Noname manuscript No.
(will be inserted by the editor)

Scheduling of healthcare professionals using Bayesian Optimization

Guido Lagos · Francisco Jara-Moroni · René Lagos · Diego Machado · Bastián Saavedra · Tomás Torres

Received: date / Accepted: date

Abstract In this paper we present a Bayesian optimization framework that iteratively "learns" good schedules for healthcare professionals of outpatient healthcare in a hospital, that minimize the overall number of patients in queue — we understand that a patient in schedule is one in queue. The hospital has several medical specialties and each is modeled as having two activities: admissions and checkups. At each activity, patients wait in queue for medical attention, and after their service they are either discharged, scheduled for a checkup, or referred to another specialty. Importantly, we consider that each doctor can serve over several specialties, so the optimization has to take into account all specialties simultaneously. We propose a modular methodology that first forecasts patients' demand, then proposes a schedule for healthcare professionals, then simulates the performance of the hospital under such schedule, and finally a Bayesian optimization module iterates through the previous steps and "learns" which are the best schedules. Furthermore, each module is fairly simple and easy to modify and adapt to the hospital's constraints, and can even be used standalone — this is an advantage for countries with low technological sophistication and penetration. We test our methodology on data from year 2019 of a small-sized public hospital in Chile with 15 medical specialties. Our computational experiments show a reduction in 8% of the patients' queue just by redistributing the existing allocation of medical hours within a specialty. Additionally we examine the case where extra medic hours are allowed, leading to a 25% reductions on the estimated queue length.

G. Lagos
Universidad Adolfo Ibáñez, Facultad de Ingeniería y Ciencias
E-mail: guido.lagos@uai.cl

F. Jara-Moroni
Universidad Diego Portales, Facultad de Ingeniería y Ciencias
E-mail: francisco.jara@mail.udp.cl

R. Lagos
Digital Health Unit, South East Metropolitan Health Service
E-mail: rlagosb@ssmso.cl

D. Machado, B. Saavedra & T. Torres
Universidad de Santiago de Chile, Departamento de Ingeniería Industrial
E-mail: {diego.machado, bastian.saavedra, tomas.torres} @usach.cl

Keywords Scheduling · Healthcare · Bayesian optimization · Simulation Optimization

1 Introduction

In this paper we propose a Bayesian optimization with simulation framework to tackle a problem of scheduling of healthcare professionals in an outpatient centre of a hospital. That is, we consider the setting of a hospital where, roughly speaking, there are several medical specialties (e.g., gastroenterology, cardiology, traumatology, etc.); patients arrive referred for a specialty consultation; there is a number of healthcare professionals that can work in one or more specialties; and for each patient there is a sequence of activities —sometimes across different specialties— required to treat their health problem. The problem is then to decide how many consultation hours to offer in each specialty in order to minimize the number of patients in queue while satisfying its operational constraints. Here we model a patient scheduled for attention as a patient in queue.

We use Bayesian optimization with simulation because it allows to model the stochastic and complex dynamics of the system, while providing a framework that automatically "learns" about "good" schedules of healthcare professionals. Indeed, in simple words, Bayesian optimization is an algorithmic framework where there is a black-box evaluation of the objective value of feasible solutions, and new ones are proposed by fitting a Gaussian model to the previously observed solutions and their objective values. In this way, the algorithm "learns" about the solutions based on the value of previously proposed ones, and it proposes new solutions based on what it has learned up to that stage; see [11] for a tutorial overview. In our case, a feasible solution is a combination of number of medical hours allocated to each specialty —and each combination has to satisfy a number of operational constraints—; its objective value is the overall performance of the healthcare system, measured as the total number of patients in queue after the planning horizon (usually one year); and the performance of a combination of allocated hours is computed by simulating the system under such configuration. Importantly, our Bayesian optimization with simulation framework allows to uncouple the *choice of the feasible solution* of the problem —the possible schedule for professionals—, from the *evaluation of the objective value* of such solution —the performance of the solution—; the former is carried out solving a linear programming problem, and the latter is computed by simulating the medical facility as a queueing network. This is key, as the nature of both problems (proposing a schedule and evaluating it) are very different, and in this way we can use completely distinct methodologies for each part. Moreover, Bayesian optimization has proved to be especially useful when the evaluation of the objective function of the optimization is expensive to compute, say because a real-world experiment has to be carried out or because a lengthy computational experiment is run, as is our case, where the objective value is obtained from simulating the hospital operating under a given schedule.

Our approach of using Bayesian optimization with simulation is a novel application in the literature, despite the fact that scheduling of healthcare professionals is not a new problem. Indeed, there is a vast literature covering variegated problems of scheduling in healthcare, like operating theatre planning, bed assignment, emergency services planning, etc.; see, e.g., the handbooks [15,27,28] for nice starting points. Nonetheless, to the best of the author's knowledge, there is no work in the literature tackling this problem using Bayesian optimization with simulation; see the literature review in Section 2.

We also claim that our proposed methodology can be a potent and effective management tool in public healthcare in developing countries. Surely, in general scheduling in healthcare can be a powerful tool in operations management, as the schedule can help to monitor and balance workloads, plan the offer of medical services, define and track KPI's, and so on. But in particular, the methodology we propose can be especially useful for developing countries with relatively low technological adoption: the methodology presented is modular, each module is fairly simple and easy to modify to incorporate new constraints of the system, and each module can be implemented independently of the others and be used to tackle different operational and managerial problems. For example, to forecast patients' demand; or quickly evaluate different scheduling scenarios; or estimate and evaluate (via simulation) productivity goals; or estimate the utilization of physical resources, etc. This can help with the gradual adoption of our proposed solution in hospitals were there may be managerial resistance to new technological solutions, as is the case in the public hospitals in Chile; see, e.g., [24,3].

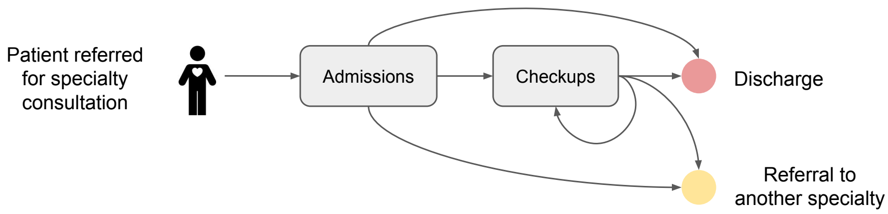

Fig. 1: Simple conceptual diagram of a patients' route in one specialty.

We model the network of specialties as follows. Each medical specialty is modeled as having two queueing *activities* or stations: first an *Admissions* queue where patients are diagnosed and start their treatment, and then a *Checkups* queue where patients are scheduled for progress checkup visits, which may be several consecutive times, until they are either discharged or referred to another specialty; see Figure 1 for a simplified diagram. Furthermore, for each patient, on its first visit to a specialty it is placed in the Admissions queue, and on subsequent visits it is placed in the Checkups queue. The routing after each patient's visit to a specialty, i.e., the decision of where a patient goes next after finishing its service, is probabilistic and independent of the patient, and only depends on the specialty. We make such a simplified representation of the network to have a relatively simple model that is inexpensive to simulate using discrete event modelling, while simultaneously being able to reflect the inter-dependencies within and across specialties.

We developed our methodology in partnership with the *Digital Health Unit*, a medical informatics unit of the *South East Metropolitan Health Service* (SEMHS onwards, or SSMSO in Spanish), which is a regional administrative division of the Chilean *Ministry of Health*. The SEMHS overlooks the management of the public health of the most populated portion of the *Metropolitan Region* in Chile.

In fact, it provides healthcare to approx. 5.9% of the Chilean population (to approx. 1,113,447 persons). It is in charge of 56 healthcare facilities, consisting of 5 hospitals, including *Hospital Sótero Del Río* the largest *high-complexity* hospital in Chile, 13 mental health centers (*Centros de Salud Mental*), and 39 community health centers (*Centros de Salud Familiar*).

It is worth mentioning that in Chile, in order to attend a specialist, patients have to be referred to a hospital by a general practitioner. There are 80 health conditions that receive prioritized attention as part of the policy of Explicit Health Guarantees (*GES*, for its acronym in Spanish). GES pathologies have a time limit to be attended and have special economic coverage. On the other hand, referrals for non-GES pathologies have to wait in queue for treatment and are placed in the national registry for waiting lists (*Sistema Informático de Gestión de Tiempos de Espera*, SIGTE). In this work we consider only non-GES pathologies.

Importantly, the partnership with the SEMHS for this work is developed as a joint effort to help alleviate the critically long waiting times to receive medical treatment of non-GES patients in public healthcare, prevalent in most of Chile and particularly at the centers overlooked by the SEMHS. For example, according to the official government data GLOSA-06 [25], by the end of 2020 SEMHS had 7.3% of the total national waiting lines for non-GES specialties (with 120,736 patients out the nation's 1,648,945 total), 5.8% of the nation's total waiting lines for surgeries (with 14,833 patients out of the nation's 254,529 total). Also, at SEMHS hospitals the average waiting time for new admissions to a specialty is of 571 days (the national average is of 501 days), with 61.8% of new admissions taking more than a year to be completed; and the average waiting time for surgery is of 547 days (the national average is of 525 days), with 68.5% of those surgeries waiting more than a year in queue. Moreover, according to the Chilean national repository of waiting lists, in 2019 a total of 34,305 patients at SEMHS facilities died before finishing their treatment, and actually 23,705 of them died in queue waiting to get registered in Admissions at some specialty. In fact, [23] found a statistically significant association between waiting time and mortality in Chile. Clearly, thus, there is a critical problem of insufficient provision of public healthcare in Chile, and we aim to tackle this problem with our methodology for the efficient scheduling of the working hours of healthcare professionals at SEMHS.

Main contributions. The main contributions of our work are the following.

1. We present a novel application of Bayesian optimization to a scheduling problem of healthcare professionals. To the best of the author's knowledge, Bayesian optimization has not been applied to healthcare problems in scheduling.
2. Our approach optimizes the schedule for all the specialties simultaneously, in contrast with methods in the literature that optimize each specialty separately. We do this because medical doctors can serve at one or more specialties, and specialties interact through patients that are referred or routed between them.
3. Our method allows to include several useful features, such as optimize the extra hours needed in a specialty, or to impose stability of the queues at a specialty (i.e., the queues not growing indefinitely in the long run). The simulation also allows to include realistic features such as patient and/or doctor no-shows, patient's prioritization, and so on.
4. In contrast with many scheduling methodologies in service sciences, our proposed methodology allows for a flexible and fairly easy to interpret modeling

of the problem, and for a gradual implementation. Indeed, our method is modular and allows to easily implement the operational constraints of the schedules in an optimization problem; and the performance evaluation part through simulation allows to model the dynamics between and within specialties with the desired amount of details and subtleties. Importantly, each of these modules are flexible, fairly easy to interpret for non-experts, and can be useful by themselves for management and operational purposes. This is a considerable strength for the gradual applicability of the methodology in settings where there is managerial resistance to sophisticated technological tools.
5. We show how to tackle a difficult problem of scheduling in service systems by decomposing it into a feasible scheduling part and a system simulation part. Indeed, the problem considered is a difficult one, as it considers the scheduling of professionals across several medical specialties, there are queueing stations within a specialty and derivations between specialties, and additionally there are a number of operational constraints to be satisfied by the schedules. Our methodology tackles this problem by decomposing the scheduling problem in two parts: an optimization part that proposes a feasible schedule that satisfies the operational constraints of healthcare professionals, and a simulation part that evaluates the performance of the proposed schedule.

