Proceedings of the 2021 Winter Simulation Conference
S. Kim, B. Feng, K. Smith, S. Masoud, Z. Zheng, C. Szabo and M. Loper, eds.

IMPROVING INPUT PARAMETER ESTIMATION IN ONLINE PANDEMIC SIMULATION

Daniel Garcia-Vicuña
Fermin Mallor

Smart Cities Institute
Public University of Navarre
Campus Arrosadia
Pamplona, 31006, SPAIN

ABSTRACT

Simulation models are suitable tools to represent the complexity and randomness of hospital systems. To be used as forecasting tools during pandemic waves, it is necessary an accurate estimation, by using real-time data, of all input parameters that define the patient pathway and length of stay in the hospital. We propose an estimation method based on an expectation-maximization algorithm that uses data from all patients admitted to the hospital to date. By simulating different pandemic waves, the performance of this method is compared with other two statistical estimators that use only complete data. Results collected to measure the accuracy in the parameters estimation and its influence in the forecasting of necessary resources to provide healthcare to pandemic patients show the better performance of the new estimation method. We also propose a new parameterization of the Gompertz growth model that eases the creation of patient arrival scenarios in the pandemic simulation.

1 INTRODUCTION

During pandemic waves, the demand for critical hospital resources like personnel and beds for both acute and critically ill patients greatly increases. Accurate forecasts of the required resources help their planning, which includes the procurement of equipment, and the redeployment of staff and other resources. Sometimes planning also requires the cancellation of elective surgeries. The challenge for healthcare planners in pandemic times is planning capacity to treat both pandemic and non-pandemic patient types.

Simulation emerges as a suitable analytical tool to help hospital managers since it can represent the complexity of the hospital system and the variability and uncertainty of the healthcare that patients need. Currie et al. (2020) discuss how simulation modelling could help to support decision-makers in making the most informed decisions. Garcia-Vicuña et al (2020a) develop a simulation model to predict the necessary resources during a pandemic. The accuracy of the predictions made by such simulation models depends on accurate modelling of the resources and hospitalization time required by patients for recovery, as well as on good modelling of the arrival process of new patients to the health service.

In this paper, we deal with the problem of getting reliable estimations of all the parameters that define the trajectory and length of stay (LoS) of patients in the hospital facilities. These parameters cannot be calibrated in advance by using historical data because each pandemic wave may have different characteristics, affecting different population groups with different intensity, and even change during the pandemic wave. However, during the first days, and even weeks, of the pandemic, few complete data are available since a significant proportion of patients are still admitted to the hospital. For this reason, it is essential to develop estimation methods that also take into account data coming from the patients of the current pandemic wave, even if this information is incomplete.

978-1-6654-3311-2/21/$31.00 ©2021 IEEE

In this work, we present a new estimator that, based on an expectation-maximization (EM) algorithm, estimates the branching probabilities using information from patients who do not yet know which path they will follow, as well as the parameters of the LoS probability distributions. The estimation method is based on the maximum likelihood method and uses exact and censored data. The new estimators are tested by simulating pandemic scenarios and compared to other estimators that make use of complete information only. The patient arrival process is simulated by using growth Gompertz models. To ease the creation of different pandemic waves mimicking real ones we also propose a different parameterization of the Gompertz growth model.

This paper is organized as follows. Section 2 introduces the population growth models, provides a new parameterization for the Gompertz model, and describes the patient pathways through the hospital. In Section 3 the estimation problem is set and the new estimator is proposed. The performance of the estimator is shown in Section 4. The paper ends with the conclusions in Section 5.

2 PANDEMIC SIMULATION

2.1 Population Growth Curves. The Gompertz Model.

Population Growth models (PG) provide methods for modeling the number of cumulative positive cases, hospitalizations, and other pandemic variables. Some examples of PG models that have been found in the literature are the Gompertz (Gompertz 1825), the Richards (Richards 1959), the Stannard (Stannard et al. 1985), and the logistic model (Ricker 1979). Growth curves are used in a large variety of scientific fields, including the modeling of the spread of outbreaks such as A/H1N1 and Ebola in (Liu et al. 2015), and to predict new infection cases caused by the SARS-CoV-2 virus in countries such as China (Shen 2020), India (Malavika et al. 2021), Spain (Sánchez-Villegas and Codina 2020), and other European countries (Cássaro and Pires 2020). These mathematical models depend on several parameters that can be estimated from real data sets.

The Gompertz model shows a better fit to data of daily Covid-19 new cases as well as better predictive capacity than other PG models (Garcia-Vicuña et al. 2020a). This model was used in (Garcia-Vicuña et al. 2020b) to fit daily hospitalized patient data series during a pandemic wave to predict the expected number of patients to be hospitalized each one of the next days. These expected values were used as the input of a non-homogeneous Poisson Process from which simulating sequences of patient arrivals. In the next subsection, we introduce a new parameterization easier to interpret and to use for creating pandemic scenarios over which running hospital simulation models. The new parameterization depends on three parameters that reflect the size, the temporal spread, and the temporal location of the pandemic.

2.2 New Parametrization of the Gompertz Growth Model

The original equation of the Gompertz model, proposed in (Gompertz 1825), was rewritten in (Zwietering et al. 1990) to ease the biological interpretation of its parameter as follows:

$$G(t) = A \exp\left(-\exp\left(\frac{Ke}{A}(D - t) + 1\right)\right) \tag{1}$$

where,

- $e = \exp(1)$
- $G(t)$ is the population size up to time t.
- A is the upper asymptote of the curve.
- K is the absolute growth rate of the curve at its inflection point.
- D, known as the lag time, is the time at which $G(t) = Aexp(-e)$, which means that it always occurs at the same percentage (6.6%) of the upper asymptote.

