IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received June 24, 2021, accepted July 15, 2021, date of publication July 20, 2021, date of current version July 27, 2021.
Digital Object Identifier 10.1109/ACCESS.2021.3098715

Modeling and Validation of the Hospital's Ambulatory and Inpatients Operations Using a Non-Homogenous Discrete Time Markovian Chains

KAOUTER KARBOUB[1,2,3], MOHAMED TABAA[4], (Member, IEEE), SOFIENE DELLAGI[3], FABRICE MONTEIRO[3], (Member, IEEE), ABBAS DANDACHE[3], AND FOUAD MOUTAOUAKKIL[2]
[1]FRDISI, Hassan II University of Casablanca, Casablanca 20000, Morocco
[2]LRI-EAS, ENSEM, Hassan II University of Casablanca, Casablanca 20000, Morocco
[3]LGIPM, University of Lorraine, 57000 Metz, France
[4]LPRI, EMSI Casablanca 20250, Morocco

Corresponding author: Mohamed Tabaa (m.tabaa@emsi.ma)

ABSTRACT Overcrowded hospitals in its different process's levels is a very common issue all over the globe. Research through different works have studied this problem from different angles. Most of those focused on a very specific part of the internal flows. This study presents a mathematical model aiming to reflect patient flow and resources usage. This model presents a new perspective on how it is possible to organize a hospital flow for inpatients of different pathologies. It is based on Non-Homogenous Discrete Time Markovian chains. It takes into consideration the patient's pathology, survival function and the current beds distribution and discharges. To model the different variations of a patient's flow, our work uses an actual case study to validate that our mathematical model based on Markov Chains can represent a real scenario of patients flow in a hospital and can predict the resources occupancy. We used as entry to this model the arrival distribution of patients as represented in the NHS dataset. Served patients and overflowing patients relate proportionally to admission rate based on how much time every diagnosis and preliminary treatment will take, then using the previous beds availability ration will permit, with knowledge of all present patients in the system, calculate the beds occupancy in the actual time. Only based on inputs from the NHS database, bed occupancy can be calculated using our model. Then compare this calculus with published data of the bed's occupancy in the NHS. in the period between March 2020 and February 2021. The results of the simulation model were then compared to the dataset using the chi-square goodness of fit test. Non-homogenous discrete Markovian model, survival function, patient's pathology, time-dependent, pathology, primary allocations, secondary allocations.

INDEX TERMS Non-homogenous discrete Markovian model, survival function, patient's pathology, time-dependent, pathology, primary allocations, secondary allocations.

I. INTRODUCTION

The public health determines, in accordance with existing constraints, the overall vision that aim to diseases prevention, decreasing mortality rates and increasing the well-being and services given by healthcare institutions to patients. However, according to the European Action Plan for Strengthening Public Health Capacities and Services in 2012-2020 [1], many challenges must be taken into consideration like ageing population, increasing levels of chronic diseases and the tremendous demographic growth. While in North Africa and Middle East, the access to healthcare services, which are concentrated in urban areas due to marginal infrastructures in rural areas, particularly makes the management of available resources very difficult as confirmed by the health finance and governance report [2]. Furthermore, and according to an article published November 2019 [3], England has recorded the worst ever levels of waiting times and key targets for

The associate editor coordinating the review of this manuscript and approving it for publication was Emanuele Crisostomi.

service quality satisfaction, treatment and waiting times have been missed for almost three years, this according to local facilities is due to huge demands with a recorded 4.42 million patients on the waiting line with only 83.6% of accident and emergency patients admitted within 4 hours. In the monthly published reports of the National Health Services [4], it is reported a rate of 4.9% as a mortality rate among ambulatory patients back in 2019. On a recent study published by the Economic injury [5], increased waiting times does not only have an impact on patient's condition but also on the hospital's cost to care those patients with an average of 6%. While in United States and according to the National Center for Health Statistics [6] a total number of 2 839 205 resident deaths were recorded in 2018. 73.8% of those deaths were referred to 10 leading causes like heart diseases, cancer, unintentional injuries, influenza & pneumonia, suicide, and kidney diseases. This along with extreme shortage of qualified physicians in rural areas as well as big cities-40 physicians per 100000 in rural areas and 53 per 100000 in cities- [7]. In Australia waiting times for elective surgery and emergency department care are increasing according to the Australian Institute of Health and Welfare report in December 2018 [8]. The same report announced that elective surgery waiting times have increased from 36 days back in 2013 to 40 days in 2017 and 2018. While the number of presentations to the public health institution's emergency departments reached 8 million in 2017 and 2018 –which is an average of 22000 presentations daily- 72% of them were seen by a professional on time. While 52% of those patients were assigned to three most seen categories (resuscitation, emergency and urgent) due to injury, poisoning and other external causes. The report also confirmed that these distributions depend strongly on the state and territory and available infrastructures. As recorded by the Australian government [9] there were 160 909 deaths in 2017, 46% of them are potentially avoidable if they would have received medical care on time and other pattern of death causes were encoded correctly. In 2019, Canada also reported [10] a median waiting time of approximately 21 weeks for surgical and other therapeutic treatments, while the orthopedic surgery, plastic surgery, ophthalmology, and neurosurgery seem to have the longest waiting times with 10 weeks as a maximum variation. In 2020, the United Nations-World Population Prospects published a ranking by deaths rates report [11], where Germany, Greece, Japan, France, Canada, and Singapore have reported 11.392%, 11.035%, 10.865%, 9.365%, 7.803%, and 4.752% as mortality rates, respectively. In fact, based on cohort studies and experimental works, research started looking for a determinant relationship between patients not admitted or waits too long for surgery or in the ED and their risks for eventual complications that can even lead to death.

In Ontario, Canada back in 2011 [12], a population-based cohort study included 21 925 275 visits recorded in the years between 2003 and 2007, 1487 094 were emergent patients and in the rest (more than 20 000 000) 87% were left without being seen or seen and discharged. Based on logistic regression models to statistically analyze the collected data, it was clearly shown that the risk of death notably increases with each additional hour of mean waiting times. Another study [13] focused on the emergency medical services response time and its potential causality relationship with mortality rate in case of life-threatening events. Based on a retrospective cohort study of adults in danger situations for a period of one year, the study aimed to define whether a response time of 8 minutes can be a threshold for associated mortalities. The response time was calculated from the moment the 911 call was received in a medical priority dispatch system to the moment where advanced life support team is on scene. Unsurprisingly, 7.1% patients with a response time >8 minutes died comparatively with 6.4% patients with a response time <8 minutes. In the Chilean community, a team of experts had let a cohort study [14] including patients with non-prioritized conditions while registration in hospitals. The data included a waiting list of 987497 patients in 77 healthcare institutions from 2008 to 2015. The study showed that despite geographical disparities and health centers differences, waiting time variability have significant impact on death rate. In the other hand, as the efforts to decrease the length of stay are getting bigger, the concerns of rising the readmission rates are also worrying. Premature discharge can be one of leading causes to readmission as shown by many studies based on cohort and longitudinal approaches [15], [16]. Recently, the coronavirus outbreak first appeared in Wuhan China on the 31st of December 2019 had made all healthcare institutions and administrations under scope and huge pressure. In 16 mars 2020, the World Health Organization [17] reported a total number of 167515 confirmed cases 81077 of them recorded in China 86438 outside China in more than 100 country all over the world. 6606 as total reported deaths, 48% of them in China.

Hence, a problem arises contain different goals to aim that may be contradictory and can cause if not taken from different aspects multiple penalties. First, decrease waiting times so the consequences of such factor on patient's condition and on hospital processes costs can be reduced. Second, Discharge the right patients in the right time, to avoid increasing the readmission rates, and at the same time, free enough beds for admitted patients.