Organization of this paper. This paper is organized is a follows. In Section 2 we give a literature review. In Section 3 we give the model of the system and define the optimization problem, and in Section 4 we give an overview of the methodology. In Section 5 we give details on the real world setting were we test our methodoloy, and in Section 6 we show computational experiments. Lastly, in Section 7 we show conclusions and discussions of our work.

2 Literature review

Scheduling problems in healthcare operations have a longstanding tradition. A seminal collection of works is [37,38] in the 1960's, that study the scheduling of a nursing department. Since then, there has been a plethora of scheduling applications in healthcare planning: of operation and procedure rooms [6], of nurses and other clinical staff [5], of an emergency department [27,28], of home care, of medical supplies, etc. See [15] for a fairly recent handbook on healthcare system planning, where solution approaches span from linear programming and dynamic programming methods to heuristics and metaheuristics.

Nonetheless, to the best of the authors' knowledge, currently there is no work in the literature tackling personnel scheduling problems —either in healthcare or not— using Bayesian optimization, the methodology we propose in this paper. The closest papers to ours are [22] and [26], where both use Bayesian networks to solve scheduling problems. The former paper tackles a nurse scheduling problem where the weekly schedule pattern of available nurses is selected using a Bayesian network with predefined pattern selection rules; and the latter work studies the scheduling of multiprocessor systems in computing, where the precedence constraints of tasks assigned to processors is imposed by using a Bayesian network. Still, the methodology of Bayesian networks is vastly different from Bayesian optimization: the former is a network model focused on efficiently computing conditional probabilities by imposing certain independence assumptions, while the latter is a black-box optimization method where the latest available information on the objective function is incorporated to update the belief on it; see [7] and [11], respectively, for introductory tutorials on these methodologies.

The two main methodological tools we use in this paper are *discrete event simulation* and *Bayesian optimization*. The we use the former to simulate the performance of a given schedule for the clinical staff, and the latter to automatically select the schedules that optimize the system performance. In what follows we give bibliographical references to these two techniques.

Regarding discrete event simulation, there is a longstanding tradition in Healthcare Operations Research of using simulation to model and optimize the performance of a system. Some of the earliest works available are the works of Fetter and Thompson in the 1960's, [10,34], that carry out computer simulations of a maternity leave and a hospital. Since then there has been an overabundance of works using simulation for healthcare system design, evaluation and optimization, see e.g. the fairly recent surveys [18,30,4,12]. Nonetheless, we can distinguish two fairly separate lines of work: first, works where simulation is used in healthcare for *system modeling and evaluation*, see e.g. [1], and second, works where simulation is used for *system optimization*, see e.g. [39,14]. Our work falls in the latter line, although to the best of the authors' knowledge there is no work in this field that uses simulation and Bayesian optimization in Healthcare Operations Research.

From an optimization methodology perspective, our work falls in the field of *Simulation Optimization*. More specifically, it lies within the family of methods where the objective of the optimization —the performance of the system— is computed by simulating the system and treating the simulation as a black box, see e.g. the chapters [2,16,20] of the handbook [13]. Indeed, the Bayesian optimization methodology we use in our work simulates the system to compute its performance for a given system design, and uses the output to decide the next design to evaluate; see [11] and [31] for fairly recent tutorials on Bayesian optimization from the perspectives of Operations Research and Machine Learning, respectively. Bayesian optimization was first proposed in the 1960's by Kushner [21] as a way of using Gaussian process modeling and regression to explore a parameter set to optimize a function. Other related black box methods are Response Surface Methodology [20], and especially kriging [17]; see [19] for an interesting survey.

3 Problem definition

In this section we show how we model the scheduling problem of health professionals at SEMHS-dependent hospitals, and we briefly describe how the problem is currently solved at the SEMHS.

In a nutshell, the problem consists on deciding which healthcare professional will perform which activity (Admissions or Checkups), of which specialty, how much time per week, and using which physical resources (rooms). That is, *who* will perform *what*, *how much* and *where*. In the following we thoroughly describe the problem and model.

Specialties. We consider a hospital that tends several medical specialties, say N specialties, labeled as $s_1, \ldots, s_N$. Each specialty has two stations or stages, that henceforth we call *activities*: first an *Admissions* station, where patients are screened and given an initial diagnose and treatment instructions; and second a *Checkup* station, where patients are subsequently checked, perhaps several times. Patients can be discharged after a visit to a specialty, say s, with probability $p_{s,\,\text{discharge}}$, or can be called again to the specialty, with probability $p_{s,\,s}$, and these probabilities are independent of the patient, the patient's history, or the activity it visited in the specialty.

Routing across specialties. After each patient's service at a a specialty there is a chance that the patient be derived to another specialty; for instance, it is not uncommon that patients in *Bronchopulmonary* specialty are derived to *Infectious Diseases.* We model this routing as there being a probability $p_{s_i,\,s_j}$ that the patient is derived from specialty s_i to specialty s_j , for each pair s_i and s_j. Overall, the probabilities $p_{s_i,\,\text{discharge}}$ and $p_{s_i,\,s_j}$ do not depend on the patient, its history in the system, or the activities it visited; thus, we say that the routing of patients is *Markovian.* In this way, we have $p_{s_i,\,\text{discharge}} + p_{s_i,\,s_i} + \sum_{s_j \neq s_i} p_{s_i,\,s_j} = 1$ for each specialty s_i.

Sequence of activities on subsequent visits. Importantly, we also model that the first time a patient arrives to a specialty it is put in queue at Admissions, and on subsequent visits to the specialty —e.g., right after a service in Admissions where it is called back to the same specialty, or after being routed back again to the specialty after being routed to another one— it is put in queue at Checkups.

Human and physical resources. There is a set M of medical professionals, where for each m in M there exists a non-empty set $S(m)$ of specialties that it can perform. A professional can serve in both Admissions and Checkups activities of any specialty it can work on. Also, there exists a given (finite) set of physical resources or rooms, and for each specialty and activity pair $(s,\,a)$ it can only be served at the subset $R(s,\,a)$ of physical resources — think, e.g., on the *Obstetrics and Gynaecology* specialty, that needs a room with special equipment.

Decisions and objective. The main question, thus, is for each medical professional m, how many hours $x_{m,\,(s,\,a),\,r}$ should the professional perform of specialty s and activity a in the physical resource r, in order to minimize the expected total number of patients in queue in the system after the planning horizon.

Constraints. There are a number of constraints that the decision variables have to satisfy. First, contractually, there is a minimum and maximum number of hours that can be allocated to a medic. Also, in Chile, the national public health department requires that, for each specialty, the total number of hours allocated to Admissions has to be at least 35% of the total number of hours allocated to the specialty. As mentioned above, each specialty and activity can only be performed in a given subset of physical resources, and for each such resource there is a maximum number of hours that it can be used. Additionally, there is a minimum number of hours that have to be allocated to each activity of each specialty. Lastly, in the public health system in Chile the time allocated to a patient's visit at a given specialty and activity is fixed — usually 20 or 30 minutes.

Other considerations not included. Our methodology also supports including the following considerations that could of interest, but for simplicity of exposition in this paper we do not delve into them. First, we ignore the number of contract hours that medics can use for vacations; that is, we assume that all the hours are effectively working hours. Also, it can easily be included that there is a probability that a patient misses or does not show up to its appointment; or that some times healthcare professionals need to cancel certain time schedules with little to no time in advance, so appointments in those time slots have to be rescheduled. Additionally, it can be included that when a patient is routed from a specialty to another it can actually be in queue for both (or more) specialties at the same time. These considerations can trivially be incorporated into our methodology, more specifically in the simulation module of Section 4.3 below.

3.1 Current situation at the Chilean SEMHS

Currently, the scheduling problem given above is solved at each SEMHS-managed hospital once per year. The scheduling is done in each hospital separately, and it is done by hand and/or using computer spreadsheets by a medical coordinating staff that take about a month to finish the job. The number of hours allocated to each medic is usually filled in by each medic using a paper form. Moreover, there is usually no forecast of the patient's demand for hours in each specialty, and also there is no consideration of the availability of physical resources (rooms).

4 Methodology

In this section we show the methodology we use to tackle the problem described in Section 3. For that, first we give an overview of the methodology, and then in Sections 4.1, 4.2, 4.3 and 4.4 we give details on each of the four modules of the algorithm.

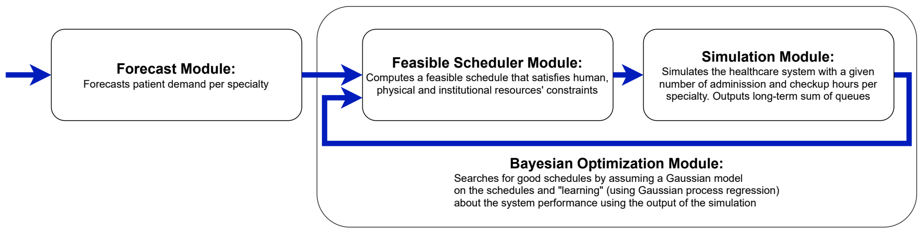

Fig. 2: Conceptual diagram of the proposed methodology.

Overview. The methodology we propose consists on four modules: (i.) a forecast module, (ii.) a feasible scheduling module, (iii.) a simulation module, and (iv.) a Bayesian optimization module that iterates between the feasible and simulation modules. A summary diagram of the methodology is shown in Figure 2.

Each module can be described in short as follows. In the *Forecast module* we use historic data to forecast the patient's demand for specialties for the planning horizon. In the *Feasible scheduler module* a linear programming problem is solved to obtain a feasible schedule for the healthcare professionals. Then, the *Simulation module* uses Discrete Event Simulation to simulate the performance of the schedule proposed by the Feasible scheduler module. Lastly, the *Bayesian optimization module* re-iterates through the feasible scheduler and the simulation modules and automatically "learns" what are the best combination of hours that give the best performance. In the following sections we give more details on each of the modules.