In the context of a pandemic mathematical modelling, parameter A has a clear meaning: the total number of infected persons or the total number of hospitalized patients at the end of the pandemic wave, when the Gompertz curve is used to model the series of new positive cases or the hospitalization process, respectively. However, the parameter K has no intuitive meaning, and health managers may have difficulties in assigning a value to it in order to create an "artificial" pandemic. To overcome this difficulty we propose to replace the parameter K by other parameter linked with the duration of the pandemic wave. The new parameterization is obtained from equation (1) by calculating the percentile t_p ($G(t_p) = pA$) as $t_p = G^{-1}(pA)$:

$$t_p = D - \frac{A}{\mathrm{K}e}[\ln(-\ln(p)) - 1]$$

Let us denote by T_{p_1,p_2} the elapsed time between t_{p_1} and t_{p_2}.

$$T_{p_1,p_2} = t_{p_2} - t_{p_1} = \frac{[-\ln(-\ln(p_2)) + \ln(-\ln(p_1))]}{e}\frac{A}{\mathrm{K}} = \frac{C_{p_1,p_2}}{e}\frac{A}{\mathrm{K}}$$

Where, C_{p_1,p_2} is a constant that depends on proportions p_1 and p_2. Then, the parameter K is equal to $K = C_{p_1,p_2}A/(eT_{p_1,p_2})$, and substituting in (1), we obtain:

$$G(t) = A\exp\left(-\exp\left(\frac{C_{p_1,p_2}}{T_{p_1,p_2}}(D-t) + 1\right)\right)$$

To simplify this expression, we consider the length of the interval time associated to the central proportion α of cases, that is, proportions p_1 and p_2 are defined as $p_1 = \alpha/2$ and $p_2 = 1 - \alpha/2$ to get the constant $C_{p_1,p_2} = C_\alpha$ and the parameter $T_{p_1,p_2} = T_\alpha$. With $\alpha = 0.1$, we obtain $C_\alpha \cong 4.0674$ and the Gompertz curve is:

$$G(t) = A\exp\left(-\exp\left(\frac{4.0674}{T_{0.1}}(D-t) + 1\right)\right) \quad (2)$$

This new parameterization determines the Gompertz curve by setting the total number of cumulative cases at the end of the outbreak, the duration of the wave, and the time at which 6.6% of the total cases are reached. Therefore, giving values to parameters A, $T_{0.1}$, and D, from equation (2), custom curves can be obtained in order to recreate an outbreak. Figure 1 shows nine different scenarios generated from the combination of the three parameters. In each row of graphs, two of the three parameters are held constant. Parameter A is modified in the first row, T_α in the second one, and D in the third one. Note that the parameter D affects only the displacement of the curve on the t-axis. Therefore, only two parameters need to be manipulated to modify the shape.

2.3 Patient Flow Model

A discrete simulation model is able to mimic the stay of each patient in the hospital facilities to get an estimation of the extra resources needed to attend to all Covid-19 patients (Garcia-Vicuña et al. 2020b). The simulation model reproduces the patient pathway outlined in Figure 2. Each patient arriving at the hospital can be admitted to the hospital ward or directly to the ICU. Besides, those patients admitted to the wards may worsen their health status and require the transfer to the ICU. From both facilities, patients can die, so they abandon the system, or they can be discharged after improving their health status. In the last situation, patients in the ICU would be transferred to the hospital ward until they get over the disease.

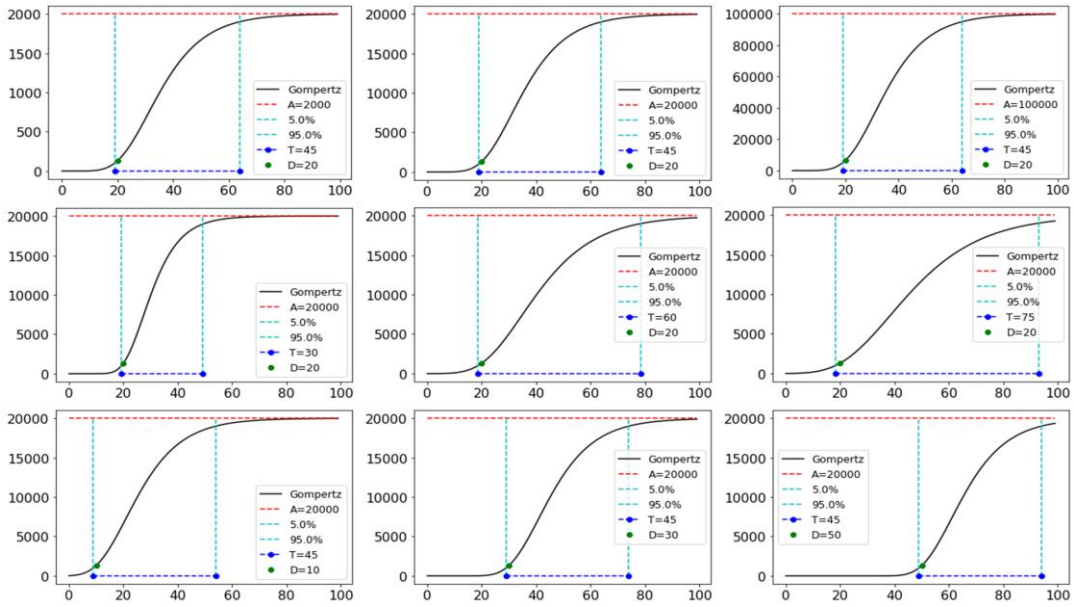

Figure 1: Nine Gompertz curves generated by fixing the three parameters of Equation 2. The effect they have on the curve can be clearly seen.

