ORIGINAL PAPER

Computational simulation of the COVID-19 epidemic with the SEIR stochastic model

Carlos Balsa[1] (ID) · Isabel Lopes[2,3] · Teresa Guarda[4,5,6] · José Rufino[7]

Abstract

A small number of individuals infected within a community can lead to the rapid spread of the disease throughout that community, leading to an epidemic outbreak. This is even more true for highly contagious diseases such as COVID-19, known to be caused by the new coronavirus SARS-CoV-2. Mathematical models of epidemics allow estimating several impacts on the population and, therefore, are of great use for the definition of public health policies. Some of these measures include the isolation of the infected (also known as quarantine), and the vaccination of the susceptible. In a possible scenario in which a vaccine is available, but with limited access, it is necessary to quantify the levels of vaccination to be applied, taking into account the continued application of preventive measures. This work concerns the simulation of the spread of the COVID-19 disease in a community by applying the Monte Carlo method to a Susceptible-Exposed-Infective-Recovered (SEIR) stochastic epidemic model. To handle the computational effort involved, a simple parallelization approach was adopted and deployed in a small HPC cluster. The developed computational method allows to realistically simulate the spread of COVID-19 in a medium-sized community and to study the effect of preventive measures such as quarantine and vaccination. The results show that an effective combination of vaccination with quarantine can prevent the appearance of major epidemic outbreaks, even if the critical vaccination coverage is not reached.

Keywords SEIR stochastic model · COVID-19 · Numerical simulations · Parallel computing

---

Extended author information available on the last page of the article

1 Introduction

An epidemic is the rapid spread of an infectious disease within a population, producing many infected individuals in a short period of time. These may ultimately die or become permanently incapacitated. Infectious diseases that can trigger an epidemic outbreak include, among other well-known diseases, HIV (May and Anderson 1987), Smallpox (Dasaklis et al. 2017), Varicella (Zha et al. 2020), SARS (Meyers et al. 2005), (H1N1) influenza (Mao and Bian 2010) and the recent COVID-19, that is responsible for the actual pandemic situation (He et al. 2020).

The mathematical modeling of an epidemic allows to describe the spreading of an infectious disease and to predict its future course. This information is of vital importance to the definition of public policies to control the spread of an epidemic. For instance, mathematical models allow to estimate several effects of an epidemic, like the total number of infected people or the duration of the epidemic, as well as the effects of possible prevention measures such as social distance, immunization or isolation, just to name the most common.

Epidemic mathematical models fall in two main broad classes: *deterministic* and *stochastic*. Typically, deterministic models are suitable to model epidemics within large communities. The "epidemic process" in these models is governed by a system of differential equations and the evolution of the process is deterministic in the sense that no randomness is allowed. The community is considered homogeneous and it is assumed that individuals mix uniformly with each other. Hence, these models do not incorporate any arbitrariness (Hethcote 2000).

Stochastic models, in turn, have been successfully applied to closed homogeneous communities. Since the spread of an infectious disease is a random process (that occurs locally through the close contact with infectious individuals), stochastic models are more realistic (Britton 2010). These models can be simulated by using Monte Carlo based methods. Essentially, this consists in the repetition of a random behaviour of the population, a large number of times. This procedure may have a high computational cost, even for small-size populations.

The Susceptible-Exposed-Infectious-Removed (SEIR) mathematical epidemic model is the most suited to describe the spread of an infectious disease with latency period, like COVID-19. In the generic SEIR model, the population is divided into four compartments that represent *susceptible*, *exposed*, *infectious* and *recovered* individuals. It is known that the SARS-CoV-2, the causative agent of COVID-19, has an incubation period [time between infection and onset of symptoms (He et al. 2020)]. Models without the exposed compartment cannot adequately address the time delay between the build-up of exposed and infectious individuals. This SEIR model also assumes that an individual is contagious only when in the infectious period. This compartmentalization does not correspond exactly to the reality of COVID-19, because the duration of periods of latency and contagion vary from person to person (He et al. 2020). Nevertheless, the SEIR model is currently the one that most closely reproduces the characteristics of the propagation of the COVID-19 disease.

The SEIR model has been largely used to model and predict the spread of COVID-19 in different regions of the globe, like China (Li et al. 2020), India (Chatterjee et al. 2020), South Africa (Mukandavire et al. 2020), Italy/Lombardia (Carcione et al. 2020) and Germany (Engbert et al. 2020). Many of these works calibrate the mathematical model with empirical data in order to have a realistic estimation of the different model parameters. However, there are certain difficulties related to the availability of representative data. For instance, most of the available information does not include undocumented infections (asymptomatic undetected) (Li et al. 2020), making it difficult to estimate realistic parameters.

The focus of this paper is the stochastic SEIR epidemic model and its computational simulation regarding the spreading of the COVID-19 infectious disease in a small-size community (20,000 individuals), using model parameters published recently. Different scenarios are explored, corresponding to hypothetical control measures based on quarantine and vaccination. Specifically, we look for the critical vaccination coverage needed to control the disease for different quarantine scenarios. In fact, in a possible scenario where a vaccine is available, but in a conditioned way (both in terms of effectiveness and in terms of quantity), the prevention of major epidemiological outbreaks can only be possible with the isolation (quarantine) of a fraction of the infectious ones.

To minimize the duration of simulations, a strategy similar to the one used in a previous work (Balsa et al. 2020) was adopted: each execution of the stochastic SEIR model, with specific parameters, accommodates randomness by running a certain number of independent simulations; these are executed in parallel in an HPC cluster; at the end of the model execution, results from individual simulations are joined; additionally, in this work, the cluster batch service was used to submit and dispatch different executions of the model, each one with different model parameters, allowing for a thorough parametric study.

The rest of this paper is structured as follows: Sect. 2 introduces a SEIR deterministic epidemic model and defines its stochastic version; Sect. 3 provides details on the computational strategy used to conduct the simulations with the stochastic SEIR model; section 4 presents and discusses the results of the computational simulations; Sect. 5 provides some final considerations.

2 SEIR epidemic model

A typical approach when modeling the spread of an infectious disease in a community is to formulate a compartmental model. The Susceptible-Exposed-Infective-Recovered (SEIR) compartmental model considers that, at each point in time, each and any individual of the population belongs to one (and only one), of the following compartments: S—an individual is Susceptible to catch the disease; E—an individual is Exposed if he is infected but does not transmit the disease, because the disease is in the incubation period; I—an individual is Infective, meaning he has got the disease and is able to infect others; R—an individual has Recovered from infection and can no longer be infected again. Thus, in the generic SEIR model, an individual can only move from compartment S to E, from compartment E to I, and then from

compartment I to R. As their name suggests, the compartments of the SEIR model are closely related to the four different stages/states of the disease: susceptible state, incubation period, infectious period, and recovered state (immunized).