4.1 Forecast module

The *Forecast module* is in charge of estimating effective arrival rates for all available medical specialties. These effective arrival rates estimations are then converted into exogenous arrival rates, for each specialty, and transition probabilities between specialties, that will feed the *Simulation module*.

The module is built upon a framework created to select the best fit among multiple forecasting choices. This framework consists of two submodules: - *method adjustment*, and - *method selection*.

Method adjustment fits various forecasting models to historic demand data, per medical specialty. These forecasting models include well known time series models such as single, double and triple exponential smoothing[29, 36], and more recent Machine Learning methods such as XGBoost[8]. For each model, several loss functions are considered such as mean absolute error (MAE), mean squared error (MSE), and mean absolute percentage error (MAPE), all of which return different optimized parameters. The data utilized for this parameter optimization comprises monthly demand form the year 2016 through 2017. The output of this stage is, then, a collection of different models, one for each method/error tuple, per medical specialty.

Method selection, as its name suggests, selects the most appropriate method to estimate the effective arrival rates per medical specialty. To do so, each model from the *method adjustment* submodule evaluates MAE, MSE and MAPE of its forecasted data contrasted against that from 2018. The forecasting method selection is based on a Multi Attribute Utility function, a standard procedure used in decision theory, to represent preferences over bundles of options.

Mathematically, for each medical specialty, *Method Adjustment* outputs a set of forecasting models $H = \{h_1, h_2, \ldots, h_m\}$, all of them adjusted to fit a certain data set X_F (demand data from 2016 to 2017). *Method selection* provides a unified score $S(h_k)$ to each function in this set. This score consists in a weighted average of normalized measures of error, $\hat{U}_i(h_k)$, with $i \in \mathcal{E} := \{MAD, MSE, MAPE\}$, evaluated using data from a set X_V (demand data from 2018). For instance, if $\bar{h}$ represents a specific model fit, using data from X_F, then $U_{MAD}(\bar{h})$ outputs the MAD of the evaluation of $\bar{h}$ using data from X_V. The normalized error for the MAD is calculated using the formula

$$\hat{U}_i(\bar{h}) = 1 - \frac{U_i(\bar{h}) - \min_{h \in H} U_i(h)}{\max_{h \in H} U_i(h) - \min_{h \in H} U_i(h)},$$

thus restricting all these scores to the interval [0, 1]. We can compute the score

$$S(\bar{h}) = \sum_{i\in\mathcal{E}} w_i \hat{U}_i(\bar{h}),$$

where w_i corresponds to a weight assigned to error measure $i \in \mathcal{E}$. In order to avoid any a priori bias over the method selection we treat $w_i = w_j$, for all $i, j \in \mathcal{E}$, although a sensitivity analysis is performed, and described in section 5.1, where we observe that the method selection is actually independent of the distribution of the weights. Finally, the selected method for the corresponding medical specialty is, then, $\arg\max_{h\in H}\{S(h_k)\}$.

4.2 Feasible scheduler module

The *Feasible scheduler module* produces a schedule satisfying the constraints put forward in Section 3.

That is, the module returns, for each healthcare professional m, how many hours $x_{m,(s,a),r}$ should she perform of specialty s and activity a (either Admissions or Checkup) in the physical resource r. The schedule satisfies the following constraints: there is a minimum and maximum number of hours that can be allocated to a medic; that there is a minimum number of hours to be allocated to a specialty; also that in Chile it is required that at least 35% of the hours assigned to a specialty are used in the *Admissions* activity; that some specialties can be performed in some subset of physical resources; that each physical resource is available at some particular times; and that in Chile each medical attention usually takes a fixed amount of time. See Appendix A for the particular formulation as a linear programming problem. Recall too that we assume that all hours of a medic are effectively working hours, and do not include vacations, days off, etc.

A linear constraint for queue stability. The module can also impose a linear constraint that implies *stability* of each queue of the network of specialties and activities, i.e., a sufficient condition for each queue of the network that guarantees that the number of patients in queue will not explode in the long term. Indeed, [32, Theorem 2.1] gives a condition on the *traffic intensity* of all the stations of a generalized Jackson network (GJN) that guarantees their stability. In our case, the network of queues for each activity and specialty is *not* a GJN, since for each patient the first time in a specialty it goes to Admissions, on subsequent times it goes to Checkup, and both queues have different service rates. Nonetheless, we can bound our network by a slower GJN and obtain the following stability condition for our system.

Proposition 1 *For each specialty s_i, let $\bar{\lambda}_i$ be its maximum (along a year) monthly exogenous arrival, p_{s_i,s_j} be the patient's routing probabilities from specialty s_i to s_j, and let $\underline{\mu}_i$ be the minimum between the monthly number of patients served at Admissions and at Checkups of specialty s_i. If the values $\bar{\gamma}_i$ defined as the unique solution of the linear system*

$$\bar{\gamma}_i = \bar{\lambda}_i + \sum_{s_j} \bar{\gamma}_j p_{s_j,s_i} \qquad \textit{for all specialty } s_i \tag{1}$$

satisfy the condition

$$\bar{\gamma}_i < \underline{\mu}_i \qquad \text{for all specialty } s_i, \tag{2}$$

then with probability one the number of patients in queue at each activity and specialty does not grow to infinity in the long run, i.e., all queues are stable.

Proof First recall that at each specialty there are two queues, Admissions and Checkup, and that for each patient, on its first pass through the specialty it goes to Admissions, and on subsequent passes it goes directly to Checkup. Assume for the time being that the monthly arrival rate at each specialty s_i is actually constant and equal to the largest monthly arrival rate, $\bar{\lambda}_i$.

Consider now an alternate system where at each specialty s_i there is only one queue, and that it serves patients at the slowest rate between the service rate at Admissions and at Checkup, i.e., at rate $\underline{\mu}_i$. Since the routing probabilities are the same for both systems (the original and the alternate) then we can think of the alternate system as the original one but where for each specialty both queues have been merged into one that serves at the slowest rate between the two original activities. Hence, we can identify each patient in the original system as one in the alternate system, but where its pass through a specialty is slower or equal than in the original one. Consequently, if the alternate system is stable then the original one will also be stable. It follows that the alternate system is a generalized Jackson network, and [32, Theorem 2.1] gives the sufficient condition (2) for the stability of the alternate network. Hence, (2) is also a sufficient stability condition for the original system.

Lastly, if the system under the maximum arrival rates $\bar{\lambda}_i$ is stable then the system with rates varying monthly will also be stable. □

Interestingly, the inequalities (2) can be easily included into the Feasible scheduler module to guarantee the stability of the proposed schedule. For that, first transform (2) into non-strict inequalities by including a small positive term to each inequality, say $\bar{\gamma}_i + 10^{-6} \leq \underline{\mu}_i$, and then write them as two linear inequalities on the total number of weekly hours allocated to each activity of the specialty: $\bar{\gamma}_i + 10^{-6} \leq \mu_{i,\text{Admissions}}$ and $\bar{\gamma}_i + 10^{-6} \leq \mu_{i,\text{Checkup}}$.

In this way, a schedule can be obtained with a simple linear program with any desired objective function. For example, one can minimize the number of healthcare professionals used, or minimize the number of additional hours needed at each activity and specialty so that inequalities (2) are satisfied and the system is stable. See Appendix A for the formulation we use of the constraints and possible objective functions.

4.3 Simulation module

In the *Simulation module* we take aggregated information of a feasible schedule for the healthcare system, and simulate its performance when it operates using this schedule.

In detail, first, the simulation module takes as input the total weekly number of hours offered by the hospital for each activity and specialty. Importantly, note that this input is only an aggregated portion of the information of the feasible schedule that is computed in the Feasible scheduler module — this detail is important for the Bayesian optimization in Section 4.4 below. With this, we simulate the system along the planning horizon and output its performance, usually measured as the total expected number of patients in queue at the end of the time horizon; that is, the sum, along all specialties, of the queue length of each specialty. It is worth mentioning that the discrete event simulation allows to output several system performance measures, e.g., the average length of stay in the system of a patient.

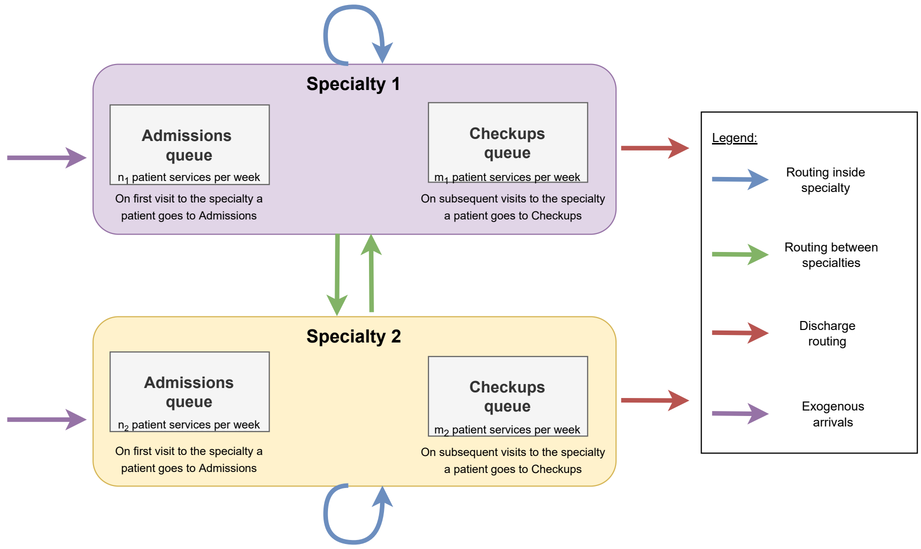

Fig. 3: Conceptual diagram of the patients' flux among specialties.

In Figure 3 we show a conceptual diagram of how we model the activities within a specialty and how they interact with other specialties. The model follows the description given in Section 3, whose main features for the simulation are the following. First, a patient first arriving to a specialty is put in the queue of Admissions, otherwise —if it previously visited the specialty— it goes directly to the queue of Checkup. Second, after a patient finishes its service at either activity of a specialty, it is either discharged from the system, or called back to the same specialty, or routed to another one. Third, the routing of the patients between specialties is Markovian, that is, the probabilities are all independent of the patient and its history, they only depend on the specialties from and to which it is derived. Also, arrival of patients from outside the system to a specialty (i.e., exogenous arrivals) follows a Poisson process with a constant rate per month, and these processes are independent between specialties. Lastly, a service in an activity and specialty takes a fixed amount of time, and in particular we only need the total number of hours allocated to an activity and specialty to compute how many services can be offered monthly.

