Multi-objective Simulation Optimization for Surgery Scheduling under Uncertainty

Fabrício Sperandio, José Borges, Bernardo Almada-Lobo

UGEI – INESC TEC, Faculty of Engineering, University of Porto, Portugal

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 schedule's performance under uncertainty. This multi-objective approach offers 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. Schedule's performance is estimated using a DES model featuring four stochastic variables: surgery duration, emergencies, cancellations and delays/advances in the start of the first surgery in each shift. The proposed approach is compared against 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

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 [1]. For instance, the average healthcare expenditure among OECD countries was rising steadily until 2010, when it fell sharply, and until now did not recover its historical growth rates [2]. 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 [3] 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 Operating Room (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 & Guido [4] concluded that the majority of published papers assume that processing times and recovery times are known in advance. In addition, Cardoen et al. [5] 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. [6] 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 [5]. 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 [4]. 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 (DES) 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. 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 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 and discussion & 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.

Paper	Problems A	B	Objectives A	B	C	D	E	Resources A	B	C	D	Constraints A	B	C	Uncertainty A	B	C	D	Approaches A	B	C	D
Addis et al. [7]	●		●					●	●			●	◔		●				●			
Bruni et al. [8]	●		◔			◔		●	●			●	◔		●	●			●			
Hans et al. [9]	◔			●		●		●	●		◔	●			●					●		
Lamiri et al. [10]	●		●			●		◔	●			●	●			●			●			
Lamiri et al. [11]	●		●			●		◔	●			◔	◔			●			●	●		
Min & Yih [12]	●					●		●	●			●			●	●			●			
Rachuba & Werners [13]	●		●	●		●		●	●			●			●	●			●			
Shylo et al. [14]	●			●					●			●			◔				●			
Batun et al. [15]		●		●		●			●	●	◔	●			●				●			
Denton et al. [16]		●		●		●	◔					●			●				●	●		
Gul et al. [17]		●				●	●		●		◔	●			●					●	●	●
Lee & Yih [18]		●		●		●	●		●		◔	●			●				●			
Mancilla & Storer [19]		●		●		●	●					●			●				●	●		
Dexter et al. [20]	●	●		●				◔	●	●		●	◔		●					●	●	
# papers (out of 14)	10	6	5	8	0	11	4	14	2	1	0	14	5	0	12	5	0	0	11	6	2	1
This paper	●	●	●	●	●	●		●		●	●		●	●	●	●		●	●	●		

Problems		**Constraints**	
- Advance scheduling	A	- Resources capacity	A
- Allocation scheduling	B	- Patient priority and waiting time	B
Objectives		- Surgeon availability and working limits	C
- Max. number of performed surgeries	A	**Sources of uncertainty**	
- Max. OR occupancy rate	B	- Surgery duration	A
- Min. number of cancelled surgeries	C	- Emergencies	B
- Min. total minutes of overtime	D	- Cancelled surgeries	C
- Min. patient and/or OR team waiting in the process flow	E	- Delays/advances starting the first surgery in each shift	D
Resources		**Solution Approaches**	
- Time blocks (OR-days)	A	- Stochastic programming, robust optimization or chance constraints	A
- Multiple ORs	B	- Heuristics or metaheuristics	C
- Multiple surgeons	C	- Discrete-event simulation	D
- Pre/post-surgical resources, surgical staff	D	- Multi-objective optimization	E

Table 1: Summary table of the literature review

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 speciality 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 speciality 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 & Demeulemeester [5], Guerriero & Guido [4] and May et al. [6]. Table 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 1 these problems have been addressed in separate. 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 a specific OR and day over a weekly planning horizon. However, there are small variations of this sub-problem. For instance, Lamiri et al. [10, 11] 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. [9] 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. [16] and Mancilla & Storer [19] 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 1 shows that 11 studies take this objective explicitly into account. In contrast, Shylo et al. [14] and Addis et al. [7] do not consider it in the objective function. However, these studies rely on robust optimization, which guarantees acceptable levels 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 [16, 17] and another on clinical resources [15, 18, 19]. We propose to take four objectives into account: (1) maximize the number of performed surgeries, (2) maximize average OR occupancy rates, (3) minimize number of cancelled surgeries and (4) minimize total minutes of overtime. The third objective was not explicitly addressed by none 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 et al. [20] consider only surgeon block time, a combination between surgeon and day, and Lamiri et al. [10, 11] consider only days of the planning horizon. Also, Rachuba & Werners [13] 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. [15] consider the intensive care unit (ICU), Lee et al. [18] address the post-anaesthesia care unit (PACU) and Gul et al. [17] look at pre/post-surgical resources (waiting area and intake/recovery rooms). Hans et al. [9] are the only to consider additional OR personnel. Finally, regarding resources, only two of the reviewed papers take surgeons explicitly into account [20, 15]. 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. An exception are OR cleaning times and surgeon turnover times. Patient urgency and waiting time limits are often addressed using penalties in the objective function. Table 1 indicates the papers which address this issue in the objective function. In addition, Rachuba & Werners [13] 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 derive from the Portuguese legislation.

The variability in surgery durations is the main source of uncertainty taken into account in the literature. Moreover, a few papers take into account the OR-time occupied by emergencies and Min & Yih [12] consider the length of stay in the intensive case unit. It is worth noting 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 [12] and sampling directly from historical data [16] to using uniform probability distributions with fixed parameters [10, 11]. In addition, Batun et al. [15] decompose the surgery duration in pre-incision, incision and post-incision and Shylo et al. [14] 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 & Yih [12] 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. [9] 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.

In which concerns 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 (SAA) method. In order to reduce the computational cost researchers have been applying decomposition approaches such as Bender's decomposition [19] and the L-Shaped method [15]. 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. [14] propose a chance-constrained model to ensure acceptable levels of overtime and Rachuba & Werners [13] apply fuzzy sets to merge the interests of different stakeholders. We propose an approach based on Simulation Optimization (SO) combining a Multi-Objective Evolutionary Algorithm (MOEA) 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 & Guido [4] 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. [21] propose a DES model for perioperative process improvement and Konrad et al. [22] 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 [23], optimization via simulation [24], simulation optimization [25]. 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 [24, 26, 25, 27, 28, 29, 30, 31], most of them in the proceedings of the Winter Simulation Conference. For instance, Fu [25] 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.

