ISSN 2281-4299

Dipartimento di Ingegneria informatica
automatica e gestionale Antonio Ruberti

Sapienza
Università di Roma

A derivative–free approach for a simulation–based optimization problem in healthcare

Stefano Lucidi
Massimo Maurici
Luca Paulon
Francesco Rinaldi
Massimo Roma

Technical Report n. 15, 2014

A derivative–free approach for a simulation–based optimization problem in healthcare

Stefano Lucidi[1], Massimo Maurici[2], Luca Paulon[2],
Francesco Rinaldi[3], Massimo Roma[1],

[1] Dipartimento di Ingegneria Informatica,
Automatica e Gestionale "A. Ruberti"
SAPIENZA, Università di Roma
via Ariosto, 25 – 00185 Roma, Italy
E-mail: lucidi@dis.uniroma1.it, roma@dis.uniroma1.it

[2] Dipartimento di Biomedicina e Prevenzione
Laboratorio di Simulazione e Ottimizzazione dei servizi del SSN
Università di Roma "Tor Vergata"
via Montpellier, 1 – 00133 Roma, Italy
E-mail: maurici@med.uniroma2.it, E-mail: luca.paulon@gmail.com

[3] Dipartimento di Matematica
Università di Padova
Via Trieste, 63 – 35121 Padova, Italy
E-mail: rinaldi@math.unipd.it

Abstract

In this work a simulation–based optimization model is considered in the framework of the management of hospital services. Given specific parameters which describe the hospital setting, the simulation model aims at reproducing the hospital processes and evaluating their efficiency. The use of a simulation–based optimization approach is necessary since the model can not be expressed as closed–form function. In order to obtain the optimal setting, we combine a derivative–free optimization method with a discrete event simulation model. The resulting framework has been tested on a real healthcare problem. More specifically, we study how to optimize the performance of an obstetric ward of a big Italian hospital, from both an economical and clinical point of view, taking into account some relevant constraints. The resulting optimization problem is a Mixed Integer Nonlinear Programming problem due to the presence of some variables constrained to be integer.

Keywords: Healthcare problems, Simulation–based optimization, Derivative–free methods.

1 Introduction

The current economic climate, despite being challenging, presents an opportunity for improvements. Particularly, more innovative approaches to health care should be developed by healthcare professionals [1, 2, 3] and by those responsible for health management [4]. Substantial attention should be placed on how to solve the cost crisis in health care [5] by considering what is value in health care [6, 7] and different stakeholders standpoints [8].

Health spending is rising. For example, in Italy, in 2011 total health spending accounted for 9.2% of the Gross Domestic Product (GDP), slightly below the average (9.3%) of the Organization for Economic Co-operation and Development (OECD) countries and significantly lower than in the United States which spent 17.7% of its GDP [9].

In recent years hospital financing has changed from a budget oriented (lump sum) system to a fee–for–service system in many National Health Services (NHS). The reimbursement for each treatment is fixed, meaning that hospitals which provide the treatment for lower costs can realize greater profits. As a consequence, hospitals should evaluate the optimal *case mix*, i.e., the optimal number of patients to treat over a period of interest, and evaluate which services could be expanded and which could be contained [10, 11].

Operations Research and related disciplines provide the necessary tools to determine, among other computation, the optimal case mix. Indeed they enable health care managers to investigate complex relationships among different parts of a hospital and to make rational administrative, economic, and medical decisions [12]. In particular, Discrete Event Simulation (DES) has been widely used over the last decades for modelling healtcare systems with a particular focus on the performance modelling of hospitals (see [13, 14, 15] for a review of the literature on simulation modelling in healtcare). The growing interest of DES models within healtcare systems is clearly witnessed by the large and increasing number of papers published in the more recent years dealing with this topic, focusing on specific healtcare applications. This is due to the fact that, in healtcare systems, like in several real–life problems, the framework of interest is represented by a stochastic model whose output is a random vector which can be sampled by computer simulation. This enables to perform a "what–if" analysis, i.e., a number of input combinations (scenarios) are simulated and the responses obtained are observed. However, often the number of combinations examined is very small due to high computational costs and anyhow, in most cases, of course, the set of possible decisions is too large to be enumerated. Nevertheless, usually in solving practical problems, the best possible combination is sought. In the most recent years the latter demand led to a merging between simulation modeling and optimization techniques and this enabled to search for the optimal input for a stochastic (discrete event) simulation model, that is to determine the best scenario according to some performance criteria. This approach is commonly known as *simulation optimization*: the optimization procedure chooses the values for the decision variables (input parameters of the simulation model), then the simulation is performed with these parameters and an estimate of the performance of the system is obtained. Then the optimization procedure uses these responses from the simulation for choosing new values of the decision variables and the loop is carried on until a stopping criterion is satisfied.

In order to realize the simulation optimization approach, many simulation softwares often include a tool which implements an optimization procedure (see, e.g. [16] for a survey). The latter procedure views the simulation model as a "black–box" used in order to evaluate the system performance. Unfortunately most of these optimization procedure available in commercial packages are heuristics. For instance, `OptQuest` developed by OptTek Systems, Inc., Boulder, Colorado, USA (see `http://www.opttek.com/OptQuest`), which is one of the most popular optimizer embedded in several simulation software, is based on the combined use of scatter search, tabu search and neural networks (see, e.g. [17]).

This points out a significant gap existing between the current research in Optimization and the optimization procedure usually used within the simulation optimization approach. In view of this, in this paper we consider a specific healtcare application where the service delivery model can not be expressed as closed–form function, thus requiring the use of simulation–based modelling approach and where the optimal configuration is sought. Instead of using the optimization procedure available within commercial simulation softwares, we propose the use of *Derivative–Free Optimization* (DFO) techniques recently proposed in Optimization literature (see [18] for a survey on DFO). The use of DFO methods is, of course, essential since all the information concerning the optimization problem are obtained via simulation as a black-box problem. More precisely, a discrete event simulation model representing the services delivery of the obstetric ward of a big Italian hospital has been constructed by using Arena simulation software [19]. Then the simulation model was connected by means of a suited interface to an external implementation of a DFO algorithm so that the simulation–optimization procedure can be executed. This enabled to use an optimizer different from OptQuest for Arena [20], which is the standard optimization engine embedded in Arena. Note that the optimization problem in hand includes some variables constrained to be integer, so that it is a *Mixed Integer Nonlinear Programming* (MINLP) problem.