In this way, the parameters of the simulation module are the following: after a visit to a specialty, the probabilities of being discharged, of being called again to Checkup, and to be routed to another specialty (see Section 3); the monthly exogenous arrival rates to each specialty, and the fixed service time of each activity and specialty.

To simulate the system we used Discrete Event Simulation over a simulation time horizon of one year, starting with either a *warm start* of the actual data of queue lengths at each specialty at the beginning of the planning horizon, or by simulating the system starting empty and giving it a ten month *warm up* period, prior to the actual twelve month simulation horizon. In the simulation we sample the arrival of each patient over the simulation horizon, and keep track of its trajectory in the system until it is discharged or the simulation ends. In this way we can easily compute several efficiency measures of interest, such as the average length of stay of a patient, the average queue length, the average utilization rate of the healthcare professionals, and so on. Nonetheless, in our methodology the output of the simulation is the expected total queue length along the system, although it can be any performance measure that we want to optimize.

In summary, the input of the simulation module is the total weekly hours offered for each specialty and activity; the output of the module is a one-dimensional performance measure of the system over a simulation horizon; and the parameters of the simulation are the monthly exogenous arrival rates to each specialty, the routing probabilities, and the fixed amount of time each specialty and activity takes.

Lastly, we recall that our framework allows to easily model patients missing their appointments with some probability, or also that some times the professionals miss their scheduled hours; in both cases the patients have to be rescheduled.

4.4 Bayesian optimization module

The *Bayesian optimization module* is the key part of our methodology, as it allows to automatically "learn" which are the best schedules. See [11,31] for two fairly recent tutorials on Bayesian optimization.

In short, Bayesian optimization (BO) is a method where the algorithm iteratively explores the set of feasible solutions while automatically learning and gaining certainty on which are the best ones. The framework is popular in settings where the objective function is a "black box" of which we have little to no information, but nonetheless have the possibility of "seeing" the value of a solution, say by simulation, a costly computation, or by conducting an experiment. The key modeling assumption of BO is that the space of solutions and its objective values follow a Gaussian process. In most cases this is an artificial assumption, but nonetheless it allows the algorithm to effectively explore the feasible set and find good solutions.

In our case, the BO uses the simulation as the "black box" objective function, say f, and a solution is a vector giving the number of hours allocated to each specialty and activity. That is, a *solution* for the BO is a vector x in $\mathbb{R}_+^{S\times A}$, where S is the set of specialties and A the set of activities (Admissions and Checkup), and $x_{s,a}$ is the number of hours assigned to specialty s and activity a. This solution x is obtained from a feasible schedule outputted by the Feasible scheduler module of Section 4.2. In turn, the objective value $f(x)$ of each solution x is the output of the Simulation module, see Section 4.3, which gives the total number of patients in queue at the end of the planning horizon while operating under the configuration of hours given by the solution x. We remark that the output of the Feasible scheduler module is not directly a solution for the BO, as the former has the information for each medic, specialty, activity and physical resource; however from such detailed schedule we can extract the information for a solution of the BO: the total number of hours assigned to each specialty and activity. Importantly then, the dimension of the solution space of the BO is two times the number of specialties — this is relevant for the BO, as it works best with lower dimensions, see [11].

In the BO setting it is assumed that the objective values $f(x)$ of the solutions x in $\mathbb{R}_+^{S\times A}$ are jointly distributed as a Gaussian process. In particular, for each solution x its objective value $f(x)$ is modeled as a normal random variable —the current "belief" on its value— with a given mean and variance. At each iteration of the BO algorithm we "see" the actual objective value of a proposed solution, in this case by simulating the system under the given configuration of hours, and with this new information we update the mean and covariance of the other solutions, therefore "learning" about the objective function based on each observation. In other words, each solution x has an actual objective value $f(x)$, however since each simulation is computationally costly to perform it is assumed *a priori* that $f(x)$ is a normal random variable, representing the belief we have on its value with the information available up to that stage.

The three main components of the BO algorithm are the *mean function*, that models the initial "belief" of the value of each solution; the covariance function, or *kernel*, that models the correlation of the objective value between different points in the space of solutions; and the *acquisition function*, that specifies how we estimate the gain in both the objective value and in information of a new (previously unexplored) solution. In the following we give further information about these parameters in our specific problem.

1. The initial mean function μ_0 we use is constant equal to zero for all solutions, i.e., $\mu_0(x) = 0$ for all x. This is a standard practice that reflects that initially the "belief" on the value of each solution is uniform because there is no information on their value.
2. The kernel function $\Sigma_0(x, x')$ gives the covariance between the normal random variables modeling the "beliefs" on the objective values of two different solutions x and x'. Some standard kernel functions are *dot product*, *rational quadratic*, *Matern*, *radial basis function* and *white kernel*; see, e.g., [35, Chapter 4]. The kernel function Σ_0 we choose is a sum and multiplication combination of standard kernel functions, also a standard practice in BO and in Gaussian Process regression. See [9, Chapter 2] for further information on kernel selection.
In this way, given the solutions $x^{(1)}, \dots, x^{(n)}$ and their respective simulated objective values $f(x^{(1)}), \dots, f(x^{(n)})$, the distribution of $f(x)$ is normal with mean

$$\mu_n(x) = \Sigma_0\left(x, x^{(1:n)}\right) \Sigma_0\left(x^{(1:n)}, x^{(1:n)}\right)^{-1} \left(f(x^{(1:n)}) - \mu_0(x^{(1:n)})\right) + \mu_0(x) \tag{3}$$

and variance

$$\sigma_n^2(x) = \Sigma_0(x, x) - \Sigma_0\left(x, x^{(1:n)}\right) \Sigma_0\left(x^{(1:n)}, x^{(1:n)}\right)^{-1} \Sigma_0\left(x^{(1:n)}, x\right), \tag{4}$$

where $x^{(1:n)}$ is the vector with the values $x^{(i)}$, $\Sigma_0(x, x^{(1:n)})$ is the vector of covariances $\Sigma_0(x, x^{(i)})$, the matrix $\Sigma_0(x^{(1:n)}, x^{(1:n)})$ is analogous; and $f(x^{(1:n)})$ and $\mu_0(x^{(1:n)})$ are the vectors with values $f(x^{(i)})$ and $\mu_0(x^{(i)})$, respectively. These are standard formulas for Gaussian process regression, see e.g. [35, Chapter 2.1].

3. The acquisition function we use is the *Expected Improvement* (EI) function, that consists on, after n steps of the algorithm and for a solution x yet "unseen" by the algorithm, quantifying by

$$EI_n(x) := \mathbb{E}\left[[f_n^* - f(x)]^+\right] \tag{5}$$

the expected improvement of both objective value and in information of "seeing" the objective value of x, where $[\cdot]^+$ is the positive part function, $[z]^+ = \max\{z, 0\}$. Here, $f(x)$ is a normal random variable —with mean $\mu_n(x)$ and variance $\sigma_n^2(x)$ given by (3) and (4), respectively— modeling the belief on the objective value of the solution x; and $f_n^* := \min\{f(x^{(1)}), \ldots, f(x^{(n)})\}$ is the best of the objective values observed by the algorithm up to the n-th iteration, where $x^{(i)}$ and $f(x^{(i)})$ are, respectively, the solution and observed objective value at the i-th iteration of the algorithm. Importantly, it holds that the EI (5) can be rewritten as the following expression that can be optimized with any off-the-shelf non-linear programming solver (see [11, Section 4.1]):

$$EI_n(x) = \Delta_n(x)\Phi\left(\frac{\Delta_n(z)}{\sigma_n(x)}\right) + \sigma_n(x)\phi\left(\frac{\Delta_n(x)}{\sigma_n(x)}\right), \tag{6}$$

where $\Delta_n(x) := f_n^* - \mu_n(x)$, and Φ and ϕ are the cumulative distribution and probability density functions, respectively, of a standard normal distribution.

Note that in this way the BO algorithm searches for the solution with best (or better) objective value not by directly minimizing the objective function f, but instead by iteratively choosing the solution that maximizes the EI acquisition function. Algorithm 1 gives a pseudo-code summarizing the BO algorithm in our methodology.

Algorithm 1 Bayesian optimization for scheduling in healthcare

Require: Start with a number n_0 of schedules (possibly obtained at random) and their simulated average queue length. Build a Gaussian model, with parameters given by (3) and (4), that interpolates the observed schedules and their simulated performance.

1: $n \leftarrow n_0$
2: **while** there is simulation budget **do**
3: Optimize the acquisition function EI_n in (5) and obtain the new schedule x_n to simulate.
4: Simulate the expected total queue length of the hospital operating under schedule x_n.
5: Update the Gaussian model using formulas (3) and (4) to include the latest schedule x_n and its simulated performance.
6: $n \leftarrow n + 1$
7: **end while**
8: **return** the schedule with best simulated performance.

We note that the optimization of the EI in step 3 of Algorithm 1 is carried out differently depending on the size of the instance. On fairly small instances the expression (5) can be optimized directly with any off-the-shelf non-linear programming solver subject the linear constraints of the Feasible scheduler module of Section 4.2. However, on larger instances this step is very expensive computationally, so we resort to sampling a large but finite pool of feasible schedules using the Feasible scheduler module and then compute the EI of each schedule. To populate the pool of feasible schedules of interest one can run the Feasible schedule module as a minimization problem with random positive weights on the medical hours of each specialty, or on the extra number of hours needed for each specialty to have a stable queue, see Proposition 1 and the paragraphs thereafter.

We also remark that each iteration of the algorithm does only one simulation of the performance of a proposed schedule, whereas it can involve the evaluation of the EI (6) of many schedules. In this way, the "simulation budget" of the algorithm is actually the number of schedules to be simulated, and the algorithm is computation-efficient when the simulation of a schedule is orders of magnitude more consuming than the computation of the EI (6) —for instance when the hospital is large—, or when the pool of schedules of possible interest is much greater than the number of schedules that can be simulated.

5 Case study at SEMHS

In this section we show how we applied our healthcare professional scheduling methodology to a hospital in SEMHS network.

We applied our methodology to the public hospital *CRS Hospital Provincia Cordillera* (*CRS Hospital* onwards), which is part of SEMHS network and is located in south-eastern Santiago, in the county of Puente Alto, the most populous county in Chile. This is a fairly new and small hospital, opened in 2016, and as of 2020 it had 221,947 registered patients, served 16 specialties with a medical staff of 118 doctors, and used approx. 45 rooms for medical attentions, see [33].

We use data of the CRS Hospital from year 2016 to the beginning of October 2019. We considered only up to that month since on October 18th 2019 in Chile it started a long period of heavy social unrest and demonstrations, where the main cities where decreed under state of emergency. Together with the COVID-19 pandemic that started in Chile in March 2020, this made the data of mid-October 2019 onwards be deemed abnormal.