The purpose of the simulation model is the short-term necessary resources prediction. Therefore, the precision of the predictions strongly depends on the model accuracy in representing the health system's initial state and the time evolution of the already admitted patients and the next patients to be admitted. Usually, when the simulation is intended to investigate the behavior of systems in the long term, in its stationary state, the mathematical modeling of the simulation input is done by using historical data which is used to fit parameters and probability distributions. Or, if no data is available, estimations can be based on data collected from an extensive questionnaire designed to elicit expert knowledge. In both cases, the estimated parameters do not change during the simulation (or they may change according to a predefined strategy). However, when the simulation is dynamically used in real-time and no historical data is available from previous experiences and expert opinions are considered not accurate enough, the estimation of the parameters should be done dynamically as new observations are collected. Our research focuses on this dynamical estimation of the input parameters of the simulation model.

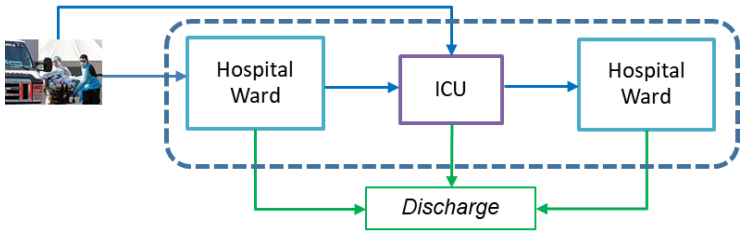

Figure 2: Representation of patient flow in the health system.

3 SIMULATION INPUT: ESTIMATION OF PARAMETERS AND PROBABIITY DISTRIBUTIONS

3.1 The Online Estimation Problem

We consider the problem of estimating the parameters associated with the pathway and LoS of patients hospitalized during a pandemic. This is a non-stationary situation, in which hospitalization parameters may

vary between different waves, between different places, and evolve during the pandemic. We propose to estimate them by using all data collected during the pandemic wave in which the simulation model is being applied, that is, from the time the first infected patient was admitted until the present time. However, the use of data associated with patients still hospitalized is a complex task, not only because of censorship but it is also unknown which event will be observed in the future and then from which variable is observed the value. Specifically, it is unknown whether a patient who has been hospitalized for some time will be finally admitted to the ICU or not, so it is unknown whether the observed value of the stay is a censored data for the variable Z, "*time until admission to the ICU*", or for the variable X, "*time to hospital discharge*". In this section, we propose an estimation method for the probability distributions of these variables, as well as the probability of admission to the ICU from the ward, p_{WI}, that uses the information of all patients admitted currently at the hospital (Figure 3, left). The same estimation methodology can be applied, to the estimation of the probability distributions of Y, "*LoS in the ICU before being transfer to hospital ward*", and Q, "*LoS in the ICU until death*", and p_{IW}, the probability of discharge to hospital ward (Figure 3, right).

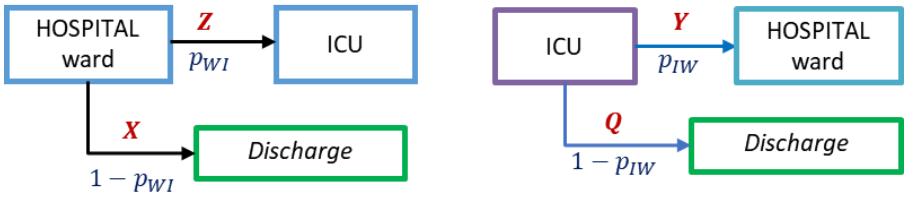

Figure 3: Parameters and probability distributions for modelling Ward-to-ICU transition (left) and ICU-to-Ward transition (right).

3.2 Data and Taxonomy of Patients

Hospital electronic health record systems provide information at the patient level that allows knowing the pathway of all patients admitted at the hospital and their current location, ward or ICU, in the case they are still admitted. Pathway and current location result from the times at hospital admission and discharge (t_{HA} and t_{HD}, respectively) and the times of ICU admission and discharge (t_{IA} and t_{ID}, respectively).

Therefore, for each patient i admitted to the hospital until time t, we suppose a known vector $\boldsymbol{u_i}(t)$ that contains these four times $\left(\boldsymbol{u_i}(t) = [t_{HA_i}, t_{HD_i}, t_{IA_i}, t_{ID_i}]\right)$. At time t, not all four times have been observed or are exactly known for all patients. For example, for an already discharged patient i from the hospital ward that did not need care in the ICU, the two components of vector $\boldsymbol{u_i}(t)$ related with the ICU are not observed and are left "empty" (we denote by the symbol $\emptyset$ this situation): $\boldsymbol{u_i}(t) = \left[t_{HA_i}, t_{HD_i}, \emptyset, \emptyset\right]$. For an admitted patient at the ICU at time t, it is known that both discharge times, from ICU and from hospital, will certainly happen but in a future time. This situation is denoted by $\boldsymbol{u_j}(t) = \left[t_{HA_j}, t, t_{IA_i}, t\right]$, and then $t - t_{IA_i}$ is a censored time for the LoS of this patient in the ICU. In turn, a patient j with $u_j(t) = \left[t_{HA_j}, t, \emptyset, \emptyset\right]$ indicates that he/she is still admitted to the hospital, and his/her LoS is censored by the value $t - t_{HA_j}$ and that so far the patient has not required admission into the ICU but it is not known if the admission will happen or not.

According to the values observed for the vector $u_i(t)$, 10 different types of patient states can be distinguished at time t. In the taxonomy of the type of patient, we use the letter H to refer to the hospital ward and the letter I to the ICU. The sequence of letters corresponds to the trajectory in the hospital facilities. The asterisk symbol $(*)$ indicates that the patient is still admitted into the facility indicated by the preceding letter, and therefore provides censored data.

- H: Patients with a full stay in the hospital ward who have not needed ICU ($[t_{HA}, t_{HD}, \emptyset, \emptyset]$).
- H^*: Patients with an incomplete stay in the hospital ward who have not needed ICU ($[t_{HA}, t, \emptyset, \emptyset]$). They do not have a discharge date and may or may not be admitted to the ICU.