The objective of the current paper is to provide hospital management system with a tested generic mathematical model with managerial insights that may serve as reference for future analytical studies that may concern decision making, statistical analyzing or operational management. Our approach uses a non-homogenous discrete time Markov chain and queuing theory properties. The core of our approach is based on dividing the treatment process into three main levels: the first level which is preliminary diagnosis then the second: treatment process (prescriptions, laboratory analysis, physical monitoring, inpatient's further diagnosis) and then in the third-place discharge. The main objective in every level is to theoretically decide the number of served patients and overflowing patients and the hospital's capacity to serve all patients. The paper is organized in three main sections; the first presents related works and techniques used in healthcare processes management and mathematical modeling. And the second depicts our proposed model with all used concepts and recommendations. While the third section includes a case study based on NHS database to validate our model. The paper is then ended with a conclusion and future insights.

II. RELATED WORKS

Modelling is a very recurrent topic back in the last century with a very bi-major orientations. First, establishing a scheme-oriented design for future healthcare facilities [18], [19] to decide the number of required beds, human resources, and medical furniture's with studying the future possibilities of augmentation and pic demands. Those studies are generally based on an agreed balance between environmental characteristics and annual targets. Second, readjusting resources usage to maximize performances of pre-existing hospitals [20], [21]. In achieving the second goal two methodological strategies are used: 1) methods to model the system in consideration and 2) methods to study and validate those models.

In general, Hospitals operations management system addresses decision making to continuously establish the most efficient resources policies to maximize its capacity and in the same time guarantee services quality.

From its first appearance in the 18th century with AK Erlang to define needed resources to meet acceptable phone service [22], queuing theory was used in many studies and works aiming in most cases to deal with delays caused by the differences between a given service demand and the ability of a system to respond to this demand.

In healthcare applications, queuing theory is based generally on three main factors; the patient's arrival rate which reflects the unplanned arrival frequency of a patient at a given time, available resources called servers and the services rate which reflects the time taken for a patient to be seen, diagnosed, and treated [23]. In the university of Tsukuba hospital in Japan, a conducted study [24], targeted obstetric patient flows in low and high-risk delivery wards. The study used hospital internal logistic data of two years. Using an (Markovian patient's arrival distribution and finite resources) M/M/m and (Markovian distribution of patient's arrival, general resources usage, infinite resources) $M/G/\infty$ to reflect the probability distribution of total patients in each ward, the study confirmed the Little's law of queuing theory. Two years later, in 2018, [25] used different queuing theory models' variations to determine operation performances like beds utilization rate, waiting times and demand behavior assessment. The model was based on six months data recorded in university hospital and concluded as results that the balance between admissions and type of pathologies can be a possible way to optimize resources usage. In the same year, a team management and chief medical officers in Lehigh Valley Health Network [26] confirmed the value of queuing model in theory as it can statistically provide alignment and focus on future insights. They also assumed the very important role of medical staff to implement such guidance in operational state.

From the other side, the appearance of Markov chains in 1906 by Andrei Andreevich Markov [27], [28] has allowed probabilistic modelling using stochastic matrices. Although, in healthcare applications, very few models were found to have focus on patient's flow using Markovian mathematical modelling as a tool. Main studies have used Markov chains in admission scheduling and resources planning by keeping track of patient pathways. [29] Considered a data driven study in London hospital based on Markov chain modelling to design an appropriate admission policy in accordance with future resources usage. The study demonstrated non-homogenous Markov chain can be an effective decision support tool for care planners and policy makers. An update of the same study came out one year later, to correct the use of a continuous time Markov chain [30]. The study considers patient's arrivals rates variations. The authors based their novel approximation on a 16-year period as a historic data from the same hospital as previous study.

III. DYNAMICS OF THE PROPOSED MODEL

A. PROBLEM DESCRIPTION

In its report [31], last updated in 2020, the Agency for Healthcare Research and Quality published a guide for hospitals to improve patient flow and reduce particularly the emergency department crowding. The report mentioned nearly half of hospitals in the US operating at or above their capacity and approximately 500 000 ambulances obliged to drive away from near hospitals because of crowding. It also reported several performances and quality measurements that most hospitals must take and update at a regular basis.

In any healthcare facility or institution, the performance of the first contact with medical care providers is very crucial.

In this study, a generic mathematical modelling where an organizational structure can iteratively change in accordance with the number of arriving patients and available resources was considered. The distribution of resources like beds, doctors and equipment might change in accordance with a discrete time nonhomogeneous Markov chain. Such is a very critical task to do especially in natural catastrophes and pandemics like Coronavirus outbreak. Even though minor or greater changes in the resource's distribution might take place daily like patient's allocation, which must be decided considering the patient's condition, survival function-reflecting a patient's chances to survive within a specified period of time- and nature of treatments. Furthermore, this decision, if considered as the best in each moment must be re-evaluated following the patient's situation updates and other arrivals. From one side, a patient's arrival to healthcare facilities occupies medical resources which can endure other patient's length of stay. From the other side, in case of admission, the patient's hospitalization is either primary or secondary. This means, that a patient is allocated to a preferred bed, and this can happen if a bed is already available, or another inpatient have been allocated to other wards. In case of secondary hospitalizations, the patient's is faced with a shortage of beds in preferred ward and can be allocated to an alternative bed waiting for adequate beds to be free. In both cases, patients in critical situations like those contaminated with the coronavirus need special care. The system under scope is a three phased or leveled system. It contains preliminary diagnosis and treatments, deeper diagnosis and treatments and a discharge phase. In the second level, the system contains different wards of different specialties and pathologies treated, surgical blocks, scans rooms and general diagnosis rooms, etc.

B. MATHEMATICAL MODELING OF THE HEALTHCARE SYSTEM

Taking all the above considerations into account, the main goal of this study is to present to crowded hospitals an efficient mathematical model. Figure 1 shows the different patient's flow considered in our model.

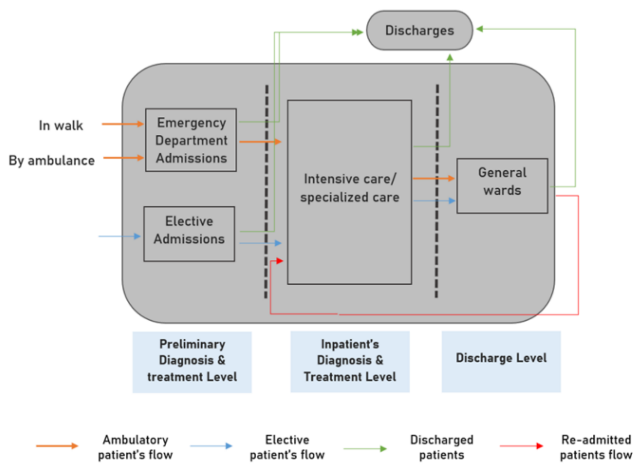


FIGURE 1. Flow chart of different possible pathways of a patient in hospital.

Therefore, it is very important to be able to model and statically quantify the performance of every new arriving patient's pathway. In the following, and as shown in figure 1, we make a distinction between elective and ambulatory patients in the first level.

As it has been proven by many works, studies and reports, the patient's arrival distribution to emergency department may be different depending on the season, the day of the week, the geographical structure and demographic patterns in the hospital's area [32], [33]. It also, according to a 2016 overview of the healthcare cost and utilization project, depends on the revenue, age, sex and patient's residence location [34]. For this reason, Time was randomly divided into different portions (that can be a day or a week depending on a hospital crowding frequency) into a D non-equilibrium time intervals denoted = {I1, I2 …, ID}. We also define L = {Level1, Level2, Level3, Level4, Level5} which reflect the degree of severity of a patient arriving in emergency department. The previous classification was made based on the saint Mary's regional medical center classification [35]. The five levels can be, Level1: resuscitation of life saving intervention, Level2: emergency, Level3: urgent, Level4: semi-urgent, level5: non-urgent. It is the determination of the severity's level that orients a patient's next interventions steps.

In the other side, the arrival of scheduled patients is a non-static parametric distribution as there is many reasons that may change or cause tardiness of a patient's decision to go to a hospital. We categorize these factors in three main blocks: financial, natural, and cultural [36]. Let us consider the following:

$P_{i,l}(I_d) \in \mathbb{N}$: Patient i arrival with a degree of severity l into the ED in d^{th} interval.