In this work, the generic SEIR model is slightly modified to reflect two important containment measures: quarantine and vaccination. A new Q compartment corresponds to infected individuals, placed in isolation (quarantine), who cannot infect others. Vaccinated individuals are included in compartment R, as they cannot be infected again. Another new compartment, D, is for individuals who died as a result of the infectious disease. The model assumes that births and natural deaths are balanced, and that the population members only decrease due to COVID-19, as dictated by the fatality rate of this disease.

Figure 1 shows a representation of the different compartments of the modified SEIR model, as well as the flow (by unit of time) of individuals between them. The meaning of the various parameters that define the flows between the compartments is presented in the following section.

2.1 Deterministic SEIR model

Let us consider a population or community of size $n \in \mathbb{N}$. Let $s(t)$, $e(t)$ $i(t)$, $q(t)$, $r(t)$ and $d(t)$ denote, respectively, the fraction of the community that belongs to the compartments S, E, I, Q, R and D, illustrated in Fig. 1, at a given instant $t \geq 0$ in time. Assuming that the population is closed, meaning that there are no births, deaths and immigration or emigration during the study period, then $s(t) + e(t) + i(t) + q(t) + r(t) + d(t) = 1$, for any t.

Assuming that the population is homogeneous, individuals mix uniformly, and $s(t)$, $i(t)$, $q(t)$, $d(t)$ and $r(t)$ are differentiable functions, then the variation of the fraction of the population in compartments S, E, I, Q, R and D, along the time t (in days), is given by the system of differential equations
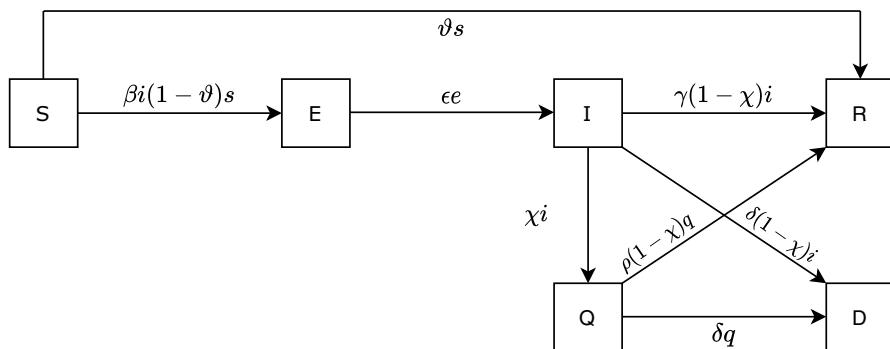

Fig. 1 SEIR model with quarantine and dead compartment

$$\begin{cases} \frac{ds}{dt} = -\beta i(1-\vartheta)s - \vartheta s \\ \frac{de}{dt} = \beta i(1-\vartheta)s - e \\ \frac{di}{dt} = e - \chi i - \gamma(1-\chi)i - \delta(1-\chi)i \\ \frac{dq}{dt} = \chi i - \delta q - \rho(1-\delta)q \\ \frac{dr}{dt} = \gamma(1-\chi)i + \vartheta s + \rho(1-\delta)q \\ \frac{dd}{dt} = \delta q + \delta(1-\chi)i \end{cases} \quad (1)$$

with initial conditions $s(0) = s_0, i(0) = i_0, q(0) = q_0, r(0) = r_0$ and $d(0) = d_0$.

In the system of ordinary differential equations (1), β is the rate of contacts between susceptible and infectious individuals, $1/\epsilon$ is the incubation period, $1/\gamma$ is the infectious period, χ is the fraction of infectious individuals insulated (quarantined), δ is the fraction of infectious individuals that died, and $1/\rho$ is the quarantine period. The system (1), along with the initial values $s(0) = 1 - i_0, e(0) = 0$, $i(0) = i_0, q(0) = 0, r(0) = 0$ and $d(0) = 0$ (where $i_0 > 0$ represents the initial fraction of infective and is assumed small), fully define the dynamics of the deterministic SEIR model considered in this work.

The computational simulation of the deterministic SEIR epidemic model is relatively straightforward. The initial value problem (1) can be rapidly solved by a numerical method like the four-order Runge–Kutta; also, the computational time does not depend on the dimension n of the population.

The parameters ϵ and γ are disease-specific for COVID-19, while the contact rate β is behaviour-specific and is different for each region, country or culture, and can vary in time to reflect societal and political actions (Linka et al. 2020). Many recent publications present values of these parameters, some obtained by fitting the SEIR model to the COVID-19 epidemiological data made available by public health entities in several countries, others from published clinical results. As shown in Table 1, there are authors who present various values for the rate of contact. This is due to the attempt to monitor the variation in the contact rate, which varies with the application of public health measures such as the use of mandatory masks, social distance or population *lockdown*.

The quarantine period $1/\rho$ is set by the health care policies of each country. In most European countries this period is 14 days. The fraction χ of infectious individuals that are insulated/quarantined includes documented infected and hospitalized individuals. Its value depends on the number of tests performed.

Table 1 Estimates of effective SEIR model parameters for the COVID-19 pandemic context

Authors/region	β	$1/\epsilon$	$1/\gamma$
Mukandavire et al. (2020)/South Africa	1.3	3.21	2.27
Chatterjee et al. (2020)/India	–	5.1	7
Li et al. (2020)/China	0.35, 0.52, 1.12	3.42, 3.6, 3.69	3.14, 3.31, 3.47
Carcione et al. (2020)/Italy	0.2, 0.75	4.24	4.02
Linka et al. (2020)/Europe	–	2.5	6.6

The mortality rate δ depends on the aggressiveness of the disease and the existence of susceptible people who are most vulnerable. COVID-19 is known to be more aggressive in older individuals, so this rate is higher in older communities. The mortality rate is related to the infection fatality rate (IFR). The IFR rate is based on all the population that has been infected along the outbreak, including the undocumented infected individuals. In terms of the global SEIR compartmental model, the IFR rate is given by

$$\text{IFR} = \frac{d_\infty}{r_\infty + d_\infty} \tag{2}$$

where ∞ refers to the end of the epidemic outbreak ($t \to \infty$), and so $r_\infty + d_\infty$ corresponds to the total individuals that have been infected, as the sum of the the recovered (r_∞) and the dead ones (d_∞). Furthermore, in the general SEIR model case, and assuming that $\delta \ll \gamma$, it is shown in Carcione et al. (2020) that

$$\text{IFR} \approx \frac{\delta}{\gamma}. \tag{3}$$

As a specific case, the IFR rate in Portugal is actually estimated around 4% (DGS 2020). Then in agreement with (3) and considering $\gamma \approx 0.23$, the corresponding death rate is $\delta \approx 0.0093$.