We test our methodology by proposing schedules to be used from January to the end of September of 2019, and then compute the total expected queue lengths at the start of October 2019. To make a fair comparison, we consider the medical hours that were actually used on that time period at the CRS Hospital, and using the Simulation module we also compute its total expected queue length, and compare it to the one of the schedule of our methodology. Note that in this way we do not use the *actual* total queue length at the start of October 2019, as that is not the *expected* total queue length, it is only one possible realization.

To tune our simulations and experiments we use data obtained from the *Hospital Production Plan and National Waitlists Repository* database from the Chilean government. In particular, in 2019 the CRS Hospital offered consultations for 15 medical specialties. In Table 1 we list the specialties, number of hours allocated to each activity, patients in queue at the beginning of January 2019, and discharge probability from each specialty. We also assign an identification label to each specialty. The table shows that most specialties had more than 250 patients in queue. Then, in Table 2 we show the transition probabilities from each specialty to each other. The transition and discharge probabilities are computed using the SEMHS database of total number of derivations and discharges on that time period. We remark that the probabilities in Tables 1 and 2 are such that each patient is discharged or not with the probabilities given in Table 1, and if it is not, then it is routed to another specialty according to the probabilities in Table 2. Note too that the sum at each row of Table 2 is one.

Id.	Specialty	Patients served per week		Patients in queue	Discharge probability
		Admissions	Checkups		
1	Cardiology	12	30	756	0.1034
2	General surgery	52	35	978	0.1533
3	Dermatology	24	36	609	0.0605
4	Endocrinology	26	52	23	0.1018
5	Gastroenterology	24	72	91	0.1619
6	Gynaecology	24	36	573	0.1249
7	Physical medicine and rehab.	52	72	1,299	0.1817
8	Internal medicine	78	192	264	0.0952
9	Nephrology	12	32	16	0.0507
10	Neurology	16	24	297	0.1135
11	Ophthalmology	54	56	1,065	0.1013
12	Otorhinolaryngology	51	52	1,141	0.1493
13	Pediatrics	36	75	274	0.0299
14	Traumatology and orthopaedics	51	76	978	0.1653
15	Urology	21	44	455	0.1283

Table 1: Specialties, patients served at each activity, number of patients in queue at the beginning of January 2019, and discharge probability at the CRS hospital.

Id. \ Id.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0	0.0264	0.0132	0.2772	0.1452	0.0132	0.1122	0.0726	0.0528	0.1254	0.0264	0.0165	0	0.0660	0.0528
2	0.2651	0	0.0066	0.0546	0.3846	0.0189	0.0102	0.0685	0.0102	0.0087	0.0109	0.0175	0	0.0830	0.0612
3	0	0.0400	0	0.2400	0.4000	0	0	0.0400	0	0.1600	0.0400	0	0	0.0800	0
4	0.2311	0.1873	0.0398	0	0.0876	0.0159	0	0.0040	0.0239	0.0637	0.0558	0.0279	0	0.2470	0.0159
5	0.3019	0.1094	0.0038	0.1132	0	0.0906	0.0075	0.0377	0.2189	0	0.0075	0.0453	0.0038	0.0075	0.0528
6	0.1326	0.0375	0.0231	0.2853	0.2075	0	0.0231	0.0720	0.0058	0.0173	0.0202	0.0086	0	0.0576	0.1095
7	0.0462	0.0018	0.0089	0.0533	0.0320	0.0107	0	0.0622	0.0249	0.0710	0.0107	0.0231	0.0515	0.5968	0.0071
8	0.1167	0.0397	0.0068	0.1343	0.1654	0.0227	0.0986	0	0.0805	0.0884	0.0725	0.0232	0.0006	0.0997	0.0510
9	0.1340	0.0101	0	0.3970	0.1608	0.0335	0.0469	0.0201	0	0	0.0101	0.0067	0	0.0402	0.1407
10	0.1108	0.0058	0.0058	0.0875	0.0117	0.0175	0.5015	0.0029	0	0	0.0612	0.0671	0	0.1166	0.0117
11	0.1242	0	0.0373	0.0932	0	0	0	0.3789	0	0.2112	0	0.1242	0.0062	0.0248	0
12	0.2152	0.0074	0.0130	0.2059	0.2115	0.0111	0.0186	0.0390	0.0445	0.0779	0.0371	0	0.0371	0.0705	0.0111
13	0.0186	0.0124	0.0464	0.0093	0.0186	0.0124	0.1981	0.0217	0.0124	0.0433	0.1548	0.2972	0	0.1548	0
14	0.0072	0.0069	0.0014	0.0076	0.0080	0.0065	0.8919	0.0098	0.0007	0.0470	0.0040	0.0047	0.0007	0	0.0036
15	0.2056	0.1682	0.0187	0	0.0935	0.0748	0.0935	0.1028	0.0374	0	0.0280	0.0093	0	0.1682	0

Table 2: Routing probabilities from each specialty (rows) to each other one (columns) used in the Simulation module. The id of the specialties are the ones listed in Table 1. The probabilities are computed using derivations data from year 2018 obtained from the databases of the SEMHS.

Corrupted demand data and estimation. Importantly, we realized that patient's demand data at the SEMHS database was inconsistent with reality: many specialties had a demand entry that was lower than the monthly patients served, but its entry for queue length was actually increasing considerably. It was then agreed with the data engineers at SEMHS to not consider that demand data they had, but to do trust and consider the data of number of services and weekly hours, with which Tables 1 and 2 were built.

Consequently, to estimate the demand per specialty we proceeded as follows. We assumed that the monthly demand on the period January to October 2019 had constant rate and then used linear regression to estimate its value from the data on initial queue length per month and specialty. Indeed, when there are patients in queue and the queue length increases, the number of patients at the end of the period is actually the initial number, minus the served patients (which we assume constant) plus the new arrivals. Thus, each month gives us an estimation of the number of patients arrived, and using linear regression we can estimate the constant demand that minimizes sum of mean squared loss. In Table 3 we show the demand values we obtain in this way.

To validate the latter procedure we perform the following. First we feed the Simulation module with the arrival rates thus estimated and the data of Tables 1 and 2; then we simulate 100 times the system on the period January to the beginning of October 2019 to estimate the mean queue length per specialty on that time period; and then we compare these numbers with the actual queue lengths in Admissions at the start of October 2019. In Table 3 we show the estimated mean demand per specialty, the estimated queue length at Admissions by the end of October 2019, and the actual queue lengths at that activity and time. We focus on Admissions activity because SEMHS did not have data for Checkups' queue lengths for that time period. Nonetheless, we observe that the overall error is pretty low, despite there being a few specialties with large error.

Id.	Specialty	Weekly demand	Admissions' queue length		Error [%]
			Simulated	Actual	
1	Cardiology	5	545	290	87.24
2	General surgery	45	785	771	1.82
3	Dermatology	28	800	867	7.73
4	Endocrinology	25	119	83	42.37
5	Gastroenterology	31	480	469	2.35
6	Gynaecology	29	865	828	4.47
7	Physical medicine and rehab.	54	1,480	1,497	1.14
8	Internal medicine	72	480	436	10.09
9	Nephrology	10	267	32	15.63
10	Neurology	23	610	626	2.56
11	Ophthalmology	44	1,081	994	8.75
12	Otorhinolaryngology	58	1,676	1,597	4.95
13	Pediatrics	27	161	130	23.85
14	Traumatology and orthopaedics	55	1,274	1,243	2.49
15	Urology	21	514	512	0.39
		Total:	10,880	10,375	5.01

Table 3: Specialties, estimated mean weekly demand, estimated mean and actual number of patients in queue at the end of October 2019, and relative error between the latter two numbers.

In this way we have estimated all the parameters for the Simulation module of Section 4.3: the specialties, the number of weekly hours offered in admissions and in checkups, the monthly arrival rates of patients, and the routing probabilities after service at each specialty.

Note that in particular in this case study we do not use the Forecast module of Section 4.1, because the demand data was corrupted and the data engineers at SEMHS deemed that it did not reflect the actual demand at the CRS Hospital. Nonetheless, to show how to apply the module, in Section 5.1 below we do a forecasting exercise on the available (corrupted) data. We include this for illustration purposes only.

As for the parameters of the Bayesian optimization module of Section 4.4, the mean function is initially zero everywhere, and for the kernel function we follow [9, Chapter 2] to obtain a kernel with good prediction power.

5.1 Forecast parameters

We use the patients' demand data from years 2016 to 2017 to tune our forecast module of Section 4.1. Recall that this module of the methodology forecasts the yearly demand per specialty, and is used in the schedule optimization we propose.

First, we group the specialties into broadly three behaviors: stationary, with trend, and stationality. Then we choose the best forecast method, per specialty, using the MAUT criterion described in Section 4.1, with data from 2018. At first, we chose same weights for all measure errors to avoid any a priori bias over our selection. Then, we performed a sensitivity analysis of the forecast method selection by computing the MAUT score with different randomized weights to assess the effect on the score calculation. Interistingly, the sensitivity analysis shows that consistently the best method for specialties showing trend behavior is the Winter-Holt's method, considering the MSE as the loss function, and for specialties with seasonal demand behavior is XGBoost with quarterly grouping.

Validation. We then validate our forecast methods by forecasting the demand for year 2019, and comparing it to the actual demand of that year. That is, first at the beginning of year 2019 we forecast the yearly demand of patients for each specialty, and then we compare each forecast with its actual monthly demand. To quantify the difference between the forecasted and actual demand we use the mean absolute deviation (MAD) of the differences and take the average over the ten month validation horizon, i.e., for each specialty, compute $\sum_{t=1}^{10} |d_t^{forecast} - d_t^{real}|/10$, where $d_t^{forecast}$ and d_t^{real} are, respectively, the forecasted and real demand for month t in that specialty. Similarly, we also compute the average mean absolute percentage error (MAPE), $\sum_{t=1}^{10} |d_t^{forecast} - d_t^{real}|/(d_t^{real} \times 10)$. In Table 4 we show these error per specialty, for the period January to October 2019.