M (1, …, k, …, M) represents the number of pathologies or specialties that can be treated in a hospital.

$E_{j,k,d} \in \mathbb{N}$: The variable reflecting the j^{th} patient appointment with pathology k in the d^{th} interval.

$Pr(E_{j,k,d}(t))$: The probability that patient of pathology k who took an appointment j will show up at t in the d^{th} interval.

$1\text{-}Pr(E_{j,k,d}(t))$: the probability of patient's absence or tardiness.

$\sum_{i\in N}\sum_{l\in\{1\ldots5\}} P_{i,l}(I_d)$: The number of patients arriving in ED with different severity levels during the time interval I_d.

$$N_{i,l}(.) = P_{i,l}(I_d) + \tilde{P}_{i,l}(I_{d-\Delta d}); \quad \forall i \in \mathbb{N},\ l \in 1,\ldots,5$$

$N_{i,l}(.)$ represents the vector representing the number of patients in the ED waiting to be served as a sum of newly arriving patients and others from previous I_d intervals.

Let us also define $\sum\sum \chi_{i,l}$ as the total number of service completions and $\sum\sum \tilde{\mathcal{X}}_{i,l}$ the sum of patients of different levels currently in service.

In the same way we denote $\lambda_{j,k}(I_d)$ as the arrival rate of patients of type k arriving as scheduled in the d^{th} interval and $\mu_{j,k}(I_{d-\Delta d})$ as service rate of patients arriving in previous intervals and still not served.

$\sum\sum Y_{j,k}$: the number of elective patients currently in service.

To model the resources consumption, we use a concept called typical treatment process. It defines the consumption amount of a pathology k on resources.

$\mathcal{R} = \{r_1, \ldots r_k, \ldots, r_M\}$: where $r_k = \frac{\text{occupied resources in department k}}{\textit{all available resources in k}}$ is the rate of a resource consumption by patients with k pathology.

The following $PS^k_{i,j} \in \mathbb{R}^{k\times k}$ matrix table represents the output from the first level, which is preliminary diagnosis and treatment, composed of; first, served and in service ambulatory and elective patients and second, waiting patients to be seen or preliminary diagnosed. The r_k is the ratio equal to used resources (beds) in a specified ward divided by total number of beds in that ward. Which means that it is always < 1. The sequence L1->L2 represents scores from

1 to 5 given to admitted patients. multiplied by these scores, overflowing patients are in most cases those with scores 1 or 2 (less urgent).

$$PS^k_{i,j}: \begin{cases} \mathcal{X}_{i,l} + N_{i,l} \\ \quad if\ r_k < 1\ with\ l \in (L1, L2, L3, L4, L5) \\ \left[\lambda_{j,k}(I_d) + \mu_{j,k}(I_{d-\Delta d}) + Y_{j,k}\right] * (1 - r_k) \\ \quad if\ r_k < 1 \mathcal{O}_d = \Big(\mu_{j,k}(I_d) * P_{i,l}(I_d) + \tilde{P}_{i,l}(I_d) \\ +(1 - \Pr(E_{j,k,d}(t))) * \mathrm{E_{j,k,d}}\Big) \\ \quad if\ r_k = 1 \end{cases}$$

where $\mathcal{O}$ represents the number of overflowing patients from time interval I_d.

This overflow is calculated considering patients in hospital and still not served and scheduled patients not arriving in time. In our model, we assume that all causes leading to overflow can be headed back to very long diagnosis times.

Let $T^k_{i,j}$ be the service time necessary to preliminary diagnosis for a patient in ED or elective patient of type k and $|I_d|$ be the length of time interval I_d. We assume that:

$$\sum_k \sum_{i,j}^{\mathbb{N}} \left(T^k_{i,j}\right) * PS^k_{i,j} > |I_d|$$

Now taking consideration of patients served and admitted to be inpatients, Let us consider $W = M$ the number of wards and $U_{q,k}(I_d) = \frac{number\ of\ patient\ of\ pathology\ k\ allocated\ to\ ward\ q}{total\ number\ of\ beds\ in\ ward\ q}$ the bed occupancy function reflecting if a ward q is occupied by a patient of pathology or condition type k. ϖ_q the number of beds in ward q. Of course, by this, we are assuming that every admitted patient occupies one bed.

Also denote $F_q(I_d) = \varpi_q - \sum_k U_{q,k}(I_d)$ the number of free beds in ward q. The $D_{q,k}$ reflects beds distribution in the time interval $\mathrm{I_d}$.

$$D_{q,k}(I_d) = \left(\begin{pmatrix} U_{11(I_d)} & \cdots & U_{1M}(I_d) \\ \vdots & \ddots & \vdots \\ U_{M1}(I_d) & \cdots & U_{MM}(I_d) \end{pmatrix}, \right. \\ \left. \times\, (F_1(I_d), F_2(I_d), \ldots, F_M(I_d))\right)$$

where $(U_{11}(I_d), U_{22}(I_d), \ldots, U_{MM}(I_d)|k = q)$ the number of patients with primary hospitalizations in interval Id.

$\alpha_k \equiv \alpha_{q,k} = \frac{\sum_{q=1}^{W}\sum_{k=1}^{M} U_{q,k}(I_d)}{\sum_{q=1}^{W}\sum_{k=1}^{M} D_{q,k}(I_d)}$: The rate of primary hospitalized patients.

$\beta_{q,k} = \frac{\sum D_{q,k}(I_d) - \sum_{q=1}^{W}\sum_{k=1}^{M} U_{q,k}(I_d) - \sum_{q=1}^{W} F_q(I_d)}{\sum_{q=1}^{W}\sum_{k=1}^{M} D_{q,k}(I_d)}$: The rate of secondary hospitalization.

The distribution of the newly arriving patient to the different ward follows the below shown process.

$$\begin{cases} PS^k_{i,j} * \alpha_k & if\ F_k > 0\ for\ k \in M \\ PS^k_{i,j} * \beta_{q,k} & if\ F_k = 0\ and\ F_{q\neq k} > 0\ for\ q \in M \end{cases}$$

In matrix above, a newly arriving patients is allocated to preferred ward in case there exist free beds. Otherwise, the patient is oriented to another ward until preferred ward is freed.

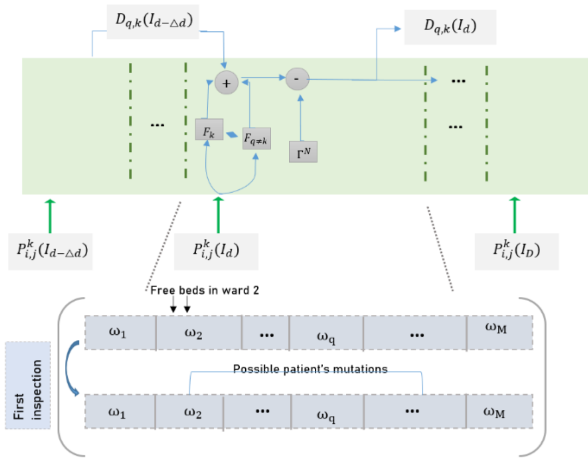


FIGURE 2. Beds distribution in function of arriving patients, internal distribution and discharge policy.

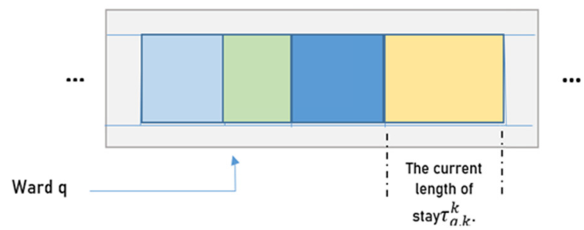


FIGURE 3. Variations in length of stay in each ward.

A patient's mutation, as shown in figure 2, can be in between two secondary hospitalizations or from a secondary hospitalization to primary one. In all cases, the patient's condition, and ability to be in a different ward must be considered. In our Markovian model, the beds distribution is beds distribution in past state plus current admitted patients minus discharges.

In the following, the time spent by a patients of type k as an inpatient in ward q denoted $\tau^k_{q,k}$. Let also denote N_d as the total number of inspections in an interval d.