The ratio

$$R_0 = \frac{\beta}{\gamma + \delta} \tag{4}$$

can be interpreted as the average number of new infections caused by an infectious individual when inserted in a community made up only of susceptible people (Carcione et al. 2020). This ratio is also referred to as the *basic* reproduction number of the general SEIR model. When an infectious disease has $R_0 > 1$ and an infected is inserted in a susceptible community, then an epidemic outbreak takes place, infecting a substantial part of the population. On the other hand, when $R_0 < 1$, there is no risk of a major epidemic outbreak (for details see Hethcote 2000).

The value of the reproduction number varies with the evolution of the epidemic, due to the variation in the contact rate. The contact rate β (the rate at which an infectious individual comes into contact and affects others), is not constant; instead, it is modulated by social behaviour and public health policies. The quantification of new infections during the pandemic, in a specific day t, is done through the *effective* reproduction number, R_t, which takes into account the contact rate at time t, $\beta(t)$. Consequently, the *effective* reproduction number enables to measure the effects of public health interventions.

The *basic* reproduction number is a measure of the potential that a disease has to trigger an epidemic outbreak, which is why it is essential to estimate its value. Since the appearance of COVID-19, many works have been published on this topic. In Liu et al. (2020) and Mukandavire et al. (2020) one can find a list of R_0 values obtained by different researchers from all over the world, where the

values collected vary widely between them, ranging from 1.95 to 6.49. Linka et al. (2020) estimated the R_t values for the 27 countries in Europe based on data provided by the published health entities in each country. The values found for the beginning of the outbreak range from 0.91 (Latvia) to 6.33 (Germany), with the general value obtained for Europe being 4.22.

As a major outbreak will not occur if $R_0 < 1$, it is then possible to estimate the fraction of the population that needs to be immunized, in order for an epidemic to be prevented. If immunization is done through vaccination, this fraction is called the *critical vaccination coverage* and is given by

$$v_c = 1 - \frac{1}{R_0}. \tag{5}$$

The critical vaccination coverage is a very important quantity: if more than this fraction of the population is vaccinated before an outbreak, then the whole community is protected from an epidemic, and not only the vaccinated, a situation called *herd* immunity (Britton and Pardoux 2019). For the *basic* reproduction number $R_0 = 4.22$, found for Europe (Linka et al. 2020), the *herd* immunity threshold would be 76%. If immunization is achieved only by contracting the disease, then *herd* immunity will only be achieved after at least 76% of the population has been infected. A country like Portugal, with a prevalence of 0.69% (Direcção 2020), is a long way from achieving *herd* immunity, even if unrecorded asymptomatic cases are included.

2.2 Stochastic SEIR model

In a homogeneous community where individuals mix uniformly with each other, the size of an epidemic outbreak is determined by the *basic reproduction number* (R_0) and by the incubation and infectious periods ($1/\epsilon$ and $1/\gamma$), and there is no incertitude or randomness in the final number of infected individuals. However, in some cases, this contradicts reality: for instance, the introduction of a small number of infected in a community may not necessarily have as a consequence a large outbreak (even if $R_0 > 1$); such situation can happen, for example, if the infectious isolate themselves from the rest of the community or, by chance, if their contacts are restricted to immune individuals. This motivates the formulation of models of a stochastic nature.

A stochastic version of the SEIR model incorporates the randomness inherent in the spread of an infectious disease. To introduce this version, we assume again a closed homogeneous and uniformly mixing community of n individuals. However, the number of susceptible, exposed, infectious, quarantined, recovered and dead individuals at time t, is now $S(t) = n \times s(t)$, $E(t) = n \times e(t)$, $I(t) = n \times i(t)$, $Q(t) = n \times q(t)$, $R(t) = n \times r(t)$ and $D(t) = n \times d(t)$, respectively.

If at time $t = 0$ there is a number of m infectious, then $S(0) = n - m$, $E(0) = 0$, $I(0) = m$, $Q(0) = 0$, $R(0) = 0$ and $D(0) = 0$. The model assumes there are no infectious contact during the latent state (E) and then, suddenly, when the latent period

ends, the rate of infectious contact becomes β until the infection period ends, when it suddenly drops down to 0 again.

Infectious individuals (I) have contact (adequate to the transmission of the disease) with other individuals randomly in time, according to a Poisson process with intensity β. Each contact is made with an individual randomly selected from the community. Any susceptible (S) that receives such contact immediately becomes exposed (E) for a random period that follows an exponential distribution with mean $1/\epsilon$. The infectious periods are also independent and exponentially distributed with a mean $1/\gamma$. The quarantine (χ) and death (δ) rates, as well as the quarantine period, are assumed to be constant.

Despite assuming a closed homogeneous community, the stochastic SEIR model is more realistic than its deterministic version. This is because the number of infectious contacts, and the latency and infectious periods, are all random, varying among individuals. Surely, it would be possible to develop an even more realistic model, allowing individuals to die from natural causes and others to be born, and accommodating more social structures (see Britton and Pardoux 2019, for instance), but this work does not yet consider these extensions.

3 Computational simulations

The generic stochastic SEIR model introduced in the previous section has some theoretical properties that allow to estimate, without performing any simulations, some consequences of the epidemic (Britton and Pardoux 2019). However, the generic model does not include the effect of the quarantine and does not provision for the prediction of the outbreak duration, or the time at which the number of simultaneously infected reaches its peak, as well as other important results. To obtain this information, it is necessary to simulate the behavior of the population a large number of times, getting the probability distribution for each of these variables. Once a large number of simulation repetitions is necessary to stabilize the distribution of frequencies, the application of the Monte Carlo method may take a lot of time and demand considerable computational effort. Furthermore, these requirements will scale up as the size of the population grows, calling for increasingly efficient computational methods.