The results obtained by using a DFO algorithm was compared with those obtained by using OptQuest, showing that our proposal enables to obtain a "better" solution requiring a much lower computational burden.

The paper is organized as follows: in Section 2 a description of the case study is reported along with a brief illustration of the discrete event simulation model. The formal statement of the optimization problem is given in Section 3, while Section 4 summarizes the DFO algorithm used. Finally, Section 5 and 6 report the results obtained, some comments and conclusions.

2 The case study

Case mix optimization is particularly relevant for hospitals which provide services for pregnant women and newborns health care. Particularly, in an efficient obstetrics ward the overall profit should be maximized while, most important, the normal rate of Cesarean sections should be lower than 10–15% of overall childbirths (due to higher risk to compromise mother or child's health in respect to the natural childbirth [21]). This is a relevant problem in many countries. Indeed, although levels of 10–15% were considered high but acceptable at the time, average Caesarean rates in most developed regions exceed 20% [22] probably due to health system factors such as the human resources and financing profile whose optimization are in contrast with respect to a low Caesarian section rate ([23]).

Thus, in this respect, we show how to perform a case mix optimization of an obstetrics ward, that is, practically, how to evaluate the number of pregnant women that would be more efficient to treat per year in respect to the actual, from both an economic and a clinical point of view, taking into account different contrasting goals and some relevant constraints. The optimized case mix would increase profits (an economic goal) while keeping the Caesarean section rate (a clinical goal) under a threshold value.

The case study was carried out in the obstetrics ward of the Fatebenefratelli San Giovanni Calibita (FBF-SGC) Hospital in Rome, one of the most important hospital of the Italian NHS ([24]) in terms of number of childbirth cases both at the regional and the national level. The results of the study were critical for the hospital, both in qualitative and quantitative terms. A research group, designated the Business Simulation for Health Care (BuS-4H) team and composed of doctors, engineers, statisticians and other experts in health care was formed. The services under evaluation were classified as DRG 371 and DRG 373 using version 24 of the Diagnosis–Related Groups (DRG) classification system [25]; they were, respectively, Cesarean section without complications or comorbidities and vaginal childbirth without complications. The top managers of the FBF-SGC Hospital, respectively, the Chief Executive Officer (CEO) and the Chief Hospital Officer (CHO), were involved in the evaluation. It was necessary to support them in defining and evaluating the actual case mix and the most important key performance indicators (KPIs) and goals related to the services under evaluation, taking into account both national and regional government recommendations regarding the most relevant indicators [26].

2.1 The conceptual model

From a system point of view, the main FBF-SGC Hospital data flow related to hospitalizations, such as the hospital discharge forms and the hospital childbirth records, as well as some other services–related data (such as costs and incomes) were imported and integrated into a single database made expressly for this study. To build useful reports, including statistical analysis, queries were then defined and used to create a sort of dashboard for the ward under evaluation. Such reports contained, among other items, the following: Caesarean section rate, profit (difference between income and costs), rate of low–length–of–stay admissions (considering the number of hospitalizations with length of stay between 0 and 2 days), bed occupation rate, overall number of supplied births, number of Caesarean sections, patient arrival time distributions, etc.

From a process point of view, pregnant women flow through the Emergency Room, i.e. phase 1 of the service delivery, and, if hospitalized, they arrive to the ward, i.e., phase 2. During each phase specific monitoring, visits, preparation and interventions occur. In particular, at the beginning of the health care nurses execute the triage after a rapid registration and verification, i.e. all patients are prioritized. In case of pregnant woman a specialistic triage is performed by obstetricians which eventually perform a fetal monitoring. Then, gynaecologists visit the patient and confirm the assigned priority and the necessary intervention, including the hospitalization or not. Instead, pregnant woman which require a scheduled Caesarian section flow directly to the ward, without performing any triage (see Figure 2.1). The hospitalization in the ward lasts about 1 day before the childbirth, and 1 day or 2 days after the childbirth for a natural childbirth or a Caesarian section, respectively.

2.2 The simulation model

The simulation model was implemented by using Arena 14 simulation software by Rockwell Software [19], one of the most popular DES software. Arena is a general–purpose simulation environment which enables the full visualization of the simulation model structure by means of flowchart modules. It allows the user to easy control the simulation parameters, providing useful input and output analysis tools, too (see e.g., [27]).

The services delivery conceptual model of the FBF-SGC Hospital obstetric ward has been implemented. The availability of an integrated database mentioned in the previous section allows us to perform an accurate *input analysis*, mainly regarding probability distribution fitting the processes operational time. The *verification* and the *validation* of the model enabled to ensure that the model is performing properly and that it is an accurate representation of the real system. An appropriate *experimental design* allows us to determine the length of the simulation run (365 days), the number of replications (10) and the warm–up period (42 days).