- HI^* : Patients with an incomplete stay in the ICU transferred from the hospital ward $([t_{HA}, t, t_{IA}, t], t_{HA} < t_{IA})$.
- HI : Patients died in the ICU after being transferred from the hospital ward $([t_{HA}, t_{HD}, t_{IA}, t_{ID}], t_{HA} < t_{IA}, t_{HD} = t_{ID})$.
- HIH^*: Patients with a full stay in the ICU after being transferred from the hospital ward. They are still admitted to the hospital ward after leaving the ICU $([t_{HA}, t, t_{IA}, t_{ID}], t_{HA} < t_{IA})$.
- HIH: Patients with a full stay in the ICU after being transferred from the hospital ward. They have been discharged from the hospital ward after leaving the ICU $([t_{HA}, t_{HD}, t_{IA}, t_{ID}], t_{HA} < t_{IA}, t_{HD} > t_{ID})$.
- I^*: Patients admitted directly to the ICU with an incomplete stay $([t_{HA}, t, t_{IA}, t], t_{HA} = t_{IA})$.
- I: Patients admitted directly to the ICU who die there $([t_{HA}, t_{HD}, t_{IA}, t_{ID}], t_{HA} = t_{IA}, t_{HD} = t_{ID})$.
- IH^*: Patients admitted directly to the ICU with a full stay. They are still admitted to the hospital ward after leaving the ICU $([t_{HA}, t, t_{IA}, t_{ID}], t_{HA} = t_{IA})$.
- IH: Patients admitted directly to the ICU with a full stay. They have been discharged from the hospital ward after leaving the ICU $([t_{HA}, t_{HD}, t_{IA}, t_{ID}], t_{HA} = t_{IA}, t_{HD} > t_{ID})$.

3.3 Maximum Likelihood Estimation at the End of the Pandemic Wave

First, we address the problem of estimating the parameters and probability distributions involved in the patient transition Ward-to-ICU when, for each patient admitted to the hospital, the values of the vector $u_i(t)$ are fully known. Therefore, the pandemic wave is over and all patients have been discharged from the hospital (classified in the types H, HI, HIH, I and IH).

For the rest of the paper we denote:

- $n_T(t)$ as the number of patients of type T (generic) at time t. To simplify the notation, and when there is no confusion, we will use n_T instead of $n_T(t)$.
- $x(t) = (x_1(t), \dots\ x_i(t), \dots\ x_{n_X(t)}(t))$ the realization of the variable X, with $x_i(t) = t_{HD_i} - t_{HA_i}$.
- $z(t) = (z_1(t), \dots\ z_i(t), \dots\ z_{n_Z(t)}(t))$ the realization of the variable Z, with $z_i(t) = t_{IA_i} - t_{HA_i}$.
- θ_V as the vector of parameters of the distribution function of variable V.
- $\hat{\theta}_V$ as the estimation of the vector of parameters θ_V.
- $L_V(\theta_V|v(t))$: is the likelihood function of sample $v(t)$ used to estimate θ_V.

The estimation of the probability distribution parameters of the LoS variables X and Z is done by the maximum likelihood method.

$$L_X(\theta_X|x(t)) = \prod_{i=1}^{n_X} f_{\theta_X}(x_i) \to \hat{\theta}_X = \arg\max_{\theta_X} L_X(\theta_X|x(t))$$

$$L_Z(\theta_Z|z(t)) = \prod_{i=1}^{n_Z} f_{\theta_Z}(z_i) \to \hat{\theta}_Z = \arg\max_{\theta_Z} L_Z(\theta_Z|z(t))$$

Where $f_{\theta_X}(x_i)$ and $f_{\theta_Z}(z_i)$ are the density functions of variables X and Z, respectively, and $n_X(t) = n_H(t)$ and $n_Z(t) = n_{HI}(t) + n_{HIH}(t)$ are the sizes of the samples for variables X and Z.

The probability p_{WI} is estimated by the observed ratio of patients that are admitted to ICU from wards:

$$\hat{p}_{WI} = \frac{n_{HI} + n_{HIH}}{n_{HI} + n_{HI} + n_{HIH}}$$

3.4 The Expectation-Maximization Algorithm for Parameter Estimation during the Pandemic Wave

In this subsection, we develop an algorithm to estimate the probability p_{WI}, and the parameters of the probability distributions of variables X and Z at any time t during the development of the epidemic wave, making use of the information of all patients that have been admitted so far at the hospital. Let us consider the variable vector $W = (S, \delta)$ with S the time spent in hospital by a patient until discharge or until admission in the ICU and δ the indicator of whether the patient is admitted to the ICU or not.

$$\delta_i = \begin{cases} 1 \; if \; patient \; i \; is \; admitted \; in \; the \; ICU \\ 0 \; if \; patient \; i \; is \; not \; admitted \; in \; the \; ICU \end{cases}$$

Variable X, the time spent in hospital by a patient that is discharged from hospital with no admission in the ICU, verifies that $X \equiv S|\delta = 0$, and Z, the time spent in hospital by a patient before his/her admission in the ICU is $Z \equiv S|\delta = 1$. At the end of the pandemic the value of δ_i is observed for each patient i admitted to the hospital. However, at a time t, when the pandemic is not over, the value of δ_i is not known for patients already admitted at the hospital ward that have not been admitted in the ICU.

Specifically, for each patient i of type H ($t_{HD_i} < t$) or types HI, HIH, HI^*, HIH^* ($t_{IA_i} < t$) the value of the indicator variable δ_i has been observed, and the vector $w_i = (s_i, \delta_i)$ provides an observation for variable X in case $\delta_i = 0$ ($x_i = s_i$), or for Z in case $\delta_i = 1$ ($z_i = s_i$). However, for each patient i of type H^* (with $t_{HA_i} < t,\ t_{HD_i} > t \; and \; t_{IA_i} > t$ or $\emptyset$) the variable δ_i has not been observed at time t, and then $t - t_{HA_i}$ is a censored time that is not known to which variable corresponds, X or Z.

