



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Genetic algorithm with cross-validation-based epidemic model and application to the early diffusion of COVID-19 in Algeria[☆]

M.T. Rouabah^{*}, A. Tounsi, N.E. Belaloui

Laboratoire de Physique Mathématique et Subatomique Frères Mentouri University Constantine - 1, Ain El Bey Road, Constantine 25017, Algeria

ARTICLE INFO

Article history:

Received 16 July 2021

Revised 23 October 2021

Accepted 10 November 2021

Editor: DR B Gyampoh

Keywords:

COVID-19

Disease spread modeling

Genetic algorithm

Cross-validation

Algeria

ABSTRACT

A dynamical epidemic model optimized using a genetic algorithm and a cross-validation method to overcome the overfitting problem is proposed. The cross-validation procedure is applied so that available data are split into a training subset used to fit the algorithm's parameters, and a smaller subset used for validation. This process is tested on Italy, Spain, Germany, and South Korea cases before being applied to Algeria. Interestingly, our study reveals an inverse relationship between the size of the training sample and the number of generations required in the genetic algorithm. Moreover, the enhanced compartmental model presented in this work has proven to be a reliable tool to estimate key epidemic parameters and the non-measurable asymptomatic infected portion of the susceptible population to establish a realistic nowcast and forecast of the epidemic's evolution. The model is employed to study the COVID-19 outbreak dynamics in Algeria between February 25th, 2020, and May 24th, 2020. The basic reproduction number and effective reproduction number on May 24th, after three months of the outbreak, are estimated to be 3.78 (95% CI 3.033–4.53) and 0.651 (95% CI 0.539–0.761), respectively. Disease incidence, CFR, and IFR are also calculated. Numerical programs developed for this study are made publicly accessible for reproduction and further use.

© 2021 The Authors. Published by Elsevier B.V. on behalf of African Institute of Mathematical Sciences / Next Einstein Initiative.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

Introduction

The outbreak of the highly infectious COVID-19 disease attributable to SARS-CoV-2 in Wuhan and other provinces in China in 2019 has become a global pandemic since the first quarter of 2020, as asserted by the World Health Organization (WHO). Epidemic modeling, along with biological and medical research [1,2], might significantly contribute to the understanding of the characteristics of the epidemic's outbreak. It is a crucial tool in predicting the inflection point and ending time and provides insights into the epidemiological situation. Such analysis can predict the potential future evolution, help

[☆] Editor: Benjamin Gyampoh

^{*} Corresponding author.

E-mail address: rouabah.taha@umc.edu.dz (M.T. Rouabah).

estimate the efficiency of already taken measures, and guide the design of alternative interventions [3,4]. The compartmental classical Susceptible Exposed Infectious Recovered (SEIR) model [5,6] has been the most widely adopted model for characterizing many historical propagating infectious diseases such as the Spanish flu [7]. The SEIR model is extensively used to study the COVID-19 pandemic in China and many other countries, with variations best suiting the region and period under study [8–10]. Since finding the correct parameters of such a dynamic model for an epidemic is essentially a curve-fitting problem, the predictive effectiveness of the model can be considerably reduced if the available data, which we will call training data, are underfitted or overfitted. Indeed, if the training data are underfitted, the model could diverge or yield overestimated numbers with very large variance. On the opposite, if the data are overfitted, the predictive curves produced by the model will be strongly influenced by the given training data and have a very low variance, thus, artificially reducing the error on the predicted numbers and eventually leading to a non-realistic forecast. Overfitting remains a major problem with epidemic dynamical models [11]. In many of them, overfitting arises because multiple parameters might fluctuate over their uncertainty ranges making their fitted values extremely susceptible to noise in the original data [12]. Therefore, restrictions have been applied to some epidemic analyses including the COVID-19 outbreak to diminish the number of free parameters and inhibit overfitting, affecting the pertinence of those studies [9]. To overcome this issue, our approach in this study consists in using a genetic fitting algorithm and cutting off the fitting process after a certain number of generations large enough to fit the training data and small enough to not go beyond overfitting limits. We will call this number the optimum generation number G_{opt} and will compute it using data from a given province or country passing through a two-samples cross-validation procedure. Thus, G_{opt} corresponds to the fitting depth that ensures the good balance between underfitting and overfitting in our model.

SARS-CoV-2 was first imported to Algeria on Feb. 17th, 2020, by an Italian national who has been confirmed positive to COVID-19 on Feb. 25th [13]. The Italian man has been repatriated via a special flight on Feb. 28th, and no contaminated individuals by this first confirmed case have been reported by the Algerian authorities [14]. As far as we know, the effective outbreak of COVID-19 in Algeria started in late Feb. 2020. Indeed, the Algerian Health Ministry (AHM) reported in a statement on March 2nd the two first confirmed cases of COVID-19 in Blida province south of the capital Algiers [15]. Since then, the spread of the virus in Algeria has gone through different epidemic phases [16]. To the best of our knowledge, few theoretical studies on the COVID-19 outbreak in Algeria have so far been achieved [17,18]. The lack of mathematical modeling and clinical studies about SARS-CoV-2 spread in Algeria exposing the actual situation and analyzing possible evolution scenarios makes the situation more confusing for the Algerian public and scientific community. Although many studies on COVID-19 specifications and dynamics around the world are published every day, some of which include the Algerian case [19,20], we believe that any analysis of the outbreak of COVID-19 in Algeria should take into consideration many specific aspects discarded in such universal studies and online simulators which use raw data accessible on many databases. Beyond the fact that most of those databases include many wrongly reported data for Algeria, data nomenclature and interpretation, as well as test methods proper to every country, should be taken into consideration for more accurate outcomes. In our analysis, instead of only relying on the official Reverse Transcriptase Polymerase Chain Reaction (RT-PCR) confirmed SARS-CoV-2 infection cases, which are strongly affected by the country's limited test capacities, we rather combine them with the official number of hospital admitted patients—due to SARS-CoV-2 infection—to deduce the effective number of new confirmed infections per day. This choice makes a significant difference not only on the cumulative number of confirmed cases, but also on the nowcast and forecast of the virus spread.

The paper is organized as follows: in the next section, we present the mathematical model we use for the dynamical modeling of COVID-19's propagation, as well as some results of the model regarding the pandemic spread in Italy, Spain, Germany, and South Korea. The third section will be devoted to applying the model to the Algerian case with the estimation of main epidemic parameters and forecast analysis. Results will be exposed and discussed in the fourth section. The concluding section includes some ideas about the future developments of this work.

Model and methods

At the very beginning of the epidemic, during the free spread phase, initial exponential growth is commonly assumed, which is distinctive of most human infectious diseases [5,21]. However, spontaneous herd immunity, protections, and lockdown measures will confront the geometrical evolution. A dynamical model is then required to describe the expansion of the disease.