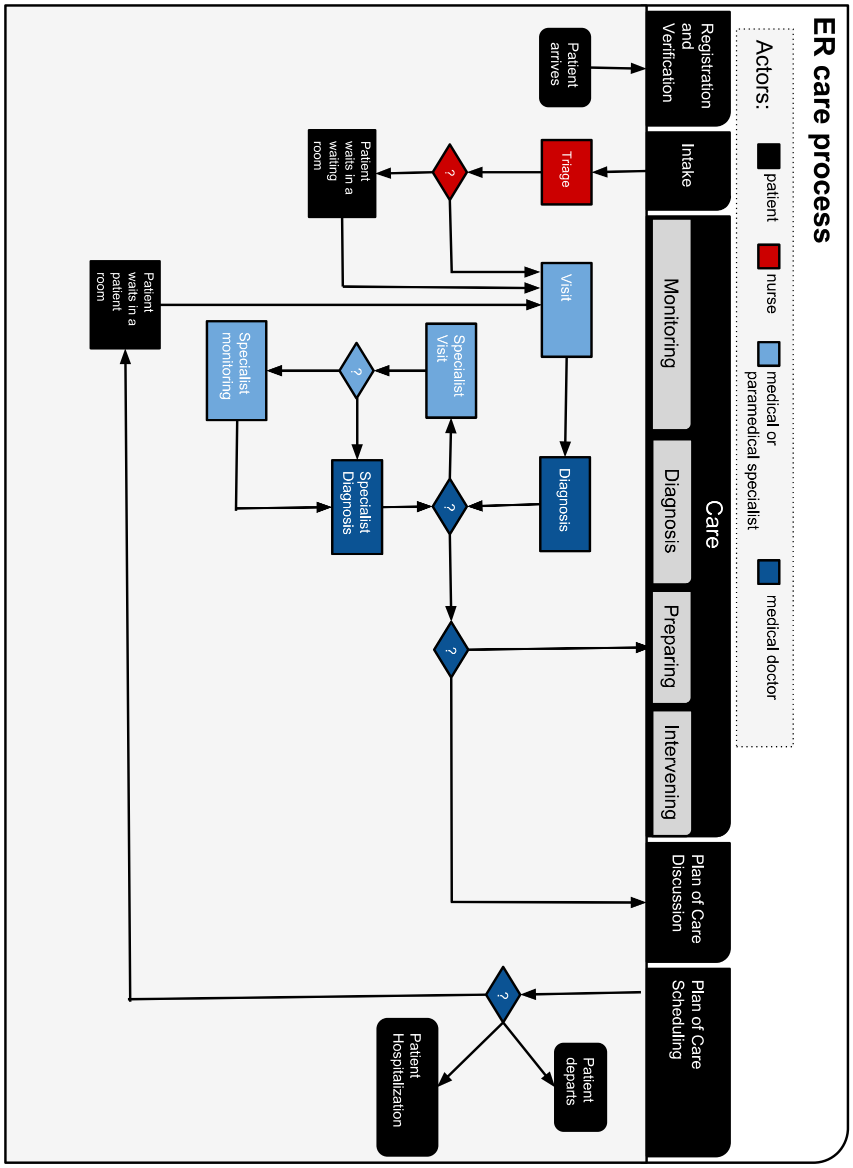

Figure 2.1: Services delivery conceptual model: pregnant woman health care in the FBF-SGC hospital is implemented by the ER (whose main activities and resources is shown in this figure) and the obstetric ward.

3 Statement of the simulation optimization problem

Starting from the case mix statement, the KPIs and goals chosen by the top managers of the ward under evaluation, the problem of finding a better case mix with respect to the actual one can be mathematically stated in the following form

$$
\begin{aligned}
&\max f(z,t,y(z,t)) \\
&g_1(z,t,y(z,t)) \leq 0 \\
&\qquad\vdots \\
&g_m(z,t,y(z,t)) \leq 0 \\
&0 \leq l_z \leq z \leq u_z \\
&0 \leq l_t \leq t \leq u_t
\end{aligned}
$$

where $z \in \mathbb{Z}^p$ and $t \in \mathbb{R}^q$ are the vectors of the services delivery decision variables and $y \in \mathbb{R}^r$, with $y_j : \mathbb{Z}^p \times \mathbb{R}^q \longrightarrow \mathbb{R}$, $j = 1, \ldots, r$, represents an estimate of the expected values of the output of the service delivery discrete–event simulation model which depends on z and t. The functions f and g_i, $i = 1, \ldots, m$ are real valued functions, $f, g_i : \mathbb{Z}^p \times \mathbb{R}^q \times \mathbb{R}^r \longrightarrow \mathbb{R}$ and $l_z, u_z \in \mathbb{Z}^p$, $l_t, u_t \in \mathbb{R}^q$. A simulation–based modeling approach is used since the service delivery model cannot be expressed as closed–form function of z and t. More precisely, z and t correspond to the *resources* of the simulation model which can be controlled by the user. The vector $y = y(z,t)$ represents the KPIs of interest obtained as output of the simulation model. In practice, the values $y_j = y_j(z,t)$, $j = 1, \ldots, r$, are obtained as an average over a certain number of independent replications of the simulation. The resulting problem is a mixed integer nonlinearly constrained problem with box constraints on the variables z and t.

In particular, in the case study the simulation model represents the services delivery of the obstetrics ward of the FBF-SGC Hospital with $p = 7$ counters z_i, $i = 1, \ldots, 7$ of allocated resources under control, $q = 1$ service demand indicator t_1 under control (in hours), and $r = 6$ case mix responses $y_j = y_j(z,t)$, $j = 1, \ldots, 6$ of the simulation model. More in detail, the component of the vectors of the decision variables are the following: z_1 is the number of stretchers, z_2 is the number of gynecologists, z_3 is number of gynecologists who discharge a patient from the hospital, z_4 is number of nurses, z_5 is number of midwives, z_6 is the number of Hospital beds, z_7 is the number of operating rooms, t_1 is the mean value of the patient interarrival times distribution. Note that, t_1 actually is not a resource. However, its value can be controlled by the hospital management due to the possibility, in some cases, to reduce or rise admissions of patients by adopting appropriate strategies.

The components of the output vector of the simulation model are the following: y_1 is number of Caesarean sections per year, y_2 is the number of vaginal childbirths per year, y_3 is the number of "extra" Caesarean sections per year, y_4 is the number of "extra" vaginal childbirths per year, y_5 is the number of hospitalized woman having as a result no childbirth, y_6 is the number of transferred woman before delivery, where "extra" means that, mainly due to the lack of resources in the ward (e.g. all stretchers or beds are busy), both woman and newborn are not hospitalized in the FBF-SGC hospital but after the delivery in the emergency room, they are transferred to another hospital.

3.1 The constraints

We derive all the constraints mainly from the actual conditions of the specific ward of the FBF-SGC Hospital under evaluation.

