Predicting the COVID-19 epidemic in Algeria using the SIR model

Mohamed Samir Boudrioua
mohamedsamirboudrioua@gmail.com
SOCIETE ORSI. 134 R6 5000. Monastir- Tunisia.

Abderrahmane Boudrioua
abderrahmaneboudrioua@gmail.com
Universiti Utara Malaysia
(

Abstract: The aim of this study is to predict the daily infected cases with Coronavirus (COVID-19) in Algeria. We apply the SIR model on data from 25 February 2020 to 24 April 2020 for the prediction. Following Huang et al [11], we develop two SIR models, an optimal model and a model in a worst-case scenario COVID-19. We estimate the parameters of our models by minimizing the negative log likelihood function using the Nelder-Mead method. Based on the simulation of the two models, the epidemic peak of COVID-19 is predicted to attain 24 July 2020 in a worst-case scenario, and the COVID-19 disease is expected to disappear in the period between September 2020 and November 2020 at the latest. We suggest that Algerian authorities need to implement a strict containment strategy over a long period to successfully decrease the epidemic size, as soon as possible.

Keywords: Prediction, COVID-19, Algeria, SIR model.

Introduction:

On 31 December 2019, a cluster of cases of pneumonia of unknown cause was detected in Wuhan City, Hubei Province of China [25]. The Chinese authorities identify on 7 January the coronavirus disease (COVID-2019) as the causative virus [25] . This epidemic continues to cause infections in other Chinese cities and in multiple countries around the world [23]. On 01 April 2020, the World Health Organization (WHO) has reported a total of 823611 confirmed cases of COVID-19 infection with 40534 death cases worldwide [26].

Algeria has reported the first case of COVID-19 on 25 February 2020, in Ouargla region and he was an Italian national [7, 1]. Two other cases were reported on 01 march 2020 in Blida region in the North of Algeria[7, 1].The epidemic continues to spread in other regions of the country. On 03 April 2020, the Algerian authorities have declared 1171 confirmed cases with 105 deaths cases [1] .

Various earlier studies have predicted the confirmed cases of COVID-19 infection in different countries. Kuniya[16] applied the well-known SEIR compartmental model to predict the epidemic peak of COVID-19 disease in Japan. He used data of daily reported cases of COVID-19, from15 January to 29 February 2020. He founds that the epidemic peak in Japan could possibly reach the early-middle summer. Roosa et al [24] forecast the covid-19 epidemic in China over a short-term period using generalized logistic growth model, the Richards growth model, and a sub-epidemic wave model. They employed data of cumulative confirmed cases between January 22, 2020 and February 9, 2020. They found that the epidemic has reached saturation in Hubei and other provinces of China. Hamidouche [7], introduce the Alg-COVID-19 Model to predict cumulative cases of COVID-19 in Algeria. This model allows predicting the incidence and the reproduction number R0 in the upcoming months [7]. According to his results, the number of infected cases in Algeria will exceed 5000 on 07 April 2020 and it will double to 10000 on 11 April 2020 [7].

This study applies the Susceptible-Infected-Removed (SIR) model without demographics (no births, deaths, or migration) [14,5,13] to predict the daily confirmed cases of COVID-19 infection in Algeria. We use data of reported confirmed cases between February 25, 2020 and April 24, 2020. We predict the epidemic peak and the end-date of this disease using an optimal SIR model and an SIR model in a worst-case COVID-19 scenario. These results could be helpful for decision makers in Algeria to implement their future counter strategies to COVID-19 outbreak.

Method:

1) Model:

The SIR model without demographics is given by the following non linear system of ordinary differential equations [14, 13, 8] :

$$\begin{aligned} \frac{dS(\mathrm{t})}{dt} &= -\beta S(\mathrm{t})I(\mathrm{t}), \\ \frac{dI(\mathrm{t})}{dt} &= \beta S(\mathrm{t})\mathrm{I}(\mathrm{t}) - \gamma I(\mathrm{t}), \\ \frac{dR(\mathrm{t})}{dt} &= \gamma I(\mathrm{t}). \end{aligned} \tag{1.1}$$

The function $S(\mathrm{t})$ represents the number of susceptible cases at time t. The function $I(\mathrm{t})$ represents the number of infected cases at time t, and the function $R(\mathrm{t})$ is the compartment of removed cases from the disease, where:

Removed cases = Death cases + Recovered cases.