We define $Q^n_{q,k}$ as the number of patients of type k in ward q right after the n^{th} inspection. We can assume that inspections and discharge patterns can be described by the same distribution and are the same from an operational perspective. The dynamics of such distribution can be cited as follow:

$$\begin{cases} Q^{n+1}_{q,k}(\alpha_k = 1) = Q^n_{q,k} + (1 - \sum_{k\epsilon M} \beta_{q,k}) * \left(P^k_{i,j}\right)^{n+1}) \\ -\xi^{n+1}_q\ for\ every\ k \\ Q^{n+1}_{q,k}(\alpha_k < 1) = Q^n_{q,k} + (1 - \alpha_k) * \left(P^k_{i,j}\right)^{n+1} - \xi^{n+1}_k \end{cases}$$

With ξ^{n+1}_q is the number of patients discharged in the $(\mathrm{n}+1)^{\mathrm{th}}$ inspection and $\Gamma^N = \sum_{n=0,\forall\, q,\forall\, k}^{N} \xi^{n+1}_q$. In the same way, a decision to discharge a patient considers the patient's pathology, condition and period already spent as an inpatient.

The main goal of expressing the number of inspections per period is to relate this factor with service rate and occupancy function. Thus, in used data, we will have to divide the 24h into specific interval and set a mean number of inspections. This will help derive a discharge rate per period which will give visibility on how many primary and secondary hospitalizations are made and how many patients to admit.

IV. STATISTICAL TESTING OF THE DTNHMC

As previously defined all needed parameters: occupancy function of beds, how many inspections for discharge, primary and secondary hospitalization rates, etc. we will be able to simulate how every department in a facility can be used. To solve the medical resources allocation problem, we model patients flows as well as "consumed" resources' density functions by a discrete non-homogenous Markov chain presented in section 2. From these density functions, specific probabilities of wards occupancy and discharges as well as the expected outcomes and next arriving patients will be derived. Such is the focus of our modeling technique.

As the presented modeling approach considers a certain number of static resources in the healthcare system, infinite number of patients -considering the limited resources- will lead to a system stagnation. Let uik be maximum number of patients of type k that can be hospitalized in ward q (wqk). Let also Lq and UPq define the lower and upper bounds on the total amount of patients hospitalized in ward q.

The non-availability probability $P(U_{qk})$ of F_k strictly determined beds determines the probability of an event in which each bed (or medical resource in general) does not allow the admission of a patient. This probability can be determined with respect to the occupancy:

$$P\left(U_{qk}\right)=\sum_{k=(w_q-t_i)}^{M} P((w_k-\sum U_{q,k})|U_k) \; with \; \sum U_{q,k} \leq w_k$$

where $P((w_k - \sum U_{q,k})|U_k)$ is the conditional non-availability probability of a selected number of beds in a selected ward, determined under the assumption that the total number of busy beds $U_{q,k}$ reserved to admit patient of k type can be used for other patients. The lower limit of the sum in Equation above determines the minimum number of beds that can induce rejection of patients.

A. CASE STUDY: HOSPITAL EPISODE STATISTICS-NHS

1) DATA DESCRIPTION

Hospital Episode Statistics (HES) is a database containing details of all admissions, A and E attendances and outpatient appointments at the National Health Service (NHS) hospitals in England.

In the year from March 2020 to February 2021 there were:

- 16.0 million finished consultant episodes (FCEs), 55.7% (8.9 million) of which included at least one procedure or intervention, and 4.7 million of which were day cases.

Monthly HES data for Outpatients

- 100.2 million outpatient appointments made, with 75.9 million (75.8%) of these attended by the patient.
- 5.6 million outpatient appointments not attended by the patient, representing 5.6% of all appointments.

Analysis of the arrival rates reveals that the arrival patterns of the most patients fit a Poisson process. The only exception was noticed for elective surgery patients. Those patients were scheduled in specific days (from Monday to Friday). While beds occupancy that can be otherwise represented by LOS (Length of Stay) which reflects for how long a given resources are used a high proportion of long LOS in different wards related to all patients. This might lead to actual consideration of outliers. Such consideration can compromise the accuracy of our model. To solve such issue and in the same keep all dataset elements, the actual relationships or correlation between LOS and age, sex and disease were analyzed. As a result, aged patients with cardiovascular diseases, strokes and trauma are the more likely to stay longer. It doesn't matter their sex.

2) BED OCCUPANCY

It is collected during a patient's time at hospital as part of the Commissioning Data Set (CDS). This is submitted to NHS Digital for processing and is returned to healthcare providers as the Secondary Uses Service (SUS) data set and includes information relating to payment for activity undertaken. It allows hospitals to be paid for the care they deliver. This same data can also be processed and used for non-clinical purposes, such as research and planning health services. Because these uses are not to do with direct patient care, they are called 'secondary uses'. All data used in the current work is collected from the NHS's official digital platform [4].

HES data covers all NHS Clinical Commissioning Groups (CCGs) in England, including:

- private patients treated in NHS hospitals.
- patient's resident outside of England
- care delivered by treatment centers (including those in the independent sector) funded by the NHS

Each HES record contains a wide range of information about an individual patient admitted to an NHS hospital, including:

- clinical information about diagnoses and operations
- patient information, such as age group, gender and ethnicity
- administrative information, such as dates and methods of admission and discharge
- geographical information such as where patients are treated and the area where they live.

So far, our calculation of the bed occupancy has taken account of the daily variations represented in NHS in form of four quarters to represent a year. Beds fill up and empty out in function of patients' arrivals (as shown in figure below), admission rate and discharge rate.

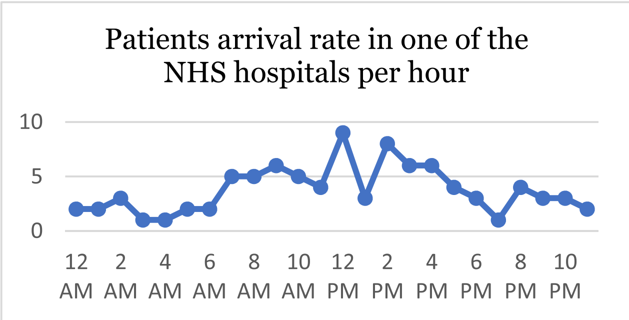


FIGURE 4. Bed occupancy-arriving patients in nine considered departments and of four-year quarter (detailed data source can be found in [4]).

TABLE 1. NHS observed bed occupancy (day and overnight in the year 2019-2020) compared to our expected bed occupancy.

Department	Elective patients	Non-elective patients	Expected bed occupancy	Chi-square goodness fit*
	Bed occupancy(day)-bed occupancy (overnight)		Bed occupancy(day)-bed occupancy (overnight)	
Neonatology	56%-65%	58.6%-54.8%	49%-33%	0.017
Pediatrics	12%-19%	16.1%-19.7%	15.7%-15%	0.0826
Maternity	70%-75%	45.9%-54.4%	45%-53.9%	0.0104
Gynecology	45%-60%	59.9%-58.9%	60%-56%	0.00571
Critical care	84%-87%	85.3%-79.1%	84.6%-78.5%	0.00432
Cardiology	78%-80%	91.4%-84.3%	90.3%-84%	0.000472
Trauma	75%-71%	87%-88.2%	86.6%-86.4%	0.0162
Elderly	80%-79%	71.2%-68.8%	71%-67.9%	0.00523
Surgery	79%-88%	49.5%-39.6%	48.2%-40.9%	0.0387
All departments	96%-93%	98.2%-93.6%	99.1%-92.9%	0.00609

A study by [37], has demonstrated a high correlation between bed occupancy and longer Emergency department waiting times. As with a nearly 100% bed occupancy, the percentage of patients in ED waiting longer than four hours increased by 9%. The study also listed: Hospital length of stays over 21 days, higher emergency admissions and lower discharges to be main factors of inconveniently higher waiting times. The study included 138 English NHS healthcare providers of which daily situation reports (Sitrep), hospital episode statistics and electronic staffing records data over 90 days between December 2016 and February 2017 were extracted.