Suppose that at time t of the pandemic there are $n(t)$ patients that have been admitted at a hospital ward. We denote as $\boldsymbol{w}(t) = (w_1(t), \dots \; w_i(t), \dots \; w_{n(t)}(t))$ the realization of the variable W in these $n(t)$ patients. The vector $\boldsymbol{w}(t)$ can be divided in two parts $\boldsymbol{w}(t) = \left(\boldsymbol{w_F}(t), \boldsymbol{w_I}(t)\right)$: $\boldsymbol{w_F}(t)$, contains the observations of patients for which the value of δ_i has been observed, and $\boldsymbol{w_I}(t)$ contains the observations of those patients with unknown value for δ_i. We have developed an iterative procedure, based on the EM (Expectation-Maximization) algorithm, to estimate the distribution functions of variables X and Z and the probability p_{WI}. First, an initial estimation of the parameters is carried out by only using the fully-known data (those observations with known value for δ_i). In the main iteration, the estimated parameters are used to update the probability of being admitted to the ICU for each one of the patients admitted in the ward. These updated probabilities are used to calculate a new likelihood function for the parameters, which is maximized to obtain a new estimation of the probability distribution parameters. These two steps (updating ICU admission probabilities and getting and maximizing new likelihood function) are repeated until stopping criteria are met. We use the following additional notation:

- $\hat{\theta}_X^{(k)}$: is the estimation of vector θ_X in the k-th iteration of the algorithm.
- $\hat{\theta}_Z^{(k)}$: is the estimation of vector θ_Z in the k-th Iteration of the algorithm.
- $\hat{F}_X^{(k)}(x) = F_X\left(x; \hat{\theta}_X^{(k)}\right)$: is the distribution function of X with parameters $\hat{\theta}_X^{(k)}$.
- $\hat{F}_Z^{(k)}(x) = F_Z\left(x; \hat{\theta}_Z^{(k)}\right)$: is the distribution function of Z with parameters $\hat{\theta}_Z^{(k)}$.
- $\hat{P}_{WI}^{(k)}$: is the estimation of the probability P_{WI} in the k-th iteration of the algorithm.
- $n_F(t)$: number of patients with full information at time t (the size of vector $\boldsymbol{w_F}(t)$).
- $n_I(t)$: number of patients with incomplete information at time t (the size of vector $\boldsymbol{w_I}(t)$).

To simplify the notation, and when there is no confusion, we will use n_F and n_I, instead of $n_F(t)$ and $n_I(t)$, respectively.

Steps of the algorithm

1. Initialization.

$k = 0$

Estimate the parameters θ_X, θ_Z and the probability P_{WI} by using the data in the vector $\boldsymbol{w_F}(t)$:

$$\hat{P}_{WI}^{(\mathrm{k})} = \frac{1}{n_F}\sum_{i=1}^{n_F} 1_{\{\delta_i=1\}}$$

$$L_X^{(\mathrm{k})}(\theta_X|\boldsymbol{w_F}(t)) = \prod_{i=1}^{n_F} f_{\theta_X}(s_i)\,(1-\delta_i) \rightarrow \hat{\theta}_X^{(\mathrm{k})} = \arg\max_{\theta_X} L_X^{(\mathrm{k})}(\theta_X|\boldsymbol{w_F}(t))$$

$$L_Z^{(\mathrm{k})}(\theta_Z|\boldsymbol{w_F}(t)) = \prod_{i=1}^{n_F} f_{\theta_Z}(s_i)\,\delta_i \rightarrow \hat{\theta}_Z^{(\mathrm{k})} = \arg\max_{\theta_Z} L_Z^{(\mathrm{k})}(\theta_Z|\boldsymbol{w_F}(t))$$

2. Repeat until the stop criteria is met.

Iteration $k+1$. From the k-th iteration, $k \geq 0$, we know the estimations $\hat{\theta}_X^{(k)}$, $\hat{\theta}_Z^{(k)}$, and $\hat{P}_{WI}^{(k)}$; of the parameters θ_{X_1}, θ_Z, and the probability P_{WI}.

The iteration is divided into two steps: in the first one the calculation of the expected value of the indicator function δ_i in each patient with incomplete data is carried out, which allows estimating the probability of admission in the ICU, P_{WI}, and the expectation of the likelihood function when all data, complete and incomplete, are considered. The second step estimates θ_X and θ_Z by maximizing the likelihood functions estimated in the previous step.

2.1 Expectation.

For each patient i, the probability of being admitted to the ICU is updated as the posterior probability given the time already spent at the ward:

$$\hat{P}_i^{(k+1)} \equiv \hat{P}^{(k+1)}\left(\delta_i = 1 \,\middle|\, \hat{P}_{WI}^{(k)}, \hat{\theta}_X^{(k)}, \hat{\theta}_Z^{(k)}\right) = \frac{\left(1-\hat{F}_Z^{(k)}(s_i)\right)\hat{P}_{WI}^{(k)}}{\left(1-\hat{F}_Z^{(k)}(s_i)\right)\hat{P}_{WI}^{(k)} + \left(1-\hat{F}_X^{(k)}(s_i)\right)\left(1-\hat{P}_{WI}^{(k)}\right)}$$