3.1 Stochastic SEIR algorithm

Algorithm 1 is the pseudo-code of the stochastic SEIR model. Its inputs (line 1) are: the initial number m of infected; a parameter β for the Poisson distribution (rate of contact between susceptible and infectious); a parameter ϵ for the exponential distribution of the individuals removal rate from compartment E (reciprocal of the incubation period); a parameter γ for the exponential distribution of the individuals removal rate from compartment I (reciprocal of the infectious period); the quarantine (χ), vaccination ϑ and death rates (δ).

Algorithm 1: Stochastic SEIR model.

---

1. **inputs**: m, β, ϵ, γ, χ, ϑ and δ
2. **for** $j \leftarrow 1 \ldots N_{sim}$
3.   choose randomly m individuals and the corresponding infectious period
4.   **while** $0 < \text{pop}(i) < 300$ for some i
5.   **for** $i \leftarrow 1 \ldots n$
6.   **if** $\text{pop}(i) > 0$ **then**
7. $\text{pop}(i) \leftarrow \text{pop}(i) - 1$
8.   **if** $\text{pop}(i) - 1 = 0$ **then** $\text{pop}(i) \leftarrow -1$ **endif**
9.   **if** $\text{pop}(i) - 1 = 100$ **then** $\text{pop}(i) \leftarrow p_I$ **endif**
10.   **if** $\text{pop}(i) - 1 = 200$ **then** $\text{pop}(i) \leftarrow -1$ **endif**
11.   **endif**
12.   generate the number of close contacts of the individual i
13.   **if** $\text{pop}(i) = 0$ **and** one of the contacted individuals is infectious **then**
14. $\text{pop}(i) \leftarrow 100 + p_E$
15.   **endif**
16.   **if** $\text{pop}(i) = 0$ **and** random value $< \vartheta$ **then** $\text{pop}(i) \leftarrow -2$ **endif**
17.   **if** $0 < \text{pop}(i) < 100$ **and** random value $< \chi$ **then** $\text{pop}(i) \leftarrow 200 + p_Q$ **endif**
18.   **if** ($0 < \text{pop}(i) < 100$ **or** $200 < \text{pop}(i) < 300$) **and** random value $< \delta$ **then**
19. $\text{pop}(i) \leftarrow -3$
20.   **endif**
21.   **end for**
22.   **end while**
23. **end for**

The algorithm repeats (line 2) the simulation N_{sim} times. In each simulation, the population is represented by a vector *pop* of n cells. Each cell *pop*(i) stores the current status of the i'th individual (with $i = 1, \ldots, n$): 0 for susceptible, > 100 and < 200 for exposed, > 0 and < 100 for infected, > 200 for quarantined, -1 for recovered, -2 for vaccinated and -3 for dead.

In the first day ($t = 0$) of a simulation (line 3), an initial number m of infected individuals are randomly selected; for each one of these individuals their specific cell in the population vector is set to p_I, where p_I is the infection period given by an exponential distribution with expected value $1/\gamma$; for the remaining individuals, that are susceptible, their cell value is set to 0.

In the next days ($t > 0$) the model dynamics repeat (while loop in line 4). At each day, the vector cell of the exposed, infected and quarantined individuals decreases one unit (line 7); furthermore: in line 8, when it reaches 0 (again susceptible) is set to -1 (recovered); in line 9, when it reaches 100 (end of the exposed period) is set to P_I (infectious), where p_I is randomly generated by an exponential distribution with parameter γ; in line 10, when it reaches 200 (end of the quarantine period) is set to -1 (recovered).

Next, for every person still susceptible, the number of close contacts with others is generated by a Poisson distribution with parameter β (line 12). The other persons are randomly selected. If they include an infected (line 13), the susceptible become exposed and their cells are set to $100 + p_E$, where p_E is randomly generated by an exponential distribution with parameter ϵ (line 14).

If an individual $pop(i)$ is still susceptible, a random number between 0 and 1 is drawn; if this number is smaller than the vaccination rate ϑ, the person becomes vaccinated; otherwise, the person stays susceptible (line 16).

If the person is or becomes infectious, a random number between 0 and 1 is drawn; if this number is smaller than the quarantine rate χ, the person becomes quarantined; otherwise, the person stays infectious (line 17).

If the person is or becomes infectious, or quarantined, a random number between 0 and 1 is drawn; if this number is smaller than the death rate δ, the person died; else, the person stays infectious or quarantined (lines 18 and 19).

The simulation stops when there are no more exposed, infectious and quarantined individuals (the condition in line 4 is false).

In order to get a probability distribution of the target variables, the number of simulations (N_{sim}) should be relatively large (more simulations lead to more accurate probability distributions). The variables targeted by this work are: the total number of infected and dead individuals, the duration of the epidemic outbreak, the maximum number of simultaneous infected individuals, the day in which this maximum happens (that is, the day corresponding to the epidemic peak), and the total number of vaccinated individuals.

3.2 Parallelization approach

Depending on the size of the population, the particular combination of the input parameters, and the number of simulations defined, the execution of Algorithm 1 may take a lot of time. But other factors, like the vaccination rate (ϑ), have also a decisive influence: the lower this rate, the higher will be the propagation of the disease, thus delaying the reaching of the stop condition.

Thus, to be able to simulate a wide range of scenarios, while generating accurate probability distributions, in a reasonable amount of time, a simple parallelization approach is applied to Algorithm 1: the initial range of simulations $(1 \ldots N_{sim})$ is fully split into mutually exclusive subranges and these are assigned to independent processors; each processor will then execute the algorithm for the subrange it was assigned; this is possible because each simulation is completely independent of the others, allowing for a SPMD (single program, multiple data) approach. Depending on the relative computing power of the processors, the partition of the initial range may be homogeneous or heterogeneous; either way, after all processors exhaust their subrange, the results of all simulations are joined to produce the final consolidated results.

Moreover, each algorithm run (execution of the stochastic SEIR model) pertains to a specific combination of parameters (that are tested through N_{sim} simulations), and there is the need to be able to perform different runs to study the impact of changes on one or more of the algorithm parameters. In order to accomplish this, the parallelization approach described above is applied to each algorithm run, and different runs are enqueued in a batch management system, awaiting for the necessary computational resources to become assigned.