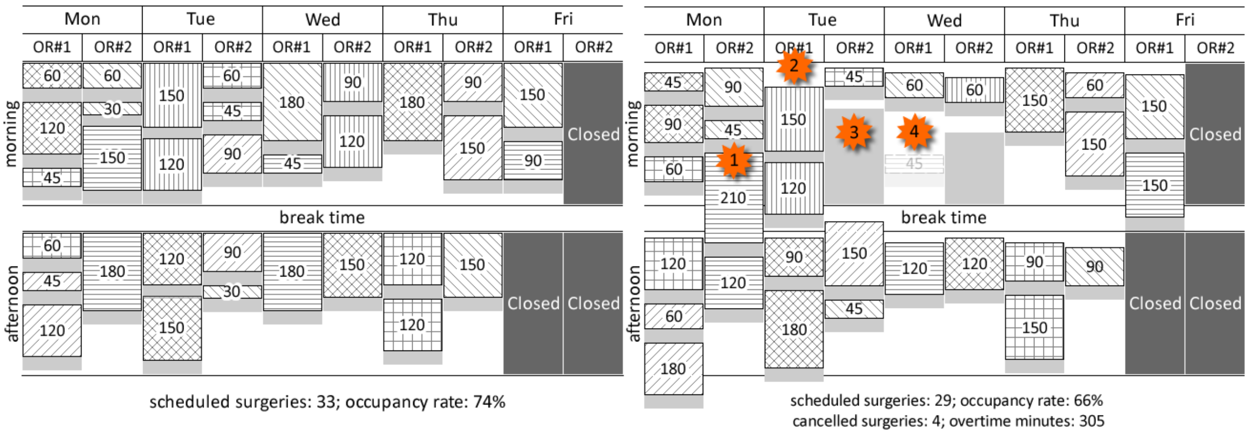
3. Problem Description

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 the problem of assigning sufficient planned slack to each working shift to deal with unveiled uncertainty.

In Portugal, the elective surgery waiting list was introduced in 2004 by the SIGIC program 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 [32] 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, Sperandio et al. [33] propose a Decision Support System to support scheduling activities in Portuguese hospitals. In order to illustrate the scheduling problem, Figure 1(a) shows a valid weekly schedule for a hypothetical surgical department.

The example schedule in Figure 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 1: Impact of uncertainty in the weekly surgery schedule

Figure 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 minutes of 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.

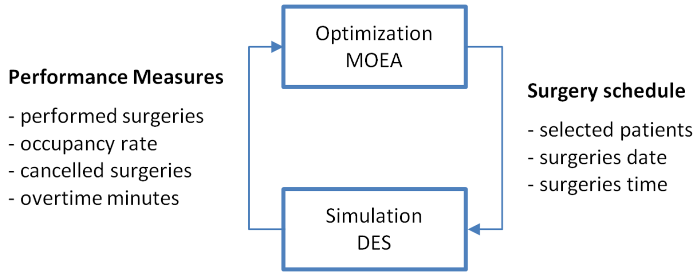


Figure 2: The Simulation Optimization loop

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 (MOEA) and the latter incorporates a Discrete-Event Simulation (DES) 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 2 illustrates the integration between simulation and optimization.

4.1. Multi-objective Evolutionary Algorithm

The optimization module implements a Multi-objective Evolutionary Algorithm (MOEA). The actual algorithm is a customized version of the NSGA-II [34] algorithm for multi-objective optimization. It 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 & Resende [35]. Preliminary results show that this approach outperforms approaches based on encoding schemes using binary variables, like the one proposed by Conforti et al. [36]. In fact, GAs were originally designed for unconstrained optimization problems, therefore they are more efficient searching in the feasible solution space only.

Figure 3 illustrates an example GA chromosome representation using real variables. For simplification purposes, in this figure as well as in Figure 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.

	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 3: An example GA chromosome representation based on real variables

Figure 4 illustrates the decoding procedure. First, Figure 4(a) shows an 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 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 4(c) shows the associated time blocks and starting times assigned by the decoding procedure and Figure 4(d) illustrates the resulting surgery schedule. For simplicity, the example schedule highlights only the five first surgeries. Moreover, it shows a solid line in 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.

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

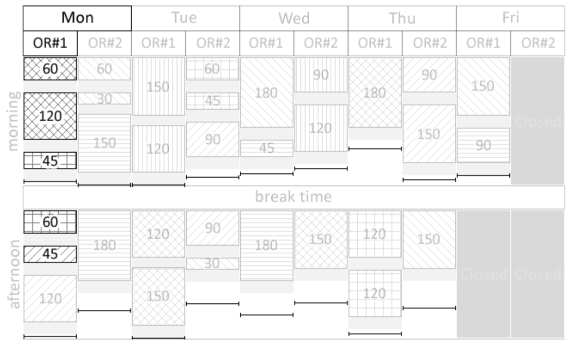
(a) Input set of surgeries with associated random numbers

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 4: Illustrative example of the decoding procedure

The algorithm to decode a chromosome into a feasible surgery schedule is composed of two phases. The first phase generates schedules meeting all the requirements but the patient priority and waiting time rules, which are tackled in the second phase. Algorithm 1 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 16-18. 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 schedule 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. Finally, if the solution meets all the requirements, the procedure completes and returns a solution, otherwise the main loop is repeated.

Algorithm 1: 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[i].timeBlock ⟵ i
                    solution[i].scheduled ⟵ true
                    solution[i].startTime ⟵ startTime
                    currentIndex ⟵ currentIndex + 1
                    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, 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 [34] for real-coded GAs.

4.2. Discrete Event Simulation

The simulation module of the integrated solution approach implements a stochastic Discrete Event Simulation (DES) model. The model structure and behaviour are based on the Adevs [37] 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 of the model is modelled with 4 random variables whose behaviour is based on historical data from a large Portuguese hospital.