The updated probabilities of being admitted in ICU for each patient of type H* allows to update the unconditional probability of admission in the ICU:

$$\hat{P}_{WI}^{(k+1)} = E\left[\delta = 1 \,\middle|\, \hat{P}_{WI}^{(k)}, \hat{\theta}_X^{(k)}, \hat{\theta}_Z^{(k)}\right] = \frac{1}{n(t)}\sum_{i=1}^{n(t)} E\left[1_{\{\delta_i=1\}}\right] = \frac{1}{n(t)}\left(\sum_{i=1}^{n_F} 1_{\{\delta_i=1\}} + \sum_{i=1}^{n_I} \hat{P}_i^{(k+1)}\right)$$

and, the likelihood functions of the sample as expected functions:

$$L_X^{(k+1)}(\theta_X|\boldsymbol{w}(t)) = E[L_X(\theta_X|\boldsymbol{w}(t))] = \prod_{i=1}^{n_F} f_{\theta_X}^{(k)}(s_i)\,(1-\delta_i)\prod_{i=1}^{n_I}\left(1-F_{\theta_X}^{(k)}(s_i)\right)(1-\hat{P}_i^{(k+1)})$$

$$L_Z^{(k+1)}(\theta_Z|\boldsymbol{w}(t)) = E[L_Z(\theta_Z|\boldsymbol{w}(t))] = \prod_{i=1}^{n_F} f_{\theta_Z}^{(k)}(s_i)\,\delta_i\prod_{i=1}^{n_I}\left(1-F_{\theta_Z}^{(k)}(s_i)\right)\hat{P}_i^{(k+1)}$$

2.2 ***Maximization***. The likelihood functions are maximized to find the estimation of the parameters.

$$\hat{\theta}_X^{(k+1)} = arg\max_{\theta_X}\left(L_X^{(k+1)}(\theta_X|\boldsymbol{w}(t))\right)$$

$$\hat{\theta}_Z^{(k+1)} = arg\max_{\theta_Z}\left(L_Z^{(k+1)}(\theta_Z|\boldsymbol{w}(t))\right)$$

3. **Stop criteria**. Repeat Step 2 until the sequence of values of the likelihood function or the values of the estimated parameters converges:

$$\left|L_X^{(k+1)}\left(\hat{\theta}_X|\boldsymbol{w}(t)\right) - L_X^{(k)}\left(\hat{\theta}_X|\boldsymbol{w}(t)\right)\right| \leq \varepsilon_1$$
$$\left|\hat{\theta}_X^{(k+1)} - \hat{\theta}_X^{(k)}\right| \leq \varepsilon_2$$

4 RESULTS

4.1 Experimental Design

A simulation experiment has been designed to validate the estimation method developed in Section 3 (named EM method) and its impact on the quality of the predictions made on the necessary resources, specifically, about the necessary ICU beds. The simulation of a pandemic wave requires the simulation of the patient arrival process, which is carried out using the Gompertz model exposed in (2), and the generation of a trajectory in the hospital for each patient, which is carried out according to the flow of patients shown in Figure 2.

Specifically, generated pandemic waves mimic those observed in reality (Garcia-Vicuña et al. 2020a) with the parameters for the arrival and hospitalization processes described below. Each pandemic wave simulation generates data over time that is used to test the performance of the proposed estimation method EM and its comparison with the other two estimation methods. The results about the estimation accuracy are shown in Section 4.2 and the impact on the precision of the predictions in Section 4.3.

Gompertz model parameters. Patient arrivals are generated according to the accumulated curve described by equation (2) with parameters $A = 2000, 20000, and\ 100000$, $T_{0.1} = 60$, and $D = 18$. That is, a pandemic wave that spreads 60 days (to account for the 90% central cases) and varying sizes, from 2000 (corresponding to a small region) to 100000 (corresponding to a medium-size country). Daily admissions are determined by the difference between the value of the Gompertz curve in two consecutive days, rounded to the nearest integer number.

Patient hospital path. Probability distributions for the LoS are assumed to be Weibull ($W(\alpha, \beta)$, where α is the scale parameter and β the shape parameter): LoS in the hospital ward of a patient not needing ICU (variable $\boldsymbol{X}$) $W(10.74, 1.25)$, the time spent by a patient in the hospital ward before transfer to the ICU (variable $\boldsymbol{Z}$) $W(5.06, 0.98)$. In addition, the LoS of a patient in the ICU $W(18.91, 1.15)$, and the LoS of a patient in the hospital ward after being discharged from the ICU $W(12.90, 1.4)$.

The probability of a patient initially admitted to a ward requiring transfer to ICU (p_{WI}) (0.073). In addition, the probability of direct admission to ICU upon arrival (0.021), and the probability of patient transfer from ICU to hospital ward (0.75).

Estimation methods. Three parameter estimation methods are compared. The first one consists of assuming that all available information is completely known (method I). That is, the probability p_{WI} is estimated by the ratio of patients that were admitted at the ICU divided by patients admitted at a hospital ward, to date. Therefore times at the hospital of patients of type H^* are considered censored times for variable X. The second method is similar to the first one, but only patients that were admitted more than 5 days before are considered for the estimation calculations (method I-5). This avoids overestimating the number of patients who will not require ICU in the future. Finally, the third method is the one developed in Section 3.3 (method EM).

4.2 Parameter Estimation Accuracy

This section assesses the accuracy in the estimation of the probability p_{WI} and the scale and shape parameters of the Weibull distribution of variables X and Z as the pandemic progresses. The simulation

model recreates the patients' arrival and their stay at the hospital according to the parameters and probability distributions fixed in Section 4.1. After simulating one day of the pandemic, the three estimation methods are applied to estimate the probability parameters and distributions. The results of these estimations are shown in Figure 4. Therefore, the differences in the estimations produced by the three methods are assigned to the differences in their prediction capability and not to the randomness of the simulation because it is controlled and the same for each method.

Figure 4 shows the evolution of the estimation of parameters p_{WI} and θ_Z over time ($\hat{\theta}_X$ values show small differences among different methods). Each pandemic scenario is simulated 10 times and the results show the estimation averages over the 10 runs. The three methods provide results that converge to the true value of the estimated parameter. However, the method EM has a fast convergence in all simulated cases, which turns to be important when the simulation model is used as a prediction tool for the resources needed in the future, as we expose in the next subsection.