Id.	Behavior	Specialty	MAUT method	Best MAD	Diff (%)	Best MAPE	Diff (%)
1	Trend	Cardiology	W-H (MSE)	W-H (MSE)	0.00%	W-H (MSE)	0.00%
2	Trend	General surgery	W-H (MSE)	W-H (MAD)	1.76%	W-H (MAD)	1.92%
3	Trend	Dermatology	W-H (MSE)	W-H (MSE)	0.00%	W-H (MSE)	0.00%
4	Trend	Endocrinology	W-H (MSE)	XGBOOST	12.10%	XGBOOST	13.04%
5	Trend	Gastroenterology	W-H (MSE)	W-H (MSE)	0.00%	W-H (MSE)	0.00%
6	Trend	Gynaecology	W-H (MSE)	XGBOOST	35.27%	XGBOOST	38.32%
7	Trend	Physical medicine and rehab	W-H (MSE)	W-H (MSE)	0.00%	W-H (MSE)	0.00%
8	Trend	Internal medicine	W-H (MSE)	W-H (MAD)	8.34%	W-H (MAD)	3.85%
9	Trend	Nephrology	W-H (MSE)	W-H (MSE)	0.00%	W-H (MAD)	4.33%
10	Trend	Neurology	W-H (MSE)	XGBOOST	19.31%	XGBOOST	13.99%
11	Trend	Ophthalmology	W-H (MSE)	W-H (MAD)	5.62%	W-H (MSE)	11.06%
12	Trend	Otorhinlaryngology	W-H (MSE)	W-H (MSE)	0.00%	XGBOOST	13.85%
13	Seasonal	Pediatrics	XGBOOST	XGBOOST	0.00%	XGBOOST	0.00%
14	Trend	Traumatology and orthopaedics	W-H (MSE)	XGBOOST	30.33%	XGBOOST	15.75%
15	Seasonal	Urology	XGBOOST	XGBOOST	0.00%	XGBOOST	0.00%

Table 4: Forecast Validation

6 Computational results

In this section we present and analyze the outcomes of our numerical experiments. First, we show the results on a small subset of specialties, followed by the results on the full set of fifteen specialties, assuming no extra medical hours are allowed into the system. Lastly, we show the outcome of an experiment performed under the assumption that hiring more medical staff (or extending current medical hours) are allowed.

The timespan of the simulation leading to our black box function is 10 months. The numerical experiments we run in a MacBook pro M1 with 8 GB of RAM.

6.1 Simple Case, Two specialties:

We run our model on the two most loaded specialties at the hospital: Traumatology and Orthopaedics (T&O), and Physical Medicine and Rehab (PMR). Currently, both specialties have two medics each, with 44 weekly hours assigned to T&O and 60 to PMR. There is a total of 5 physical resources (medical rooms) available for these specialties, and the Health Institute has required that the total number of new patient examinations does not exceed 30% of total examinations. By running the BO algorithm for only 10 iterations we obtained the outcome described in Table 5. The resulting aggregated queue length is 4,528 patients. This represents a 9.8% reduction from the current situation (estimated average queue length is 5,025 patients), which gives us a very powerful insight: just by redistributing medic hours we are able to reduce the expected queue length for these two specialties. This is a very important result in our study, since the Chilean public health system rarely determines budget increases for the purpose of staff hiring. Being able to improve the current situation with the current resources is already a big step up.

Variable	T&O	PMR
Medic 1 (hrs)	20	30
Medic 2 (hrs)	21	30
New examinations (%)	29	30
Rooms utilized	4	
Average queue length	4,528	

Table 5: BO output for two specialties

From the computational perspective, Figure 4 shows the Gaussian process estimation and the black box values per iteration. We can see that the estimations are not particularly precise, which gives us a sense that 10 iterations are not enough to get a good approximation of the objective function. Therefore, we also performed the same experiment but letting the algorithm run for 100 iterations. Figure 5 shows how the estimations become more precise when the algorithm is ran for more iterations. As for the outcome after the 100 iterations, it is very similar to the ones presented in Table 5, in terms of hours for both medics on both specialties, but with an utilization of only 3 medical resources. The resulting aggregated average queue length is 4,519.

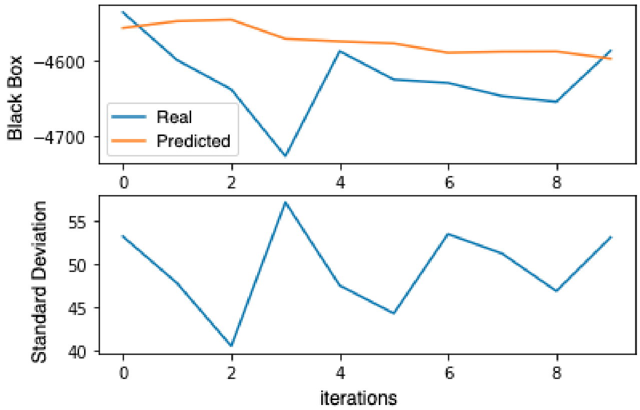

Fig. 4: BO (10 iterations): black box vs GP estimators

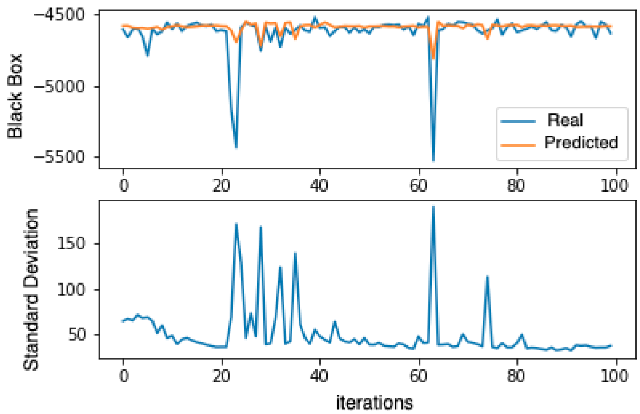

Fig. 5: BO (100 iterations): black box vs GP estimators

6.2 Full Case, Fifteen specialties

The next test involves all 15 available specialties at the hospital. We tested two settings: one which only focuses on redistributing the current medic hours, and the other which considers eventually hiring new medics or extending medic hours. Table 6 shows the current distribution of medic hours, and available medics, per medical specialty. In this setting, the current estimated average queue length is 17,687.

1. Redistribution of current medic hours

   The first numerical experiments in the full case assumes there is no hiring of new medics. Therefore, any improvement on the average queue length comes from redistributing admission and control medic hours.
   The algorithm was run for 100 iterations, and it took around 4 hours and 15 minutes. Analyzing its behavior, Figure 6 suggests that the procedure was

Id.	Specialty	Number of medics	Admission hours	Checkup hours	Total hours	Expected queue length
1	Cardiology	1	6	10	16	625.6
2	General surgery	4	14	7	21	2,017.0
3	Dermatology	2	8	9	17	1,140.0
4	Endocrinology	1	13	13	26	381.8
5	Gastroenterology	3	12	18	30	544.1
6	Gynaecology	3	8	9	17	1,140.7
7	Physical medicine and rehab	2	26	24	50	2,269.9
8	Internal medicine	8	39	64	103	885.2
9	Nephrology	1	6	8	14	71.1
10	Neurology	1	8	8	16	839.6
11	Ophthalmology	4	18	14	32	2,046.2
12	Otorhinolaryngology	3	17	13	30	2,568.5
13	Pediatrics	3	18	25	43	318.1
14	Traumatology and orthopaedics	5	17	19	36	2,041.7
15	Urology	3	7	11	18	797.4
					Total	17,686.97

Table 6: Current medic staff, hours distribution and expected queue length

mainly focused on exploitation with a few attempts of exploring different areas of the feasible region, around iterations 30, 50, 70 and 80. This is further supported by Figure 7 which shows the ℓ_2 distance between each iterate (lighter is closer, darker is further). Around the same iterations we can see dark blue bands, suggesting that the algorithm explored far away from most of the iterates.

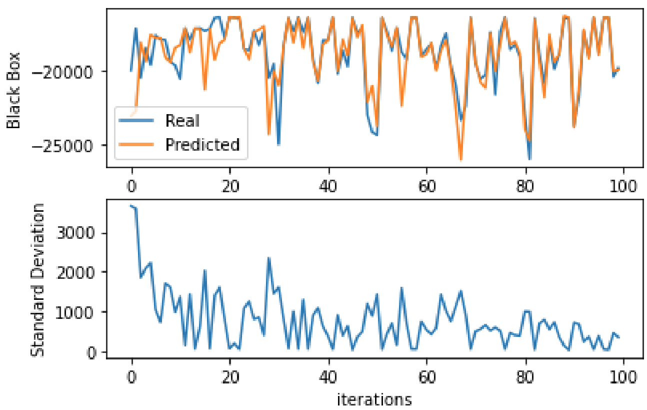

Fig. 6: BO (100 iterations): black box vs GP estimators

The obtained solution is described in Table 7. All specialties, but Cardiology, Gastroenterology and Urology, changed their hours distribution between admissions and checkups. This redistribution leads to a decrease in the average estimated queue length of 8%. In terms of utilization of medical boxes, the

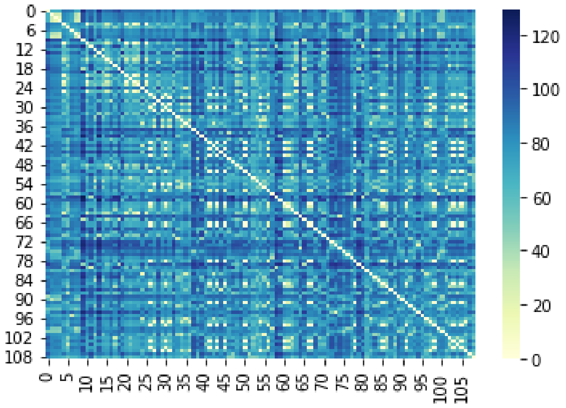

Fig. 7: BO (100 iterations): ℓ_2 distance between iterates

optimal solution suggests that only 11 out of the 20 available should be used, as compared to the 15 medical boxes utilized under the current setting. As for the average queue length per specialty, every single one of them experimented a decrease. The lowest percentage decrease happened in Dermatology (3.9%) while the largest happened in Endocrinology (30%). An interesting observation is that these decreases included those specialties where the hour distribution remained unchanged. This is easily explained because of derivations. For example, most of the referrals from General Surgery are derived to either Cardiology and Gastroenterology (aggregated 65%). Therefore changes in the medic hours assigned to the former will have an impact in the arrival rates of the latter, affecting their corresponding average queue lengths, even though their medic hours did not change.