In particular, regarding the decision variables and by considering their actual values, the box constraints for z_i , $i = 1\ldots,7$ are mainly due to budget and logistic limits, while for t_1 it is due to specific clinical and managerial limits on patients admission. They are the following:

$8 \le z_1 \le 15$	stretchers,
$2 \le z_2 \le 7$	gynecologists,
$1 \le z_3 \le 3$	gynecologists who discharge a patient,
$1 \le z_4 \le 5$	nurses,
$2 \le z_5 \le 9$	midwives,
$33 \le z_6 \le 45$	beds,
$1 \le z_7 \le 3$	operating rooms,
$2.0 \le t_1 \le 4.0$	mean value of patients interarrival times distribution (in hours),

with $z_6 = 3\ell$, $\ell \in \mathbb{Z}$ number of rooms in the ward. Note that the actual value of t_1 is an estimation of the patients interarrival times obtained from a statistical analysis of the hospital database.

The general constraints are the following:

- a lower bound of the number of Caesarean sections per year, in order to avoid that it decreases too much under the actual condition:

$$y_1(z,t) \ge 500$$

- a lower bound of the overall number of childbirths per year, in order to avoid that it decreases too much under the actual condition:

$$y_1(z,t) + y_2(z,t) \ge 3500$$

- a lower bound on the overall patient occupation rate (defined as the ratio between the effective overall length of the patients stay and the theoretical stay available) in order to avoid the underutilization of the ward:

$$\frac{3.3\,((y_2(z,t) - y_4(z,t)) + 5.0\,(y_1(z,t) - y_3(z,t)) + 5.0 y_5(z,t)}{365(z_1 + z_6)} \ge 0.75$$

- an upper bound on the number of transferred woman before delivery in order to minimize clinical risks:

$$y_6(z,t) \le 0.25\,(y_1(z,t) + y_2(z,t))$$

- an upper bound on the rate of Cesarean sections which is a better (lower) value, i.e. 25% as discussed in Section 2, with respect to the actual value (44%):

$$\frac{y_1(z,t) - y_3(z,t)}{y_1(z,t) - y_3(z,t) + y_2(z,t) - y_4(z,t)} \le 0.25 \qquad (3.1)$$

3.2 The objective function

The optimal case mix which maximizes the overall net profit objective (difference between income and costs) is sought. In particular, the aim is to obtain the values of the decision variables which determine such a case mix improvement, comparing them with the actual conditions

$$(z^0, t^0) = (z_1^0,\ z_2^0,\ z_3^0,\ z_4^0,\ z_5^0,\ z_6^0,\ z_7^0,\ t_1^0) = (10,\ 5,\ 1,\ 1,\ 6,\ 42,\ 1,\ 2.4). \tag{3.2}$$

The resulting objective function can be stated as follows:

$$\begin{aligned} f(z,t,y) \;=\;& 382\,(y_1(z,t) - y_3(z,t)) + 309\,(y_2(z,t) - y_4(z,t)) - 4500\max\{0, z_1 - z_1^0\} \\ & -10352\max\{0, z_2 - z_2^0\} - 10352\max\{0, z_3 - z_3^0\} - 9589\max\{0, z_4 - z_4^0\} \\ & -9589\max\{0, z_5 - z_5^0\} - 5000\max\{0, z_6 - z_6^0\} - 50000\max\{0, z_7 - z_7^0\} \\ & -2737 z_1 - 14600 z_6 \end{aligned}$$

where the first two terms correspond to the profit (in euros) due to Caesarean sections and vaginal childbirths, the terms of the form $c_i \max\{0, z_i - z_i^0\}$ correspond to set up costs and the last two terms correspond to some additional costs for stretchers and beds utilization. Its simulated actual value is $400,876.00$ euros.

4 The optimization algorithm

In this section, we describe the algorithmic framework used to deal with the problem described in the previous section. First, we note that, since the calculation of the objective function and constraint functions is obtained via numerical simulations, and some of the variables in the model are constrained to be integer, the given problem is basically a *Black–Box* Mixed Integer Nonlinear Programming problem which we re–write in the following form

$$\begin{aligned} \min\ & f(x) \\ & g_1(x) \le 0 \\ & \vdots \\ & g_m(x) \le 0 \\ & l \le x \le u \\ & x_i \in \mathbb{Z}, \quad i \in I_z, \end{aligned} \tag{4.1}$$

where $x, l, u \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$ and $g_j : \mathbb{R}^n \to \mathbb{R}$, $j = 1, \ldots, m$ and $I_z \subset \{1, \ldots, n\}$ is the set of the indices of integer variables. Moreover, we denote by I_c the set of the index of continuous variables, $I_c = \{1, \ldots, n\} \setminus I_z$. The objective function f and the general nonlinear constraints function $g = (g_1, \ldots, g_m)^T$, are assumed to be continuously differentiable with respect to x_i, $i \in I_c$, even thought the derivatives actually are not used.

In solving this problem we face with a twofold difficulty. On one hand, the objective and constraint functions are of the black-box type, so that first order derivatives are not available (see [18] for a recent survey on Derivative–Free methods). On the other hand, the presence of discrete variables requires an ad-hoc treatment.

In literature, various Derivative–Free methods have been proposed for solving MINLP problems (see e.g. [28, 29, 30, 31, 32, 33, 34, 35, 36]). In particular, we adopted the Derivative–Free Linesearch (DFL) algorithm for MINLP problems described in [37].

In order to give a detailed description of the procedure, we need to report some definitions and theoretical results from [37]. We start by the definition of the following sets:

$$\mathcal{X} := \{x \in \mathbb{R}^n : l \le x \le u\}, \quad \mathcal{F} = \{x \in \mathbb{R}^n : g(x) \le 0\} \cap \mathcal{X}, \quad \mathcal{Z} := \{x \in \mathbb{R}^n : x_i \in \mathbb{Z}, i \in I_z\}.$$