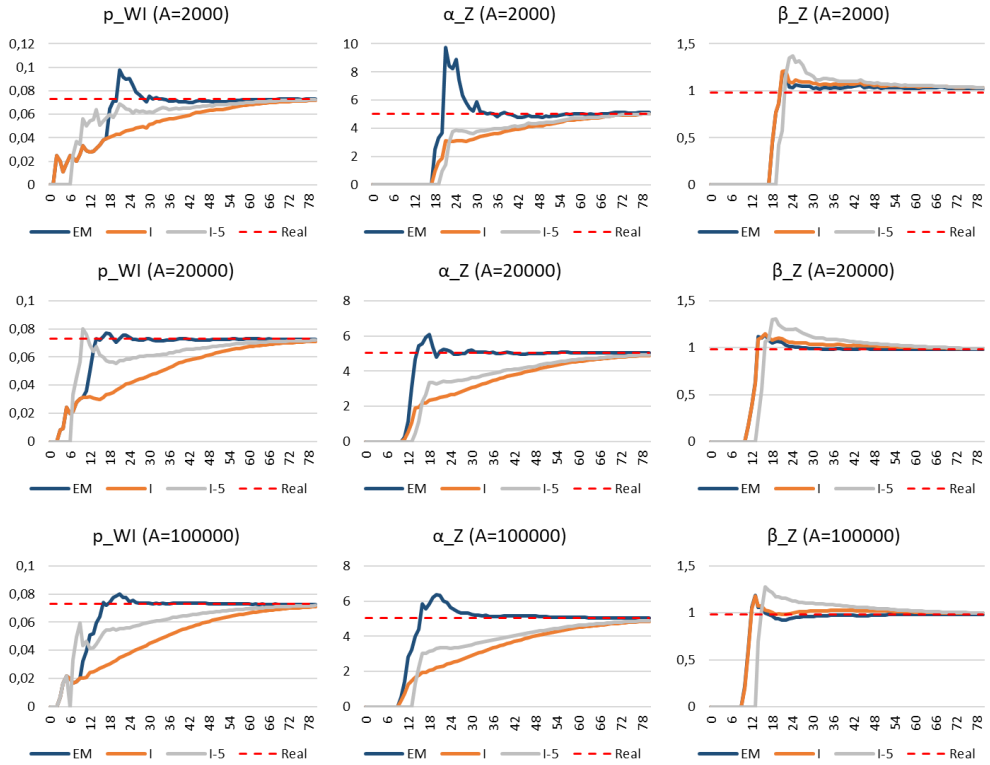

Figure 4: Estimation of parameters p_{WI} and θ_Z over time (the horizontal axes represent the time during the pandemic) with the 3 methods (I, I-5, and EM) and the real values. Results are shown for different values of parameter A of the Gompertz curve (2000, 20000, and 100000).

4.3 Impact on the Simulation Output. Bed Occupancy Prediction Accuracy

The objective of the simulation model is to predict the future bed occupancy level during the course of the pandemic wave. The predictions of the simulation model are obtained by the statistical analysis of the output of many simulation model runs. In this section, we evaluate the quality of the predictions made with the simulation model with each of the three estimation methods. For each estimation method, predictions are obtained at different times of the pandemic evolution by simulating patient pathways and LoS by using the respective estimated branching probabilities and probability distribution parameters. The results obtained from each method are compared with those obtained by simulating using the true value of the parameters and probabilities.

Once the prediction day is set, many simulations are run with each method and the predictions obtained are compared with those made from the actual parameters. Figure 5 shows nine predictions of ICU bed occupancy made with all methods from 3 different days (20th, 25th, and 30th). Note that these days are quite far away from the peak occupancy. The green line in each graph represents the evolution of the simulated pandemic up to the Simulation Starting Point (SSP), which is represented by a black dot. For each prediction, the 5th percentile (P5) and the 95th percentile (P95) are plotted. As the pandemic progresses, the predictions of occupation become closer to reality. But in all cases, the *EM* method is the closest.

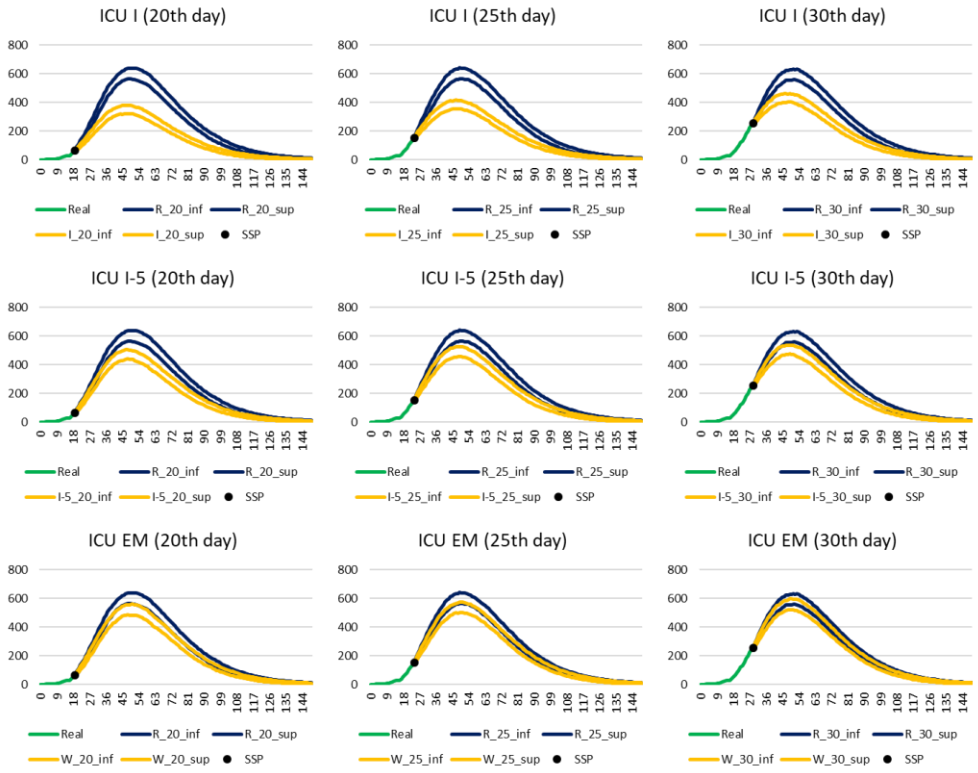

Figure 5: Prediction of ICU bed occupancy on the 20th, 25th, and 30th days of the pandemic with the 3 methods (I, I-5, and EM) compared to prediction with actual parameters.