2. Extra medic hours incorporated

The second experiment on the full set allows for medics to be hired in any specialty. The maximum number of medic hours that can be added to the staff is equivalent to one full medic contract for each specialty, that is, 660 hours. Incorporating this condition into the model described in Appendix A requires adding a new set of variables $y_{e,a}$, corresponding to the number of extra hours allowed in specialty $e \in E$ for activity $a \in A$, and constraint

$$\sum_{a \in A} \sum_{e \in E} y_{e,a} \leq b \tag{7}$$

which ensures the extra medic hours do not exceed the predefined budget b. In this case, the value of b is set to 660 extra medic hours.
Additionally, we want to encourage the network to be stable. We can do so by trying to maximize the throughput for every specialty $e \in E$ and activity $a \in A$, which is defined as given in equation (8).

$$t_{e,a} = \min \left\{ \lambda_e, \left(\sum_{m \in ME_e} \sum_{r \in R} x_{m,a,r} + y_{e,a} \right) \mu_{e,a} \right\} \tag{8}$$

Id.	Specialty	Admission hours	Checkup hours	Expected queue length
1	Cardiology	6 (37%)	10 (63%)	503.2
2	General surgery	8 (38%)	13 (62%)	1,910.2
3	Dermatology	6 (35%)	11 (65%)	1,095.0
4	Endocrinology	10 (40%)	16 (60%)	267.4
5	Gastroenterology	12 (35%)	18 (65%)	442.4
6	Gynaecology	6 (40%)	11 (60%)	1,085.7
7	Physical medicina and rehab	20 (40%)	30 (60%)	2,060.3
8	Internal medicine	41 (36%)	62 (64%)	815.0
9	Nephrology	5 (38%)	9 (37%)	57.5
10	Neurology	6 (38%)	10 (37%)	751.9
11	Ophthalmology	12 (40%)	20(60%)	1,917.6
12	Otorhinolaryngology	12 (40%)	18 (60%)	2,428.0
13	Pediatrics	17 (40%)	26 (60%)	245.4
14	Traumatology and orthopaedics	14 (39%)	22 (61%)	1,956.5
15	Urology	7 (39%)	11 (61%)	753.5
			Total	16,289.6

Table 7: Optimal hours distribution

where λ_e is the effective arrival rate to the corresponding specialty and $\mu_{e,a}$ is the service time for specialty $e \in E$ and activity $a \in A$.
Hence, we add the corresponding variable $t_{e,a}$ and add constraints

$$t_{e,a} \leq (\sum_{m \in ME_e} \sum_{r \in R} x_{m,a,r} + y_{e,a})\mu_{e,a} \quad \forall e \in E, a \in nAE_e \tag{9}$$

and

$$t_{e,a} \leq \lambda_e \quad \forall e \in E, a \in nAE_e. \tag{10}$$

Finally, an objective functions is added to the model used in the *Feasible scheduler module*

$$\max \sum_{a \in nAE_e} \sum_{e \in E} w_{e,a} t_{e,a} - q_{e,a} y_{e,a} \tag{11}$$

which seeks to maximize the *throughput*, with the least number of extra medic hours. The parameters w and q correspond to weights that allowed the model to obtain diverse feasible medic hours distributions, by randomizing their values. Weights q maybe interpreted as the cost invested in hiring new medics.
The algorithm ran for approximately 3 hours and 44 minutes performing 100 iterations. Figures 8 and 9 show that, again, the algorithm was mainly focused in exploiting, although it aimed to explore around iterates 15 and 70. Notice that in those iterations the predicted function value drops to zero, hence encouraging exploration. The reason for these sudden drops in the predicted function lies on the fact that the Gaussian process has two key elements as a priori information: the mean function and the covariance function. The latter is usually set to zero at initialization, since it is supposed to adjust itself along further iterations. It can be interpreted as a "by default" value for the substitute function when not enough information is available. The real values during that exploration showed that it should go back to exploitation.

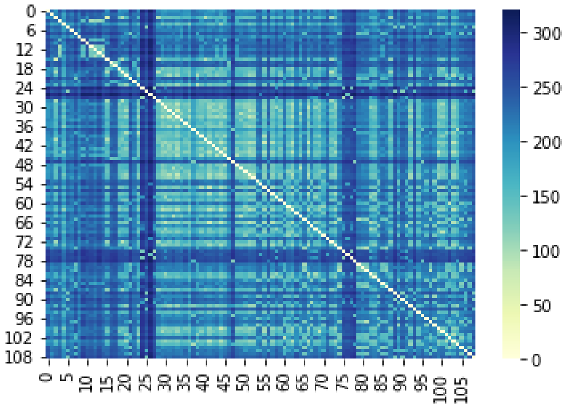

Fig. 8: BO: ℓ_2 distance between iterates - extra hours

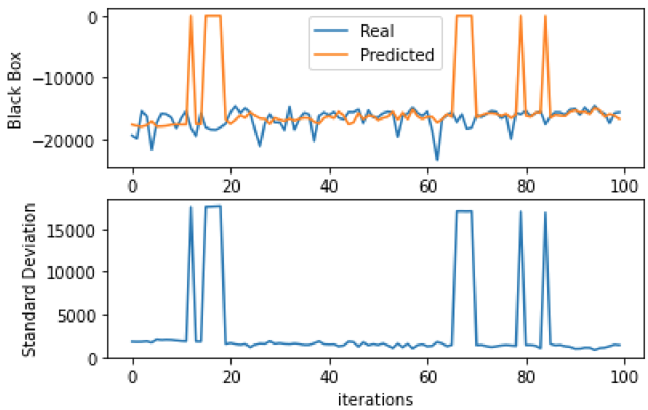

Fig. 9: BO: Black box vs GP estimators - extra hours

The results of this new experiment are summarized in Table 8. Allowing 660 extra medic hours into the system can decrease the total expected queue length in about 24%. Notice that the optimal solution proposes to add extra medic hours only on five specialties: Dermatology, Physical Medicine and Rehab, Internal Medicine, Ophthalmology, and Otorhinolaryngology, at the expense of increased queue lengths in several other specialties like Cardiology, Gastroenterology, and Urology, to name a few. Interestingly, although 103 extra medic admission hours were added to Internal Medicine, its expected queue length also increased. This may be explained due to the increased throughput in Ophthalmology - since 132 extra medic hours are now assigned to this specialty -, together with the fact that approximately 36% of its patients are derived to Internal Medicine. Hence, the extra medic hours for Internal Medicine can be interpreted as a consequence of this output rate increase in Ophthalmology, in order to keep the utilization rate below 1. This is consistent with the result stated in Proposition 1. It is important to mention that this increase in medic hours still left 6 out of the 20 available boxes unused.

Id.	Specialty	Admission hours	Extra Admission hours	Checkup hours	Extra Checkup hours	Expected queue length
1	Cardiology	6	0	10	0	510.8
2	General surgery	8	0	13	0	1,906.3
3	Dermatology	6	78	11	0	897.6
4	Endocrinology	6	0	20	0	857.7
5	Gastroenterology	12	0	18	0	459.2
6	Gynaecology	6	0	11	0	1,098.1
7	Physical medicine and rehab	6	108	44	0	1,227.5
8	Internal medicine	41	103	62	0	843.8
9	Nephrology	6	0	8	0	354.1
10	Neurology	6	0	10	0	748.9
11	Ophthalmology	6	132	26	0	1,222.7
12	Otorhinolaryngology	6	174	24	65	13.7
13	Pediatrics	17	0	26	0	249.0
14	Traumatology and orthopaedics	14	0	22	0	1,935.1
15	Urology	6	0	12	0	1,210.0
					Total	13,534.6

Table 8: Optimal hours distribution with hiring allowed

7 Conclusions and discussion

In this paper we have presented a modular methodology for the problem of scheduling healthcare professionals for outpatient healthcare in a hospital. The methodology is developed in partnership with the South-Eastern Metropolitan Health Service (SEMHS), a regional division of the Chilean national healthcare ministry, and we give computational experiments on real data from the SEMHS. In the following we give further information and discussion.

Methodology. We present a modular methodology consisting of four relatively simple and easily modifiable modules; see Figure 2. Apart from the Bayesian optimization module, the other ones are relatively simple to interpret, modify, and adapt to the special needs and constraints of a hospital, and can be implemented independently of the other modules to provide partial solutions.

The Forecast module forecasts the monthly exogenous arrival of patients to each specialty and considers four forecasting methods: time series with single, double and triple exponential smoothing, and the XGBoost method. The Feasible schedule module outputs a schedule that satisfies the operational and institutional constraints of the system. Some salient features of this module are that we can impose a condition that guarantees that the patient's queues remain stable in the long term, and that we can minimize the number of extra medical hours needed to stabilise the queues. The latter can be useful to determine the specialists' gap in each specialty. The Simulation module uses discrete event simulation to simulate the performance of the system over the planning horizon, and it can include key features of the system's dynamics, such as patient or doctor no-shows, service policies, and so on. Lastly, the Bayesian optimization module ties the previous modules and iteratively proposes and simulates a feasible schedule, automatically learning which have better performance.

From a theoretical perspective, this methodology is the first to use Bayesian optimization for a scheduling problem in general, and in particular for a scheduling problem in healthcare management.

Practitioner's perspective. From a practical point of view, our methodology can be seen more as a toolkit than a single tool, where each module help provide partial managerial solutions and can be implemented independently. For instance, it can help estimate the patient's demand, determine if there are enough physical resources to carry out a given schedule, estimate the queue lengths and other KPI's of a particular schedule, and so on. This can help in the gradual adoption of the methodology in settings where there is low technological penetration, such as the public healthcare sector in Chile, where we have applied our algorithm.

Our methodology can also be a useful decision making tool for staffing and training at a hospital, health service or at the national level. Indeed, the doctor's "flexibility" between specialties and the queue stabilization feature of our method are specially relevant in the long-term planning horizon, and can help design staffing and training policies.

We claim that it is feasible to implement the model in all public hospitals in Chile, as the data required to run our algorithm is regularly gathered and reported to the Ministry of Health by all public hospitals in Chile. Particularly, referrals can be obtained from the national repository of waiting lists, and medical hours for admissions and checkups are defined in quarterly hospital programs.

Collaboration with South-Eastern Metropolitan Health Service (SEMHS). This paper is the result of two years of collaboration with the SEMHS (a regional division of the national health ministry in Chile) to tackle a relevant healthcare problem using Operations Research tools. The team established a close collaboration relationship with SEMHS and hospital planners, helping them with data gathering and cleansing tasks, and participating in the implementation of their current hospital planning platform. This helped us understand in depth the outpatient process, the data and the hospital planning process that we aim to improve.

Computational experiments and insights. We present computational experiments on data from a small-to-medium public hospital in Chile managed by the SEMHS. We run the experiments for the planning horizon of January 2019 to October 2019, as that was the last period with relatively normal operation conditions, i.e., before the "social eruption" period in Chile that started on October 18th 2019, and before the COVID-19 pandemic began.

The first experiment considers only two specialties with the highest derivation rates between them, according to SEMHS data: *Traumatology and Orthopaedics* and *Physical Medicine and Rehabilitation.* The second experiment considers the full set of specialties available at the hospital, with no hiring of extra medics allowed. Our experiments showed that this reduction can go up to 8%, and it reduces the expected queue length on all specialties individually. This is a remarkable achievement because it reveals that the currently available medic hours can be redistributed in order to reduce the total expected queue length. The third experiment is an extension of the second experiment, now allowing up to 660 extra medic hours to be hired, equivalent to one full medic per specialty. The obtained solution required assigning these extra hours to only five specialties, resulting in a 25% improvement over the current scenario.