Taking the previous assumptions into consideration, we apply the actual probability distribution on a switched single time-interval (day on -night off and then day off-night on). Such assumption will allow to first consider the system as inactive at night, then as inactive during day. Table 1 represents the different beds resources occupancy for every department. Thus, our Markov chain can be considered as steady state process [38], [39]. Let π be the steady state probability distribution of the Discrete Time [40], [41] Markovian chain that fills in the following balance equation:

$$\pi Q = 0$$

where $\sum_{xi\in S} \pi_i = 1$ and Q is the infinitesimal generator matrix. Such equation can be transposed and rewritten in the form of a linear equation: Ax = b. Thus we get $Q^T \pi^T = 0 \rightarrow Q^T \pi^T = (D - X - Y)\pi^T = 0$ where D, X, and Y are the diagonal, lower and upper strictly triangular matrices of Q^T. Where Q can be written as $Q = A_1 * u' + A_2$ with A_1 and A_2 can be expressed as follow:

$$A_1 = \begin{bmatrix} -1 & 0 & \cdots & 0 \\ 1 & -1 & \ddots & \vdots \\ 0 & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & -1 & 0 \\ -1 & -1 & \cdots & -2 \end{bmatrix}$$

$$A_2 = \begin{bmatrix} & 0 & 0 & \cdots & 0 \\ & 0 & -2\mu & 2\mu & \vdots \\ & \ddots & -3\mu & 3\mu & 0 \\ & 0 & \ddots & \ddots & 0 \\ & \vdots & -S\mu & S\mu & 0 \\ & 0 & \ddots & \ddots & 0 \\ & 0 & \cdots & 0 & S\mu \\ 0 & \cdots & 0 & S\mu & -S\mu \end{bmatrix}$$

With S is the number of servers. And μ the current number of patients waiting to be served. And μ' the total number of patients to be served from both elective and emergency paths. Using overrelaxation technique, the following values were determined the values of π using python programming. We found: Π = [0.53,0.41, 0.67, 0.88, 0.79, 0,91, 0.93, 0.77, 0.99].

We conducted a statistical test to assess our model's fitness with observations on ward occupancy. To our knowledge, there exists no standard technique to test the fitness of a DTMC with a complexity as considered in our work. Thus, this section presents a discrete event simulation approach that combines a simulation of the DTMC behavior and compares this to hospital data on ward occupancy. To begin with, our null hypothesis is that the observed values are generated by the DTMC process. If that is the case, we would expect the observed frequency of occupied beds to be quite like the marginal distributions of π for each ward. A standard approach would be to test the observed frequencies against the corresponding expected frequencies from the DTMC. We conducted a simulation approach on each separated department. The total number of used resources by all patient types combined (elective and patients from emergency department) was then deduced from the total number of resources (beds in this case). The proportion of expected used beds can be seen in figure below.

TABLE 2. Studies using DES, Markov chains and queuing theory to model patients flow.

Ref	Model technique	Objective	Performance measurements	Findings
[51]	Agent based modeling and Queuing model	Resource's optimization in the emergency department using agent-based modeling and evaluate waiting times using queuing theory.	Number of patients waiting and how long they wait to be seen.	No quantitative analysis was performed. Qualitatively, the agent-based model was shown to be adaptable to many scenarios.
[52]	Queuing model	Description of patients' flow inside the hospital and optimization of hospital resources usage.	The average cost per unit of time. Number of rejected patients.	An average of 629, 827, 951, and 1049 cost per unit time for 140, 150, 155, and 160 beds respectively.
[53]	Discrete event simulation	Studying the impact of medical resources alteration on the hospital's performance indicators.	Total waiting time before starting treatment for both minor and major patients.	More flexible admission service is a key factor for reducing patients waiting time.
[54]	Queuing model	Hospital's capacity management.	Utilization rate, waiting probability, estimated bed occupancy, capacity simulations and demand behavior assessment.	The study proved that balancing admissions and type of patients over the week represents a very powerful policy to optimize resources utilization.
[30]	Nonhomogeneous discrete time Markov chains	Admission scheduling, resource requirement forecasting and resources allocation.	Theoretically estimate the number of needed resources and to help allocating resources and admission resources.	No simulation or model validation was performed.
[55]	Discrete event simulation	Improvement of medical resources utilization by altering inpatients and resources constraints and process improvement scenarios.	Patients' length of stay and rate of patients who leave without being seen.	Adding more 6 beds to receive inpatients reduced LOS by 8%.

Bed occupancy rates are often arbitrary and depend on multiple factors, including the number of admissions from the emergency department and other hospital wards, scheduled surgery, and patient types and their LOS in the different department. The rationalization of beds is a difficult daily task that becomes even more complicated when resources are limited. There is no consistent evidence regarding the outcome of current official patient admission and discharge guidelines, which must be adapted to individual wards and hospital contexts [42], [43]. Admission and discharges policies are important not only in resource management terms but also in terms of patient care outcome. Shortage of beds is an independent factor when deciding not to admit patients to any of the available wards [44]. Several studies show that in the event of bed shortage, admissions and discharges are triaged. This increases the number of rejected admission requests, increases the severity threshold for admission, and shortens LOS [45], [46]. However, assuming the same degree of severity, the prognosis is better for patients admitted as a primary hospitalization than for those with secondary hospitalization. Other consequences of an excessive bed occupation rate are scheduled surgery cancellations, transfers to other centers, or wards oversizing. Therefore, a simulation model with the capacity to mimic real bed requirements taking all implied factors into account (type of hospital, unit-specific admission and discharge criteria, adjustments considering the requirement/availability ratio) constitutes a useful tool for effective hospital bed management. The stochastic nature of patient flow could lead to an underestimation of the resources required in busy units. In the past, 85% mean bed occupancy used to be considered optimum. However, the mean could be misleading, because, unless the daily occupancy distribution is considered, it fails to capture long periods in which high bed occupancy may result in an unacceptable rate of rejected admissions because of shortage of beds. Thus, to cover potential activity peaks, several empty beds should be available [47]. While simulation has been extensively used in other healthcare problems [48], [49], including ICU bed management [50]. However, most studies are approaching mathematical modeling and simulation as two different approaches. Table 2 represents a comparative review of some of these studies.

Thus, the objective of this study was to develop a reliable model to reflect the daily bed occupancy rate in the different wards based on a high accuracy Markovian model, and thus avoid the bed availability problem and below-optimum occupancy rates and the unnecessary costs these entail. Our model is based on a complex statistical analysis that, despite its complexity, uses properly defined, easily collectible, common wards management variables. In methodologic terms, we think that this model offers several advantages over other commonly used models (as shown in table 2). We modeled admissions by patient type using variations of the Poisson distribution. It also offers accurate resources usage prediction using fewer inputs.

Table 1 represents the actual resulting mean bed occupancy of all wards considered.

The results obtained for the simulated daily occupancy rate can be summarized as follows. Patients with high emergency are urgent situations are the first to be served as shown with Figure 5. a, and Figure 5.b the system tries to serve patients with high of risk of mortality first. As the number of patients increases, the waiting time also increases as limited resources constraints were defined while simulating.