3.3 Implementation details

The computational environment used for this work consisted on 8 homogeneous nodes of a small Linux cluster, with 32 CPU-cores each (2 × AMD EPYC 7351 16-core CPUs per node), for a total of 256 cores. A single MATLAB installation, network-shared by all nodes, was also used (no distributed computing facilities of MATLAB were used, due to the lack of the proper toolbox). The implementation of Algorithm 1, and of its parallelization strategy, were both made in order to take maximum advantage of such environment.

Algorithm 1 was implemented as a MATLAB (MathWorks 2012) program, but with the lower and upper limits of the simulations loop (line 2) as input parameters, thus restraining the loop to a subrange of simulations. The lower limit was used to seed the MATLAB random number generator, ensuring a different strain of random numbers per subrange. Each running instance of the program saves the results of its subrange of simulations in specific CSV files. These are later consolidated by a BASH script, which further explores MATLAB (through a secondary MATLAB program) to produce histograms and related graphics.

To generate jobs for the cluster batch management system, another BASH script was used. This script generates subranges of simulations, evenly dividing the total N_{sim} simulations by the 256 CPU-cores of the cluster; then, for each subrange, the script submits a specific job to the batch manager of the cluster; when assigned to one of the 256 CPU-cores, each job executes the MATLAB implementation of Algorithm 1, specifically parameterized.

Finally, to allow for thorough parametric studies, namely on the effect of varying vaccination and quarantine rates, the MATLAB implementation of the algorithm was changed to support those rates as user parameters. This is leveraged by two additional BASH scripts: one to automate the generation of different sets of 256 jobs (each set for a specific combination of input parameters); another to gather and process all results produced in this way.

4 Results and discussion

4.1 Parameters of the simulations

This section presents the results obtained with the stochastic SEIR model, by applying Algorithm 1 for the spread of COVID-19 in a community with a dimension similar to a small city ($n = 20,000$ individuals), considering that a unique infectious individual ($m = 1$) is introduced initially in the community. The model parameters $1/\epsilon$ and $1/\gamma$ that were used are the mean of the values presented in Table 1 ($1/\epsilon = 3.7$ and $1/\gamma = 4.3$). Concerning the infectious contact rate β, simulations assumed $\beta = 0.5$, that, according to Eq. (4), corresponds approximately to the basic reproduction number $R_0 \approx 2.1$. This value aims to represent a situation in which there is some care to prevent the transmission of the disease by the community, but

without mandatory measures of social distancing. Also, it was considered a constant death rate $\delta = 0.0093$, in agreement with the IFR verified in Portugal (DGS 2020).

For this scenario, it was analyzed the effect of quarantine and vaccination, by varying ϑ from 0.0 to 0.01 in steps of 0.001, combined with varying χ from 0.0 to 0.2 in steps of 0.02, for a total of $11 \times 11 = 121$ executions of the algorithm model. Each model execution incurred in $N_{s\,im} = 10,000$ simulations, for a total of 1,210,000 simulations and 56 hours of continuous compute time, fully using all the 256 CPU-cores of the cluster in parallel.

4.2 Results of the simulations

Figure 2 shows the empirical distribution of the (i) total number of infected individuals during the epidemic outbreak, (ii) maximum number of infected simultaneously, (iii) duration of the epidemic, and (iv) higher incidence day, for a vaccination rate $\vartheta = 0$ and a quarantine rate $\chi = 0$. Figure 3 shows the same results, but just for the simulations leading to a major epidemic outbreak.

The bimodal distribution, which is characteristic of the stochastic SIR method (Britton 2010), is also observed here for the case of the SEIR method. Moreover, the simulation data reveals that a large number of simulations resulted in a small
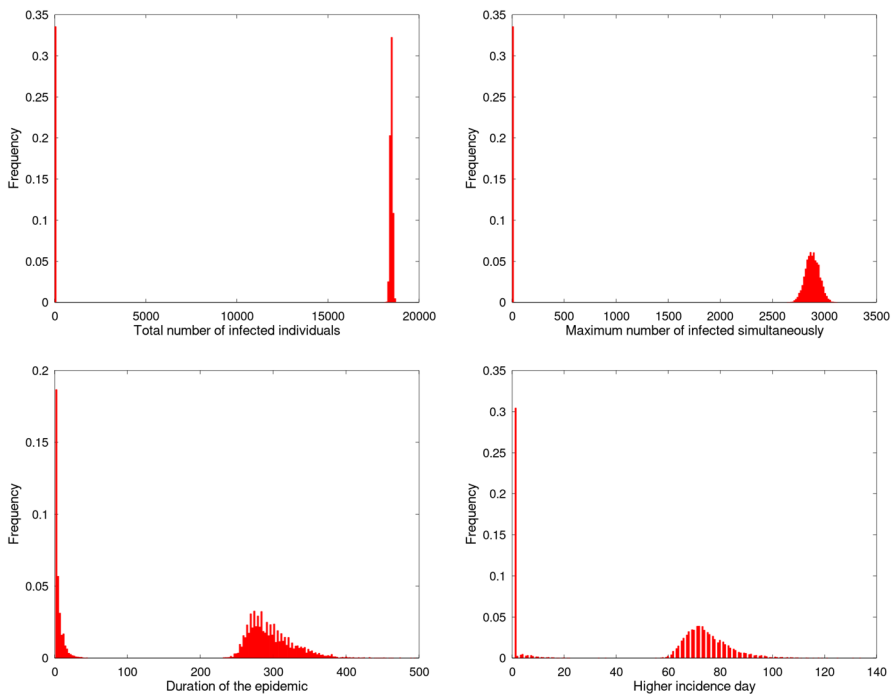