5 CONCLUSIONS

In this work, we have proposed a new method suitable for the online estimation of hospital simulation-model input parameters. The method is based on an EM algorithm which allows the use of the data provided by all patients admitted so far. This characteristic is a big advantage when only a small ratio of patients have been discharged and can provide full information about their hospital pathway and LoS. Simulation tests have shown a better performance than other estimation methods that use only complete information. Poor estimation of the parameters and probabilities leads to poorer estimations of the output variables of interest as the number of beds necessary to attend to all pandemic coming patients.

The estimation method has been applied to the estimation of the admission probability to the ICU from the ward and to the parameters of the probability distributions of variables LoS in the hospital ward and the time to admission to the ICU. The same stochastic situation occurs for the admitted patients to the ICU, since it is unknown whether at the end of their stay they will be transferred back to the ward, due to improved health, or will leave the ICU due to death. The application of the estimation method to this case allows estimating the probability of recovery, and the parameters of the probability distributions of the time until transfer to the ward and the time until death.

Note that, in addition to the probabilities and parameters set forth above, it is only necessary to estimate the probability of direct admission to the ICU and the LoS in the ward after transfer from the ICU to complete the estimation of all the variables and parameters that describe the randomness of the pathways and LoS of patients through the hospital, as it is described in Figure 2. The probability of direct admission to the ICU is estimated by means of the observed proportion, which obviously uses the information of all the patients admitted so far. The probability distribution parameters of the LoS in the hospital ward after ICU can be estimated by maximum likelihood: those patients already discharged provide an exact value, while those who are still hospitalized provide a censored value.

ACKNOWLEDGMENTS

This paper was supported by the COVID grant of Navarre's Government 0011-3597-2020-000003 and the grant PID2020-114031RB-I00 (AEI, FEDER EU).

REFERENCES

Cássaro, F.A.M., and L.F. Pires. 2020. "Can we predict the occurrence of COVID-19 cases? Considerations using a simple model of growth". *Sci Total Environ* 728:138834.

Currie, C.S.M., J.W. Fowler, K. Kotiadis, et al. 2020. "How simulation modelling can help reduce the impact of COVID-19". *J Simul* 14(2):83–97.

Garcia-Vicuña, D., L. Esparza, and F. Mallor. 2020a. "Hospital preparedness in epidemics by using simulation. The case of COVID-19 Authors". medRxiv. https://doi.org/10.1101/2020.08.12.20173328

Garcia-Vicuña, D., F. Mallor, and L. Esparza. 2020b. "Planning Ward And Intensive Care Unit Beds For Covid-19 Patients Using A Discrete Event Simulation Model". In: Proceedings of the 2020 Winter Simulation Conference, edited by Bae, K.; Feng, B.; Kim, S.; Lazarova-Molnar, S.; Zheng, Z.; Roeder, T.; and Thiesing, R., editor(s), pages 759–770. Piscataway New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Gompertz, B. 1825. "On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies". *Philos Trans R Soc London B Biol Sci* 182:513–585.

Liu, W., S. Tang, and Y. Xiao. 2015. "Model Selection and Evaluation Based on Emerging Infectious Disease Data Sets including A/H1N1 and Ebola". *Comput Math Methods Med* 2015:207105.

Malavika, B., S. Marimuthu, M. Joy, et al. 2021. "Forecasting COVID-19 epidemic in India and high incidence states using SIR and logistic growth models". *Clin Epidemiol Glob Heal* 9:26–33.

Richards, F.J. 1959. "A flexible growth function for empirical use". *J Exp Bot* 10(2):290–301.

Ricker, W.E. 1979. "Growth rates and models". *Fish Physiol* 8:677–743.

Sánchez-Villegas, P., and A.D. Codina. 2020. "Modelos predictivos de la epidemia de COVID-19 en España con curvas de Gompertz". Gac Sanit. https://doi.org/10.1016/j.gaceta.2020.05.005

Shen, C.Y. 2020. "Logistic growth modelling of COVID-19 proliferation in China and its international implications". *Int J Infect Dis* 96:582–589.

Stannard, C.J., A.P. Williams, and P.A. Gibbs. 1985. "Temperature/growth relationships for psychrotrophic food-spoilage bacteria". *Food Microbiol* 2(2):115–122.

Zwietering, M.H., I. Jongenburger, F.M. Rombouts, and K. van 't Riet. 1990. "Modeling of the bacterial growth curve". *Appl Environ Microbiol* 56(6):1875–1881.

AUTHOR BIOGRAPHIES

DANIEL GARCIA-VICUÑA studied industrial engineering at the Public University of Navarre, Spain. Currently, he is a Ph.D. student at the Institute of Smart Cities of the Public University of Navarre. His research interests lie in the field of complex real problems simulation modelling. His email address is daniel.garciadevicuna@unavarra.es.

FERMIN MALLOR is Full Professor of Statistics and Operations Research at the Public University of Navarre, Spain. He holds a Ph.D. and MSc in Mathematics. His research interests include applications of simulation-optimization (classical and metaheuristics methods) in health, energy, logistics, and production. He is the director of the research group q-UPHS (quantitative methods for Uplifting the Performance of Health Services, www.unavarra.es/quphs). His email address is mallor@unavarra.es.