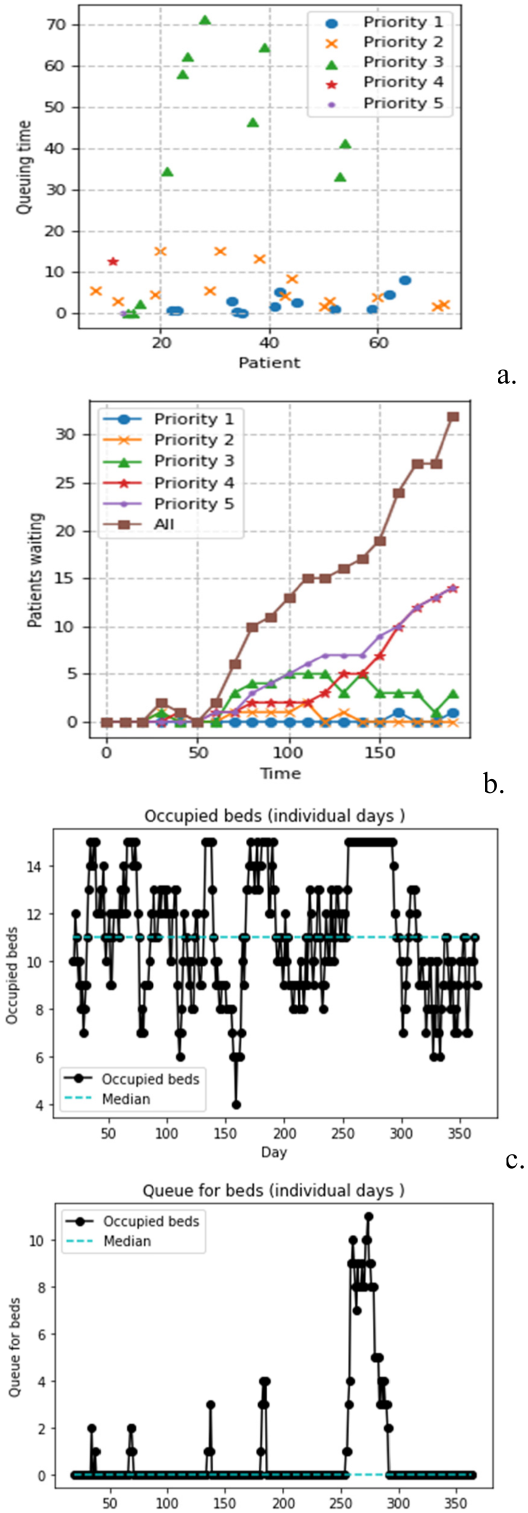


FIGURE 5. Simulation results. a. represents the arrival rate of every previously defined emergency levels following a Poisson process. b. the cumulative function of admitted patients with assumption of no lost patients. c. represents the average number of occupied beds patients in all departments. d. is the number of patients in queue for a bed.

While studying such complex systems with limited resources, it is important to keep good fluctuation near the median value. Suh is the case of beds in hospitals with high admission rate. Figure 5.c represent the results of our DES based on the previous mentioned Markov chain. It is important to mention that in previous of high admission rate, the triage of patients with high priority to be admitted is very important. Figure 5.d suggests how our system must deal with such case. The number of patients in queue represent the number of patients with secondary hospitalizations and waiting to be relocated to the ward specialty to treat their disease which referred to as primary hospitalization.

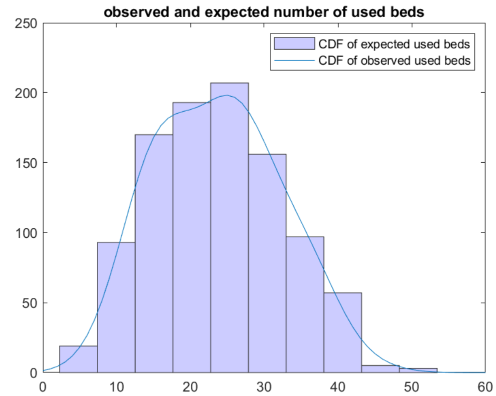


FIGURE 6. Cumulative density function of the observed and expected bed occupancy.

As it might be deduced from Table 2 and Figure 6, our Markovian model with the pre-defined probability distributions seem to be of good fit to real world scenarios.

V. CONCLUSION

We have demonstrated how Markov chains and queuing theory can be an effective to model patient's flow and to forecast needed beds in different hospital's wards. The study used a Nonhomogeneous Discrete time Markov chain to represent the different flow variations. Thus, to validate this model, discrete event simulations were used. it gives as results the frequencies of a server and resources utilization. The results of the simulation were then compared to the average bed occupancy in the National Health Services database using the chi-square technique.

The methodology developed in this work enables estimation of the main characteristics of access to service of patients and hospital managers as well as medical care providers: forecasting of the arrival rate of patients and bed occupancy in function of patient's arrival and their priority.

We assumed that the hospital can be described with Nonhomogeneous Discrete time Markov chain. We also assumed this process can be compared to a steady state distribution of a M/M/s Queuing system. The results of the global equation of the steady state distribution were solved using arrival and service rate. The two parameters follow an exponential distribution. The resulting vector was then used to find the transition matrix coefficients. the results after simulation have shown enormous correlation between arrival rate, level of emergency and occupancy rate of beds as shown in figure 5. Then the coefficients of the transition matrix were compared to the actual (observed) rates of the real data records of the NHS using chi-square goodness fit. Managers and healthcare providers must have look then to ''bottle neck'' departments such critical care, cardiology, and trauma. These are the most crowded wards with very high bed occupancy rates. A very good policy can be to set a median value of admission rate to keep smooth fluctuation around the mean value of occupation rate. This might help stabilize the system's dynamics and deliver better medical care to patients.

REFERENCES

[1] *Regional Committee for Europe Sixty-Second Session European Action Plan for Strengthening Public Health Capacities and Services Regional Committee for Europe European Action Plan for Strengthening Public Health Capacities and Services*. Accessed: Mar. 10, 2020. [Online]. Available: http://www.euro.who.int/__data/assets/pdf_file/0005/171770/RC62wd12rev1-Eng.pdf

[2] *Trends in Health Financing and the Private Health Sector in the Middle East and North Africa*. Accessed: Mar. 10, 2020. [Online]. Available: https://www.shopsplusproject.org/

[3] *Hospital Waiting Times at Worst-Ever Level—BBC News*. Accessed: Mar. 11, 2020. [Online]. Available: https://www.bbc.com/news/health-50397856

[4] *Summary Hospital-Level Mortality Indicator (SHMI) Deaths Associated With Hospitalization, England, October 2018 September 2019—NHS Digital*. Accessed: Mar. 11, 2020. [Online]. Available: https://www.england.nhs.uk

[5] L. Woodworth and J. F. Holmes, ''Just a minute: The effect of emergency department wait time on the cost of care,'' *Econ. Inquiry*, vol. 58, no. 2, pp. 698–716, 2019, doi: 10.1111/ecin.12849.

[6] (2048). *Centers for Disease Control and Prevention- National Center for Health Statistics. Mortality in the United States*. Accessed: Mar. 12, 2020. [Online]. Available: https://www.cdc.gov/nchs/products/databriefs/db355.htm

[7] *The Commonwealth Fund // 2018 Annual Report*. Accessed: Mar. 12, 2020. [Online]. Available: https://annualreport.commonwealthfund.org/2018/

[8] *Waiting Times for Emergency Department and Elective Surgery on the Rise—Australian Institute of Health and Welfare*. Accessed: Mar. 12, 2020. [Online]. Available: https://www.aihw.gov.au/news-media/media-releases/2018/december/waiting-times-for-emergency-department-and-elective

[9] (Mar. 13, 2020). *Australian Institute of Health and Welfare 2019. Deaths in Australia. Cat. no. PHE 229. Canberra: AIHW Viewed*. [Online]. Available: https://www.aihw.gov.au/reports/life-expectancy-death/deaths-in-australia

[10] B. Barua and M. Moir, *Waiting Your Turn: Wait Times for Health Care in Canada, 2019 Report*. Accessed: Mar. 12, 2020. [Online]. Available: https://www.fraserinstitute.org/sites/default/files/waiting-your-turn-2019-rev17dec.pdf

[11] *World Population Prospects—Population Division—United Nations*. Accessed: Mar. 12, 2020. [Online]. Available: https://population.un.org/wpp/DataQuery/

[12] A. Guttmann, M. J. Schull, M. J. Vermeulen, and T. A. Stukel, ''Association between waiting times and short term mortality and hospital admission after departure from emergency department: Population based cohort study from Ontario, Canada,'' *BMJ*, vol. 342, Jun. 2011, Art. no. d2983, doi: 10.1136/bmj.d2983.