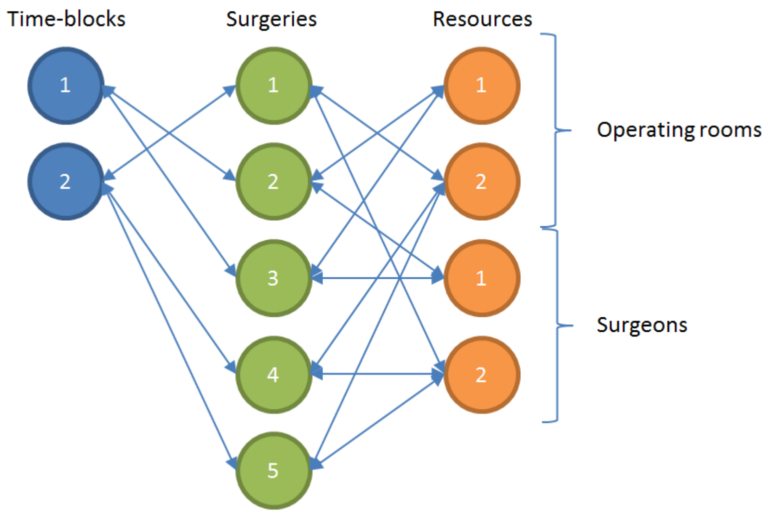


Figure 5: A simple network model showing the three types of components and the connections among them

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 2: Set of sequential states of each atomic model

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 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 sends a message to the Resource components to release them.

Table 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 'request' and send messages to the two required resources (surgeon and operating room). Both resources receive the messages, change state to 'not available' and send a message back to the surgery. The surgery receives both messages, change state to 'work' and schedules the next event to the end of the simulated surgery. When the scheduled event is triggered, the surgery changes state to 'release' and sends messages to the resources been used to release them. Both resources receive the messages, change state to 'available' 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

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

Table 3: Attributes used for modelling the behaviour of stochastic variables

cancelled surgeries would be presumably much higher.

The stochastic behaviour of the simulation model is modelled by 4 stochastic variables: (1) duration of surgical procedures; (2) cancelled surgeries; (3) total time (in minutes) occupied by emergencies; (4) delays/advances (in minutes) on the start of the first surgery in each shift. Table 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 a 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 minutes of 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. Computational Experiments

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 the estimate performance measures for each solution. Six alternative configurations are evaluated, each with the following number of replications each: 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.