Fig. 2 Empirical distribution of the main epidemic variables: total number of infected individuals, maximum number of infected simultaneously, duration of the epidemic, and day of maximum infected simultaneously (peak of the epidemic), for $\chi = 0$ and $\vartheta = 0$
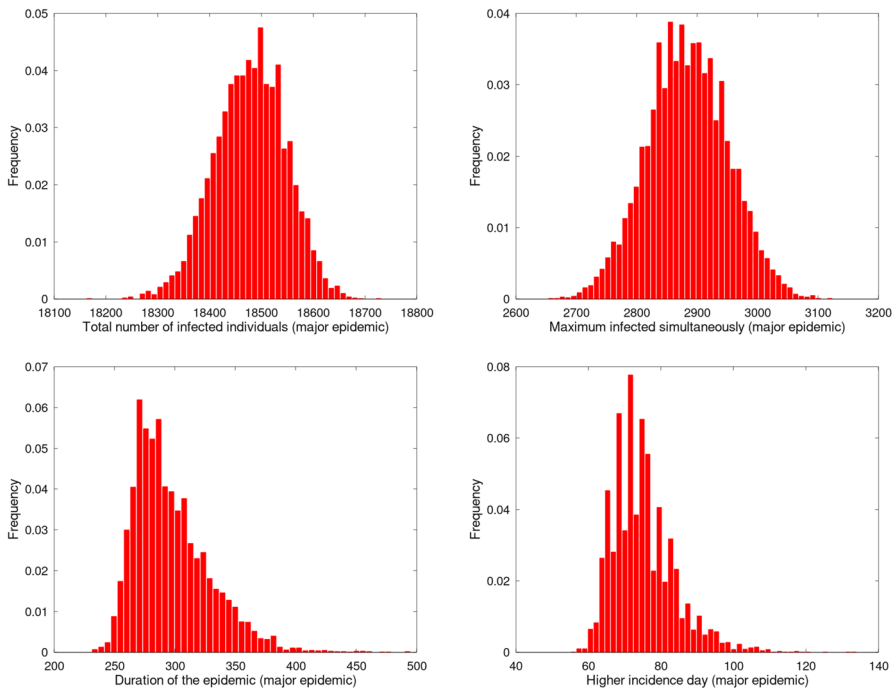
Fig. 3 Empirical distribution of the main epidemic variables from 6645 simulation of the stochastic SEIR epidemic model that lead to a major outbreak, with $\chi = 0$ and $\vartheta = 0$

epidemic outbreak, and another portion of the simulations ended up in a major epidemic outbreak. The proportion of simulations that result in a small outbreak is approximately 0.34, corresponding to 3355 simulations (in this case, a very small number of individuals are infected); conversely the proportion of simulations that results in a major outbreak is 0.66, corresponding to 6645 simulations (in which case, a large number of individuals become infected).

Looking closely at the major epidemic outbreak, depicted in Fig. 3, it can be observed that the total number of infected people has a distribution approximately normal, with mean 18, 479. The maximum number of infected individuals simultaneously also follows a normal distribution with a mean equal to 2883. The distribution of the duration of the epidemic and of the maximum incidence day, have means equal to 296 and 75 days, respectively. In short and broadly speaking, without any preventive measures there is a 66% probability of a major epidemic outbreak lasting 296 days, during which 92% of the population becomes infected, of which 14% is infected simultaneously on the 75rd day after the start of the pandemic. Considering an IFR $\approx$ 4%, this scenario corresponds to 736 deaths.

Figure 4 shows the effect of the quarantine. The images shown from left to right, for $\chi = 0, 0.05, 0.1, 0.15$ and 0.20, indicate that, as the quarantine rate increases, there is a progressive reduction in the probability of a major outbreak. By increasing the quarantine rate to $\chi = 0.2$, the probability of a major outbreak, which would affect an average of 5468 individuals, is just 0.1.
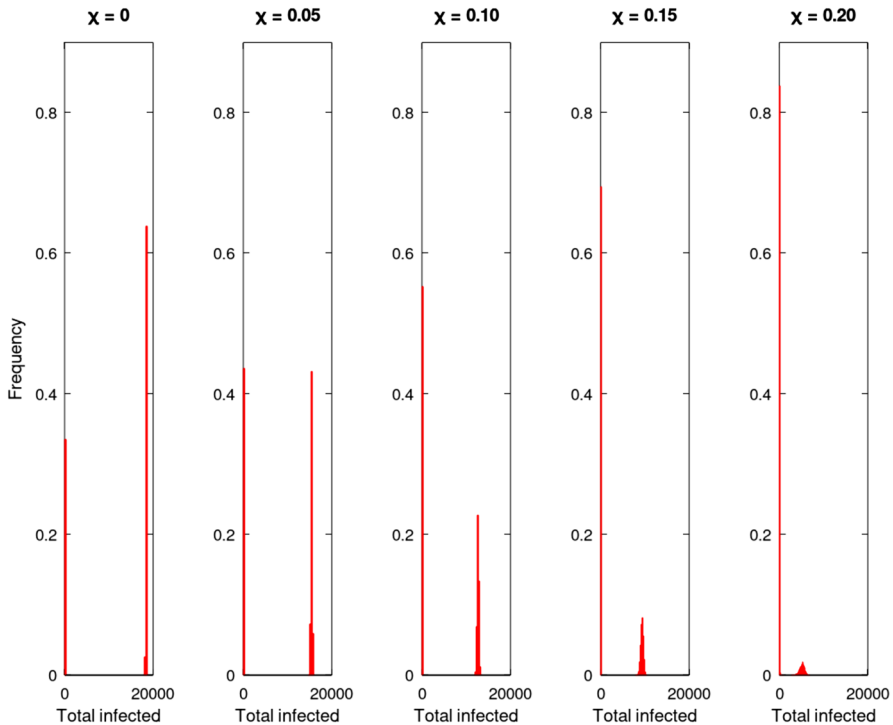
Fig. 4 Total infected individuals for different quarantine rates (without vaccination)

Figures 5 and 6 show that combining vaccination and quarantine can prevent a large infectious outbreak. As can be observed in Fig. 5, the daily isolation of a fraction of $\chi = 0.2$ from infectious individuals reduces the probability of a major outbreak to values near 0.2, which corresponds to a total number of infected individuals close to 5500 (see Fig. 6). However, if the same quarantine rate is applied with a vaccination rate equal to $\vartheta = 0.003$, the same probability is reduced to a value close to zero. For higher vaccination rates, this reduction is achieved for lower quarantine rates. For example, for $\vartheta = 0.01$ and $\chi = 0.1$ the probability of a major outbreak is almost nil.

