07121E-35	1,0227E-51		Yes	
50	1,45139E-49	1,10982E-28	5,96886E-45	1,8202E-128	3,2074E-110	9,4421E-134	1,6897E-104	6,04174E-89	3,3181E-110		Yes	
75	2,10206E-77	2,41661E-52	4,6356E-72	2,017E-134	4,505E-116	1,0027E-139	2,1867E-111	1,23707E-95	3,7875E-117		Yes	
100	4,94489E-75	3,02691E-50	9,79716E-70	3,2232E-106	1,45208E-88	2,2091E-111	8,93514E-89	8,11261E-74	2,5934E-94		Yes	
150	3,90381E-68	3,20743E-44	5,48038E-63	3,97251E-85	2,01789E-68	4,44283E-90	1,93889E-73	2,70997E-59	9,80234E-79		Yes	

Orthopaedics

	Epsilon Indicator			Hypervolume Indicator			R Indicator			Overall result		
	0	10%	20%	0	10%	20%	0	10%	20%	0	10%	20%
5	2,50903E-10	0,86252	1	1,61083E-07	0,999643	1	1,93822E-07	0,999944	1			
25	1,69771E-43	1,48011E-12	1	3,01954E-50	8,30004E-13	0,877367	8,95201E-46	2,79696E-09	1			-
50	8,24851E-59	3,94751E-22	1	5,92943E-81	1,11292E-32	4,17049E-05	2,99579E-70	2,76931E-23	0,998473			-
75	1,39368E-50	8,79393E-17	1	3,38836E-74	7,63497E-28	0,00201225	1,09161E-61	6,72337E-18	0,999994			
100	9,60877E-48	4,82125E-15	1	9,14819E-72	3,69001E-26	0,00642289	4,96729E-59	2,52546E-16	0,999999			
150	7,53316E-40	1,53135E-10	1	3,58561E-59	5,66014E-18	0,343025	2,55566E-49	5,09588E-11	1			

Urology

	Epsilon Indicator			Hypervolume Indicator			R Indicator			Overall result		
	0	10%	20%	0	10%	20%	0	10%	20%	0	10%	20%
5	9,69372E-08	1	0,991994	2,22853E-15	0,0014572	0,00963062	5,99135E-13	0,951387	0,0538791			
25	8,04299E-47	0,699404	1,34437E-13	7,94904E-86	2,88607E-55	1,19439E-51	6,78018E-72	6,4611E-25	4,25673E-40		-	Yes
50	2,07882E-77	1,79003E-06	3,12806E-34	7,8346E-110	5,20244E-77	5,36831E-73	1,0395E-96	4,48736E-43	2,34886E-61			
75	3,02317E-68	0,000739916	1,65673E-27	3,78987E-89	3,31975E-58	1,5963E-54	1,75135E-78	1,87409E-29	1,42696E-45			
100	1,85406E-56	0,114564	1,70439E-19	6,41762E-69	9,7681E-41	1,63289E-37	1,63358E-61	2,8886E-18	8,96165E-32			
150	1,37858E-30	0,999915	1,76732E-05	1,61724E-36	8,64703E-16	1,01931E-13	5,05939E-33	0,000262312	1,23713E-11			

Otolaryngology

	Epsilon Indicator			Hypervolume Indicator			R Indicator			Overall result		
	0	10%	20%	0	10%	20%	0	10%	20%	0	10%	20%
5	3,29869E-11	0,999991	0,073893	1,72774E-21	0,692096	0,0033313	5,16729E-18	0,991146	0,00082187			
25	2,67402E-18	0,982575	0,000134971	5,69024E-59	1,51941E-12	5,55914E-24	3,83484E-40	0,00368726	3,07997E-16			
50	4,80471E-41	0,00119646	2,5239E-18	9,82022E-98	3,19543E-37	2,15196E-54	1,79152E-74	7,89183E-18	2,65012E-42		Yes	
75	3,51803E-67	1,88378E-14	2,03813E-38	3,1666E-109	5,22906E-46	2,29692E-64	1,13696E-93	7,1573E-30	9,34076E-59		Yes	
100	7,66953E-88	9,02384E-27	4,77667E-56	1,0817E-118	1,24781E-53	8,02413E-73	3,7852E-106	9,33489E-39	6,24945E-70			
150	6,14969E-80	9,64886E-22	3,65054E-49	1,1911E-103	1,11972E-41	1,74656E-59	4,70174E-93	1,8787E-29	3,26307E-58			

General surgery 1

	Epsilon Indicator			Hypervolume Indicator			R Indicator			Overall result		
	0	10%	20%	0	10%	20%	0	10%	20%	0	10%	20%
5	4,52863E-11	0,073976	0,853744	1,93875E-18	0,000646235	0,0490961	1,66771E-20	0,000503501	0,112596			
25	9,04949E-25	4,84548E-08	0,00209897	1,74879E-52	9,2888E-25	1,23479E-18	1,81072E-51	4,60236E-22	1,37334E-14		Yes	Yes
50	6,60374E-50	5,57776E-25	1,39884E-15	8,58669E-92	7,96161E-57	5,25508E-48	3,50125E-89	8,25854E-52	1,23917E-40			
75	8,37352E-63	4,38751E-35	6,46008E-24	6,01297E-95	1,33516E-59	1,25309E-50	2,09317E-92	1,40888E-54	3,59228E-43			
100	4,5412E-52	1,26675E-26	6,67334E-17	1,4558E-84	1,47277E-50	4,04648E-42	7,74666E-82	1,30027E-45	5,26928E-35			
150	7,50519E-53	3,18445E-27	2,18021E-17	2,96436E-63	5,49039E-33	5,91377E-26	1,29589E-63	5,15495E-31	4,07878E-22			