The third analysis is not based on new computational experiments, but consists in comparing the best configurations of the deterministic and the simulation optimization approaches. 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 2: 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()
```

Surgical speciality	# Patients	# Surgeons	# Procedures	# Time blocks	Surgery duration Avg. length	Surgery duration 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
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

Table 4: Characteristics of the testing instances

5.2. Testing Instances

The computational experiments are performed over a set of 8 testing instances built with real data from a large Portuguese hospital. Each instance concerns a different surgical speciality and represents different testing settings in terms of the size of the problem and the degree of uncertainty. Table 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 guide.

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. [38] and Zitzler et al. [39]. 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 2 shows the steps performed to compute the performance measures for all specialities and alternative configurations. Note that, for each configuration, the RunExperiment

	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 6: Illustrative example of the analysis of the performance assessment measures

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 need 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 [40] 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 6 illustrates the assessment procedure showing the results of different configurations of the deterministic approach applied to the Ophthalmology speciality. In this example, configurations 1, 2 and 3 represent 0, 10% 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 against configuration 1 in the Epsilon indicator. In this case, as the result of the statistical test is lower than the significance level (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 were implemented in C++ and the computational experiments were carried out on an Intel Xeon

	0	10%	20%
0	-	0	0
10%	6	-	2
20%	5	0	-

Table 5: Summary of the comparison of different configurations of the deterministic approach

	5	25	50	75	100	150	better
5	-	0	0	0	0	2	5%
25	8	-	2	2	3	3	45%
50	7	3	-	0	3	5	45%
75	6	3	0	-	3	5	43%
100	5	1	0	0	-	3	23%
150	5	1	1	0	0	-	18%
worse	78%	25%	9%	6%	28%	50%	

Table 6: Summary of the comparison between different configurations of the simulation optimization approach

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.

5.4. Experimental Results

5.4.1. Deterministic Approach

Table 5 shows a summary of the statistical tests applied to quality indicators across all specialities. 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, none of the statistical tests indicates that it is better. On the other hand, comparing 2 against 1, in 6 out of 8 instances the tests indicate that 2 is better. Furthermore, comparing 3 against 1, in 5 out of 8 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 none. 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.

5.4.2. 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 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 5% of the total comparisons. Also, in 78% 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 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 18% of the cases, but most of them are comparisons against 5 replications (only 2 are higher than 5). The results also show that in 50% 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.

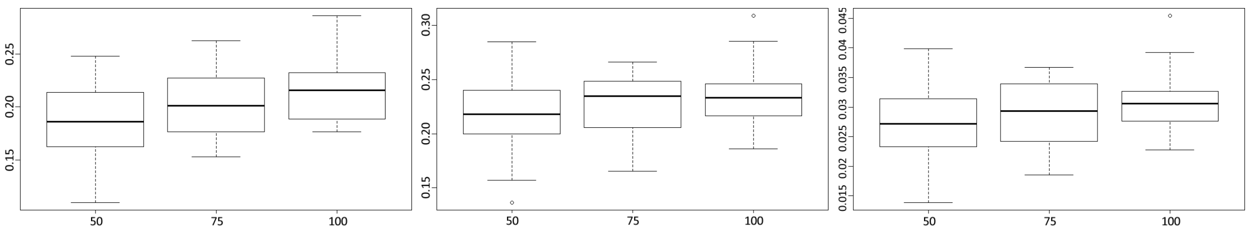
When the number of replications is between 5 and 150 the aggregated results are not so clear and can be misleading. The results indicate that 50 replications are better than other options in 45% of the comparisons. However, it clearly depends on the characteristics of the instance, such as the problem size and degree of uncertainty. For instance, Table 7 marks with "X" the best configurations for each surgical speciality. In Urology and General surgery 1, 25

	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
Urology		X				
Otolaryngology			X	X		
General surgery 1		X				
General surgery 2				X	X	X

Table 7: Configurations in which it is not able to determine a better configuration

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 2.75 ties. In order to further characterize the influence of the number of replications in the algorithm performance it is necessary to analyse indicators data.

Figure 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

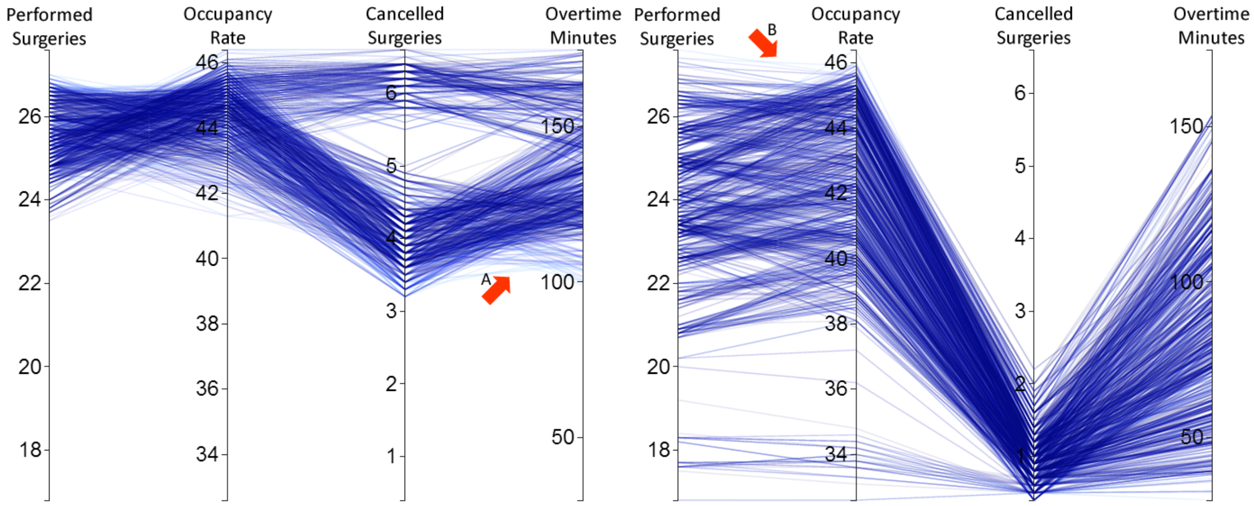
Figure 7: Descriptive statistics of the indicators data for Vascular surgery