Moreover, for any vector $v \in \mathbb{R}^n$, $v_c \in \mathbb{R}^{|I_c|}$ and $v_z \in \mathbb{R}^{|I_z|}$ denote the subvectors

$$v_c = [v_i]_{i \in I_c}, \qquad v_z = [v_i]_{i \in I_z}.$$

Futhermore, since the characterization of local minimizers in mixed problems strongly depends on the particular neighborhood used, we need to report different definitions of neighborhoods that correspond to variations of continuous and discrete variables. Hence, for any point $\bar{x} \in \mathbb{R}^n$ and $\rho > 0$, the following definitions are given:

$$\begin{aligned} \mathcal{B}_c(\bar{x}, \rho) &= \{x \in \mathbb{R}^n : x_z = \bar{x}_z,\ \|x_c - \bar{x}_c\|_2 \le \rho\}, \\ \mathcal{B}_z(\bar{x}) &= \{x \in \mathcal{Z} : x_c = \bar{x}_c,\ \|x_z - \bar{x}_z\|_2 = 1\}. \end{aligned}$$

Now we are ready to report the definition of local minimizer:

Definition 4.1 *A point $x^\star \in \mathcal{F} \cap \mathcal{Z}$ is a* local minimizer *of Problem (4.1) if, for some $\epsilon > 0$,*

$$\begin{aligned} f(x^\star) \le f(x), &\quad \text{for all} \quad x \in \mathcal{B}_c(x^\star; \epsilon) \cap \mathcal{F}, \\ f(x^\star) \le f(x), &\quad \text{for all} \quad x \in \mathcal{B}_z(x^\star) \cap \mathcal{F}. \end{aligned}$$

Under standard assumptions (see [37] for a complete description of the assumptions adopted), it is possible to give *stationary conditions* for Problem (4.1). The latter conditions make use of the Lagrangian function associated to Problem (4.1), namely $L(x, \lambda) = f(x) + \sum_{i=1}^m \lambda_i g_i(x)$. The following proposition (see [37]), reports the *necessary optimality conditions* for Problem (4.1). Here the notation $\nabla_c L(x, \lambda)$ is used to denote the gradient of the function L with respect to the continuous variables.

Proposition 4.2 *Let $x^\star \in \mathcal{F} \cap \mathcal{Z}$ be a local minimizer of Problem (4.1). Then there exists a vector $\lambda^\star \in \mathbb{R}^m$ such that*

$$\nabla_c L(x^\star, \lambda^\star)^T (x - x^\star)_c \ge 0, \qquad \text{for all} \quad x \in \mathcal{X} \tag{4.2}$$

$$(\lambda^\star)^T g(x^\star) = 0 \qquad \lambda^\star \ge 0 \tag{4.3}$$

$$f(x^\star) \le f(x) \qquad \text{for all} \qquad x \in \mathcal{B}_z(x^\star) \cap \mathcal{F}. \tag{4.4}$$

Finally, we report the definition of *stationary point* for Problem (4.1).

Definition 4.3 *A point $x^\star \in \mathcal{F} \cap \mathcal{Z}$ is a stationary point of Problem (4.1) if a vector $\lambda^\star \in \mathbb{R}^m$ exists such that the pair $(x^\star, \lambda^\star)$ satisfies (4.2), (4.3) and (4.4).*

Now, in order to give a description of the Derivative–Free algorithm we use, we report the penalty function used to handle the general constraints. As in [38], the following sequential penalty function

$$P(x; \epsilon) := f(x) + \frac{1}{\epsilon} \sum_{i=1}^m \max\{0, g_i(x)\}^q, \quad \text{with} \quad q > 1$$

is used and the original problem is solved by means of a sequence of penalty problems of the form

$$\min P(x;\epsilon)$$
$$x \in \mathcal{X} \cap \mathcal{Z},$$

where penalization of constraint violation is progressively increased.

Now we report the basic scheme of the Derivative–Free framework for MINLP problems we use.

A Derivative–Free MINLP framework

Input: an initial point $x_0 \in \mathcal{X}$, a decrease parameter $\xi_0 > 0$, a penalty parameter $\epsilon_0 > 0$, a set of stepsizes $\alpha_0^i > 0,\ i = 1, \dots, n$ and a set of search directions $d_0^i = e^i,\ i = 1, \dots, n$.

Output: a *stationary point* of Problem (4.1).

Set $k = 0$.
repeat
  Set $y_k^1 = x_k$
  for $i = 1, 2, \dots, n$ **do**
    if i-th variable is continuous
    then compute an α continuous stepsize along the i-th search direction enforcing
      $(\alpha_k^i)^2$-sufficient decrease by **Continuous search(**$\alpha_k^i, y_k^i, d_k^i; \alpha$**)**
    else compute an α discrete stepsize along the i-th search direction enforcing
      ξ_k-sufficient decrease by **Discrete search(**$\alpha_k^i, y_k^i, d_k^i, \xi_k; \alpha$**)**
    end if
    Set new point $y_k^{i+1} = y_k^i + \alpha d_k^i$ and update α_{k+1}^i.
  end for
  Find $x_{k+1} \in X \cap \mathcal{Z}$ s.t. $P(x_{k+1}, \epsilon_k) \leq P(y_k^{n+1}, \epsilon_k)$.
  Use updating rule to obtain ϵ_{k+1} and ξ_{k+1}.
  Set $k = k + 1$.
until convergence

The described method, like many other Derivative–Free techniques, is based on a suitable sampling strategy along a set of directions. This strategy is able to get, in the limit, sufficient knowledge of the problem functions (by using the Continuous and Discrete Search) to recover both first order information for the continuous variables, and some sort of local optimality for the discrete ones. Anyway, since we are in a constrained context, we need to take also care of the penalty parameter (i.e. the penalty parameter has to be updated and, as we said before, progressively driven to zero), by somehow connecting it to the sampling technique. Roughly speaking, the penalty parameter must converge to zero more slowly than the maximum stepsize used by the sampling scheme. This is the reason why we need, other than the updating rules for the stepsizes α^i and for the updating of ξ (the parameter driving the sufficient decrease in the Discrete Search), a rule for the updating of the parameter ϵ.

Summarizing, the main features of the algorithm are four:

1. the Continuous Search, which performs a classic Derivative–Free linesearch (see e.g. [39]) guaranteeing a sufficient decrease of the objective function;
2. the Discrete Search, which performs a Derivative–Free linesearch in a "discrete fashion";
3. the updating rule for the stepsizes α^i;
4. the updating rule for the penalty parameter ϵ and the parameter ξ.

All these ingredients are needed to guarantee convergence of the algorithm to stationary points of the original problem.