[13] I. E. Blanchard, C. J. Doig, B. E. Hagel, A. R. Anton, D. A. Zygun, J. B. Kortbeek, D. G. Powell, T. S. Williamson, G. H. Fick, and G. D. Innes, ''Emergency medical services response time and mortality in an urban setting,'' *Prehospital Emergency Care*, vol. 16, no. 1, pp. 142–151, Jan. 2012, doi: 10.3109/10903127.2011.614046.

[14] D. A. Martinez, H. Zhang, M. Bastias, F. Feijoo, J. Hinson, R. Martinez, J. Dunstan, S. Levin, and D. Prieto, ''Prolonged wait time is associated with increased mortality for chilean waiting list patients with non-prioritized conditions,'' *BMC Public Health*, vol. 19, no. 1, pp. 1–11, Dec. 2019, doi: 10.1186/s12889-019-6526-6.

[15] P. J. Kaboli, J. T. Go, J. Hockenberry, J. M. Glasgow, S. R. Johnson, G. E. Rosenthal, M. P. Jones, and M. Vaughan-Sarrazin, ''Associations between reduced hospital length of stay and 30-day readmission rate and mortality: 14-year experience in 129 veterans affairs hospitals,'' *Ann. Internal Med.*, vol. 157, no. 12, p. 837, Dec. 2012, doi: 10.7326/0003-4819-157-12-201212180-00003.

[16] I. Chopra, T. L. Wilkins, and U. Sambamoorthi, ''Hospital length of stay and all-cause 30-day readmissions among high-risk medicaid beneficiaries,'' *J. Hospital Med.*, vol. 11, no. 4, pp. 283–288, 2015, doi: 10.1002/jhm.2526.

[17] *Coronavirus Disease 2019 (COVID 19) Situation Report-56*. Accessed: Mar. 17, 2020. [Online]. Available: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200316-sitrep-56-covid-19.pdf?sfvrsn=9fda7db2_6

[18] R. S. Ulrich. *Ph.D. Effects of Healthcare Environmental Design on Medical Outcomes*. Accessed: Mar. 14, 2020. [Online]. Available: https://pdfs.semanticscholar.org/6d4d/cb6d3f8d55a9eac8a109799f61d37a570e4e.pdf

[19] H. Salonen, M. Lahtinen, S. Lappalainen, N. Nevala, L. D. Knibbs, L. Morawska, and K. Reijula, ''Design approaches for promoting beneficial indoor environments in healthcare facilities: A review,'' *Intell. Buildings Int.*, vol. 5, no. 1, pp. 26–50, Jan. 2013, doi: 10.1080/17508975.2013.764839.

[20] M. Khalifa and I. Zabani, ''Utilizing health analytics in improving the performance of healthcare services: A case study on a tertiary care hospital,'' *J. Infection Public Health*, vol. 9, no. 6, pp. 757–765, Nov. 2016, doi: 10.1016/j.jiph.2016.08.016.

[21] P. C. Smith, E. Mossialos, I. Papanicolas, and S. Leatherman. *Performance Measurement for Health System Improvement Experiences*. Challenges and Prospects. Accessed: Mar. 15, 2020. [Online]. Available: http://www.euro.who.int/__data/assets/pdf_file/0003/135975/E94887_Part_V.pdf?ua=1

[22] R. Mehandiratta, ''Applications of queuing theory in health care,'' *Int. J. Comput. Bus. Res.*, vol. 2, no. 2, p. 2229, 2011.

[23] *Using Queuing Theory to Reduce Wait, Stay in Emergency Department*. Accessed: Mar. 16, 2020. [Online]. Available: https://www.physicianleaders.org/news/queuing-theory-reducing-wait-stay

[24] H. Takagi, Y. Kanai, and K. Misue, ''Queueing network model for obstetric patient flow in a hospital,'' *Health Care Manage. Sci.*, vol. 20, no. 3, pp. 433–451, 2017, doi: 10.1007/s10729-016-9363-5.

[25] O. Bittencourt, V. Verter, and M. Yalovsky, ''Hospital capacity management based on the queueing theory,'' *Int. J. Productiv. Perform. Manage.*, vol. 67, no. 2, pp. 224–238, Feb. 2018, doi: 10.1108/ijppm-12-2015-0193.

[26] N. M. Chowdhury, L. Riddles, and R. Mackenzie. *Using Queueing Theory to Reduce Wait, Stay in Emergency Department*. Accessed: Mar. 18, 2020. [Online]. Available: https://www.physicianleaders.org/news/queuing-theory-reducing-wait-stay

[27] Eugene Seneta. *Markov and the Creation of Markov Chains*. Accessed: Mar. 19, 2020. [Online]. Available: http://www.maths.usyd.edu.au/u/eseneta/senetamcfinal.pdf

[28] S. Y. Grodzensky, *Andrei Andreevich Markov*. Moscow, Russia: Nauka, 1987.

[29] L. Garg, S. McClean, B. Meenan, and P. Millard, ''Optimal control of patient admissions to satisfy resource restrictions,'' in *Proc. 21st IEEE Int. Symp. Comput. Based Med. Syst.*, Jun. 2008, pp. 512–517, doi: 10.1109/cbms.2008.33.

[30] L. Garg, S. McClean, B. Meenan, and P. Millard, ''A non-homogeneous discrete time Markov model for admission scheduling and resource planning in a cost or capacity constrained healthcare system,'' *Health Care Manage. Sci.*, vol. 13, no. 2, pp. 155–169, Jun. 2010, doi: 10.1007/s10729-009-9120-0.

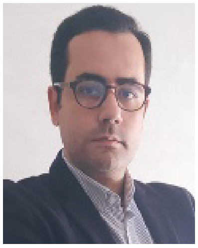
[31] (Jul. 2018). *Improving Patient Flow and Reducing Emergency Department Crowding: A Guide for Hospitals*. Content. Agency for Healthcare Research and Quality, Rockville, MD, USA. [Online]. Available: https://www.ahrq.gov/research/findings/final-reports/ptflow/index.html

[32] L. Anselmi, R. Meacock, S. R. Kristensen, T. Doran, and M. Sutton, ''Arrival by ambulance explains variation in mortality by time of admission: Retrospective study of admissions to hospital following emergency department attendance in england,'' *BMJ Qual. Saf.*, vol. 26, no. 8, pp. 613–621, Aug. 2017, doi: 10.1136/bmjqs-2016-005680.

[33] A. Möller, L. Hunter, L. Kurland, S. Lahri, and D. J. van Hoving, ''The association between hospital arrival time, transport method, prehospital time intervals, and in-hospital mortality in trauma patients presenting to Khayelitsha hospital, cape town,'' *Afr. J. Emergency Med.*, vol. 8, no. 3, pp. 89–94, Sep. 2018, doi: 10.1016/j.afjem.2018.01.001.