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 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 from 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.

Figure 8(a) shows that the deterministic approach generates solutions with an excessive number of cancelled surgeries and overtime minutes. On the other hand, Figure 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 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.

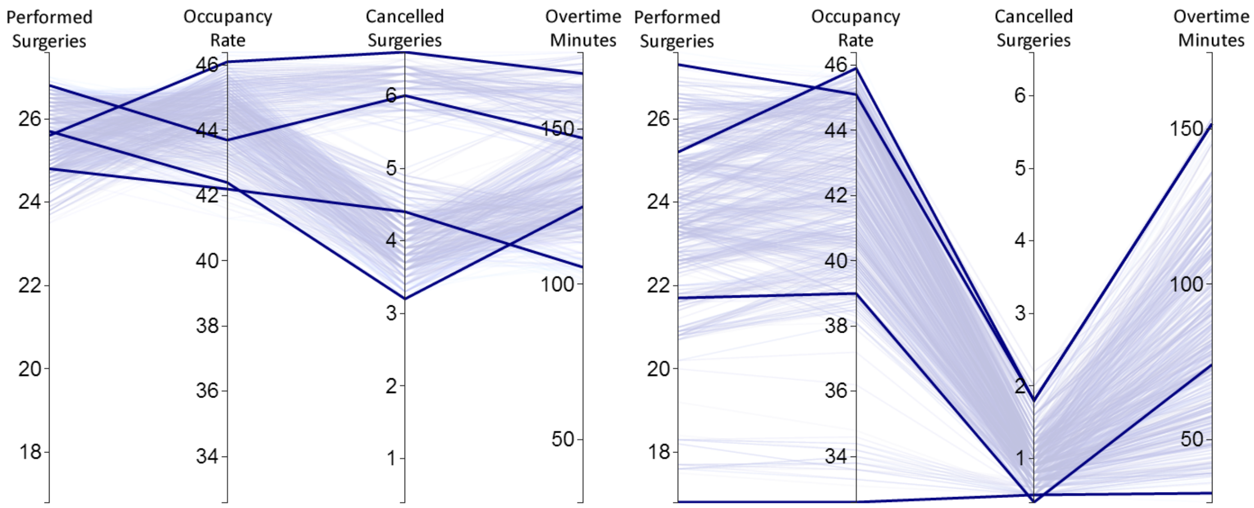
Similar results were obtained to other instances. For instance, Figure 10 shows the points delimiting attainment surfaces for the Ophthalmology department. It is a quite demanding instance type due to the short duration of the surgeries and high variance. The deterministic approach with a 10% planned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



(a) Deterministic approach

(b) Simulation optimization approach

Figure 8: Points delimiting the attainment surfaces of multiple runs of the MOEA for the Vascular Surgery department



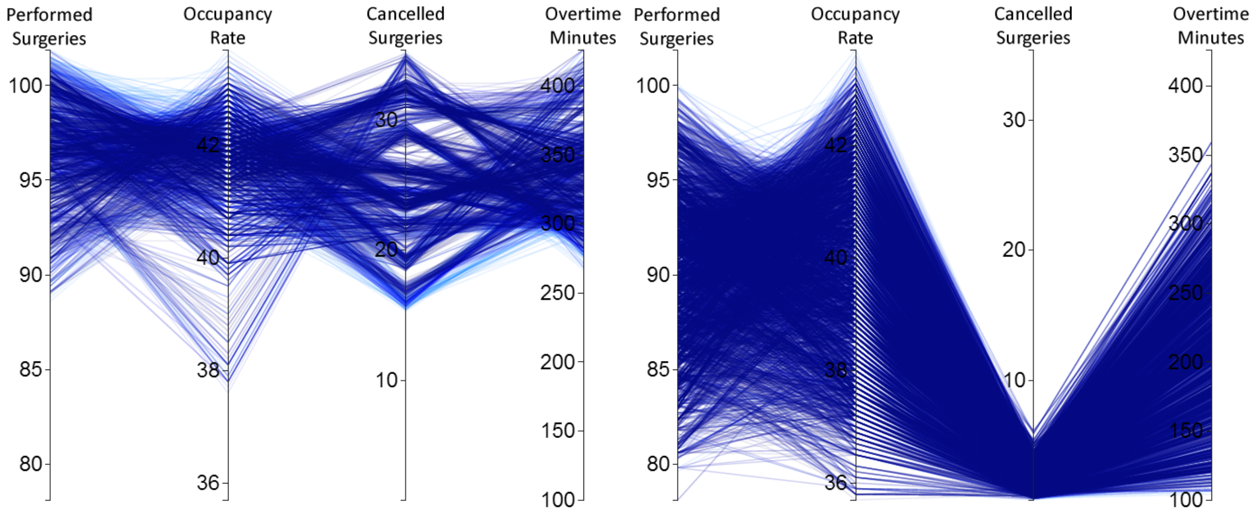
(a) Deterministic approach

(b) Simulation optimization approach

Figure 9: Best solutions in each objective for the Vascular Surgery department

tions with a reduced number of cancelled surgeries and overtime minutes, while keeping high numbers of performed surgeries and occupancy rates.

Figure 11 shows a set of matrices comparing the best configurations of the deterministic MOEA against the best configurations of the simulation optimization approach based on the results of the Kruskal-Wallis statistical testing procedure (at a 95% confidence level) in three different quality indicators. These results indicate that the simulation optimization approach is better than the deterministic MOEA in all tested instances. The consistency across the quality indicators means the Pareto sets generated by the SO approach are better than the ones generated by the deterministic approach in every aspect considered by each quality indicator, therefore it is safe to conclude that the SO approach is better. In addition, the testing instances are based on real data and represent different characteristics pertinent to each surgical specialty, such as: number of patients, number of surgeons, number of working shifts and degree of uncertainty. It means the SO approach is suitable to deal with different surgical specialties working in a large hospital.



(a) Deterministic approach

(b) Simulation optimization approach

Figure 10: Points delimiting the attainment surfaces of multiple runs of the MOEA for the Ophthalmology department

6. Discussion & Future Work

This paper presents a multi-objective simulation optimization approach for the surgery scheduling problem under uncertainty. The aim is to generate 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 against 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 the 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 of the estimation noise, but should not be too high because of the estimation cost.

In a 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

Vascular surgery

Epsilon Indicator

	0	10%	20%
5	2,85037E-10	4,06438E-07	1,08103E-07
25	2,87107E-35	3,06179E-29	2,12041E-30
50	7,64016E-32	4,36068E-26	3,43359E-27
75	5,78677E-37	8,24502E-31	5,37967E-32
100	1,07983E-48	8,51541E-42	3,90264E-43
150	1,50881E-52	1,90253E-45	7,91066E-47

Hypervolume Indicator

	0	10%	20%
5	2,00331E-10	2,81651E-07	4,14325E-07
25	2,35374E-42	6,55596E-36	1,53899E-35
50	5,57859E-52	5,43829E-45	1,38118E-44
75	2,01351E-63	6,40125E-56	1,75208E-55
100	3,76866E-79	4,05434E-71	1,19966E-70
150	1,4696E-95	3,88369E-87	1,21808E-86

R Indicator

	0	10%	20%
5	6,26118E-10	6,99654E-07	6,80291E-07
25	7,42892E-41	1,58785E-34	1,49126E-34
50	8,70102E-43	2,46743E-36	2,31427E-36
75	1,97534E-51	1,7208E-44	1,60552E-44
100	6,47948E-66	2,43792E-58	2,25895E-58
150	9,20613E-80	9,74107E-72	8,98155E-72

Overall result
Is line i better than column j?

	0	10%	20%
5			
25			
50	Yes	Yes	Yes
75	Yes	Yes	Yes
100	Yes	Yes	Yes
150			

Oral and maxillofacial surgery

Epsilon Indicator

	0	10%	20%
5	5,50538E-07	7,07355E-05	0,000052035
25	9,84323E-25	1,97212E-20	1,01089E-20
50	1,39552E-31	9,33206E-27	4,38364E-27
75	2,1244E-52	1,89848E-46	7,38024E-47
100	4,31946E-69	1,49976E-62	5,27486E-63
150	1,05092E-75	5,53824E-69	1,88845E-69

Hypervolume Indicator

	0	10%	20%
5	7,62589E-10	9,08872E-07	3,60961E-08
25	2,78445E-41	7,97755E-35	7,00674E-38
50	4,32615E-56	8,24255E-49	2,83414E-52
75	4,05537E-52	4,93401E-45	2,11922E-48
100	1,03497E-55	1,89246E-48	6,6413E-52
150	6,44981E-51	6,79803E-44	3,13508E-47

R Indicator

	0	10%	20%
5	1,40949E-09	1,60639E-06	3,52036E-08
25	7,76867E-33	6,30064E-27	3,13191E-30
50	3,67554E-49	3,74768E-42	4,27752E-46
75	9,29739E-52	1,30755E-44	1,23729E-48
100	1,29928E-63	6,42864E-56	2,91728E-60