Open questions and future research. An interesting open question is to derive a sufficient stability constraint that is tighter than Proposition 1. Indeed, the latter result bounds the stochastic behavior of the system by merging both activities (Admissions and Checkups) of a specialty into a single queue, and using the service rate of the slowest of the two activities. This can be a loose bound, and a tighter one may expand the pool of feasible solutions provided by the Feasible schedule. Nonetheless, a more heuristic approach to increase the space of feasible solutions is to loosen the traffic intensity inequality (2), say from traffic intensity less or equal than 1, to 1.5 or 2.5.

Another open question is how to reduce the dimension of the input of the objective function of the Bayesian optimization, as it is known that the latter works best with lower dimensions. We explored using singular value decomposition (SVD) and working on a lower-dimensional space, however it was not clear that it provided better solutions.

Another interesting question is to explore other objective functions to minimize, especially more conservative measures of the performance of the system. Indeed, here we have tackled the expected total number of patients in queue at the end of the planning horizon, however the expected value is a risk-neutral measure. Hence, one can explore more conservative measures, such as the probability that the total number of patients in queue exceeds a certain value, or the expected exceedance above a certain value. Other options are to explore optimizing the expected length of stay (LOS) of a patient, however through Little's Law it could be connected to a weighted sum of queue lengths.

Appendix

A Feasible schedule optimization problem

The formulation of the Feasible Schedule optimization problem is as follows. We let M, A, R, and E correspond to the sets of medics, activities, physical resources, and specialties, indexed by their corresponding lower-case letters, m, a, r, and e, respectively. We let $x_{m,a,r}$ be the weekly amount of hours that medic $m \in M$ is assigned to perform activity $a \in A$ in resource $r \in R$. We can now state the constraints of our problem.

Each resource $r \in R$ has a certain amount of available hours within a week. This limit is represented by the parameter O_r.

$$\sum_{m \in M} \sum_{a \in A} x_{m,a,r} \leq O_r \quad \forall r \in R \tag{12}$$

Every medic $m \in M$ has a contract with a predetermined maximum amount of hours, C_m, they must work.

$$\sum_{a \in A} \sum_{r \in R} x_{m,a,r} \leq C_m \quad \forall m \in M \tag{13}$$

Some activities $a \in A$ have a minimum amount of hours h_a that need to be satisfied.

$$\sum_{m\in M}\sum_{r\in R} x_{m,a,r} \geq h_a \quad \forall a \in A \tag{14}$$

Some medic-specialty tuples $(m, e) \in M \times E$ cannot perform some specific activites $a \in nAE_e$, since they are not allowed by specialty $e \in E$. The set nAE_e indexes all $a \in A$ that are unavailable for specialty $e \in E$. The set ME_e indexes all medics $m \in M$ who do not belong to specialty $e \in E$.

$$\sum_{m\in ME_e}\sum_{a\in nAE_e}\sum_{r\in R} x_{m,a,r} = 0 \quad \forall e \in E \tag{15}$$

Some activities $a \in A$ related to specialty $e \in E$ are not allowed to be performed in resource $r \in R$. The set nAR_r gathers these activities $a \in A$ for every resource $r \in R$.

$$\sum_{m\in ME_e}\sum_{a\in nAR_r} x_{m,a,r} = 0 \quad \forall r \in R, e \in E \tag{16}$$

The ratio between the number of activities labeled as *admission* and the total number of activities must not exceed a certain parameter value $p \in (0, 1)$. The set CN_e contains all activities $a \in A$ labeled as *admission* in specialty $e \in E$. The ratio parameter $p \in \{0, 1\}$ is usually imposed by the Chilean Health Department.

$$\sum_{m\in ME_e}\sum_{a\in CN_e}\sum_{r\in R} x_{m,a,r} - p \sum_{m\in ME_e}\sum_{a\in A}\sum_{r\in R} x_{m,a,r} \leq 0 \quad \forall e \in E \tag{17}$$

Finally, all variables $x_{m,a,r}$ must be non-negative.

$$x_{m,a,r} \geq 0 \quad \forall m, a, r \tag{18}$$

Acknowledgements Guido Lagos acknowledges the financial support of ANID FONDECYT grant 3180767, and of CORFO InnovaChile project 14ENI2-26865. The authors are grateful to Luz María Durango for support at the CRS Hospital, and Mariola Mestanza and Juan Cristóbal Morales for overall support at SEMHS.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Abo-Hamad, W., Arisha, A.: Simulation-based framework to improve patient experience in an emergency department. European Journal of Operational Research **224**(1), 154–166 (2013)
2. Andradóttir, S.: A review of random search methods. Handbook of Simulation Optimization pp. 277–292 (2015)
3. Ayers, D.J., Menachemi, N., Ramamonjiarivelo, Z., Matthews, M., Brooks, R.G.: Adoption of electronic medical records: the role of network effects. Journal of Product & Brand Management (2009)
4. Barnes, S., Golden, B., Price, S.: Applications of agent-based modeling and simulation to healthcare operations management. In: Handbook of healthcare operations management, pp. 45–74. Springer (2013)
5. Burke, E.K., De Causmaecker, P., Berghe, G.V., Van Landeghem, H.: The state of the art of nurse rostering. Journal of scheduling **7**(6), 441–499 (2004)
6. Cardoen, B., Demeulemeester, E., Beliën, J.: Operating room planning and scheduling: A literature review. European journal of operational research **201**(3), 921–932 (2010)
7. Charniak, E.: Bayesian networks without tears. AI magazine **12**(4), 50–50 (1991)

8. Chen, T., Guestrin, C.: Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785–794 (2016)
9. Duvenaud, D.: Automatic model construction with gaussian processes. Ph.D. thesis, University of Cambridge (2014)
10. Fetter, R.B., Thompson, J.D.: The simulation of hospital systems. Operations research **13**(5), 689–711 (1965)
11. Frazier, P.I.: Bayesian optimization. In: Recent Advances in Optimization and Modeling of Contemporary Problems, pp. 255–278. INFORMS (2018)
12. Friesen, M.R., McLeod, R.D.: A survey of agent-based modeling of hospital environments. IEEE Access **2**, 227–233 (2014)
13. Fu, M.C.: Handbook of simulation optimization. Springer (2015)
14. Ghanes, K., Wargon, M., Jouini, O., Jemai, Z., Diakogiannis, A., Hellmann, R., Thomas, V., Koole, G.: Simulation-based optimization of staffing levels in an emergency department. Simulation **91**(10), 942–953 (2015)
15. Hall, R.W.: Handbook of Healthcare System Scheduling, vol. 168. Springer Science & Business Media (2012)
16. Hu, J.: Model-based stochastic search methods. Handbook of Simulation optimization pp. 319–340 (2015)
17. Huang, D., Allen, T.T., Notz, W.I., Zeng, N.: Global optimization of stochastic black-box systems via sequential kriging meta-models. Journal of Global Optimization **34**(3), 441–466 (2006)
18. Jun, J.B., Jacobson, S.H., Swisher, J.R.: Application of discrete-event simulation in health care clinics: A survey. Journal of the Operational Research Society **50**(2), 109–123 (1999)
19. Kleijnen, J.P.: Design and analysis of simulation experiments. In: International Workshop on Simulation, pp. 3–22. Springer (2015)
20. Kleijnen, J.P.: Response surface methodology. In: Handbook of Simulation Optimization, pp. 81–104. Springer (2015)
21. Kushner, H.J.: A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise. Journal of Basic Engineering **86**(1), 97–106 (1964)
22. Li, J., Aickelin, U.: Boa for nurse scheduling. In: Scalable Optimization via Probabilistic Modeling, pp. 315–332. Springer (2006)
23. Martinez, D.A., Zhang, H., Bastias, M., Feijoo, F., Hinson, J., Martinez, R., Dunstan, J., Levin, S., Prieto, D.: Prolonged wait time is associated with increased mortality for chilean waiting list patients with non-prioritized conditions. BMC public health **19**(1), 1–11 (2019)
24. Meskó, B., Drobni, Z., Bényei, É., Gergely, B., Győrffy, Z.: Digital health is a cultural transformation of traditional healthcare. Mhealth **3** (2017)
25. Minsal: Lista de espera no ges, garantías de oportunidad ges retrasadas y análisis epidemiológico y causas de muerte de personas fallecidas estando en lista de espera no ges (2020)
26. Muhuri, P.K., Biswas, S.K.: Bayesian optimization algorithm for multi-objective scheduling of time and precedence constrained tasks in heterogeneous multiprocessor systems. Applied Soft Computing **92**, 106274 (2020)
27. Mustafee, N.: Operational Research for Emergency Planning in Healthcare: Volume 1. Springer (2016)
28. Mustafee, N.: Operational Research for Emergency Planning in Healthcare: Volume 2. Springer (2016)
29. Nahmias, S., Olsen, T.L.: Production and operations analysis. Waveland Press (2015)
30. Rais, A., Viana, A.: Operations research in healthcare: a survey. International transactions in operational research **18**(1), 1–31 (2011)
31. Shahriari, B., Swersky, K., Wang, Z., Adams, R.P., De Freitas, N.: Taking the human out of the loop: A review of bayesian optimization. Proceedings of the IEEE **104**(1), 148–175 (2015)
32. Sigman, K.: The stability of open queueing networks. Stochastic Processes and their Applications **35**(1), 11–25 (1990)
33. SSMSO: Presidenta inaugura crs hospital provincia cordillera (2017)
34. Thompson, J.D., Fetter, R.B.: The economics of the maternity service. The Yale journal of biology and medicine **36**(1), 91 (1963)
35. Williams, C.K., Rasmussen, C.E.: Gaussian Processes for Machine Learning, vol. 2. MIT press Cambridge, MA (2006)

36. Winston, W.L., Goldberg, J.B.: Operations research: applications and algorithms, vol. 3. Thomson Brooks/Cole Belmont (2004)
37. Wolfe, H., Young, J.P.: Staffing the nursing unit: Part i. controlled variable staffing. Nursing Research **14**(3), 236–242 (1965)
38. Wolfe, H., Young, J.P.: Staffing the nursing unit part ii. the multiple assignment technique. Nursing Research **14**(4), 299–303 (1965)
39. Yeh, J.Y., Lin, W.S.: Using simulation technique and genetic algorithm to improve the quality care of a hospital emergency department. Expert Systems with Applications **32**(4), 1073–1083 (2007)