The parameters β and γ are called the contact rate and the removed rate, respectively. The initial condition of this model $S(0), I(0), R(0)$ must satisfy the condition $S(0)+I(0)+R(0)=N$, where N is the population size. [8]

The model can be written also as follows:

$$\begin{aligned} \frac{dS(\mathrm{t})}{dt} &= -\beta \frac{S(\mathrm{t})}{N} I(\mathrm{t}), \\ \frac{dI(\mathrm{t})}{dt} &= \beta S(\mathrm{t}) \frac{I(\mathrm{t})}{N} - \gamma I(\mathrm{t}), \\ \frac{dR(\mathrm{t})}{dt} &= \gamma I(\mathrm{t}). \end{aligned} \tag{1.2}$$

With: $\frac{S(\mathrm{t})+I(\mathrm{t})+R(\mathrm{t})}{N}=1$ and $\mathrm{S}(0)>0, \mathrm{I}(0)>0, R(0)=0$ [13].

The basic reproductive ratio R_0 is'' the average number of secondary cases arising from an average primary case in an entirely susceptible population, and it's given by β/γ'' [13].

We integrate numerically our non-linear system using the Livermore Solver for Ordinary Differential Equations (LSODA) algorithm [9, 20]. This algorithm handles stiff and non stiff systems of size N, with general form $dy/dt = f(t,y)$. It has an automatic method selection where it uses Adams methods with the non stiff problems and Backward Differentiation Formula (BDF) methods with the stiff problems [9].

Following Huang et al [11], we consider two SIR models: The first one is an optimal model, where we suppose that the initial population size N is equal to the total of confirmed cases in Algeria.

In the second SIR model we assume that whole real population in Algeria is susceptible, with high contact rate β and high removed rate γ (either recovered or died). This is the worst-case scenario, and this model could be used to predict the largest infections of the epidemic [11].

Python software version 3.6.4 was used to implement the SIR model and solve it.

2) Maximum likelihood estimation:

After resolving the SIR system we obtain the approximate functions $\tilde{I}(\mathrm{t}), \tilde{\mathrm{S}}(\mathrm{t}), \tilde{\mathrm{R}}(\mathrm{t})$ .We are interested in the function $\tilde{I}(\mathrm{t})$ the approximate of the function of the number of infections $I(\mathrm{t})$ Since we have count data y_t (the real numbers of daily infected cases of COVID-19 in Algeria), we suppose that y_t follows a Poisson distribution with parameter $\lambda = p\tilde{I}(\mathrm{t})$ [15] :

$$y_t \sim \mathrm{Poisson}\left(\lambda = p\tilde{I}\left(t\right)\right) .$$

Where the parameter p is a constant reflects a combination of sampling efficiency and the detectability of infections [15].We ignore the constant p in this study.

Thus the likelihood function is defined as follows:

$$\begin{aligned} L(\lambda) &= \prod_{t=0}^{n} f(y_t \mid \lambda) \\ &= \prod_{t=0}^{n} \frac{\lambda^{y_t} e^{-\lambda}}{y_t!} \end{aligned}$$

Then the log likelihood function is:

$$LL(\lambda) = \sum_{t=0}^{n} \log \frac{\lambda^{y_t} e^{-\lambda}}{y_t!} = \sum_{t=0}^{n} (y_t \log \lambda - \lambda - \log y_t!)$$

Ignoring the constant involving $\log y_t!$, and substituting $\lambda = \tilde{I}(\mathrm{t})$, we get:

$$LL(\lambda = \tilde{I}(\mathrm{t})) = \sum_{t=0}^{n} [y_t \log(\tilde{I}(\mathrm{t})) - \tilde{I}(\mathrm{t})]$$

The objective is estimating the parameters β and γ that give us the optimum solution $\tilde{I}(\mathrm{t})$. The Nelder-Mead algorithm [17] was used to find the minimum of the negative log likelihood function, and therefore maximize the log likelihood function. This algorithm is used for solving the unconstrained optimization problem based on the iterative update of a simplex:

$$\min f(\mathrm{x}).$$

Where $f:\mathbb{R}^n \to \mathbb{R}$ is the objective function and n the dimension [6, 2].