150	8,62171E-57	2,15145E-49	1,45678E-53

Overall result
Is line i better than column j?

	0	10%	20%
5			
25			
50	Yes	Yes	Yes
75	Yes	Yes	Yes
100	Yes	Yes	Yes
150	Yes	Yes	Yes

Neurosurgery

Epsilon Indicator

	0	10%	20%
5	6,70594E-34	9,69755E-13	5,83607E-12
25	5,67185E-77	2,55674E-45	6,37472E-44
50	1,71402E-61	1,20125E-32	2,01242E-31
75	7,16338E-55	1,67494E-27	2,28972E-26
100	1,08641E-41	6,69634E-18	5,66863E-17
150	2,09117E-32	8,34166E-12	4,67397E-11

Hypervolume Indicator

	0	10%	20%
5	1,17076E-21	0,00077831	0,000280402
25	1,0678E-77	5,20602E-40	1,73838E-41
50	2,11363E-73	1,35286E-36	5,07396E-38
75	3,27699E-64	1,53938E-29	7,60855E-31
100	1,03411E-56	4,55663E-24	2,90994E-25
150	2,73938E-43	4,50739E-15	4,95058E-16

R Indicator

	0	10%	20%
5	1,55811E-26	2,00215E-05	4,60442E-06
25	5,49485E-88	3,72823E-47	5,78369E-49
50	3,2077E-82	1,97353E-42	3,58421E-44
75	7,36218E-72	3,04855E-34	7,63106E-36
100	8,35757E-61	5,32176E-26	1,99371E-27
150	1,27745E-44	4,0378E-15	3,1387E-16

Overall result
Is line i better than column j?

	0	10%	20%
5			
25		Yes	Yes
50		Yes	Yes
75		Yes	Yes
100			
150			

Ophthalmology

Epsilon Indicator

	0	10%	20%
5	0,0980582	0,998004	0,329606
25	7,84054E-10	0,028012	9,74373E-08
50	1,45139E-49	1,10982E-28	5,96886E-45
75	2,10206E-77	2,41661E-52	4,6356E-72
100	4,94489E-75	3,02691E-50	9,79716E-70
150	3,90381E-68	3,20743E-44	5,48038E-63

Hypervolume Indicator

	0	10%	20%
5	0,0421761	0,821366	0,00652258
25	3,73796E-68	1,37678E-52	7,75747E-73
50	1,8202E-128	3,2074E-110	9,4421E-134
75	2,017E-134	4,505E-116	1,0027E-139
100	3,2232E-106	1,45208E-88	2,2091E-111
150	3,97251E-85	2,01789E-68	4,44283E-90

R Indicator

	0	10%	20%
5	0,0901484	0,840149	0,0146829
25	4,31163E-47	4,07121E-35	1,0227E-51
50	1,6897E-104	6,04174E-89	3,3181E-110
75	2,1867E-111	1,23707E-95	3,7875E-117
100	8,93514E-89	8,11261E-74	2,5934E-94
150	1,93889E-73	2,70997E-59	9,80234E-79

Overall result
Is line i better than column j?

	0	10%	20%
5			
25		Yes	
50		Yes	
75		Yes	
100		Yes	
150		Yes	

Urology

Epsilon Indicator

	0	10%	20%
5	9,69372E-08	1	0,991994
25	8,04299E-47	0,699404	1,34437E-13
50	2,07882E-77	1,79003E-06	3,12806E-34
75	3,02317E-68	0,000739916	1,65673E-27
100	1,85406E-56	0,114564	1,70439E-19
150	1,37858E-30	0,999915	1,76732E-05

Hypervolume Indicator

	0	10%	20%
5	2,22853E-15	0,0014572	0,00963062
25	7,94904E-86	2,88607E-55	1,19439E-51
50	7,8346E-110	5,20244E-77	5,36831E-73
75	3,78987E-89	3,31975E-58	1,5963E-54
100	6,41762E-69	9,7681E-41	1,63289E-37
150	1,61724E-36	8,64703E-16	1,01931E-13

R Indicator

	0	10%	20%
5	5,99135E-13	0,951387	0,0538791
25	6,78018E-72	6,4611E-25	4,25673E-40
50	1,0395E-96	4,48736E-43	2,34886E-61
75	1,75135E-78	1,87409E-29	1,42696E-45
100	1,63358E-61	2,8886E-18	8,96165E-32
150	5,05939E-33	0,000262312	1,23713E-11

Overall result
Is line i better than column j?

	0	10%	20%
5			
25		-	Yes
50			
75			
100			
150			

Otolaryngology

Epsilon Indicator

	0	10%	20%
5	3,29869E-11	0,999991	0,073893
25	2,67402E-18	0,982575	0,000134971
50	4,80471E-41	0,00119646	2,5239E-18
75	3,51803E-67	1,88378E-14	2,03813E-38
100	7,66953E-88	9,02384E-27	4,77667E-56
150	6,14969E-80	9,64886E-22	3,65054E-49

Hypervolume Indicator

	0	10%	20%
5	1,72774E-21	0,692096	0,0033313
25	5,69024E-59	1,51941E-12	5,55914E-24
50	9,82022E-98	3,19543E-37	2,15196E-54
75	3,1666E-109	5,22906E-46	2,29692E-64
100	1,0817E-118	1,24781E-53	8,02413E-73
150	1,1911E-103	1,11972E-41	1,74656E-59

R Indicator

	0	10%	20%
5	5,16729E-18	0,991146	0,00082187
25	3,83484E-40	0,00368726	3,07997E-16
50	1,79152E-74	7,89183E-18	2,65012E-42
75	1,13696E-93	7,1573E-30	9,34076E-59
100	3,7852E-106	9,33489E-39	6,24945E-70
150	4,70174E-93	1,8787E-29	3,26307E-58

Overall result
Is line i better than column j?

	0	10%	20%
5			
25			
50		Yes	
75		Yes	
100			
150			

General surgery 1

Epsilon Indicator

	0	10%	20%
5	4,52863E-11	0,073976	0,853744
25	9,04949E-25	4,84548E-08	0,00209897
50	6,60374E-50	5,57776E-25	1,39884E-15
75	8,37352E-63	4,38751E-35	6,46008E-24
100	4,5412E-52	1,26675E-26	6,67334E-17
150	7,50519E-53	3,18445E-27	2,18021E-17

Hypervolume Indicator

	0	10%	20%
5	1,93875E-18	0,000646235	0,0490961
25	1,74879E-52	9,2888E-25	1,23479E-18
50	8,58669E-92	7,96161E-57	5,25508E-48
75	6,01297E-95	1,33516E-59	1,25309E-50
100	1,4558E-84	1,47277E-50	4,04648E-42
150	2,96436E-63	5,49039E-33	5,91377E-26

R Indicator

	0	10%	20%
5	1,66771E-20	0,000503501	0,112596
25	1,81072E-51	4,60236E-22	1,37334E-14
50	3,50125E-89	8,25854E-52	1,23917E-40
75	2,09317E-92	1,40888E-54	3,59228E-43
100	7,74666E-82	1,30027E-45	5,26928E-35
150	1,29589E-63	5,15495E-31	4,07878E-22

Overall result
Is line i better than column j?

	0	10%	20%
5			
25		Yes	Yes
50			
75			
100			
150			

General surgery 2

Epsilon Indicator

	0	10%	20%
5	0,0132926	1	0,525264
25	2,20091E-21	0,0256988	2,07231E-13
50	5,49383E-31	0,000013521	2,59957E-21