General surgery 2

	Epsilon Indicator			Hypervolume Indicator			R Indicator			Overall result		
	0	10%	20%	0	10%	20%	0	10%	20%	0	10%	20%
5	0,0132926	1	0,525264	0,993523	1	1	0,0181759	1	1			
25	2,20091E-21	0,0256988	2,07231E-13	1,84639E-14	0,128778	0,00316091	2,09734E-29	0,120521	0,000387195			
50	5,49383E-31	0,000013521	2,59957E-21	7,37705E-27	5,45247E-06	1,18283E-09	2,10359E-46	2,27584E-06	8,36615E-12			
75	1,45862E-37	1,42063E-08	5,9065E-27	1,55796E-44	8,55461E-16	1,29181E-21	5,00499E-67	3,60704E-16	1,88113E-24		Yes	Yes
100	1,04244E-34	3,21246E-07	1,73633E-24	2,17972E-44	1,05791E-15	1,65399E-21	1,16455E-64	7,20837E-15	7,34072E-23		Yes	Yes
150	9,43411E-45	2,39496E-12	2,78984E-33	2,34004E-63	8,06411E-29	2,62057E-36	3,54444E-85	2,33374E-27	1,33439E-37		Yes	Yes

General surgery 3

	Epsilon Indicator			Hypervolume Indicator			R Indicator			Overall result		
	0	10%	20%	0	10%	20%	0	10%	20%	0	10%	20%
5	0,000119445	1	1	0,00820666	1	1	0,00890449	1	1			
25	4,92936E-07	1	1	0,000103524	1	1	0,000429246	1	1			
50	1,58836E-13	0,999943	1	2,79267E-22	1	0,999979	1,95904E-16	1	1		-	-
75	6,96564E-21	0,960495	0,999928	6,12509E-36	1	0,848853	9,42114E-28	1	0,998864		-	-
100	2,17703E-18	0,992319	0,999996	2,26112E-55	0,99991	0,00444112	7,77343E-47	0,999029	0,203443		-	-
150	5,23309E-21	0,957666	0,999919	3,86959E-81	0,318889	6,18659E-12	7,87282E-72	0,118664	1,82723E-07		-	-

Figure 5.11: Results of the Kruskal-Wallis test comparing runs of the deterministic and simulation optimization versions of the MOEA with different configurations

surgery schedules able to maximize the number of performed surgeries and occupancy rate, as well as to minimize the number of cancelled surgeries and minutes of overtime. Schedule's performance is evaluated using a simulation model which takes into account 4 sources of uncertainty: surgery duration, emergencies, cancellations and delays/advances. The proposed approach is compared with a standard deterministic approach based on planned slacks, a traditional way to tackle the problem. The performance assessment of both approaches, as well as the comparison between them relies on a comprehensive methodology for performance assessment of multi-objective optimizers.

The proposed approach outperforms the deterministic one in the majority of cases. Planned slacks are effective in reducing the impact of uncertainty. However, they also reduce the number of performed surgeries and occupancy rate. On the other hand, the simulation optimization is not only able to generate solutions with a high number of performed surgeries and occupancy rate but also with low cancellations and overtime minutes. To achieve these results the number of replications should be properly set according to the characteristics of each instance. It should not be too low due to the estimation noise, but should not be too high because of the estimation cost.

In future work, the idea of an adaptive number of replications could be further explored. Moreover, the behaviour of each stochastic variable within the simulation model could be characterized more precisely. For instance, the surgery duration could take into account more characteristics of the procedures being performed as well as of the patient and members of the surgical team. It would help to reduce the variability among simulation replications of the same surgery schedule contributing to reduce the required number of replications. Finally, the proposed approach could be applied to other surgery management problems, such as the master surgery scheduling problem.

Acknowledgements

The authors would like to thank Andreia P. Guerreiro and Professor Carlos M. Fonseca from the Department of Informatics Engineering (DEI) of the University of Coimbra for the kindness in providing the required software for computing the empirical attainment functions with 4 dimensions which helped us to assess the performance of the simulation optimization approach. This work is financed by the ERDF – European Regional Development Fund through the COMPETE Programme (operational programme for competitiveness) and by National Funds through the FCT – Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) within project ≪FCOMP-01-0124-FEDER-037281≫.

Chapter 6

Conclusions

6.1 Summary and conclusions

Based on identified research gaps and on the needs of hospital managers, this thesis develops new methods and techniques for operating room management problems. In particular, the work focuses on the elective surgery scheduling problem at the operational decision level which is considered a complex and challenging decision problem due to its combinatorial, multi-objective and stochastic characteristics. It balances theory and practice contributing both to scientific community as well as to society.

The methodological framework followed in the development of this thesis allowed: (1) to create a solid base consisting of a literature review on operating room management problems and a series of workshops and interviews with hospital managers; (2) to develop a decision support system to aid the elective surgery scheduling process; and (3) to develop new and progressively more complex scheduling methods to address large scale and stochastic problems.

From a management perspective, the developed DSS represents a powerful planning tool based on quantitative methods for decision making, in contrast to ad hoc planning methods based on empiricism which still proliferate in healthcare institutions. It is a must have requirement in order to benefit from the overwhelming amount of data generated from modern information systems. The development of the DSS represents a good example of research linked to companies, or public institutions, as it was developed in close collaboration with hospital managers and physicians. In addition, the scheduling methods proposed in chapters 4 and 5 are examples of innovative research, since both explored new solution methods which were not found in the operating room management literature before.