Compartmental SEIQRDP model

Regarding the novelty of the time course of infection shown by the disease and the preconized protection measures, to simulate COVID-19 spread we use a SEIQRDP model in which at time t the population is split between different compartments representing different stages in the course of the disease [9,22]. The susceptible portion of the population, namely individuals yet to be infected, is represented by the compartment $S(t)$. $P(t)$ represents the effectively protected population, mainly individuals who tend to strictly follow the advised protection measures such as wearing masks, washing hands, physical distancing, etc. Hence, this part of the population is not susceptible to be infected. Introducing this compartment is crucial to reflect increasing awareness within the major part of the population as the pandemic evolves and allows to take into account control measures taken by authorities to fight against the pandemic such as closing public areas, suspending

public transportation, and lockdown. $E(t)$ covers individuals that have been exposed to the virus but are not infectious yet. This compartment represents a latent state in which individuals are infected but cannot infect other individuals, whereas $I(t)$ represents individuals that are currently infectious. The asymptomatic exposed and infectious portions of the population are not detectable and hence non-measurable. The proportion of this part of the population can appear only through theoretical modeling of the disease. $Q(t)$ represents quarantined individuals considered as active cases, $R(t)$ corresponds to the portion of the population that has recovered from the disease and is supposed to be no longer involved in the virus propagation, and $D(t)$ represents closed cases or deaths. $N = S(t) + E(t) + I(t) + Q(t) + R(t) + D(t) + P(t)$ is the total population at time t considered constant at the timescale of the epidemic evolution. The SEIQRDP model represents the virus propagation by a collection of ordinary differential equations associating a set of transition parameters to the movement of individuals among the population compartments defined above:

$$\dot{S} = -\beta S(t)I(t)/N - \alpha S(t) \quad (1)$$

$$\dot{E} = \beta S(t)I(t)/N - \gamma E(t) \quad (2)$$

$$\dot{I} = \gamma E(t) - \delta I(t) \quad (3)$$

$$\dot{Q} = \delta I(t) - \lambda Q(t) - \kappa Q(t) \quad (4)$$

$$\dot{R} = \lambda Q(t) \quad (5)$$

$$\dot{D} = \kappa Q(t) \quad (6)$$

$$\dot{P} = \alpha S(t) \quad (7)$$

where \dot{S} refers to the time derivative of S . The positive rate α called the protection rate is introduced into the model assuming that the susceptible population is steadily decreasing as a result of increasing population awareness and public health authorities actions [9]. All the other parameters depend on the evolution of the epidemic, testing, and health care capacities and are calculated based on the official daily confirmed cases, deaths, and recoveries numbers. The transmission rate β represents the ability of an infected individual to infect others (depending on the population density, the toxicity of the virus, etc.), and $\beta S(t)I(t)/N$ is the incidence of the disease, i.e., the number of newly infected individuals yielding in unit time at time t [23]. γ^{-1} is the average latent time that an individual spends incubating the virus to become infectious (infected but not yet infectious), and δ^{-1} is the average infectiousness time, i.e., time for an infectious individual to get symptoms and get detected and quarantined. λ is the cure rate and κ is the mortality rate, while λ^{-1} and κ^{-1} represent respectively quarantine to recovery time and quarantine to death time. These transition parameters are used to define a time-dependent number of secondary cases generated by a primary infectious individual, so-called the effective reproduction number $\mathcal{R}_t = \beta \delta^{-1} S(t)/N$. It is an important parameter to analyze an epidemic outbreak and yields a measure for the severity of interventions necessary to overcome the virus spread. In general, if $\mathcal{R}_t > 1$, which mathematically corresponds to $\dot{E} + \dot{I} > 0$, the disease propagates epidemically, and when $\mathcal{R}_t < 1$, the disease is vanishing. At the beginning of the epidemic matching a situation of a fully susceptible population, this quantity is known as the basic reproduction number $\mathcal{R}_0 = \beta \delta^{-1}$ and is obtained by the next-generation matrix method [24].

Even though many COVID-19 studies try to calculate universal mean values of the reproduction number and transition parameters in some specific spots of the outbreak, they remain strongly related to local data and may change from one country to another and even from one region to another within the same country. Such parameters are the kind of valuable information this model could provide, in addition to approximate peak times of the disease (infection peak time and active cases peak time) and numbers of the non-measurable asymptomatic cases, active cases, recoveries, and deaths cases. An a priori knowledge of those numbers, though approximative, could help to optimize human and material resources on the global and local scales of a country. In this SEIQRDP model, key parameters are extracted from official numbers of cumulative confirmed cases, recoveries, and deaths available at a given period of the epidemic. The parameters obtained either by direct calculation or by a fitting algorithm are used to construct the variables curves that fit the initial data. Those curves are then extrapolated to a longer period, thus forecasting the evolution of the epidemic.

Fitting with a genetic algorithm

To calibrate our model's parameters and fit the data originating from a specific region of the world, many fitting methods are available, most of which are widely used in epidemiological studies and machine learning models. Single-stage problems

such as calibrating the parameters of our model are usually solved with modified deterministic optimization methods such as the L-BFGS-B method [25,26]. However, a stochastic method would have the advantage of considering the diversity of the possible calibrations scenarios. Evolutionary genetic algorithms are one of those stochastic methods that have a good reputation for solving optimization problems. In the following section, we discuss the advantages that the genetic algorithm method yields to our study.

Definition

A genetic algorithm (GA) is an optimization approach inspired by Darwinian evolution in which an initial set of candidate solutions called initial population, each represented by a sequence of values forming a genome, evolve by breeding and reproducing while being subject to random mutations [27]. The key mechanism in a GA is that only the best-performing solutions get to reproduce and pass on their genes following Darwinian natural selection. The evolution process is finally stopped after a certain number of generations when a defined stopping condition is verified.

Application

In our case, applying a GA to find the best fitting for the SEIQRD model parameters is a straightforward application of the above definition. Namely, the genome is the set of parameters itself, the breeding consists of creating a new set from two parent-sets by randomly selecting genes from either one of the parents, and mutation is a random alteration of one of the genes of the resulting new genome. Explicitly, the genome which is subject to evolution in our case is the set of parameters $\{\alpha, \beta, \gamma, \delta, I_0, E_0\}$. With the recovery and fatality rates, $\{\lambda, \kappa\}$ being computed directly from the training data. Each element in the former set is initiated with a random value which is constrained to vary within a given range of possible values. This is to prevent the algorithm from exploring values that do not make sense epidemiologically, e.g., negative values or too large values. Moreover, we can speed up the process by constraining the randomly generated initial population to be somewhat around already published values for the COVID-19 epidemic's parameters [28].

The best performing set of parameters is the one that produces the curves that match the best original data, by minimizing the normalized root mean squared error (nRMSE) [29]

$$F = \frac{\sqrt{\langle (rcQ_i - cQ_i)^2 \rangle}}{\langle rcQ_i \rangle}, \quad (8)$$

where rcQ_i is the real cumulative quarantined cases number for the day i (from the training data), and cQ_i is the predicted cumulative quarantined cases number for that same day. This fitness function is the normalized mean-square deviation between the real data and computed fitting data. It is applied only on the cumulative quarantined cases. This latter is more accurate than the daily quarantined cases number because it doesn't implicitly include recoveries and deaths, which are not always reported accurately, especially in Algeria. The number of cumulative quarantined individuals is widely used as the parameter to be fitted [10,18] given that the epidemic will end only if that number no longer increases with all active cases being closed. Validation of this method through its application to different cases will be presented in Section 2.4.

Starting with a set of 100 random solutions, which make up the initial population, we select at each generation an elite of 10 solutions to pass directly into the new population and generate 90 offspring by breeding. They add up to form a total of 100 solutions that constitute the population of the next generation. Each offspring, resulting from the breeding of two parent solutions, inherits its genes from both parents with an equal probability. That is, the set $\{\alpha, \beta, \gamma, \delta, I_0, E_0\}$ of the offspring solution is populated by randomly picking one element after the other from either parent 1 or 2. In this breeding stage, there is a 40% chance for each gene to mutate when passed on from a parent to an offspring. Thus, each element of the set $\{\alpha, \beta, \gamma, \delta, I_0, E_0\}$, can randomly shift to a lower or higher value within the same bounds to which these values were initially constrained. This breeding and mutation process ensures that the whole population improves throughout the successive generations. For other applications, the size of the population, the size of the elite, the mutation probability, and the maximum number of generations to be computed can all be adjusted.

The data we use in the fitting process for our work is the daily cumulative number of COVID-19 patients, i.e., the total number of infections since the start of the epidemic, on any given day. Therefore, the model's parameters, $\alpha, \beta, \gamma, \delta, \lambda, \kappa$, are expressed in days⁻¹.