Now, we give some details about those features. The updating rule for the stepsizes α^i is very simple as either it sets the stepsize to the α given by the related search in case a sufficient decrease is obtained, or it shrinks the stepsize in case of failure. The updating rule for the parameters ϵ and ξ works as follows: if no discrete variable has been updated and all the tentative steps along discrete coordinates are equal to one, the sufficient reduction parameter is decreased, and the procedure further checks if the penalty parameter has to be updated.

We also report the detailed schemes of Continuous and Discrete Search below.

Continuous search($\tilde{\alpha}, y, d; \alpha$).

Data. $\gamma > 0$, $\delta \in (0, 1)$.

Step 1. Compute the largest $\bar{\alpha}$ such that $y + \bar{\alpha} d \in X \cap \mathcal{Z}$. Set $\alpha = \min\{\bar{\alpha}, \tilde{\alpha}\}$.

Step 2. If $\alpha > 0$ and $P(y + \alpha d) \leq P(y) - \gamma \alpha^2$ **then** go to Step 6.

Step 3. Compute the largest $\bar{\alpha}$ such that $y - \bar{\alpha} d \in X \cap \mathcal{Z}$. Set $\alpha = \min\{\bar{\alpha}, \tilde{\alpha}\}$.

Step 4. If $\alpha > 0$ and $P(y - \alpha d) \leq P(y) - \gamma \alpha^2$ **then** set $d = -d$ and go to Step 6.

Step 5. Set $\alpha = 0$ and **return**.

Step 6. While $\left(\alpha < \bar{\alpha} \text{ and } P\left(y + \frac{\alpha}{\delta} d\right) \leq P(y) - \gamma \frac{\alpha^2}{\delta^2} \right)$

$\alpha \leftarrow \alpha / \delta$.

Step 7. Set $\alpha \leftarrow \min\{\bar{\alpha}, \alpha\}$ and **return**.

Discrete search($\tilde{\alpha}, y, d, \xi; \alpha$).

Step 1. Compute the largest $\bar{\alpha}$ such that $y + \bar{\alpha} d \in X \cap \mathcal{Z}$. Set $\alpha = \min\{\bar{\alpha}, \tilde{\alpha}\}$.

Step 2. If $\alpha > 0$ and $P(y + \alpha d) \leq P(y) - \xi$ **then** go to Step 6.

Step 3. Compute the largest $\bar{\alpha}$ such that $y - \bar{\alpha} d \in X \cap \mathcal{Z}$. Set $\alpha = \min\{\bar{\alpha}, \tilde{\alpha}\}$.

Step 4. If $\alpha > 0$ and $P(y - \alpha d) \leq P(y) - \xi$ **then** set $d = -d$ and go to Step 6.

Step 5. Set $\alpha = 0$ and **return**.

Step 6. While ($\alpha < \bar{\alpha}$ and $P(y + 2\alpha d) \leq P(y) - \xi$)

$\alpha \leftarrow 2\alpha$.

Step 7. Set $\alpha \leftarrow \min\{\bar{\alpha}, \alpha\}$ and **return**.

The Continuous search procedure is defined by specifying values for parameters γ and δ which are used, respectively, in the sufficient reduction criterion and for the expansion of the step. The main distinguishing feature of the Discrete search procedure with respect to the Continuous search consists in the sufficient decrease criterion which employs the decrease parameter ξ instead of the usual squared stepsize which, for a discrete variable, is bounded away from zero. Indeed, we say that the new trial point $(y \pm \alpha d)$ guarantees a sufficient decrease of the objective function value when its value is better than $P(y) - \xi$.

As regards the convergence properties of the algorithm, we now report the main theoretical result concerning the global convergence (see [37]).

Theorem 4.4 *Let $\{x_k\}$ and $\{\epsilon_k\}$ be the sequences generated by the algorithm. Let*

$$K_\xi = \{k : \xi_{k+1} < \xi_k\} \subseteq \{1, 2, \ldots\} \quad \text{and} \quad K_\epsilon = \{k : \xi_{k+1} < \xi_k, \epsilon_{k+1} < \epsilon_k\} \subseteq K_\xi$$

Then, the sequence $\{x_k\}$ admits limit points

(i) if $\lim_{k \to \infty} \epsilon_k = \bar{\epsilon}$, every limit point of $\{x_k\}_{k \in K_\xi}$ is stationary for Problem (4.1);

(ii) if $\lim_{k \to \infty} \epsilon_k = 0$, every limit point of $\{x_k\}_{k \in K_\epsilon}$ is stationary for Problem (4.1).

5 Results and discussion

In order to determine an optimal solution of the MINLP problem in hand described in Section 3, we used a Fortran 90 implementation of the DFO algorithm described in Section 4. Of course, it was necessary to create an interface between the fortran code and Arena simulation software. To this aim, we use the Visual Basic for Applications (VBA) tool included in Arena which enables to build custom user interfaces to Arena models and to transfer data to/from Arena.

The procedure used is the following: the DFO algorithm selects the values for the decision variables (z, t) which represents the input parameters of the simulation model. These values are transferred to Arena model and the simulation is run, for the prefixed number of independent replications, in order to obtaining an estimate of the system performance, namely the component of the output vector y. The DFO algorithm uses these responses from Arena to choose the next set of values for the decision variables. The loop is carried on until the stopping criterion is satisfied.

We also used `OptQuest` for solving the problem we are considering. For both the algorithms we used as starting point the one corresponding to the actual condition (z^0, t^0) reported in (3.2). It is important to notice that such point is infeasible for the problem in hand since the constraint (3.1), imposing an upper bound on the Cesarean sections rate, is not satisfied by the values of variables corresponding to the actual condition. As regards the tolerance used in the stopping criterion we use the same (10^{-6}) for both the algorithms. We monitored the computational burden by counting the *number of simulations* needed by an algorithm for satisfying the stopping criterion.