According to chapter 2, the literature on operations research applied to operating room management still presents some research gaps. In fact, there is a mismatch between the characteristics of the problems and the features included in the majority of the solutions. Problems are described as combinatorial, multi-objective, subject to strong uncertainty and to the availability of multiple resources. However, the majority of the solutions is deterministic and considers only a limited number of objectives and constraints. Most of these gaps are associated to the high computational cost required for tackling complex scheduling problems. This issue forces researchers to apply simplified approaches. The uncertainty inherent to healthcare management problems is one of the main drivers of such high computational cost. There are stochastic approaches to surgery scheduling problems, but most of them rely on more traditional stochastic programming approaches, which lead to high computational costs and are not suitable to address multi-objective optimization problems.

Throughout the development of the decision support system (DSS) it became evident that it can foster efficiency gains in the elective surgery planning process. The complexity of the operating room management problems requires a paradigm shift from decision processes based on the empirical experience of the decision makers to decision processes based on quantitative decision support methods. Chapter 3 showed that regarding both surgery scheduling and duration estimation, the results are significantly better than the reality of surgical services and can provide the end-user a great advantage when planning, compared to the traditional planning methods. In addition, this method depends on the quality of the available data, so that hospital managers should implement rigorous processes to collect high quality data.

The two new scheduling methods proposed in chapter 4 can generate far better schedules (in terms of quality of solutions and required computational times) than the method proposed in chapter 3, which is commonly found in the literature. In fact, the new scheduling methods are very competitive and can generate high quality results. The results of the heuristic approach depend heavily on the additional local search procedures applied to each solution generated by the genetic algorithm. These approaches are recommended in a scenario with low uncertainty. For instance, in ambulatory services, the operating rooms are dedicated exclusively for elective cases and the surgery durations show low variability.

In general surgery services, where the surgical process is subject to strong uncertainty, the simulation optimization approach proposed in chapter 5 is preferable. This approach is more realistic as it takes into account four sources of uncertainty and evaluates the quality of solutions based on the estimated performance of the execution of the plans. It can effectively reduces the deviation between the planned and actual performance of surgery schedules. As a result, it helps to control undesirable effects such as excessive overtime and cancelled surgeries. However, such approach also depends on the quality of the available data, as the prediction methods applied in chapter 3.

In conclusion, the decision support system and the new scheduling methods presented can improve the efficiency of surgical services. Therefore, they contribute to tackle an important societal issue. In order to take advantage of these tools, the collaboration of physicians and hospital managers is needed. In this sense, it is imperative to promote the application of operations research tools in healthcare institutions. In addition, the proposed methods represent original contributions to the scientific literature on operations research applied to operating room management. These advances can help researchers to bring innovations to society.

6.2 Contributions of the thesis

This thesis contributed to the field of operations research applied to operating room management, namely to elective surgery scheduling problems in deterministic and stochastic settings. Innovative models were developed to address practical issues, as described in the following topics that summarize the main contributions of the thesis:

1. The decision support system presented in chapter 3: The solution presented is mainly directed to the effective management of the operating theater, where data mining and optimization components are added to allow for more efficient scheduling. To the best of our knowledge this work is the first to combine the aforementioned techniques to reduce surgery uncertainty and to achieve a better utilization of the existing resources through scheduling optimization within decision support systems.

2. The exact model with a continuous representation of time presented in chapter 4: This study proposed a new modelling approach for the integrated (advance and allocation) surgery scheduling problem using a continuous representation of time, thus providing a more accurate

representation of the problem and potentially a higher resource utilization.

3. The meta-heuristic and local search procedures presented in chapter 4: This study proposed an original heuristic solution method aiming to find near optimal solutions within a reduced amount of time. The proposed approach is based on the biased random-key genetic algorithm (BRKGA)(Gonçalves and Resende, 2011) framework and on an efficient decoding procedure to translate each individual in the population into a high quality schedule. The results of the computational experiments emphasized the value of well tailored heuristics.

4. The simulation optimization framework presented in chapter 5: This study proposed a multi-objective simulation optimization approach to the surgery scheduling problem under uncertainty. To the best of our knowledge, it is the first multi-objective optimization approach to tackle the general stochastic surgery scheduling problem. In addition, it is the first approach to take into account four important sources of uncertainty arising in a large Portuguese hospital and to model surgery duration considering its main determinant attributes.

6.3 Directions for future research

Throughout the development of this thesis additional studies were carried out but were not published yet. In addition, the comprehensive research of operating room scheduled problems generated other promising ideas that were not developed. For instance, concerning deterministic scheduling methods, a constraint programming (CP) model was developed and demonstrated good performance on preliminary computational experiments. In the future, the performance of scheduling models based on CP needs to be compared against the scheduling models proposed in chapter 4. In addition, concerning stochastic solution approaches, it is worth to compare the performance of the approach proposed in chapter 5, based on a combination of a multi-objective genetic algorithm with a discrete-event simulation model, against more traditional models based on stochastic programming (SP).

Concerning new problem settings, the scheduling methods proposed in chapter 4 should be evaluated in a rolling horizon scenario. This would require small modifications of the problems, e.g. including the previously scheduled surgeries and an additional objective function term to minimize their rescheduling. This approach would be closer to what happens in reality. Furthermore, it would reduce the required time to generate a new schedule, since only a small modification of a previous schedule would be required. In a rolling horizon framework the impact of using different objective functions can be evaluated. For instance, different objective functions aiming to maximize the number of scheduled surgeries, or to maximize the utilization of ORs, or to minimize the under utilization of ORs (considering that ORs are occupied during cleaning procedures). In addition, such framework would help to assess the impact of additional SIGIC rules, such as the maximum number of days a patient can be in the waiting list without being scheduled, the minimum time between the date in which the patient is notified of the surgery date and the actual surgery date, and the maximum distance in days between the scheduling dates of two patients (to preserve the equity in the waiting list).