The initial conditions for the Eqs. (1)–(7) are taken from the starting point of the data to be fitted, except for $E(0)$ and $I(0)$, which have to be found by the GA.

Notice that different runs of the GA give slightly different solutions, whose distribution allows a measurement of the prediction error.

Computation of the optimum generations' number using cross-validation

Cross-validation is a procedure where an original data set is divided into training and validation subsets, and where the model is trained on the first subset and tested for the second one [30]. In our case, the original data set is the whole available data on the COVID-19 epidemic for a given country or region, for n days. Those data are then cut into a training subset containing the first $n - \nu$ days' data and a validation subset with the last ν days' data. The ratio ν/n depends on the number of adjustable parameters in the regression problem. Following Guyon [31], since we have 8 parameters in the model, this ratio turns out to be around 1/4. However, it may also depend non-trivially on other parameters, such as the size of the training data set.

To avoid underfitting or overfitting of the training data, we need to determine an optimum number of generations, G_{opt} , at which the GA must stop. This number should be large enough to fit the data appropriately, but as small as possible to avoid overfitting. To do so, we apply the GA on the training subset while measuring the fitness of the best solution on the validation subset. If the GA runs for too few generations, the training data set will not be well-fitted, yielding poor fitness in the validation part. If it runs for too long, it overfits the training data, de novo producing a bad fitness for the validation data set. The number of optimum generations, G_{opt} , which sits between the previous two cases, and produces the best fitness for the validation set, is selected. This procedure will determine when the GA needs to stop when applied for predictive purposes.

Computational tools

To allow the readers to take advantage of the fitting algorithm and the cross-validation method we presented in this study for other epidemic cases, a tailored set of Python programs developed by the authors have been gathered in a Python package and made accessible online with all the necessary instructions for installation and efficient use [32]. This package is adapted for parallel computation and includes tools to download data from online repositories, determine the optimum fitting depth for a given city, region, or country using the cross-validation method, calibrate the SEIQRDP model by fitting the data with the genetic algorithm, and solve the system of ODEs to produce the forecast. Hence, our generic programs might be easily applied to study outbreaks for which a compartmental analysis is adequate in any region of the world, provided that a sufficient amount of epidemic data is available.

Model validation

Provided reasonably accurate data, our model successfully reproduces the evolution of COVID-19 in different spots for which a sufficient amount of data is available. In this section, we present the results obtained using the enhanced SEIQRDP model to estimate the active cases evolution in Italy, Spain, Germany, and South Korea. For these countries, we use publicly available confirmed cases, recoveries, and deaths numbers from online raw data sets [33,34]. We pick the training data starting from the date for which all confirmed cases, recoveries, and deaths numbers take non-zero values to avoid computational bugs and optimize parameters fitting. The active cases curve is then reproduced for six months following that date. To highlight the model's efficiency, we use only an early part of the official data to train the model, instead of all the available data. We have used 30 days of training data for Italy, 45 days for Spain, 60 days for Germany, and 90 days for South Korea. Fig. 1 shows that the results obtained using our model are in good accordance with official statistics for the number of active cases in those countries. All the fittings have a coefficient of determination $R^2 > 0.9$, and as our fitting method is based on a non-linear regression algorithm, we use a normalized standard error S_N of the estimate to evaluate the goodness of the fits.

We remind that to avoid overfitting the training data and obtain the best forecast, we look for the optimum fitting corresponding to G_{opt} rather than the best one. For Italy, the model can reproduce to a good approximation ($S_N = 17\%$) the active cases' curve with only 30 days of training data. Fig. 1 reveals that the larger the training data sample is, the lower is the optimum number of generations G_{opt} used by the genetic algorithm to fit the data with a lower S_N . Germany and South Korea are specific cases and need profound analysis that goes beyond the scope of this paper, but one can observe the quicker decrease in their active cases after the epidemic peak time due to particular strategies to control the epidemic.

Indeed, the lower panel of Fig. 1 suggests that countries with stronger protective measures, in this case Germany and South Korea, necessitate larger training data to fit the model properly. This highlights the fact that the linear protection rate (Eq.(7)) in the model may not be well suited for such cases.

We emphasize that the curve of the active case is one of the most pertinent in our opinion, as it reflects the amplitude of the epidemic outbreak and the efficiency of the measures applied to control it. Moreover, the epidemic will end only if all the active cases are closed.

Model estimation for Algeria

Data

For COVID-19's dynamics study in Algeria, we use the official public data provided by the Algerian Health Ministry (AHM) [35,36]. The specificity of our analysis that makes its predictive results for Algeria more accurate than different studies where Algeria is presented as an example [19] is the fact that instead of relying on official numbers of RT-PCR-confirmed COVID-19 cases, we deduce the effective number of confirmed infections per day by considering the number of hospitals admitted patients. Given that the RT-PCR numbers are strongly affected by the country's limited testing capacities. This number, which is unfortunately no longer provided by the AHM since May 25th, 2020, is considered as the effective number of active cases in our study. Then, we deduce the effective confirmed cases number for a given date by adding computed tomography (CT) scans confirmed cases to the official RT-PCR confirmed cases (see Appendix A).