75	1,45862E-37	1,42063E-08	5,9065E-27
100	1,04244E-34	3,21246E-07	1,73633E-24
150	9,43411E-45	2,39496E-12	2,78984E-33

Hypervolume Indicator

	0	10%	20%
5	0,993523	1	1
25	1,84639E-14	0,128778	0,00316091
50	7,37705E-27	5,45247E-06	1,18283E-09
75	1,55796E-44	8,55461E-16	1,29181E-21
100	2,17972E-44	1,05791E-15	1,65399E-21
150	2,34004E-63	8,06411E-29	2,62057E-36

R Indicator

	0	10%	20%
5	0,0181759	1	1
25	2,09734E-29	0,120521	0,000387195
50	2,10359E-46	2,27584E-06	8,36615E-12
75	5,00499E-67	3,60704E-16	1,88113E-24
100	1,16455E-64	7,20837E-15	7,34072E-23
150	3,54444E-85	2,33374E-27	1,33439E-37

Overall result
Is line i better than column j?

	0	10%	20%
5			
25			
50			
75		Yes	Yes
100		Yes	Yes
150		Yes	Yes

Figure 11: Results of the Kruskal–Wallis test comparing runs of the deterministic and simulation optimization versions of the MOEA with different configurations

FCT Funda£o para a Ciłncia e a Tecnologia (Portuguese Foundation for Science and Technology) within project FCOMP-01-0124-FEDER-037281.

References

[1] C. de la Maisonneuve, J. O. Martins, Public spending on health and long-term care: A new set of projections, Main Paper, Economic Policy Paper 6, OECD, Paris (jun 2013). doi:10.1787/5k44t7jwwr9x-en.

[2] OECD, Health spending continues to stagnate, says oecd, on-line, accessed 25 October 2013 (jun 2013).
URL `http://www.oecd.org/newsroom/health-spending-continues-to-stagnate-says-oecd.htm`

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

[4] F. Guerriero, R. Guido, Operational research in the management of the operating theatre: a survey, Health Care Management Science 14 (1) (2011) 89–114. doi:10.1007/s10729-010-9143-6.

[5] B. Cardoen, E. Demeulemeester, J. Beliën, Operating room planning and scheduling: A literature review, European Journal of Operational Research 201 (3) (2010) 921–932. doi:10.1016/j.ejor.2009.04.011.

[6] J. H. May, W. E. Spangler, D. P. Strum, L. G. Vargas, The Surgical Scheduling Problem: Current Research and Future Opportunities, Production and Operations Management 20 (3) (2011) 392–405. doi:10.1111/j.1937-5956.2011.01221.x.

[7] B. Addis, G. Carello, E. Tànfani, A robust optimization approach for the operating room planning problem with uncertain surgery duration, in: A. Matta, J. Li, E. Sahin, E. Lanzarone, J. Fowler (Eds.), Proceedings of the International Conference on Health Care Systems Engineering, Vol. 61 of Springer Proceedings in Mathematics & Statistics, Springer International Publishing, 2014, pp. 175–189. doi:10.1007/978-3-319-01848-5_14.
URL `http://dx.doi.org/10.1007/978-3-319-01848-5_14`

[8] M. E. Bruni, P. Beraldi, D. Conforti, A stochastic programming approach for operating theatre scheduling under uncertainty, IMA Journal of Management Mathematicsdoi:10.1093/imaman/dpt027.

[9] E. Hans, G. Wullink, M. van Houdenhoven, G. Kazemier, Robust surgery loading, European Journal of Operational Research 185 (3) (2008) 1038–1050. doi:10.1016/j.ejor.2006.08.022.

[10] M. Lamiri, X. Xie, A. Dolgui, F. Grimaud, A stochastic model for operating room planning with elective and emergency demand for surgery, European Journal of Operational Research 185 (3) (2008) 1026–1037. doi:10.1016/j.ejor.2006.02.057.

[11] M. Lamiri, F. Grimaud, X. Xie, Optimization methods for a stochastic surgery planning problem, International Journal of Production Economics 120 (2) (2009) 400–410. doi:10.1016/j.ijpe.2008.11.021.

[12] D. Min, Y. Yih, Scheduling elective surgery under uncertainty and downstream capacity constraints, European Journal of Operational Research 206 (3) (2010) 642–652. doi:10.1016/j.ejor.2010.03.014.

[13] S. Rachuba, B. Werners, A robust approach for scheduling in hospitals using multiple objectives, Journal of the Operational Research Society 65 (4) (2014) 546–556. doi:10.1057/jors.2013.112.

[14] O. V. Shylo, O. A. Prokopyev, A. J. Schaefer, Stochastic Operating Room Scheduling for High-Volume Specialties Under Block Booking, INFORMS Journal on Computing 25 (4) (2012) 682–692. doi:10.1287/ijoc.1120.0530.

[15] S. Batun, B. T. Denton, T. R. Huschka, A. J. Schaefer, Operating Room Pooling and Parallel Surgery Processing Under Uncertainty, INFORMS Journal on Computing 23 (2) (2010) 220–237. doi:10.1287/ijoc.1100.0396.

[16] B. Denton, J. Viapiano, A. Vogl, Optimization of surgery sequencing and scheduling decisions under uncertainty, Health Care Management Science 10 (1) (2007) 13–24. doi:10.1007/s10729-006-9005-4.

[17] S. Gul, B. T. Denton, J. W. Fowler, T. Huschka, Bi-criteria scheduling of surgical services for an outpatient procedure center, Production and Operations Management 20 (3) (2011) 406–417. doi:10.1111/j.1937-5956.2011.01232.x.