In Table 5.1 we report, for each algorithm, the optimal value of the decision variables, the optimal objective function value (in euros) and the number of simulations needed. For a comparison with the actual operating condition we also report the value of the variables corresponding to this actual condition along with the value of the objective function. In Table 5.2 we report the values of the responses obtained by the simulation model corresponding to the three configurations detailed in Table 5.1.

	z_1	z_2	z_3	z_4	z_5	z_6	z_7	t_1	f	number of simulations
Actual values (z^0, t^0)	10	5	1	1	6	42	1	2.4	400,876.00	—
`OptQuest`	14	5	2	1	5	42	1	1.738	548,672.00	1777
DFO Algorithm	15	5	1	1	6	39	1	1.822	565,368.00	215

Table 5.1: Resources and objective function values corresponding to the actual operating conditions (z^0, t^0) and to the optimal value obtained by the two algorithms along with the number of simulations needed.

	y_1	y_2	y_3	y_4	y_5	y_6
Simulated actual values y^0	883.40	2514.70	12.80	220.60	1080.00	551.70
`OptQuest`	944.60	3266.00	24.10	428.10	945.40	949.20
DFO Algorithm	909.30	3176.10	21.80	395.20	961.90	881.00

Table 5.2: Corresponding responses of the simulation model.

By observing Table 5.1 it can be clearly pointed out that the use of the DFO Algorithm allows us to obtain a better solution in terms of objective function value (the net profit) with respect to the one obtained by `OptQuest`. Moreover, DFO Algorithm clearly outperforms `OptQuest` in terms of computational effort required. Indeed `OptQuest` needs 1777 simulations against 215 simulations required by the DFO Algorithm. As regards the optimal solution determined by DFO Algorithm, by comparing the optimal values of the decision variables with respect to those corresponding to the actual condition, it can be observed that some of them remains unchanged (z_2, z_3, z_4, z_5, z_7). They concern human resources (number of gynecologist, nurses, midwives) and a structural resource (the number of operating rooms) whereas changes are expected in the number of stretchers and beds.

Moreover, the optimal values of t_1 corresponds to an increase of the average number of patients arriving in a day. In fact, the interarrival time between two subsequent patients passes from a value of 2.4 hours to 1.822 hours. As already discussed, the hospital management can control the value of this parameter by means of appropriate strategies. On the overall, it is important to note that the optimal solution obtained by the DFO Algorithm is easy to adopt in practice since it only requires few changes with respect to the actual condition and these changes do not regard human resources. On the opposite, `OptQuest` suggests to increase the number of gynecologists who discharge patients and to decrease the number of obstetricians.

As regards the values of the responses, Table 5.2 evidences that by adopting the solutions obtained (by both the DFO Algorithm or `OptQuest`) a significant increase of y_2 (number of natural childbirths) is expected with respect to the actual situation. But the most interesting point in both solutions is the increasing of y_3, y_4, y_6 (which is slightly lower for the solution obtained by DFO Algorithm, probably due to different human resource allocation). In any case the solution we obtained suggests to the top managers to improve the ER emergency activities related to childbirth in order to satisfy, among others, the constraint of a good rate of Caesarean sections (less than 25%), still improving the profit. This is a very interesting average condition (in terms of clinical risk and economical benefits both for patients, hospitals and for the NHS) between hospitalization (i.e. the common politics in Italy) and assisted childbirth at home which is the novel politics proposed by the Lazio Region of Italy.

6 Concluding remarks

In this work we propose the use of a Derivative–Free Optimization within the Simulation Optimization framework. With reference to a particular problem arising in the management of hospital systems, we showed that the new approach we propose is effective and outperforms the standard use of heuristic methods usually embedded within simulation software packages both in terms of quality of the solution provided and in terms of efficiency. On the overall, the results obtained on this case study indicate that the use of Derivative–Free Optimization algorithms within Simulation Optimization is very promising. Future works concerns the use of a multiobjective formulation of healthcare management problems, since very often contrasting goals arise in this context.

Acknowledgement The authors are grateful, among others, to SiliconDev s.r.l. (namely Dr. A. Di Gregorio, Dr. S. Valore, Dr. M. Cenerelli, Dr. S. Tomasiello, Dr. G. Iamele, Dr. E. Di Landa) and FILAS s.p.a. for granting the *BuS-4H* project, to the Department of Biomedecine and Prevention at the University of Rome "Tor Vergata" (namely Prof. L. Palombi, Prof. E. Franco, Dr. F. Lisena and Dr. N. Toschi) in supporting the startup of the "Laboratorio di Simulazione e Ottimizzazione dei servizi del SSN", to the FBF-SGC hospital and the FBF Foundation (namely Dr. C. Cellucci, Dr. M. Ferrante, Dr. M.G. Pellegrini, Dr. F. Lagana, Dr. D. Vassallo, Prof. P.M. Rossini, Dr. P. Pasqualetti, Dr. D. Milazzo, Dr. A. Campolongo, Dr. L. Lagana, Dr. A. Giordani, Mrs. E. Berrettoni) in supporting mainly the simulation modeling of the case study, to the Department of Basic and Applied Science at SAPIENZA, University of Rome (namely Prof. F. Pitolli, Prof. B. Vantaggi and Dr. V. Bruni) in supporting the formulation of the *BuS-4H* proposal.

References

[1] M. Maurici, L. Paulon, A. Campolongo, C. Meleleo, C. Carlino, A. Giordani, F. Perrelli, S. Sgricia, M. Ferrante, E. Franco, et al. Quality measurement and benchmarking of HPV vaccination services: A new approach. *Human vaccines & immunotherapeutics*, 10(1):35–42, 2013.

[2] M. Maurici, C. Meleleo, A. Campolongo, C. D'Anna, M.L. Mangia, S. Sgricia, L. Serino, L. Paulon, E. Franco, and M. Ferrante. Application of the QuaVaTAR model to vaccination services in Latium, Italy. *Igiene e sanità pubblica*, 66(6):793–801, 2009.

[3] L. Paulon. *Mathematics in Health Care with Applications*. Phd thesis, SAPIENZA, Università di Roma, 2013.

[4] G. Quaglio, T. Karapiperis, L. Van Woensel, E. Arnold, and D. McDaid. Austerity and health in Europe. *Health Policy*, 113(1-2):13–19, 2013.