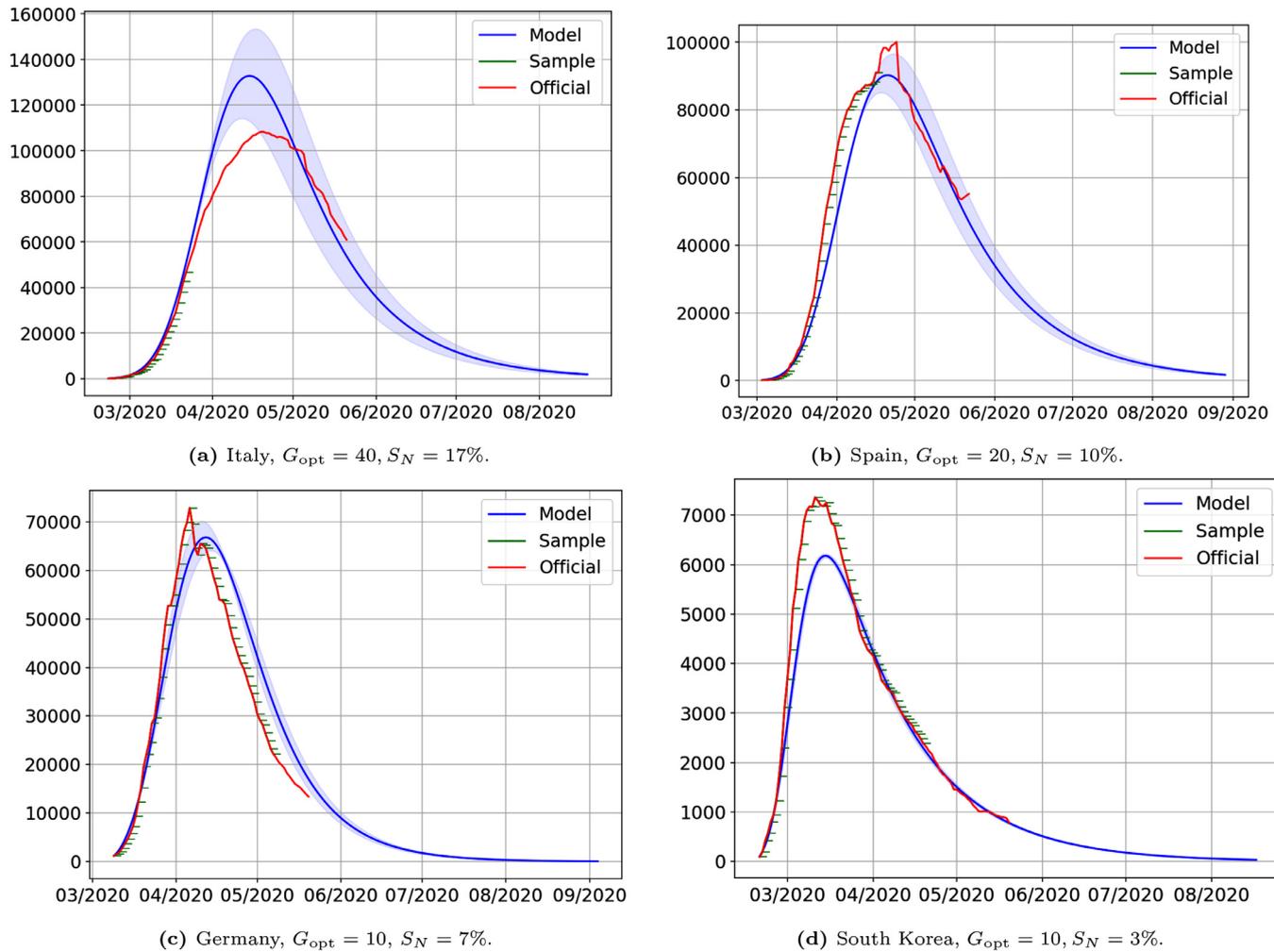


Fig. 1. Results of the SEIQRDP model forecast for the COVID-19 outbreak in (a) Italy, (b) Spain, (c) Germany, and (d) South Korea. Training sets of 30, 45, 60, and 90 days of official data (green marker lines), respectively, are initially used to fit the model's parameters. The forecast curves (blue lines) are calculated using the SEIQRDP model and compared to the real active cases curves (red lines) for each country. Light blue shading represents 95% confidence intervals of the model estimate. Even with only four or six weeks of training data, the model can produce a realistic forecasting estimate. All fits have a coefficient of determination $R^2 > 0.9$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Epidemic parameters

Besides the more exciting forecasting use of the SEIRQDP model, this latter is particularly efficient for nowcasting. Indeed, fitting the official data allows us to estimate key epidemic parameters of the early stage spread of COVID-19 in Algeria. Even though hundreds of studies have estimated these parameters for COVID-19 in different regions of the world, a local estimation is of major importance as their values are strongly related to the local population's discipline, public health capacities, and the severity of local containment measures at the very beginning and during the epidemic period.

During one month after the first confirmed case of COVID-19 in Algeria on Feb. 25th, 2020, the disease has undergone a practically free propagation phase. On March 12th, universities, schools, and nurseries were closed. On March 19th, all trips between Algeria and European countries have been canceled by Algerian authorities that have decided the first strong containment measures against COVID-19 spread on March 24th. A total lockdown of Blida province and partial lockdown in many other provinces have been applied. Coffee shops, restaurants, and all non-essential shops have been closed, public transportation suspended and grouping of more than two persons forbidden. On April 24th, the authorities decided a partial release of lockdown measures in Blida and other provinces and allowed many commercial activities to resume. This date coincided with the starting of the holy month of Ramadan, resulting in a brutal increase in social and commercial activities. Due to low respect for physical distancing and protection measures, the number of new confirmed cases increased significantly, and shops have been closed again in many provinces since May 7th. In the light of this chronology of measures, we have estimated the intermediate mean values of the epidemic parameters during the free-propagation phase (Feb. 25th–Mar. 25th) and then every two weeks till May 24th. Those intermediate mean values exposed in Fig. 2 provide valuable information revealing the evolution of the epidemic in Algeria during its three first months and the impact of the applied control measures.

Forecast

To forecast the evolution of COVID-19 in Algeria, we implement the SEIRQDP with a training data period from Feb. 25th to May 24th. We apply the cross-validation method script on the first 70 days of the data set and test it on the 20 remaining to calculate the optimum number of generations. For the chosen data sample, we obtain $G_{opt} = 20$. Then, the genetic algorithm and the rest of the SEIRQDP set of programs are applied to the whole training data to calculate the optimum fit and reproduce the SEIRQDP variables curves using the fit parameters obtained. We present in this paper a forecast of the COVID-19 outbreak dynamics until the end of September 2020, the time for which the reopening of schools and universities is scheduled. That step might represent a turning point in the disease's epidemic evolution and requires a specific analysis.

Results and discussion

Our model estimates that on Feb. 25th, in addition to the first confirmed SARS-CoV-2 infected case in Algeria, at least seven other individuals had been infected without showing any symptoms. On March 2nd, when the two first cases have been confirmed in Blida, we estimate that the number of asymptomatic infectious people had already reached ten individuals, and at least ten others have been in a latent period. One week later, the number of asymptomatic infected persons had already exceeded seventeen, following our estimations. Officially, twenty of them have been confirmed at that time.

The epidemic parameters model estimates for the first three months of COVID-19 in Algeria are in good agreement with the on-the-ground evolution of the outbreak. The estimated basic reproduction number on Feb. 25th is $\mathcal{R}_0 = 3.78$ (95% CI 3.033–4.53), while the value of \mathcal{R}_t on May 24th is estimated to be 0.651 (95% CI 0.539–0.761), and the mean effective reproduction number during the first three months of the epidemic is evaluated to be 1.74 (95% CI 1.55–1.92). The notable decline in \mathcal{R}_t during this period might reflect the outbreak control efforts and increasing consciousness of COVID-19. By the same token, we distinguish a significant rise of the protection rate α after the first control measures on March 24th, jumping from 0.0041 during the free propagation phase before March 25th to 0.0089 on the period of March 26th–April 10th and doubling again to 0.021 on the next period (see Fig. 2 upper-left corner). Interestingly, the protection rate curve reflects the softening of the containment measures and lower respect of protection measures in the period between April 27th and May 12th, resulting in a decline of α during the next period. The protection rate's mean value of the overall study period is estimated to be 0.015 (95% CI 0.014–0.017). The increase of the transmission rate shown in Fig. 2, lower-left corner, is reasonable due to the continuous propagation of the virus and the apparition of many clusters in densely populated provinces. In addition, the low number of daily tests and the relatively long test-to-result time of the used testing technology increase the probability that an asymptomatic infectious individual spreads the virus before being quarantined. The mean value of the transmission rate is estimated to be 0.64 (95% CI 0.62–0.66). The mean latent time is 2.7 (95% CI 2.6–2.8) days, and the mean infectious time is 5.9 (95% CI 5.7–6.1) days. The mean incubation time (latent time + infectiousness time) has a mean value of 8.6 (95% CI 8.3–8.9) days. One remarkable point that can be observed in Fig. 2 middle panel is that besides the first period, the incubation time remains relatively stable, taking values within the range [7.9–8.6] days. That reflects the global features of the evolution of the hidden variables representing the exposed and the infectious portions of the population. The decrease of the incubation time after the first period of the study might be a consequence of a better