Figure 7 represents the variation of the average of the total number of vaccinated individuals, for different values of the vaccination rate ϑ. These values vary approximately linearly with ϑ, from 1179 with $\vartheta = 0.001$, to 10,927 with $\vartheta = 0.01$. All of these values, except for the last one, are lower than the critical vaccination coverage, defined by Eq. (5), which for the case under study is close to 10,500 individuals.
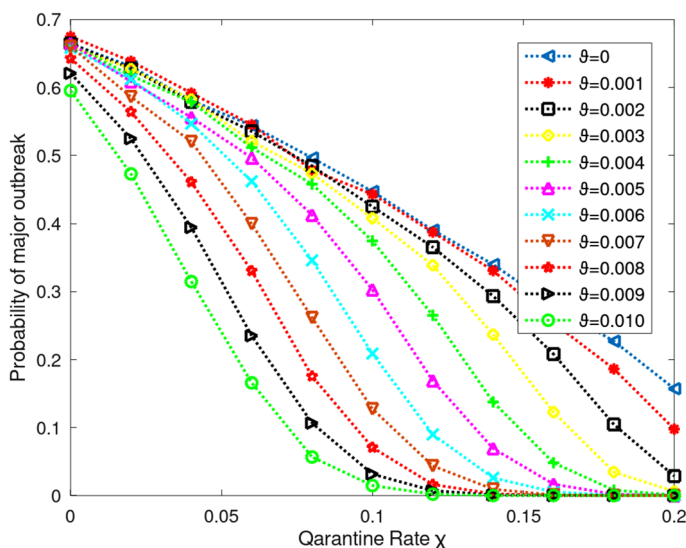
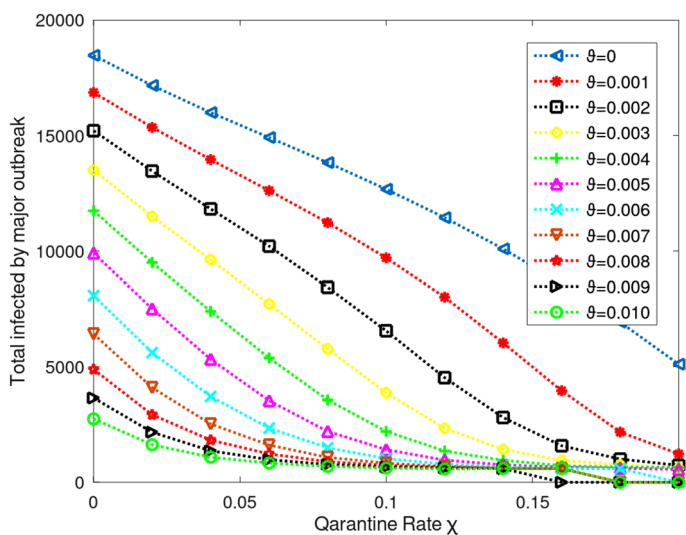**Fig. 5** Effects of the quarantine and vaccination rates on the major outbreak probability

Fig. 6 Effects of the quarantine and vaccination rates on the total number of infected individuals

4.3 Discussion

Figure 5 shows that the reduction in the probability of a major outbreak depends mainly on the rate of vaccination. Vaccination acts directly on the compartment of the susceptibles, reducing its number. As a result, there will be fewer people
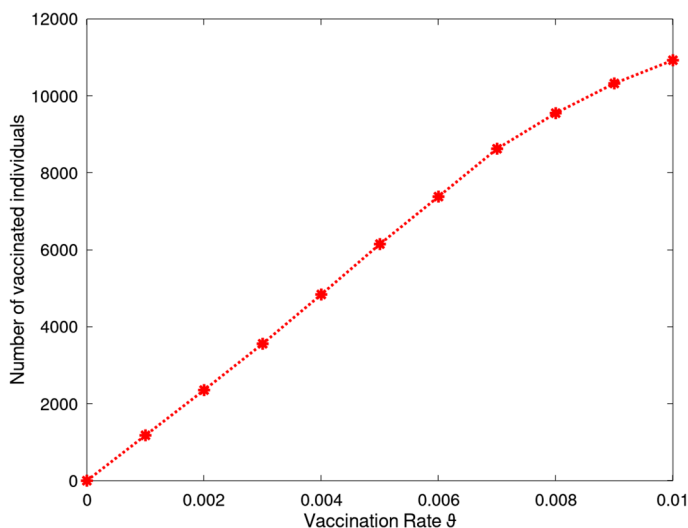
Fig. 7 Total number of vaccinated individuals for a vaccination rate from $\vartheta = 0$ to $\vartheta = 0.01$

infected. Quarantine acts only on infectious ones, isolating them and thus preventing them from infecting others. The higher the quarantine rate, the less likely it is that a susceptible person will come into contact with an infectious person and thus contracts the disease. In a real situation, quarantine includes hospitalized people and people with a positive test. However, as there are infected individuals who are not detected, it is not possible to raise the quarantine rate to levels that prevent a large epidemic outbreak. Evidently, the increase in the number of daily tests would contribute to increase this rate.

However, the results presented in Fig. 6 show also that the combination of quarantine with vaccination, even with modest rates, allows to significantly reduce the total number of infected. Therefore, in a scenario in which a vaccine (with an efficacy that is not complete), is made progressively available to the community, the isolation of the detected infectious can decisively contribute to a reduction in the risk of occurring a major epidemic outbreak.

The results also show that it is possible to prevent a major epidemic outbreak with less than critical vaccination coverage. According to Eq. (5), this rate is approximately 52%, which corresponds to vaccinating about 10,500 individuals. As can be seen in Fig. 5, this is possible even for vaccination rates below 0.01, as long as there is an adequate quarantine rate. For example, with a vaccination rate of $\vartheta = 0.003$, which involves vaccinating approximately 3559 people (see Fig. 6), the probability of a major outbreak is reduced to around 0.02.

The results presented are valid for any community with 20,000 individuals, as long as the disease transmission rate is the same ($\beta = 0.5$).

We did a parametric study of the joint effect of the two major containment measures (quarantine and vaccination) of an epidemic, based on a stochastic SEIR Model. As far as we know, this kind of empirical study is not common, which makes a comparison

with other results difficult or impossible to achieve. We are planning to do the same study with other values of transmission rate and population size in order to determine the effects of these two variables.

5 Conclusion

A stochastic SEIR epidemic model allows to realistically simulate the spread of COVID-19 in a given community. Through the usage of this model, it is possible to conclude that a single infected person can cause a major epidemic outbreak. For the reference case of a population with 20,000 individuals, there is a probability of 66% that a single infectious individual becomes responsible for an epidemic outbreak that will infect a total of 18,489 individuals on average. This risk can be reversed by the application of preventive measures that increase the number of immunized individuals (vaccination), or reduce the risk of contagious through insulating infectious individuals (quarantine). The combination of these two measures makes it possible to prevent the appearance of a major epidemic outbreak, by vaccinating only a small fraction of the population, well below the critical vaccination coverage.