[18] S. Lee, Y. Yih, Reducing patient-flow delays in surgical suites through determining start-times of surgical cases, European Journal of Operational Research (2014) –Article in Press. doi:10.1016/j.ejor.2014.03.043.

[19] C. Mancilla, R. Storer, A sample average approximation approach to stochastic appointment sequencing and scheduling, IIE Transactions 44 (8) (2011) 655–670. doi:10.1080/0740817X.2011.635174.

[20] F. Dexter, A. Macario, R. D. Traub, M. Hopwood, D. A. Lubarsky, An operating room scheduling strategy to maximize the use of operating room block time: computer simulation of patient scheduling and survey of patients' preferences for surgical waiting time., Anesthesia & Analgesia 89 (1) (1999) 7–20. doi:10.1213/00000539-199907000-00003.

[21] S. Azari-Rad, A. Yontef, D. M. Aleman, D. R. Urbach, A simulation model for perioperative process improvement, Operations Research for Health Care 3 (1) (2014) 22–30. doi:10.1016/j.orhc.2013.12.003.

[22] R. Konrad, K. DeSotto, A. Grocela, P. McAuley, J. Wang, J. Lyons, M. Bruin, 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) (2013) 66–74. doi:10.1016/j.orhc.2013.04.001.

[23] A. Shapiro, Simulation-based optimization-convergence analysis and statistical inference, Communications in Statistics. Stochastic Models 12 (3) (1996) 425–454. doi:10.1080/15326349608807393.

[24] M. C. Fu, Optimization via simulation: a review, Annals of Operations Research 53 (1994) 199–248. doi:10.1007/BF02136830.

[25] M. C. Fu, Optimization for Simulation: Theory vs. Practice, INFORMS Journal on Computing 14 (3) (2002) 192–215. doi:10.1287/ijoc.14.3.192.113.

[26] F. Glover, J. Kelly, M. Laguna, New advances for wedding optimization and simulation, in: Proceedings of the 1999 Winter Simulation Conference, Vol. 1, 1999, pp. 255–260. doi:10.1109/WSC.1999.823081.

[27] J. April, F. Glover, J. P. Kelly, M. Laguna, Practical introduction to simulation optimization, in: Proceedings of the 2003 Winter Simulation Conference, Vol. 1, 2003, pp. 71–78. doi:10.1109/WSC.2003.1261410.

[28] M. C. Fu, F. W. Glover, J. April, Simulation optimization: a review, new developments, and applications, in: Proceedings of the 2005 Winter Simulation Conference, 2005, p. 13 pp. doi:10.1109/WSC.2005.1574242.
[29] M. C. Fu, C.-H. Chen, L. Shi, Some topics for simulation optimization, in: Proceedings of the 2008 Winter Simulation Conference, 2008, pp. 27–38. doi:10.1109/WSC.2008.4736053.
[30] L. J. Hong, B. L. Nelson, A brief introduction to optimization via simulation, in: Proceedings of the 2009 Winter Simulation Conference, 2009, pp. 75–85. doi:10.1109/WSC.2009.5429321.
[31] G. Figueira, B. Almada-Lobo, Hybrid simulationoptimization methods: A taxonomy and discussion, Simulation Modelling Practice and Theory 46 (0) (2014) 118 – 134, simulation-Optimization of Complex Systems: Methods and Applications. doi:http://dx.doi.org/10.1016/j.simpat.2014.03.007.
[32] I. Marques, M. E. Captivo, M. Vaz Pato, Scheduling elective surgeries in a portuguese hospital using a genetic heuristic, Operations Research for Health Care (2013) –Article in Press. doi:10.1016/j.orhc.2013.12.001.
[33] F. Sperandio, C. Gomes, J. Borges, A. Carvalho Brito, B. Almada-Lobo, An Intelligent Decision Support System for the Operating Theater: A Case Study, IEEE Transactions on Automation Science and Engineering 11 (1) (2014) 265–273. doi:10.1109/TASE.2012.2225047.
[34] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, A fast and elitist multiobjective genetic algorithm: NSGA-II, in: IEEE Transactions on Evolutionary Computation, Vol. 6, 2002, pp. 182–197. doi:10.1109/4235.996017.
[35] J. Gonçalves, M. Resende, Biased random-key genetic algorithms forcombinatorial optimization, Journal of Heuristics 17 (5) (2011) 487–525. doi:10.1007/s10732-010-9143-1.
[36] D. Conforti, F. Guerriero, R. Guido, A multi-objective block scheduling model for the management of surgical operating rooms: New solution approaches via genetic algorithms, in: 2010 IEEE Workshop on Health Care Management (WHCM), 2010, pp. 1–5. doi:10.1109/WHCM.2010.5441264.
[37] J. J. Nutaro, Discrete-event systems, in: Building Software for Simulation: Theory and Algorithms, with Applications in C++, 1st Edition, Wiley, 2010, Ch. 4, pp. 100–182.
[38] J. Knowles, L. Thiele, E. Zitzler, A Tutorial on the Performance Assessment of Stochastic Multiobjective Optimizers, 214, Computer Engineering and Networks Laboratory (TIK), ETH Zurich, Switzerland (2006).
[39] E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca, V. G. da Fonseca, Performance assessment of multiobjective optimizers: an analysis and review, Evolutionary Computation, IEEE Transactions on 7 (2) (2003) 117–132. doi:10.1109/TEVC.2003.810758.
[40] A. P. Guerreiro, Efficient algorithms for the assessment of stochastic multiobjective optimizers, Master's thesis, IST, Technical University of Lisbon (2011).