European Journal of Operational Research 271 (2018) 420–435

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Discrete Optimization

Robust allocation of operating rooms: A cutting plane approach to handle lognormal case durations

Guillaume Sagnol[a,b,*], Christoph Barner[c], Ralf Borndörfer[b], Mickaël Grima[d], Matthes Seeling[c], Claudia Spies[c], Klaus Wernecke[c]

[a] Institut für Mathematik, Sekr., MA 5-2, TU Berlin, Straße des 17. Juni 136, 10623 Berlin, Germany
[b] Zuse Institute Berlin, Takustr. 7, 14195 Berlin, Germany
[c] Charité-Universitätsmedizin Berlin, Charité pl., Berlin, 1, 10117, Germany
[d] Fakultät für Mathematik, Boltzmannstr., Garching bei München, 3, 85748, Germany

ARTICLE INFO

Article history:
Received 9 May 2016
Accepted 9 May 2018
Available online 18 May 2018

MSC:
90-B36
90-C47
90-B50

Keywords:
Parallel machines scheduling
Extensible bin packing
OR in medicine
Robust optimization
Lognormal durations
Hilbert's projective metric

ABSTRACT

The problem of allocating operating rooms (OR) to surgical cases is a challenging task, involving both combinatorial aspects and uncertainty handling. We formulate this problem as a parallel machines scheduling problem, in which job durations follow a lognormal distribution, and a fixed assignment of jobs to machines must be computed. We propose a cutting-plane approach to solve the robust counterpart of this optimization problem. To this end, we develop an algorithm based on fixed-point iterations that identifies worst-case scenarios and generates cut inequalities. The main result of this article uses Hilbert's projective geometry to prove the convergence of this procedure under mild conditions. We also propose two exact solution methods for a similar problem, but with a polyhedral uncertainty set, for which only approximation approaches were known. Our model can be extended to balance the load over several planning periods in a rolling horizon. We present extensive numerical experiments for instances based on real data from a major hospital in Berlin. In particular, we find that: (i) our approach performs well compared to a previous model that ignored the distribution of case durations; (ii) compared to an alternative stochastic programming approach, robust optimization yields solutions that are more robust against uncertainty, at a small price in terms of average cost; (iii) the *longest expected processing time first* (LEPT) heuristic performs well and efficiently protects against extreme scenarios, but only if a good prediction model for the durations is available. Finally, we draw a number of managerial implications from these observations.

1. Introduction

The operating theater (OT) is one of the most expensive hospital resources. Recent studies indicate that in certain hospitals, surgical interventions concentrate up to 70% of all patient admissions, and as much as 40% of the total expenses (Denton, Viapiano, & Vogl, 2007). The management of the OT is a very complex task, which involves several hierarchical decision levels and combinatorial aspects for many different types of resources (operating rooms, surgeons, nurses, anesthesiologists, etc.), all this in an uncertain environment (surgical durations, emergency cases, availability of recovery beds). For this reason, there has been a considerable effort to develop optimization procedures to improve the management of resources in the operating theater; we refer the reader to Guerriero and Guido (2011) for a comprehensive review of the operations research literature on OT management.

This paper focuses on the problem of allocating operating rooms to elective patients, typically on the day prior to operation. More precisely, the goal is to assign operating rooms (OR) to a list of *patient blocks*, that is, groups of elective patients to be operated one after another by the same surgical team. This is a crucial planning step for the so-called *block-scheduling* system (cf. Guerriero & Guido, 2011), in which individual surgeons or surgery specialties have predefined slots of OR-time allocated in a periodic schedule (the *Master surgery schedule*, MSS), and cases must be booked within these slots.

The problem of managing the operating rooms is characterized by a very strong stochasticity; see e.g. Tancrez, Roland, Cordier, and Riane (2009). In particular, it is well known that durations exhibit a close fit with the lognormal distribution, see e.g. Stepaniak, Heij,

* Corresponding author at: Institut für Mathematik, Sekr., MA 5-2, TU Berlin, Straße des 17. Juni 136, 10623 Berlin, Germany.
E-mail addresses: sagnol@math.tu-berlin.de, sagnol@zib.de (G. Sagnol).

https://doi.org/10.1016/j.ejor.2018.05.022
0377-2217/

Mannaerts, de Quelerij, and de Vries. (2009), Kayiş, Khaniyev, Suermondt, and Sylvester (2014) and the references therein. This uncertainty frequently leads to operations that exceed the planned OR-time in one slot of the MSS. The excess of OR-time is known as *overtime*, and induces high costs for the hospital. The need to take into account the uncertainty, and to exploit distributional information on durations is thus extremely important, all the more so as lognormal variables are heavy-tailed. This suggests the use of robust optimization techniques, which aim at protecting against extreme scenarios. The problem we study is a variation of a robust optimization problem introduced in Denton, Miller, Balasubramanian, and Huschka (2010). The only difference is in the model: we specifically take into account that durations are lognormal, which allows us to define a natural uncertainty set in terms of likely scenarios.