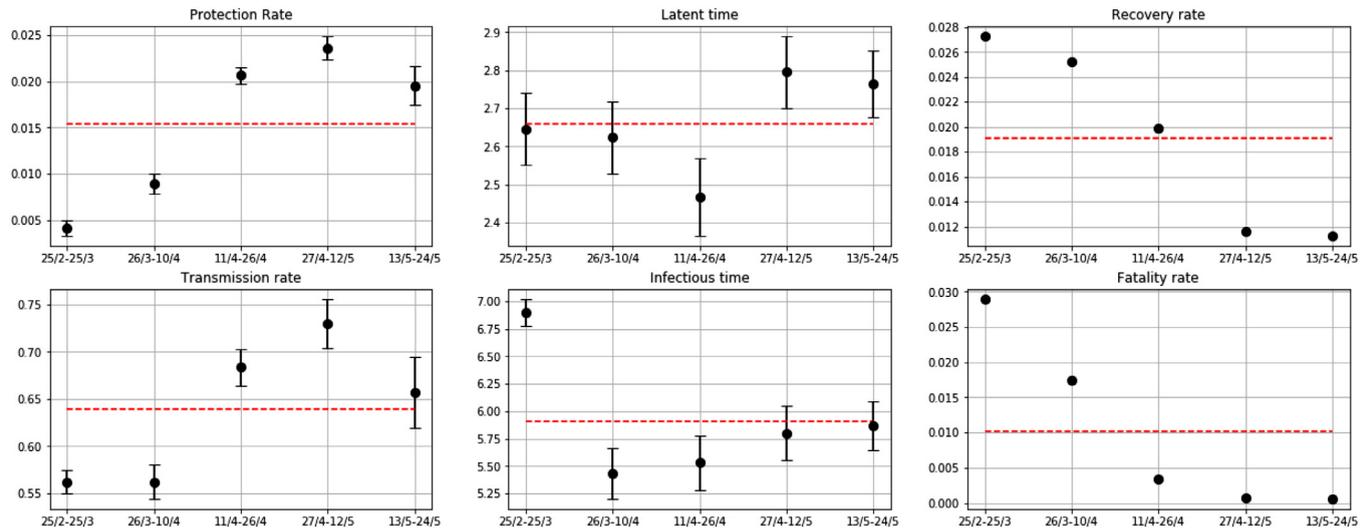


Fig. 2. Estimation of main epidemic parameters during the early stage of the COVID-19 outbreak in Algeria (Feb. 25th–May 24th, 2020). Intermediate values are calculated for five different periods corresponding to specific phases of the virus propagation with peculiar circumstances and authorities measures severity. The intermediate values for each parameter are compared to its mean value on the three months (dashed red line). The protection rate, transmission rate, latent time, and infectious time are estimated using the SEIQRDP model, while the recovery and fatality rates are calculated from official data. Error bars represent 95% confidence intervals of the model estimates. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1

Summary of the SEIQRDP parameters' mean estimates for the early stage of the COVID-19 outbreak in Algeria compared to Wuhan (China). \mathcal{R}_0 is estimated on Feb. 25th, while the mean values of the other parameters are calculated for the three first months of the outbreak.

Parameter	Definition	Value for Algeria (95% CI)	Value for Wuhan (95% CI)	Reference (Wuhan)
α	Protection rate (days ⁻¹)	0.015 (0.014–0.017)	0.085	Peng et al. [9]
β	Transmission rate (days ⁻¹)	0.64 (0.62–0.66)	0.99	Peng et al. [9]
γ^{-1}	Latent time (days)	2.7 (2.6–2.8)	2	Peng et al. [9]
δ^{-1}	Infectious time (days)	5.9 (5.7–6.1)	7.4	Peng et al. [9]
\mathcal{R}_0	Basic reproduction number	3.78 (3.033–4.53)	6.47 (5.71–7.23)	Tang et al. [22]

detection scheme. A high diagnosis capacity allowing a large-scale testing strategy and efficient tracking are an essential tool to reduce the onset to quarantine (incubation) period since early and quick detection of infectious individuals enables authorities to quarantine them before showing symptoms, hence limiting the number of their contacts. Moreover, this will help diminish the effective reproduction number \mathcal{R}_t and then better control the disease spread.

In contrast to other epidemic parameters, we calculate recovery and fatality rates shown in the right panel of Fig. 2 directly from official data. The recovery rate varies in the range [1.1–2.7%] with a mean value of 1.9%. Interestingly, the fatality rate initially estimated as the highest in the world at the time falls under 0.5% since mid-April with a mean value estimated to 1.02%. The significant decrease of fatality rate, even though affected by the growing test capacities after the number of RT-PCR daily tests has been increased, and the CT-scan diagnostic of COVID-19 adopted at the beginning of April could also be interpreted as the consequence of better medical care. The fatality rate seems to stabilize during the last month of the study (0.071% on April 27th–May 12th and 0.058% on May 13th–May 24th), as newly deployed RT-PCR test capacities were reaching their limits again. The epidemic analysis of the parameters' values is beyond the scope of this paper as it requires information to which we do not have access. Nevertheless, we notice that the key parameters' values obtained through our model for Algeria fall within the values ranges estimated for the Chinese city of Wuhan, where SARS-CoV-2 first appeared [9,22,28], as shown in Table 1.

The forecast simulations (Fig. 3), based on the available official data, estimate that the infection peak time corresponding to the maximum incidence occurred on April 24th–26th, with 387 (95% CI 267–509) new infections per day as shown in Fig. 3d. The effective reproduction number continuously decreased, reflecting a better control of the disease spread and crossed the line $\mathcal{R}_t = 1$ by May 1st (see Fig. 3c). At that crucial point, the disease entered the attenuation phase. The SEIQRDP model evaluates the active cases peak time for the first wave of COVID-19 outbreak in Algeria, corresponding to the active cases' maximum, to be on the period between May 20th and May 30th with 9794 (95% CI 8770–1024) active cases (see Fig. 3a).

We estimate that the number of new infections will vanish by mid-September. At that time, the number of active quarantined cases will still be above 500. Assuming that the epidemic will remain ongoing as long as all active cases have not been closed yet, the model predicts the first wave of the outbreak to end no earlier than October 2020, with an estimated total quarantined individuals of 24,021 (95% CI 20768–27274), 15,291 (95% CI 13272–17310) recovered and 8172 (95% CI 7093–9251) deaths as shown on Fig. 3b. Notice that the predicted total number of deaths appears to be particularly overestimated compared to official numbers (blue dashed line). A solution to this technical issue is under investigation [37]. That can partly be explained by the fact that official COVID-19 deaths are those confirmed by PCR tests only, whereas our model deals with both PCR and CT scan diagnosed individuals. We emphasize that the numbers we present in this forecast are only estimations that could be seriously affected by the quality of the available data.

Another important piece of information that could be extracted from the official public data is the Case Fatality Rate (CFR) corresponding to the ratio of deaths to effective confirmed cases. Furthermore, the Infected Fatality Rate (IFR), often confused with CFR, is the ratio of deaths to infected cases, including asymptomatic cases which are non-measurable. For that reason, we calculate the CFR based on the official data, while the IFR is calculated through the ratio of the official cumulative deaths to the cumulative number of infected individuals obtained from the SEIQRDP model (see Fig. 4). The mean CFR on the period Feb. 25th–May 24th is estimated to be 5.3% while the mean value of IFR on the same period is 2.9% (95% CI 1.7–3.9%). Notice that the mean IFR for the three first months of the outbreak in Algeria is higher than the global value estimated to 1.4% by a recent study using cumulative COVID-19 data from 139 countries [38].

It is worth knowing that compartmental models, including the SEIQRDP model, work perfectly when some conditions on the studied population are assumed. Indeed, the SEIQRDP model requires a well-mixed and homogeneous population. A well-mixed population means that all individuals in the population have the same chance to be infected by an infectious one. Homogeneity means that all individuals behave likely toward the disease and thus are governed by the same rules of transitions probabilities between different population compartments. Consequently, all calibrated parameters in this study should be seen as a statistical average over the population. Moreover, the SEIQRDP model is fundamentally not additive, i.e., the sum of different SEIQRDP models applied to different provinces of a given country is not necessarily equivalent to the SEIQRDP model applied to the whole country. Because of the previous considerations altogether, it would be interesting to apply our study to different major infected cities of the country separately.