The approach described in the previous paragraph consists in the on-line surgery scheduling problem. According to the literature review presented in chapter 2, it is one of the less studied operating room management problems. Therefore, it should be addressed in the future as an extension of the scheduling methods presented in this thesis. In addition, OR management problems on the tactical and strategic decision levels are also not as well addressed as the off-line problem on the operational level. These problems should be addressed in the future in order to complement the decision support system. They have the potential to promote high efficiency gains and usually imply less organizational changes with associated lower risks and implementation costs than operational problems.

Concerning the solution approach proposed in chapter 3, more advanced simulation allocation rules (SAR) may be evaluated in the future. The literature review on chapter 2 showed that they are an essential component of simulation optimization approaches. In preliminary computational experiments the optimal budget computing allocation (OCBA) approach showed worse results than simple allocation rules. These results may be linked to the cost of the ranking and selection procedure and the simulation cost. In the particular case of chapter 5, the model runs very fast, but in the case of larger instances or in the case of adding additional resources it may be worth to try more advanced SAR rules.

Finally, in future work, the decision support system proposed in chapter 3 should be implemented in other hospitals. In order to achieve this goal the top management of hospitals should be convinced of the benefits that the application of operations research methods can bring to their organizations. In addition, new hospital information systems should be designed with integrated decision support models. May this thesis help to achieve these goals.

References

ACSS (2014). Relatório síntese da atividade cirúrgica programada - ano 2013. Technical report, ACSS - Unidade Central de Gestão de Inscritos para Cirurgia.

Addis, B., Carello, G., and Tànfani, E. (2014). A robust optimization approach for the operating room planning problem with uncertain surgery duration. In Matta, A., Li, J., Sahin, E., Lanzarone, E., and Fowler, J., editors, *Proceedings of the International Conference on Health Care Systems Engineering*, volume 61 of *Springer Proceedings in Mathematics and Statistics*, pages 175–189. Springer International Publishing.

Alvarez, R., Bowry, R., and Carter, M. (2010). Prediction of the time to complete a series of surgical cases to avoid cardiac operating room overutilization. *Canadian journal of anaesthesia*, 57(11):973–979.

Andradóttir, S. (2006). Chapter 20 an overview of simulation optimization via random search. In Henderson, S. G. and Nelson, B. L., editors, *Simulation*, volume 13 of *Handbooks in Operations Research and Management Science*, pages 617–631. Elsevier.

April, J., Glover, F., Kelly, J. P., and Laguna, M. (2003). Practical introduction to simulation optimization. *Proceedings of the 2003 Winter Simulation Conference.*

Augusto, V., Xie, X., and Perdomo, V. (2010). Operating theatre scheduling with patient recovery in both operating rooms and recovery beds. *Computers & Industrial Engineering*, 58(2):231–238.

Azari-Rad, S., Yontef, A., Aleman, D. M., and Urbach, D. R. (2014). A simulation model for perioperative process improvement. *Operations Research for Health Care*, 3(1):22–30.

Barton, R. R. and Meckesheimer, M. (2006). Chapter 18 metamodel-based simulation optimization. In Henderson, S. G. and Nelson, B. L., editors, *Simulation*, volume 13 of *Handbooks in Operations Research and Management Science*, pages 535 – 574. Elsevier.

Batun, S., Denton, B. T., Huschka, T. R., and Schaefer, a. J. (2010). Operating Room Pooling and Parallel Surgery Processing Under Uncertainty. *INFORMS Journal on Computing*, 23(2):220–237.

Bean, J. C. (1994). Genetic algorithms and random keys for sequencing and optimization. *ORSA Journal on Computing*, 6(2):154–160.

Beliën, J. and Demeulemeester, E. (2007). Building cyclic master surgery schedules with leveled resulting bed occupancy. *European Journal of Operational Research*, 176(2):1185–1204.

Better, M., Glover, F., and Laguna, M. (2007). Advances in analytics: Integrating dynamic data mining with simulation optimization. *IBM Journal of Research and Development*, 51(3):477–487.

Bjornberg, A., Garrofe, B. C., and Lindblad, S. (2009). Euro health consumer index. Technical report, Health Consumer Power House.

Blake, J. and Carter, M. (1997). Surgical process scheduling: a structured review. *Journal of the Society for Health Systems*, 5(3):17–30.

Blake, J. T. and Carter, M. W. (2002). A goal programming approach to strategic resource allocation in acute care hospitals. *European Journal Of Operational Research*, 140(3):541–561.

Blake, J. T. and Donald, J. (2002). Mount Sinai Hospital Uses Integer Programming to Allocate Operating Room Time. *Interfaces*, 32(2):63–73.

Boesel, J., Nelson, B. L., and Ishii, N. (2003). A framework for simulation-optimization software. *IIE Transactions*, 35(3):221–229.

Brailsford, S. C., Harper, P., Patel, B., and Pitt, M. (2009). An analysis of the academic literature on simulation and modelling in health care. *Journal of Simulation*, 3(3):130–140.

Brailsford, S. C. and Vissers, J. (2011). OR in healthcare: A European perspective. *European Journal of Operational Research*, 212(2):223–234.

Branke, J., Chick, S. E., and Schmidt, C. (2007). Selecting a selection procedure. *Management Science*, 53(12):1916–1932.

Breiman, L. (1996). Bagging predictors. In *Machine Learning*, pages 123–140.