[5] R. S. Kaplan and M. E. Porter. How to solve the cost crisis in health care. *Harvard Business Review*, 89(9):47–64, 2011.

[6] M. E. Porter. What is value in health care? *New England Journal of Medicine*, 363(26):2477–2481, 2010.

[7] J. Y. Kim, P. Farmer, and M. E. Porter. Redefining global health-care delivery. *Lancet*, 382(9897):1060–1069, 2013.

[8] R. M. Kaplan and Y. M. Babad. Balancing influence between actors in healthcare decision making. *Bmc Health Services Research*, 11:14, 2011.

[9] OECD. OECD Health Statistics 2013. How does Italy compare. Report, OECD, 2013.

[10] K. Brink. Case–mix–optimization in a German hospital network. In *Annual International Conference of the German-Operations-Research-Society (GOR)*, Operations Research Proceedings, pages 85–90, 2011.

[11] P.T. Vanberkel, R.J. Boucherie, E.W. Hans, and J.L. Hurink. Optimizing the strategic patient mix combining queueing theory and dynamic programming. *Computers & Operations Research*, 43:271–279, 2014.

[12] M. Lagergren. What is the role and contribution of models to management and research in the health services? A view from Europe. *European Journal of Operational Research*, 105(2):257–266, 1998.

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

[14] M.M. Günal and M. Pitt. Discrete event simulation for performance modelling in health care: a review of the literature. *Journal of Simulation*, 4:42–51, 2010.

[15] B. Mielczarek and J. Uzialko-Mydlikowska. Application of computer simulation modeling in the health care sector: a survey. *Simulation-Transactions of the Society for Modeling and Simulation International*, 88(2):197–216, 2012.

[16] H. Eskandari, E. Mahmoodi, H. Fallah, and C.D. Geiger. Performance analysis of commercial simulation–based optimization packages: OptQuest and Witness Optimizers. In S. Jain, R.R. Creasey, J. Himmelspach, K.P. White, and M. Fu, editors, *Proceedings of the 2011 Winter Simulation Conference*, pages 2363–2373, 2011.

[17] J.P.C. Kleijnen and J. Wan. Optimization of simulated systems: OptQuest and alternatives. *Simulation Modelling Practice and theory*, 15:354–362, 2007.

[18] A.R. Conn, K. Scheinberg, and L.N. Vicente. *Introduction to derivative–free Optimization*. MOS/SIAM Series on Optimization. SIAM, Philadelphia, MA, 2009.

[19] Rockwell Automation. *Arena User's guide*, 2010. Allen–Bradley, Rockwell Software.

[20] Rockwell Automation. *OptQuest for Arena User's guide*, 2012. Allen–Bradley, Rockwell Software.

[21] B. Moore. Appropriate-technology for birth. *Lancet*, 2(8458):787–787, 1985.

[22] OECD. Caesarean sections per 100 live births, 2009 and change between 2000 and 2009. Report, 2011.

[23] J.A. Lauer, A.P. Betràn, M. Merialdi, and D. Wojdyla. Background paper, No 29. Determinants of Caesarean section rates in developed countries: supply, demand and opportunities for control. Report, 2010.

[24] G. France, F. Taroni, and A. Donatini. The Italian health-care system. *Health Economics*, 14:S187–S202, 2005.

[25] Italian Health Ministry. DRG Medicare versione 24.0. Report, Italian National Health Service. (In Italian).

[26] Italian Health Ministry. Direzione generale e della programmazione sanitaria, dei livelli essenziali di assistenza e dei principi etici di sistema - Il dipartimento. Report, Italian National Health Service, 2005. (In Italian).

[27] W.D. Kelton, R.P. Sadowski, and D.T. Sturrock. *Simulation with Arena*. McGraw–Hill, Fourth edition, 2007.

[28] M.A. Abramson, C. Audet, J.W. Chrissis, and J.G. Walston. Mesh adaptive direct search algorithms for mixed variable optimization. *Optimization Letters*, 3(1):35–47, 2009.

[29] C. Audet and J.E. Dennis Jr. Pattern search algorithms for mixed variable programming. *SIAM Journal on Optimization*, 11(3):573–594, 2001.

[30] M. Kokkolaras, C. Audet, and J. Dennis Jr. Mixed variable optimization of the number and composition of heat intercepts in a thermal insulation system. *Optimization and Engineering*, 2(1):5–29, 2001.

[31] L.N. Vicente. Implicitly and densely discrete black-box optimization problems. *Optimization Letters*, 3:475–482, 2009.

[32] S. Lucidi, V. Piccialli, and M. Sciandrone. An algorithm model for mixed variable programming. *SIAM Journal on Optimization*, 14:1057–1084, 2005.

[33] U.M. García-Palomares, E. Costa-Montenegro, R. Asorey-Cacheda, and F.J. Gonzáles-Castaño. Adapting derivative free optimization methods to engineering models with discrete variables. *Optimization and Engineering*, 13:579–594, 2012.

[34] J. Müller, C.A. Shoemaker, and R. Piché. SO–MI: A surrogate model algorithm for computationally expensive nonlinear mixed-integer black-box global optimization problems. *Computers and Operations Research*, 40(5):1383–1400, 2013.

[35] M. Laguna, F. Gortázar, M. Gallego, A. Duarte, and R. Martí. A black-box scatter search for optimization problems with integer variables. *Journal of Global Optimization*, 2013.

[36] G. Liuzzi, S. Lucidi, and F. Rinaldi. Derivative-free methods for bound constrained mixed-integer optimization. *Computational Optimization and Applications*, 53:505–526, 2012.

[37] G. Liuzzi, S. Lucidi, and F. Rinaldi. Derivative-free methods for mixed–integer constrained optimization problems. *Journal of Optimization Theory and Applications*, 2014. To appear.

[38] G. Liuzzi, S. Lucidi, and M. Sciandrone. Sequential penalty derivative-free methods for nonlinear constrained optimization. *SIAM Journal on Optimization*, 20:2614–2635, 2010.

[39] S. Lucidi and M. Sciandrone. A derivative-free algorithm for bound constrained optimization. *Computational Optimization and Applications*, 21(2):119–142, 2002.