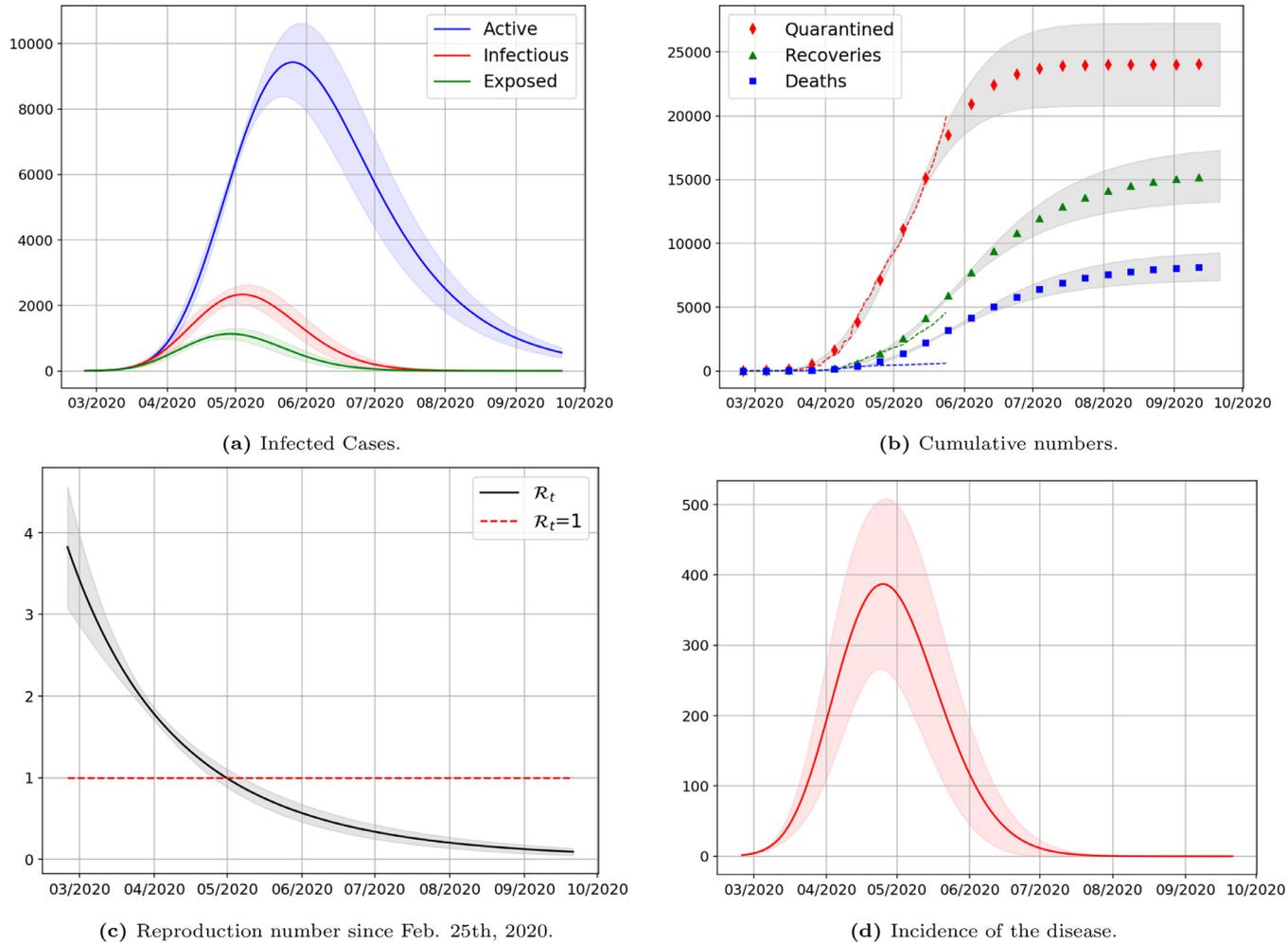


Fig. 3. SEIQRD model's forecast. (a) Number of exposed (infected not yet infectious), infectious (asymptomatically infectious), and active (quarantined) cases. The figure shows the epidemic's peak time, corresponding to the maximum active cases, to occur on the period of May 20th - May 30th, with roughly ten thousand active cases. (b) Total quarantined, recoveries, and deaths. Real data are represented with the red, green, and blue dashed lines, respectively. (c) Time-dependent reproduction number. The $R_t = 1$ point is in perfect accordance with the exposed, infected, and active cases inflection points. (d) The number of newly infected individuals per day. Light shadings represent 95% confidence intervals of the model estimates. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

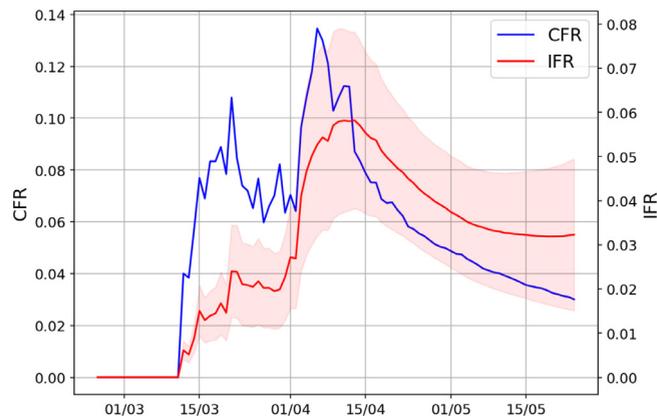


Fig. 4. Case to Fatality Rate (CFR) and Infection to Fatality Rate (IFR) for the COVID-19 outbreak in Algeria between Feb. 25th and May 24th, 2020. Light red shading represents 95% confidence intervals of the model estimate. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Conclusion

In this paper, we have presented an enhanced compartmental SEIQRDP model for epidemics in which we introduced a protection rate and where the noteworthy compartments of quarantined and protected populations have been added compared to the most widely used SEIR models. Our approach is based on a genetic fitting algorithm and uses cross-validation to overcome the overfitting problem. To determine the optimum number of generations for the genetic algorithm yielding the best parameters, we applied the cross-validation procedure to the data of various countries in such a way that the whole available data on the COVID-19 epidemic for a given country for n days is split into a training subset containing the data for the first $n - \nu$ days, and a validation subset for the last ν days. The ratio ν/n depends on the number of adjustable parameters in the regression problem, and turns out to be around 1/4 for our model. We tested the procedure on Italy, Spain, Germany, and South Korea cases before applying it to Algeria. Remarkably, our study emphasizes an inverse relationship between the size of the training sample and the number of generations required in the genetic algorithm. As more data becomes available for a given country, the optimum number of generations decreases. Therefore, optimization should be re-adjusted at different points in the outbreak period to ensure the most accurate results are obtained using the model. We have designed a generic open-source package containing all computational tools used in our analysis [32]. This package includes tools to pick up online data, calculate the optimum fitting depth using the cross-validation method, fit the data with the genetic algorithm to estimate the SEIQRDP parameters, and produce forecast curves. Also, the package includes parallel computation functionality. That allows the programs to be deployed on high-performance computers for better accuracy. We have neatly prepared this package to be easily implementable to study epidemics for which a compartmental analysis is adequate in any region of the world.