Brito, A. E. d. S. C. and Teixeira, J. M. F. (2001). *Simulação por computador: fundamentos e implementação de código em C e C++*. Publindústria.

Bruni, M. E., Beraldi, P., and Conforti, D. (2014). A stochastic programming approach for operating theatre scheduling under uncertainty. *IMA Journal of Management Mathematics*.

Cardoen, B. (2009). Operating room planning and scheduling: solving a surgical case sequencing problem. *4or*, 8(1):101–104.

Cardoen, B., Demeulemeester, E., and Beliën, J. (2009). Sequencing surgical cases in a day-care environment: An exact branch-and-price approach. *Computers & Operations Research*, 36(9):2660–2669.

Cardoen, B., Demeulemeester, E., and Beliën, J. (2010a). Operating room planning and scheduling: A literature review. *European Journal of Operational Research*, 201(3):921–932.

Cardoen, B., Demeulemeester, E., and Beliën, J. (2010b). Operating room planning and scheduling problems: a classification scheme. *International Journal of Health Management and Information*, 1(1):71–83.

Chaabane, S., Meskens, N., Guinet, A., and Laurent, M. (2008). Comparison of two methods of operating theatre planning: Application in belgian hospital. *Journal of Systems Science and Systems Engineering*, 17(2):171–186.

Chen, C.-H. and Lee, L. H. (2010). *Stochastic simulation optimization: an optimal computing budget allocation*, volume 1. World Scientific Pub Co Inc.

Commission, A. (2003). Operating theatres — Review of national findings. Technical report, National Health Service in England & Wales, London.

Conforti, D., Guerriero, F., and Guido, R. (2010). A multi-objective block scheduling model for the management of surgical operating rooms: New solution approaches via genetic algorithms. *2010 IEEE Workshop on Health Care Management WHCM*, pages 1–5.

de la Maisonneuve, C. and Martins, J. O. (2013). Public spending on health and long-term care: A new set of projections. Main Paper, Economic Policy Paper 6, OECD, Paris.

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197.

Denton, B., Viapiano, J., and Vogl, A. (2007). Optimization of surgery sequencing and scheduling decisions under uncertainty. *Health Care Management Science*, 10(1):13–24.

Denton, B. T., Miller, A. J., Balasubramanian, H. J., and Huschka, T. R. (2010). Optimal allocation of surgery blocks to operating rooms under uncertainty. *Operations Research*, 58:802–816.

Dexter, F., Epstein, R. H., Lee, J. D., and Ledolter, J. (2009). Automatic updating of times remaining in surgical cases using bayesian analysis of historical case duration data and "instant messaging" updates from anesthesia providers. *Anesthesia & Analgesia*, 108(3):929–40.

Dexter, F. and Macario, A. (1999). Decrease in case duration required to complete an additional case during regularly scheduled hours in an operating room suite: a computer simulation study. *Anesthesia & Analgesia*, 88(1):72–76.

Dexter, F., Traub, R. D., and Qian, F. (1999). Comparison of statistical methods to predict the time to complete a series of surgical cases. *Journal of Clinical Monitoring and Computing*, 15:45–51.

Eijkemans, M. J. C., Van Houdenhoven, M., Nguyen, T., Boersma, E., Steyerberg, E. W., and Kazemier, G. (2010). Predicting the unpredictable: a new prediction model for operating room times using individual characteristics and the surgeon's estimate. *Anesthesiology*, 112(1):41–49.

Escoval, A., Lopes, M., Ferreira, P. L., and Sakellarides, C. (2012). Relatório de Primavera 2012: Crise & saúde. Um país em sofrimento. Technical report, Observatório Português dos Sistemas de Saúde.

Fei, H., Chu, C., and Meskens, N. (2008a). Solving a tactical operating room planning problem by a column-generation-based heuristic procedure with four criteria. *Annals of Operations Research*, 166(1):91–108.

Fei, H., Chu, C., Meskens, N., and Artiba, A. (2008b). Solving surgical cases assignment problem by a branch-and-price approach. *International Journal of Production Economics*, 112(1):96–108.

Fei, H., Meskens, N., and Chu, C. (2010). A planning and scheduling problem for an operating theatre using an open scheduling strategy. *Computers & Industrial Engineering*, 58(2):221–230.

Figueira, G. and Almada-Lobo, B. (2014). Hybrid simulation-optimization methods: A taxonomy and discussion. *Simulation Modelling Practice and Theory*, 46:118–134.

Fu, M., Glover, F., and April, J. (2005). Simulation optimization: a review, new developments, and applications. In *Simulation Conference, 2005 Proceedings of the Winter*, pages 83–95.

Fu, M. C. (1994). Optimization via simulation: A review. *Annals of Operations Research*, 53(1):199–247.

Fu, M. C. (2002). Optimization for simulation: Theory vs. practice. *INFORMS Journal on Computing*, 14(3):192–215.

Fu, M. C. (2006). Chapter 19 gradient estimation. In Henderson, S. G. and Nelson, B. L., editors, *Simulation*, volume 13 of *Handbooks in Operations Research and Management Science*, pages 575–616. Elsevier.

Fu, M. C., Andradóttir, S., Carson, J. S., Glover, F., Harrell, C. R., Ho, Y.-C., Kelly, J. P., and Robinson, S. M. (2000). Integrating optimization and simulation: research and practice. *Proceedings of the 2000 Winter Simulation Conference*, pages 610–616.

Fu, M. C., Chen, C.-h., and Shi, L. (2008). Some topics for simulation optimization. In *Proceedings of the 40th Winter Simulation Conference*, pages 27–38. Winter Simulation Conference.