[34] *Overview of U.S. Hospital Stays in 2016: Variation by Geographic Region #246*. Accessed: Feb. 21, 2020. [Online]. Available: https://www.hcup-us.ahrq.gov/reports/statbriefs/sb246-Geographic-Variation-Hospital-Stays.jsp
[35] *Five Steps of Emergency Care | St. Mary's Regional Medical Center*. Accessed: Feb. 21, 2020. [Online]. Available: https://www.stmarysregional.com/services/emergency-services/emergency-care-what-to-expect
[36] M. Bahadori, E. Teymourzadeh, R. Ravangard, A. Nasiri, M. Raadabadi, and K. Alimohammadzadeh, "Factors contributing towards patient's choice of a hospital clinic from the patients' and managers' perspective," *Electron. Physician*, vol. 8, no. 5, pp. 2378–2387, May 2016, doi: 10.19082/2378.
[37] S. Paling, J. Lambert, J. Clouting, J. González-Esquerré, and T. Auterson, "Waiting times in emergency departments: Exploring the factors associated with longer patient waits for emergency care in England using routinely collected daily data," *Emergency Med. J.*, vol. 37, no. 12, pp. 781–786, Sep. 2020, doi: 10.1136/emermed-2019-208849.
[38] A. Dudin, A. Nazarov, and A. Moiseev, "[Communications in computer and information Science] information technologies and mathematical modelling," in *Proc. 17th Int. Conf., (ITMM) 12th Workshop Retrial Queues Rel. Topics, (WRQ)* in Steady State Probabilistic Characteristics of the On/Off Production Rate Control Production-Inventory System with MMPP Demand Arrivals, A. A. F. Terpugov, Ed. Tomsk, Russia, Sep. 2018, Ch. 20, pp. 248–262, doi: 10.1007/978-3-319-97595-5_20.
[39] S. Gilmore, D. Reijsbergen, and A. Vandin, "Transient and steady-state statistical analysis for discrete event simulators," in *Proc. Int. Conf. Integr. Formal Methods*. Cham, Switzerland: Springer, Sep. 2017, pp. 145–160.
[40] Y. Shi, L. Jin, S. Li, J. Li, J. Qiang, and D. K. Gerontitis, "Novel discrete-time recurrent neural networks handling discrete-form time-variant multi-augmented Sylvester matrix problems and manipulator application," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Oct. 19, 2020, doi: 10.1109/tnnls.2020.3028136.
[41] Y. Zhang and Y. Zhang, "New discrete-time models of zeroing neural network solving systems of time-variant linear and nonlinear inequalities," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 50, no. 2, pp. 565–576, Feb. 2020, doi: 10.1109/TSMC.2017.2751259.
[42] M. Rodríguez-Carvajal, D. Mora, A. Doblas, M. García, P. Domínguez, A. Tristancho, and M. Herrera, "Impact of the premature discharge on hospital mortality after a stay in an intensive care unit," *Medicina Intensiva (English Ed.)*, vol. 35, no. 3, pp. 143–149, Jan. 2011.
[43] A. Egol, "Guidelines for intensive care unit admission, discharge, and triage," *Crit. Care Med.*, vol. 27, pp. 633–638, 1999.
[44] C.-P. Heidegger, M. M. Treggiari, and J.-A. Romand, "A nationwide survey of intensive care unit discharge practices," *Intensive Care Med.*, vol. 31, no. 12, pp. 1676–1682, Dec. 2005.
[45] T. Eldabi, R. J. Paul, and T. Young, "Simulation modelling in healthcare: Reviewing legacies and investigating futures," *J. Oper. Res. Soc.*, vol. 58, no. 2, pp. 262–270, Feb. 2007.
[46] J. B. Jun, S. H. Jacobson, and J. R. Swisher, "Application of discrete-event simulation in health care clinics: A survey," *J. Oper. Res. Soc.*, vol. 50, no. 2, pp. 109–123, Feb. 1999.
[47] M. A. Metcalfe, A. Sloggett, and K. McPherson, "Mortality among appropriately referred patients refused admission to intensive-care units," *Lancet*, vol. 350, pp. 7–11, Jul. 1997.
[48] K. Seung-Chul, H. Ira, Y. Karl K, and B. Thomas A, "Flexible bed allocation and performance in the intensive care unit," *J. Oper. Manage.*, vol. 18, no. 4, pp. 427–443, Jun. 2000.
[49] W. Cahill and M. Render, "Dynamic simulation modeling of ICU bed availability," in *Proc. WSC Winter Simulation Conf. Simulation Bridge Future*, Dec. 1999, pp. 1573–1576.
[50] T. Sinuff, K. Kahnamoui, D. J. Cook, J. M. Luce, and M. M. Levy, "Rationing critical care beds: A systematic review," *Crit. Care Med.*, vol. 32, no. 7, pp. 1588–1597, Jul. 2004.
[51] K. B. Ahsan, M. R. Alam, D. G. Morel, and M. A. Karim, "Emergency department resource optimisation for improved performance: A review," *J. Ind. Eng. Int.*, vol. 15, no. S1, pp. 253–266, Dec. 2019, doi: 10.1007/s40092-019-00335-x.
[52] P. Bhattacharjee and P. K. Ray, "Patient flow modelling and performance analysis of healthcare delivery processes in hospitals: A review and reflections," *Comput. Ind. Eng.*, vol. 78, pp. 299–312, Dec. 2014, doi: 10.1016/j.cie.2014.04.016.
[53] D. Botje, G. ten Asbroek, T. Plochg, H. Anema, D. S. Kringos, C. Fischer, C. Wagner, and N. S. Klazinga, "Are performance indicators used for hospital quality management: A qualitative interview study amongst health professionals and quality managers in The Netherlands," *BMC Health Services Res.*, vol. 16, no. 1, p. 574, Dec. 2016, doi: 10.1186/s12913-016-1826-3.
[54] L. V. Green, "Capacity planning and management in hospitals," in *Operations Research and Health Care*. Boston, MA, USA: Springer, 2005, pp. 15–41.
[55] Y. Simwita and B. Helgheim, "Simulation analysis of resource flexibility on healthcare processes," *J. Multidisciplinary Healthcare*, vol. 9, pp. 519–528, Oct. 2016, doi: 10.2147/JMDH.S112713.

KAOUTER KARBOUB received the degree in engineering from the National Higher School of Electricity and Mechanics, Morocco, in 2018, with a focus on industrial and logistics sciences. She is currently pursuing the international dual Ph.D. degree with the University of Lorraine, France, and the Foundation of Research, Development and Innovation in Engineering Sciences, Morocco. Her thesis concerns the application of AI approaches in the medical field. Her research interests include data analysis, machine learning, and semantic technologies.



MOHAMED TABAA (Member, IEEE) received the degree in engineering from the Moroccan School of Engineering Science (EMSI), Casablanca, Morocco, in 2011, the master's degree in radiocommunication, and embedded electronic systems from the University of Paul Verlaine, Metz, France, and the Ph.D. and H.D.R. Diploma degrees in electronics systems from the University of Lorraine, Metz, France, in 2014 and 2020, respectively. Since 2015, he has been the Director of the LPRI Private Laboratory attached to the EMSI. His research interests include array of digital signal processing for wireless communications, the IoT, digitalization, renewable energy, and embedded systems. He served on the Organizing Committee and Technical Program Committee of several international conferences, including IEEE ICM, IEEE REPS&GIE, IEEE Systol, INTIS, ASD, JDSI, and FIoE.

SOFIENE DELLAGI received the Ph.D. degree in industrial engineering from the University of Lorraine, France. He has been an Associate Professor in industrial engineering with the University of Lorraine, since 2008, and responsible for the MPM Team with LGIPM Laboratory, since 2015. His research contributions include the publication of several articles and a book (13 articles and one book). His main expertise and research interests include systems reliability and maintenance, and production management. He has been the Co-Editor of a special issue of the *International Journal of Production Research* in 2008.

FABRICE MONTEIRO (Member, IEEE) received the Ph.D. degree in microelectronics from the University of Montpellier, France, in 1992. He has been an Associate Professor and a Professor with the University of Lorraine, France (formerly, University of Metz), since 1994. He has been involved in several academic and industrial research projects targeting interdisciplinary topics in the areas of telecommunications, mechatronics, and sensor networks. His current research interests include fault tolerant digital circuits and systems, coding theory and related high throughput circuitry for communication applications, new MPSoC, and NoC paradigms.

ABBAS DANDACHE received the Ph.D. degree both in computer sciences and microelectronics from the INPG, Grenoble, in 1983 and 1986, respectively, and the Habilitation degree from the University of Metz, Metz, France, in 2000. He has been a Professor with the University of Lorraine, Metz, since 2001. Since 2007, he has been the Director of the LICM Laboratory. He has been a principal investigator of several research contracts with the French telecommunications industry. He has authored or coauthored several publications in the domain. His current research and teaching interests include electronics design are related to embedded systems and smart sensors.

FOUAD MOUTAOUAKKIL received the degree in engineering with a focus on industrial automation, the Ph.D. degree in engineering science, and the Diploma of accreditation to supervise research (H.D.R.) degree in engineering science from the National Higher School of Electricity and Mechanics (ENSEM), Hassan II University of Casablanca, Casablanca, Morocco, in 2005, 2010, and 2015, respectively. In 2006, he joined the System Architecture Team, ENSEM, Hassan II University of Casablanca, where he has been a Full Professor in automatic production and computer sciences, since 2011.