Based on the official cumulative recoveries, cumulative deaths, and deduced effective cumulative confirmed cases—including approximated CT-scan diagnosed cases—this model allowed us to estimate the epidemic parameters for COVID-19 outbreak in Algeria (basic reproduction number, protection rate, transmission rate, infectious time, latent time...). We have estimated intermediate mean values of key epidemic parameters between Feb. 25th and May 24th. These intermediate values exposed in Fig. 2 permit us to evaluate the epidemic situation in the country and the effect of the different phases of control measures during the first three months of the outbreak. We recapitulated in Table 1 the basic reproduction number estimated on Feb. 25th and the calculated mean values of key epidemic parameters for the whole period Feb. 25th–May 24th, which we compared to recently published results for the COVID-19 epidemic in Wuhan. Such parameters estimations might be of high interest for further epidemic studies of the virus spread in Algeria and the African continent. Using our SEIQRDP model, we have predicted the disease's effective reproduction number time evolution (\mathcal{R}_t) which is considered an essential indicator of the epidemic's situation. Fig. 3c exposes the evolution of \mathcal{R}_t since the beginning of the outbreak and an approximate period on which this parameter has gone below one. Our simulations suggest the basic reproduction number on Feb. 25th, 2020 to be $\mathcal{R}_0 = 3.78$ (95% CI 3.033–4.53), while the value of \mathcal{R}_t on May 24th is estimated to be 0.651 (95% CI 0.539–0.761). The mean effective reproduction number during the first three months of the epidemic is evaluated to 1.74 (95% CI 1.55–1.92). Moreover, we have been able to provide a valuable approximate estimation of the daily evolution of the non-measurable asymptomatic exposed and infectious cases, in addition to the daily active cases from the beginning until an advanced stage of the COVID-19 outbreak in Algeria (Fig. 3a and b). We have estimated the periods in which these numbers have been at their highest peak, and approximated the maximum values they could reach for the outbreak's first stage. The model predicts that this wave of the COVID-19 epidemic in Algeria to end no earlier than October 2020, with an estimated total quarantined individuals of 24,021 (95% CI 20768–27274), 15,291 (95% CI 13272–17310) recovered, and 8172 (95% CI 7093–9251) deaths. We have also estimated the time at which the number of new infections will eventually vanish

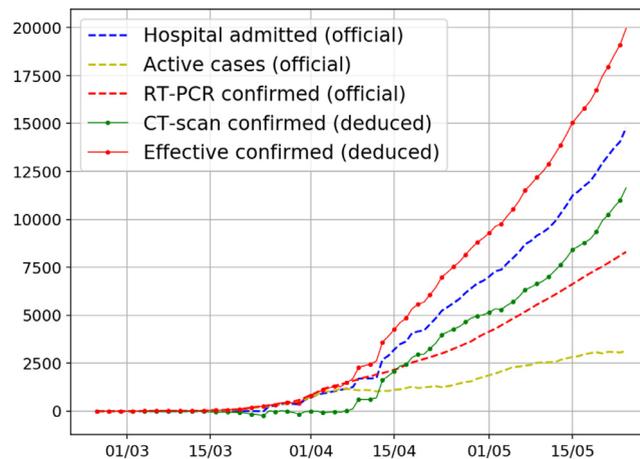


Fig. A.1. The estimated effective confirmed cases curve (dotted red) compared to the official RT-PCR confirmed cases curve (dashed red) between Feb. 25th and May 24th, 2020. We deduce the number of CT-scan confirmed cases (green) for a given date by subtracting the number of official active cases (dashed yellow) from the number of hospital admitted patients (dashed blue) for that date. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(Fig. 3d). Even though the SEIQRDP model we presented, like many of the SEIR derivatives, is effective in different contexts, we are still studying the Algerian case carefully since the reported COVID-19 epidemic's evolution in Algeria quickly reached the country's maximum capacity of diagnosis which is well reflected in the linear form of official confirmed cases data. Furthermore, we should note that, in a basic way, the SEIQRDP model is well established to simulate outbreaks in a well-mixed and closed population, although being very sensitive to data accuracy. In this instance, we stress the fact that our estimations depend strongly on the publicly available data at the time of the study, and we emphasize the specificity of our study in considering an effective cumulative confirmed cases number, which includes approximated CT-scan diagnosed SARS-CoV-2 infections, deduced from the official numbers of hospital admitted patients as illustrated in Appendix A.

We are investigating many possibilities to optimize our model to fit the COVID-19 evolution in Algeria and elsewhere with more ingenious methods. Additionally, a completely different epidemic agent-based model is already in an advanced development stage and will be used to tackle the virus spread from a different perspective. A comparison of updated epidemic parameters in Algeria using more recent data to recent studies performed on African countries [39] might also be of great interest.

We hope this study can serve as a useful guideline for scientists and governments and efficiently contribute to the fight against the COVID-19 pandemic on the national and international scales.

Declaration of Competing Interest

Authors declare that they have no conflict of interest.

CRediT authorship contribution statement

M.T. Rouabah: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft, Supervision. **A. Tounsi:** Conceptualization, Methodology, Software, Formal analysis, Writing – review & editing. **N.E. Belaloui:** Conceptualization, Methodology, Software, Formal analysis, Writing – review & editing.

Acknowledgments

This study presents results of curiosity-driven research, which has been achieved only through the personal resources of the authors.

Appendix A. Effective confirmed cases

In Fig. A.1, we expose the official number of hospital admitted patients due to COVID-19 in Algeria (blue dashed line) and the official number of active cases (yellow dashed line) computed by subtracting the official numbers for recoveries and deaths from the official RT-PCR-confirmed cases. Notice that before April 6th, 2020, the two curves are perfectly superimposed. The number of hospital admitted patients start to significantly increase compared to the official number of active cases after April 6th. Since that time, the Algerian health authorities have adopted the CT-scan diagnosis of COVID-19 in addition to the RT-PCR tests. Hence, we believe that the difference between the two curves represents the number of CT-scan

confirmed cases (green line). The deduced number of CT-scan confirmed cases is added to the official RT-PCR-confirmed cases to obtain the effective number of confirmed cases (red dotted line). Notice that the curve of the active cases displays a plateau at the beginning of April due, in our opinion, to the fact that the country has reached its maximum capacity of RT-PCR testing. This curve started to increase again in the second half of April 2020, after many hospitals and biology research entities started performing RT-PCR tests.