Glover, F., Kelly, J., and Laguna, M. (1999). New advances for wedding optimization and simulation. In *Simulation Conference Proceedings, 1999 Winter*, volume 1, pages 255–260. IEEE.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1st edition.

Goldsman, D., Nance, R., and Wilson, J. (2010). A brief history of simulation revisited. In *Simulation Conference (WSC), Proceedings of the 2010 Winter*, pages 567–574.

Gomes, P., Paiva, N., and Simões, B. (2009). Análise da Viabilidade Económica das Aplicações SAM e SAPE. Technical report, GANEC Gabinete de Análise Económica, Universidade Nova de Lisboa, Lisboa, Portugal.

Gonçalves, J. and Resende, M. (2011). Biased random-key genetic algorithms for combinatorial optimization. *Journal of Heuristics*, 17(5):487–525.

Gonçalves, J. F., Mendes, J. J. M., and Resende, M. G. C. (2005). A hybrid genetic algorithm for the job shop scheduling problem. *European Journal of Operational Research*, 167(1):77 – 95.

Gonçalves, J. F., Mendes, J. J. M., and Resende, M. G. C. (2008). A genetic algorithm for the resource constrained multi-project scheduling problem. *European Journal of Operational Research*, 189(3):1171 – 1190.

Gonçalves, J. F. and Resende, M. G. C. (2013). A biased random-key genetic algorithm for a 2d and 3d bin packing problem. *International Journal of Production Economics*, 145:500–510.

Gonçalves, J. F., Resende, M. G. C., and Costa, M. D. (2014a). A biased random-key genetic algorithm for the minimization of open stacks problem. *International Transactions in Operational Research*, published online 2 July 2014.

Gonçalves, J. F., Resende, M. G. C., and Toso, R. F. (2014b). An experimental comparison of biased and unbiased random-key genetic algorithms. *Pesquisa Operacional*, 34:143 – 164.

Gordon, T., Paul, S., Lyles, A., and Fountain, J. (1988). Surgical unit time utilization review: Resource utilization and management implications. *Journal of Medical Systems*, 12(3):169–179.

Guerlain, S., Brown, D., and Mastrangelo, C. (2000). Intelligent decision support systems. In *Systems, Man, and Cybernetics, 2000 IEEE International Conference on*, volume 3, pages 1934–1938.

Guerreiro, A. P. (2011). Efficient algorithms for the assessment of stochastic multiobjective optimizers. Master's thesis, IST, Technical University of Lisbon.

Guerriero, F. and Guido, R. (2011). Operational research in the management of the operating theatre: a survey. *Health Care Management Science*, 14(1):89–114.

Guinet, A. and Chaabane, S. (2003). Operating theatre planning. *International Journal of Production Economics*, 85(1):69–81.

Gul, S., Denton, B. T., Fowler, J. W., and Huschka, T. (2011). Bi-criteria scheduling of surgical services for an outpatient procedure center. *Production and Operations Management*, 20(3):406–417.

Hachicha, W., Ammeri, A., Masmoudi, F., and Chachoub, H. (2010). A comprehensive literature classification of simulation optimisation methods. Mpra paper, University Library of Munich, Germany.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The weka data mining software: An update. *SIGKDD Explor. Newsl.*, 11(1):10–18.

Hans, E., Wullink, G., Van Houdenhoven, M., and Kazemier, G. (2008). Robust surgery loading. *European Journal of Operational Research*, 185(3):1038–1050.

Henderson, S. G. and Nelson, B. L. (2006). Chapter 1 stochastic computer simulation. In Henderson, S. G. and Nelson, B. L., editors, *Simulation*, volume 13 of *Handbooks in Operations Research and Management Science*, pages 1 – 18. Elsevier.

HFM (2003). Achieving operating room efficiency through process integration. *Healthcare financial management : journal of the Healthcare Financial Management Association*, 57(3):suppl 1–7 following 112.

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems.* The University of Michigan Press.

Holmes, G., Hall, M., and Frank, E. (1999). Generating rule sets from model trees. In *Proceedings of the 12th Australian Joint Conf. on Artificial Intelligence*, pages 1–12. Springer-Verlag.

Hong, L. and Nelson, B. (2009). A brief introduction to optimization via simulation. In *Simulation Conference (WSC), Proceedings of the 2009 Winter*, pages 75–85.

Hong, L. J. and Nelson, B. L. (2005). The tradeoff between sampling and switching: New sequential procedures for indifference-zone selection. *IIE Transactions*, 37(7):623–634.

Hulshof, P. J. H., Kortbeek, N., Boucherie, R. J., Hans, E. W., and Bakker, P. J. M. (2012). Taxonomic classification of planning decisions in health care: a structured review of the state of the art in OR/MS. *Health Systems*, 1(2):129–175.

Jao, C. and Hier, D. (2010). Clinical decision support systems: An effective pathway to reduce medical errors and improve patient safety. *Decision Support Systems*, pages 121–138.

Jaspers, M., Smeulers, M., Vermeulen, H., and Peute, L. (2011). Effects of clinical decision-support systems on practitioner performance and patient outcomes. *Journal of American Medical Informatics Association*, 18:327–334.

Jebali, A., Hadj Alouane, A. B., and Ladet, P. (2006). Operating rooms scheduling. *International Journal of Production Economics*, 99(1-2):52–62.

Kim, S. and Nelson, B. (2007). Recent advances in ranking and selection. In *Proceedings of the 2007 Winter Simulation Conference*, pages 162–172.