Moreover, the efficacy of quarantine is highly dependent on the number of infectious cases detected. The more cases detected, the more individuals will be quarantined and thus prevented from transmitting the disease. Therefore, it is very important to have a high number of tests performed in a daily basis.

The study was carried out for a population of 20,000 people, considering a fixed rate of contagion. In future work, the same study will be conducted for populations with other dimensions, taking into account other contagion rates.

The developed computational method allows to realistically simulate the spread of COVID-19 in a medium-size community and to study the effect of the level of preventive measures such as quarantine and vaccination.

Despite its realism, stochastic SEIR methods demand considerable computational resources to ensure that the empirical frequency distribution reflects all probabilities of occurrence of the various possible scenarios. This is highlighted by the computing time consumed by the parametric study conducted in this work, covering only 11 different values of the vaccination rate and of the quarantine rate. Increasing the number of parameter combinations, or the size of the population, will further augment the simulation times, in some cases substantially. This kind of constraints, along with the need to further expand this research domain in the current pandemic situation, provide the necessary context and motivation to develop efficient simulation methodologies.

References

Balsa C, Lopes I, Rufino J, Guarda T (2020) An exploratory study on the simulation of stochastic epidemic models. In: Trends and innovations in information systems and technologies. Springer, Cham, pp 726–736. https://doi.org/10.1007/978-3-030-45688-7_71
Britton T (2010) Stochastic epidemic models: a survey. Math Biosci 225(1):24–35

Britton T, Pardoux E (2019) Stochastic epidemics in a homogeneous community. Springer, Cham. https://doi.org/10.1007/978-3-030-30900-8

Carcione JM, Santos JE, Bagaini C, Ba J (2020) A simulation of a COVID-19 epidemic based on a deterministic SEIR model. Front Public Health 8:230. https://doi.org/10.3389/fpubh.2020.00230

Chatterjee K, Chatterjee K, Kumar A, Shankar S (2020) Healthcare impact of COVID-19 epidemic in India: a stochastic mathematical model. Med J Armed Forces India 76(2):147–155. https://doi.org/10.1016/j.mjafi.2020.03.022

Dasaklis TK, Rachaniotis N, Pappis C (2017) Emergency supply chain management for controlling a smallpox outbreak: the case for regional mass vaccination. Int J Syst Sci Oper Logist 4:27–40

DGS - Direcção Geral de Saúde: Ponto da Situação atual em Portugal (25 September 2020). https://covid19.min-saude.pt/pontode-situacao-atual-emportugal

Engbert R, Rabe MM, Kliegl R, Reich S (2020) Sequential data assimilation of the stochastic SEIR epidemic model for regional COVID-19 dynamics. Bull Math Biol 10000:1000. https://doi.org/10.1101/2020.04.13.20063768

He X, Lau EHY, Wu P, Deng X, Wang J, Hao X, Lau YC, Wong JY, Guan Y, Tan X, Mo X, Chen Y, Liao B, Chen W, Hu F, Zhang Q, Zhong M, Wu Y, Zhao L, Zhang F, Cowling BJ, Li F, Leung GM (2020) Temporal dynamics in viral shedding and transmissibility of Covid-19. Nat Med 26:672–675. https://doi.org/10.1038/s41591-020-0869-5

Hethcote HW (2000) The mathematics of infectious diseases. SIAM Rev 42(4):599–653

Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, Shaman J (2020) Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). Science 368:489–493. https://doi.org/10.1126/science.abb3221

Linka K, Peirlinck M, Kuhl E (2020) The reproduction number of COVID-19 and its correlation with public health interventions. Comput Mech 1000:1000. https://doi.org/10.1007/s00466-020-01880-8

Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J (2020) The reproductive number of COVID-19 is higher compared to SARS coronavirus. J Travel Med 27(2):taaa021. https://doi.org/10.1093/jtm/taaa021

Mao L, Bian L (2010) Spatial-temporal transmission of influenza and its health risks in an urbanized area. Comput Environ Urban Syst 34(3):204–215

MathWorks (2012) MATLAB and Statistics Toolbox Release

May RM, Anderson RM (1987) Transmission dynamics of HIV infection. Nature 326:137–142. https://doi.org/10.1038/326137a0

Meyers LA, Pourbohloul B, Newman MEJ, Skowronski DM, Brunham RC (2005) Network theory and SARS: predicting outbreak diversity. J Theor Biol 232:71–81

Mukandavire Z, Nyabadza F, Malunguza NJ, Cuadros DF, Shiri T, Musuka G (2020) Quantifying early COVID-19 outbreak transmission in South Africa and exploring vaccine efficacy scenarios. PLoS ONE 15(7):e0236003. https://doi.org/10.1371/journal.pone.0236003

Zha WT, Pang FR, Zhou N, Wu B, Liu Y, Du YB, Hong XQ, Lv Y (2020) Research about the optimal strategies for prevention and control of varicella outbreak in a school in a central city of China: based on an SEIR dynamic model. Epidemiol Infect 1000:1000. https://doi.org/10.1017/s0950268819002188

Authors and Affiliations

Carlos Balsa[1] ⬤ **· Isabel Lopes[2,3] · Teresa Guarda[4,5,6] · José Rufino[7]**

✉ Carlos Balsa
  balsa@ipb.pt

  Isabel Lopes
  isalopes@ipb.pt

  Teresa Guarda
  tguarda@gmail.com

  José Rufino
  rufino@ipb.pt

[1] Research Centre in Digitalization and Intelligent Robotics (CeDRI), Instituto Politécnico de Bragança, Campus de Santa Apolónia, 5300-253 Bragança, Portugal

[2] Applied Management Research Unit (UNIAG), Instituto Politécnico de Bragança, Campus de Santa Apolónia, 5300-253 Bragança, Portugal

[3] Centro ALGORITMI, Escola de Engenharia, Universidade do Minho, Campus Azurém, 4800-058 Guimarães, Portugal

[4] Universidad Estatal Península de Santa Elena - UPSE, La Libertad, Ecuador

[5] Universidad de las Fuerzas Armadas-ESPE, Sangolqui, Quito, Ecuador

[6] Centro ALGORITMI, Escola de Engenharia - Universidade do Minho, Campus Azurém, 4800-058 Guimarães, Portugal

[7] Research Centre in Digitalization and Intelligent Robotics (CeDRI), Instituto Politécnico de Bragança, Campus de Santa Apolónia, 5300-253 Bragança, Portugal