References

- [1] J. Sun, W.-T. He, L. Wang, A. Lai, X. Ji, X. Zhai, G. Li, M.A. Suchard, J. Tian, J. Zhou, M. Veit, S. Su, COVID-19: epidemiology, evolution, and cross-disciplinary perspectives, *Trends Mol. Med.* 26 (2020) 483–495, doi:[10.1016/j.molmed.2020.02.008](https://doi.org/10.1016/j.molmed.2020.02.008).
- [2] H.A. Rothan, S.N. Byrareddy, The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak, *J. Autoimmun.* 109 (2020) 102433, doi:[10.1016/j.jaut.2020.102433](https://doi.org/10.1016/j.jaut.2020.102433).
- [3] M. Ajelli, B. Gonçalves, D. Balcan, V. Colizza, H. Hu, J.J. Ramasco, S. Merler, A. Vespignani, Comparing large-scale computational approaches to epidemic modeling: agent-based versus structured metapopulation models, *BMC Infect. Dis.* 10 (2010), doi:[10.1186/1471-2334-10-190](https://doi.org/10.1186/1471-2334-10-190).
- [4] G. Chowell, A. Tariq, J.M. Hyman, A novel sub-epidemic modeling framework for short-term forecasting epidemic waves, *BMC Med.* 17 (2019), doi:[10.1186/s12916-019-1406-6](https://doi.org/10.1186/s12916-019-1406-6).
- [5] M. Anderson, R.M. May, *Infectious Diseases of Humans, Dynamics and Control*, OXFORD University Press, 1992, doi:[10.1016/0169-4758\(92\)90018-W](https://doi.org/10.1016/0169-4758(92)90018-W).
- [6] S. Meloni, N. Perra, A. Arenas, S. Gómez, Y. Moreno, A. Vespignani, Modeling human mobility responses to the large-scale spreading of infectious diseases, *Sci. Rep.* 1 (2011) 62, doi:[10.1038/srep00062](https://doi.org/10.1038/srep00062).
- [7] G. Chowell, H. Nishiura, L.M. Bettencourt, Comparative estimation of the reproduction number for pandemic influenza from daily case notification data, *J. R. Soc. Interface* 4 (2007) 155–166, doi:[10.1098/rsif.2006.0161](https://doi.org/10.1098/rsif.2006.0161).
- [8] J.T. Wu, K. Leung, G.M. Leung, Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study, *Lancet* 395 (2020) 689–697, doi:[10.1016/S0140-6736\(20\)30260-9](https://doi.org/10.1016/S0140-6736(20)30260-9).
- [9] L. Peng, W. Yang, D. Zhang, C. Zhuge, L. Hong, Epidemic analysis of COVID-19 in China by dynamical modeling, *medRxiv* (2020). 10.1101/2020.02.16.20023465
- [10] J. Labadin, B.H. Hong, Transmission dynamics of 2019-nCoV in Malaysia, *medRxiv* (2020). 10.1101/2020.02.07.20021188
- [11] W. Yang, D. Zhang, L. Peng, C. Zhuge, L. Hong, Rational evaluation of various epidemic models based on the COVID-19 data of China, *medRxiv* (2020). 10.1101/2020.03.12.20034595
- [12] S. Basu, J. Andrews, Complexity in mathematical models of public health policies: a guide for consumers of models, *PLoS Med.* 10 (2013) 1–6, doi:[10.1371/journal.pmed.1001540](https://doi.org/10.1371/journal.pmed.1001540).
- [13] Algerian-Health-Ministry, Situation report of 25/02/2020, <http://covid19.sante.gov.dz/fr/2020/02/25/25-fevrier-2020/>.
- [14] Algerian-Health-Ministry, Situation report of 29/02/2020, <http://covid19.sante.gov.dz/fr/2020/03/03/point-de-situation-au-29-02-2020/>.
- [15] Algerian-Health-Ministry, Situation report of 02/03/2020, <http://covid19.sante.gov.dz/fr/2020/03/03/point-de-situation-au-02-03-2020-2/>.
- [16] M. Lounis, COVID-19-in-Algeria-chronology-and-evaluation-of-preventive-actions, *Eur. J. Med. Educ. Technol.* 13 (2020), doi:[10.30935/ejmet/8012](https://doi.org/10.30935/ejmet/8012).
- [17] S. Bentout, A. Chekroun, T. Kuniya, Parameter estimation and prediction for coronavirus disease outbreak 2019 (COVID-19) in Algeria, *AIMS Public Health* 7 (2020) 306–318, doi:[10.3934/publichealth.2020026](https://doi.org/10.3934/publichealth.2020026).
- [18] S. Belkacem, COVID-19 data analysis and forecasting: Algeria and the world, 2020, ArXiv preprint.
- [19] Z. Zhao, X. Li, F. Liu, G. Zhu, C. Ma, L. Wang, Prediction of the COVID-19 spread in African countries and implications for prevention and control: a case study in South Africa, Egypt, Algeria, Nigeria, Senegal and Kenya, *Sci. Total Environ.* 729 (2020) 138959, doi:[10.1016/j.scitotenv.2020.138959](https://doi.org/10.1016/j.scitotenv.2020.138959).
- [20] J. Luo, When will COVID-19 end? Data-driven prediction, 2020, <https://ddi.sutd.edu.sg/>.
- [21] J. Yuan, M. Li, G. Lv, Z.K. Lu, Monitoring transmissibility and mortality of COVID-19 in Europe, *Int. J. Infect. Dis.* 95 (2020) 311–315, doi:[10.1016/j.ijid.2020.03.050](https://doi.org/10.1016/j.ijid.2020.03.050).
- [22] B. Tang, X. Wang, Q. Li, N.L. Bragazzi, S. Tang, Y. Xiao, J. Wu, Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions, *J. Clin. Med.* 9 (2020) 462, doi:[10.3390/jcm9020462](https://doi.org/10.3390/jcm9020462).
- [23] Z. Ma, Y. Zhou, J. Wu, *Modeling and Dynamics of Infectious Diseases*, Co-Published with Higher Education Press, 2009, doi:[10.1142/7223](https://doi.org/10.1142/7223).
- [24] O. Diekmann, H. Heesterbeek, T. Britton, *Mathematical Tools for Understanding Infectious Diseases*, Princeton Series in Theoretical and Computational Biology, Princeton University Press, 2013, doi:[10.23943/princeton/9780691155395.001.0001](https://doi.org/10.23943/princeton/9780691155395.001.0001).
- [25] L.A. Hannah, Stochastic optimization, in: *International Encyclopedia of the Social & Behavioral Sciences*, Elsevier, 2015, pp. 473–481, doi:[10.1016/b978-0-08-097086-8.42010-6](https://doi.org/10.1016/b978-0-08-097086-8.42010-6).
- [26] A. Comunian, R. Gaburro, M. Giudici, Inversion of a sir-based model: a critical analysis about the application to COVID-19 epidemic, *Phys. D* 413 (2020) 132674, doi:[10.1016/j.physd.2020.132674](https://doi.org/10.1016/j.physd.2020.132674).
- [27] L. Chambers, *Practical Handbook of Genetic Algorithms, Applications*, vol. 1, CRC-Press, Inc., 1995, doi:[10.1201/9780429128332](https://doi.org/10.1201/9780429128332).
- [28] Q. Lin, S. Zhao, D. Gao, Y. Lou, S. Yang, S.S. Musa, M.H. Wang, Y. Cai, W. Wang, L. Yang, D. He, A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action, *Int. J. Infect. Dis.* 93 (2020) 211–216, doi:[10.1016/j.ijid.2020.02.058](https://doi.org/10.1016/j.ijid.2020.02.058).
- [29] K.N. Nabi, M.T. Tahmid, A. Rafi, M.E. Kader, M.A. Haider, Forecasting COVID-19 cases: a comparative analysis between recurrent and convolutional neural networks, *Results Phys.* 24 (104137) (2021) 104137, doi:[10.1016/j.rinp.2021.104137](https://doi.org/10.1016/j.rinp.2021.104137).
- [30] M.W. Browne, Cross-validation methods, *J. Math. Psychol.* 44 (2000) 108–132, doi:[10.1006/jmps.1999.1279](https://doi.org/10.1006/jmps.1999.1279).
- [31] I. Guyon, *A scaling law for the validation-set training-set size ratio*, AT & T Bell Laboratories, 1997.
- [32] M.T. Rouabah, A. Tounsi, N.E. Belaloui, Python package for epidemic SEIQRDP model using genetic fitting algorithm with cross-validation to overcome overfitting problem, <https://github.com/Taha-Rouabah/COVID-19>.
- [33] Data-Packaged-Core-Datasets, <https://raw.githubusercontent.com/datasets/covid-19/master/data/time-series-19-covid-combined.csv>.
- [34] H. Ritchie, Our world in data, <https://ourworldindata.org/coronavirus-source-data>.
- [35] Algerian-Health-Ministry-EADN, Epidemic map COVID-19 in Algeria, <http://covid19.sante.gov.dz/carte/>.
- [36] CDTA, Epidemic map COVID-19 in Algeria, <https://covid19.cda.dz/dashboard/production/index.php#>.
- [37] M. Nikolaou, A fundamental inconsistency in the sir model structure and proposed remedies, *medRxiv* (2020). 10.1101/2020.04.26.20080960
- [38] R. Grewelle, G. De Leo, Estimating the global infection fatality rate of COVID-19, *medRxiv* (2020). 10.1101/2020.05.11.20098780
- [39] K. Amouzouvi, K.A. Assamagan, S. Azote, S.H. Connell, J.B.F. Fankam, F. Fanomezana, A. Guga, C.E. Haliya, T.S. Mabote, F.F. Macucule, D. Mathebula, A. Muronga, K.C.C. Mwale, A. Njeri, E.F. Onyie, L. Rakotondravohitra, G. Zimba, A model of COVID-19pandemic evolution in african countries, 2021, [arXiv:2104.09675](https://arxiv.org/abs/2104.09675)