Kim, S., Pasupathy, R., and Henderson, S. (2015). A guide to sample average approximation. In Fu, M. C., editor, *Handbook of Simulation Optimization*, volume 216 of *International Series in Operations Research & Management Science*, pages 207–243. Springer New York.

Knowles, J., Thiele, L., and Zitzler, E. (2006). A Tutorial on the Performance Assessment of Stochastic Multiobjective Optimizers. 214, Computer Engineering and Networks Laboratory (TIK), ETH Zurich, Switzerland.

Konrad, R., DeSotto, K., Grocela, A., McAuley, P., Wang, J., Lyons, J., and Bruin, M. (2013). Modeling the impact of changing patient flow processes in an emergency department: Insights from a computer simulation study. *Operations Research for Health Care*, 2(4):66–74.

Lamiri, M. and Augusto, V. (2008). Patients scheduling in a hospital operating theatre. *Automation Science and Engineering, 2008. CASE 2008. IEEE International Conference on*, pages 627–632.

Lamiri, M., Dreo, J., and Xie, X. (2007). Operating Room Planning with Random Surgery Times. In *Automation Science and Engineering, 2007. CASE 2007. IEEE International Conference on*, pages 521–526. IEEE.

Lamiri, M., Grimaud, F., and Xie, X. (2009). Optimization methods for a stochastic surgery planning problem. *International Journal of Production Economics*, 120(2):400–410.

Law, A. and McComas, M. (2011). How the ExpertFit distribution-fitting software can make your simulation models more valid. In *Proceedings of the 2011 Winter Simulation Conference*, pages 63–69.

Lee, L., Chew, E., Teng, S., and Chen, Y. (2008). Multi-objective simulation-based evolutionary algorithm for an aircraft spare parts allocation problem. *European Journal of Operational Research*, 189(2):476–491.

Lee, L. H., Chew, E. P., and Li, H. (2011). Multi-objective compass for discrete optimization via simulation. *Proceedings of the 2011 Winter Simulation Conference*, pages 4070–4079.

Lee, S. and Yih, Y. (2014). Reducing patient-flow delays in surgical suites through determining start-times of surgical cases. *European Journal of Operational Research*, 238(2):620–629.

Li, M., Li, G., and Azarm, S. (2008). A Kriging Metamodel Assisted Multi-Objective Genetic Algorithm for Design Optimization. *Journal of Mechanical Design*, 130(3):031401.

Li, Y., Zhang, S., Baugh, R. F., and Huang, J. Z. (2009). Predicting surgical case durations using ill-conditioned CPT code matrix. *IIE Transactions*, 42(2):121–135.

López, M. J., Pérez, J. A. R., and Zubiaurre, M. Z. (2008). A goal programming model to support the case-mix planning in state hospitals. *Cuaderno de Gestión*, 8(1):73–88.

Ma, G. and Demeulemeester, E. (2012). A multilevel integrative approach to hospital case mix and capacity planning. *Computers & Operations Research*, pages 1–10.

Macario, A. (2009). Truth in scheduling: is it possible to accurately predict how long a surgical case will last? *Anesthesia & Analgesia*, 108(3):681–685.

Mancilla, C. and Storer, R. (2011). A sample average approximation approach to stochastic appointment sequencing and scheduling. *IIE Transactions*, 44(8):655–670.

Marques, I., Captivo, M. E., and Pato, M. V. (2012). An integer programming approach to elective surgery scheduling. *OR Spectrum*, 34(2):407–427.

Marques, I., Captivo, M. E., and Pato, M. V. (2014). Scheduling elective surgeries in a portuguese hospital using a genetic heuristic. *Operations Research for Health Care*, 3(2):59 – 72. Special Issue of the 2012 conference of the {EURO} working group Operational Research Applied To Health Services (ORAHS).

May, J. H., Spangler, W. E., Strum, D. P., and Vargas, L. G. (2011). The surgical scheduling problem: Current research and future opportunities. *Production and Operations Management*, 20(3):392–405.

Mayer, E., Faiz, O., Athanasiou, T., and Vincent, C. (2008). Measuring and enhancing elective service performance in NHS operating theatres: an overview. *Journal of the Royal Society of Medicine*, 101(6):273–277.

Mendes, J. J. M., Gonçalves, J. F., and Resende, M. G. C. (2009). A random key based genetic algorithm for the resource constrained project scheduling problem. *Computers & Operations Research*, 36(1):92–109.

Meskens, N., Duvivier, D., and Hanset, A. (2013). Multi-objective operating room scheduling considering desiderata of the surgical team. *Decision Support Systems*, 55(2):650–659.

Min, D. and Yih, Y. (2010). Scheduling elective surgery under uncertainty and downstream capacity constraints. *European Journal of Operational Research*, 206(3):642–652.

Ministério da Saúde (2011). Manual do SIGIC - MGIC 2011 - Volume I – Princípios Gerais. Technical report, Unidade Central de Gestão de Inscritos para Cirurgia, Administração Central do Sistema de Saúde, IP, Lisboa, Portugal.

Nelson, B. L. (2010). *Optimization via Simulation Over Discrete Decision Variables*, chapter 10, pages 193–207.

Nutaro, J. J. (2010). Discrete-event systems. In *Building Software for Simulation: Theory and Algorithms, with Applications in C++*, chapter 4, pages 100–182. Wiley, 1 edition.

OECD (2013). Health at a glance 2013.

OECD (2014). Society at a glance 2014.

Ogulata, S. and Erol, R. (2003). A hierarchical multiple criteria mathematical programming approach for scheduling general surgery operations in large hospitals. *Journal of Medical Systems*, 27(3).