The efficiency of the OT can be measured by a combination of the number of under-utilized hours and the number of over-utilized hours in the operating rooms (Dexter & Traub, 2002). However, a few days before the day of surgery, the staff has already been scheduled, and so (Dexter & Traub, 2002) claims that under-utilized time does not cause a loss of revenue for the surgical suite. This is consistent with (McIntosh, Dexter, & Epstein, 2006), where it is shown that on a short-term perspective, the goal is solely to minimize the overtime in the OR. Also, we consider a fixed cost for *opening an OR*, as proposed in the model of Denton et al. (2010). However, under-utilization of the OR can still have indirect costs on a rather short-term perspective (even with a fixed staffing). This happens, e.g., when the load of surgeries is not balanced uniformly among several operating days. In this case, it might be well-suited to postpone some patients to a later day. Our model can easily be adapted to the situation in which the decision maker may cancel some jobs, or on the contrary when he may accept more jobs than initially planned, by using a rolling horizon with deferral costs. This situation naturally occurs in settings where the OR allocation problem is to be solved for a sequence of several planning periods; in this case, the ability to postpone (or bring forward) a job to a later (or earlier) planning period can help to balance the overtime over the whole planning horizon.

Since the end of the 90's, many papers have demonstrated the benefits of robust optimization (RO) to handle uncertainty. In many cases, the method offers tractable mathematical formulations which are much easier to solve than their stochastic programming counterparts (Ben-Tal & Nemirovski, 1998; Bertsimas & Sim, 2004). Moreover, RO offers the possibility to tune the *budget of uncertainty* to choose the tradeoff between performance and robustness.

Another traditional selling point for these approaches is that no distributional information for the uncertain parameters is required. In the context of surgery scheduling however, we already mentioned that lognormality of the durations can be assumed, and we want to take advantage of this. We point out that most statistical studies on prediction models for the distribution of surgery durations, such as Kayiş et al. (2014), Stepaniak, Heij, and De Vries. (2010), focus on the procedure time only. In our approach however, the relevant duration is the total duration of patient blocks, which consists of the sum of the procedure times, set-up times, and clean-up times of all patients in this block. In practice, most patient blocks contain between one and three surgical cases, and we think that the lognormal model is still a good model for the whole block duration. Indeed, it has been proposed to approximate the sum of lognormal distributions by the lognormal variable that matches its first two moments, which is the well-known Fenton–Willkinson approximation (Fenton, 1960). It is used routinely in financial engineering and other fields, such as signal processing, and provides a reasonable approximation for a range of lognormal parameters (Cobb, Rumi, & Salmerón, 2012). Another possibility would be to match three moments of the 3-parameters lognormal (which has an additional shift parameter), which has also been proposed in Stepaniak et al. (2009) to model surgery durations; it would be straightforward to extend our robustness model to the case of shifted lognormal distributions.

The fat-tail behaviour of the lognormal makes it likely that standard uncertainty models, such as the ellipsoidal uncertainty model of Ben-Tal and Nemirovski (1998) or the cardinality-constrained uncertainty model of Bertsimas and Sim (2004), will offer a rather poor model of the real-life setting. To remedy this problem, we propose to use robust optimization with an uncertainty model which protects against all scenarios in a confidence region of the lognormal distribution. This turns robust optimization into a risk assessment technique (cf. Proposition 3.5), as other approaches relying on the conditional value-at-risk (CVaR), cf. (Rockafellar & Uryasev, 2000), or the ordered weighted average (OWA); see Kasperski, Kurpisz, and Zieliński (2012).

To the best of our knowledge, one of the first papers to consider the problem of allocating operating rooms to a list of surgical procedures is Ozkarahan (1995), who proposed a mixed integer programming (MIP) formulation to minimize the under- and overutilization of the ORs. This paper makes the assumption that all the procedures of a given practitioner are performed in the same OR. This is also the approach that we adopt here (patients to be operated by the same surgical team are grouped in a block), for two main reasons:

1. When a surgeon performs two procedures in two different ORs, there is a risk that the first procedure takes longer than expected, which induces waiting time in both ORs and generates overtime. In contrast, it is known that planning all the procedures of one surgeon in a single OR is a guarantee of stability, (cf. Dexter, Traub, & Lebowitz, 2001), a feature desired by many OR planners, in particular at the Charité hospital in Berlin.
2. Mathematical formulations allowing a practitioner to change the OR within a day are much harder to solve, because we need to take synchronization issues into account. The resulting problem is a stochastic RCPSP (Resource constrained project scheduling problem). While several MIP formulations are available for the deterministic RCPSP (Koné, Artigues, Lopez, & Mongeau, 2011), in the stochastic setting precedence models must be used. These models suffer from relying on *big-M* constraints to avoid that two procedures performed by the same surgeon take place at the same time, which leads to weak relaxations and very long computing times.

We are aware that in some cases, in particular when the number of surgeons is a bottleneck for the planning, it might be better to let practitioners alternate between two rooms, so they can perform a surgery in room *B* while room *A* is being cleaned-up and prepared for the next patient. This is the approach used for example in Pham and Klinkert (2008); Pulido, Aguirre, Ortega-Mier, García-Sánchez, and Méndez (2014), where MIPs are proposed to solve a deterministic resource constrained scheduling problem. There are also stochastic programming approaches for the problem with room-changing surgeons: (Mancilla & Storer, 2013) scheduled one surgeon operating in two ORs, and Batun, Denton, Huschka, and Schaefer (2011) used the L-shaped method to solve a stochastic MIP model (with big-M's), for the case where surgeons may change room but have a predefined sequence of patients to operate in a given order.

The popularity of robust optimization techniques can also be observed in the literature on OT management. A non-exhaustive list of recent contributions using robust optimization follows: in Hans, Wullink, Houdenhoven, and Kazemier (2008), RO is used to allocate slack times in each OR to reduce the risk of overtime; an RO model is proposed in Addis, Carello, and Tánfani (2014) to allocate patients to OR-blocks (in a block-scheduling