Olafsson, S. (2006). Chapter 21 metaheuristics. In Henderson, S. G. and Nelson, B. L., editors, *Simulation*, volume 13 of *Handbooks in Operations Research and Management Science*, pages 633–654. Elsevier.

Olafsson, S. and Kim, J. (2002). Simulation Optimization. In *Proceedings of the 2002 Winter Simulation Conference*, pages 79–84.

Ozkarahan, I. (1995). Allocation of surgical procedures to operating rooms. *Journal of medical systems*, 19(4):333–52.

Persson, M. J. and Persson, J. A. (2010). Analysing management policies for operating room planning using simulation. *Health Care Management Science*, 13(2):182–191.

Pham, D.-N. and Klinkert, A. (2008). Surgical case scheduling as a generalized job shop scheduling problem. *European Journal Of Operational Research*, 185(3):1011–1025.

Rachuba, S. and Werners, B. (2014). A robust approach for scheduling in hospitals using multiple objectives. *Journal of the Operational Research Society*, 65(4):546–556.

Ribeiro, J. A. M., Barreto, A. X. D., Varanda, J. A. V., Penedo, J. M. V. d. S., Caiado, J. C. F., Nunes, J. M., Santos, N. d. S., Boto, P. A. F., Ivo, R. d. S., and dos Santos, T. A. C. (2011). Relatório Final do Grupo Técnico para a Reforma Hospitalar. Technical report, Ministry of Health, Lisbon.

Riise, A. and Burke, E. (2011). Local search for the surgery admission planning problem. *Journal of Heuristics*, 17(4):389–414.

Roland, B., Di Martinelly, C., Riane, F., and Pochet, Y. (2010). Scheduling an operating theatre under human resource constraints. *Computers & Industrial Engineering*, 58(2):212–220.

Santana, R. (2003). O financiamento hospitalar e a definição de preços. *Revista Portuguesa de Saúde Pública*, pages 93–118.

Santibáñez, P., Begen, M., and Atkins, D. (2007). Surgical block scheduling in a system of hospitals: an application to resource and wait list management in a British Columbia health authority. *Health Care Management Science*, 10(3):269–282.

Shapiro, A. (1996). Simulation-based optimization-convergence analysis and statistical inference. *Communications in Statistics. Stochastic Models*, 12(3):425–454.

Shylo, O. V., Prokopyev, O. A., and Schaefer, A. J. (2012). Stochastic Operating Room Scheduling for High-Volume Specialties Under Block Booking. *INFORMS Journal on Computing*, 25(4):682–692.

Sieber, T. J. and Leibundgut, D. L. (2002). Operating room management and strategies in switzerland: results of a survey. *European Journal of Anaesthesiology*, 19(6):415–423.

Siebers, P. O., Macal, C. M., Garnett, J., Buxton, D., and Pidd, M. (2010). Discrete-event simulation is dead, long live agent-based simulation! *Journal of Simulation*, 4(3):204–210.

Spangler, W. E., Strum, D. P., Vargas, L. G., and May, J. H. (2004). Estimating procedure times for surgeries by determining location parameters for the lognormal model. *Health care management science*, 7(2):97–104.

Spears, V. M. and Jong, K. A. D. (1991). On the virtues of parameterized uniform crossover. In *In Proceedings of the Fourth International Conference on Genetic Algorithms*, pages 230–236.

Stepaniak, P., Heij, C., Mannaerts, G., de Quelerij, M., and de Vries, G. (2009). Modeling procedure and surgical times for current procedural terminology-anesthesia-surgeon combinations and evaluation in terms of case-duration prediction and operating room efficiency: a multicenter study. *Anesthesia & Analgesia*, 109(4):1232–45.

Strum, D. P., May, J. H., and Vargas, L. G. (2000). Modeling the Uncertainty of Surgical Procedure Times: Comparison of Log-normal and Normal Models. *Anesthesiology*, 92(4):1160–7.

Tànfani, E. and Testi, A. (2009). A pre-assignment heuristic algorithm for the Master Surgical Schedule Problem (MSSP). *Annals of Operations Research*, 178(1):105–119.

Testi, A., Tanfani, E., and Torre, G. (2007). A three-phase approach for operating theatre schedules. *Health Care Management Science*, 10(2):163–172.

Toso, R. F. and Resende, M. G. C. (2014). A c++ application programming interface for biased random-key genetic algorithms. *Optimization Methods and Software*, published online 13 March 2014.

Turban, E. (1982). Decision support systems in hospitals. *Health Care Management Review*, 7(3):35–42.

van Oostrum, J., Van Houdenhoven, M., Hurink, J., Hans, E., Wullink, G., and Kazemier, G. (2008). A master surgical scheduling approach for cyclic scheduling in operating room departments. *OR spectrum*, 30(2):355–374.

Vijayakumar, B., Parikh, P. J., Scott, R., Barnes, A., and Gallimore, J. (2013). A dual bin-packing approach to scheduling surgical cases at a publicly-funded hospital. *European Journal of Operational Research*, 224(3):583–591.

Xu, J., Nelson, B. L., and Hong, J. L. (2010). Industrial Strength COMPASS: A Comprehensive Algorithm and Software for Optimization via Simulation. *ACM Transactions on Modeling and Computer Simulation*, 20(1):1–29.

Zhong, L., Luo, S., Wu, L., Xu, L., Yang, J., and Tang, G. (2014). A two-stage approach for surgery scheduling. *Journal of Combinatorial Optimization*, 27(3):545–556.

Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C. M., and da Fonseca, V. G. (2003). Performance assessment of multiobjective optimizers: an analysis and review. *Evolutionary Computation, IEEE Transactions on*, 7(2):117–132.