For more details about the Nelder-Mead algorithm, see [6, 2, 17].

The procedure of estimation can be resumed in the following two steps [18]:

Step 1: Select the initial parameters β, γ and initialize the SIR model and the Nelder-Mead method,

Step 2: Run the Nelder-Mead algorithm to get new parameters which will be used in the simulation and the prediction of the COVID-19 epidemic.

All the calculations were conducted as previously stated by Python software version 3.6.4

Empirical results:

The data were obtained from Johns Hopkins University, Center for systems science and engineering [10]. We used the number of daily new confirmed cases for the COVID-19 epidemic in Algeria, from 25 February 2020 to 24 April 2020.

Firstly we fit the optimal SIR model. We used equation (1.2) with initial parameters $\beta = 0.3$, $\gamma = 0.1$ and initial conditions:

$$\begin{aligned} I_0 &= \frac{I(0)}{N} = \frac{1}{N} \\ S(0) &= 1 - I_0 \\ R(0) &= 0 \end{aligned}$$

With $N = 3127$, equal to the total confirmed cases till 24 April 2020.

Solving the SIR system and following the procedure of estimation cited previously, we get the new parameters: $\hat{\beta} = 0.18$ and $\hat{\gamma} = 0.033$. Thus the basic reproduction number R_0 is given by:

$$R_0 = \frac{\beta}{\gamma} = \frac{0.18}{0.033} \approx 5.45$$

We substitute the new parameters in the model to simulate and predict the COVID-19 epidemic in Algeria.

Figure 1 shows the predictions of the number of daily new infected cases with COVID-19 in Algeria, by fitting the optimal SIR model. The mean absolute error (MAE) is used to evaluate the accuracy of this model, where:

$$MAE = mean(|y_t - \tilde{I}_t|).$$

y_t is the real reported value at time t and $\tilde{I}_t$ is the predicted value at the same time t [12, 3]. We get $MAE = 14$ as accuracy of the optimal SIR model for the prediction of new infected cases.

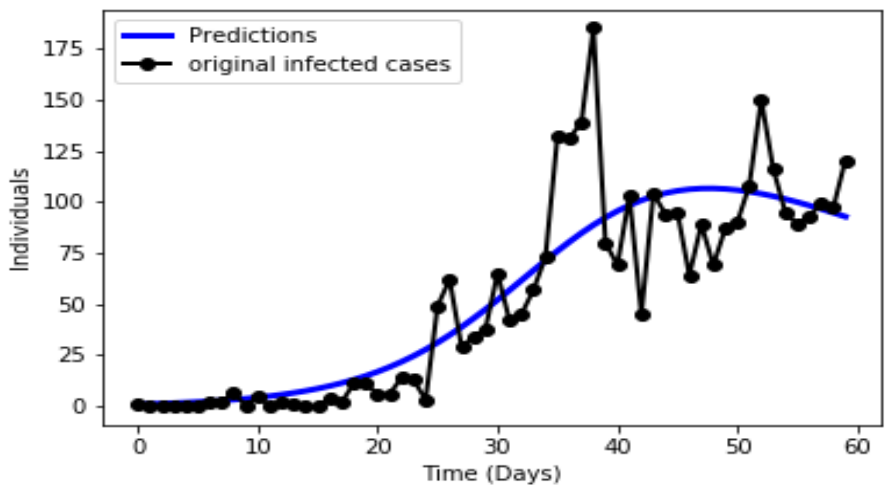

Figure 1: Fit of the optimal SIR model for Algeria infected cases with covid-19, from 25 February 2020 to 24 April 2020

Concerning the second model in a worst-case scenario COVID-19, we reset the initial parameters as follow:

$\beta = 0.95$,as a high contact rate.

$\gamma = 0.75$,as a high removed rate (either recovered or death cases).

$N = 43851044$ The real population size of Algeria[27].

We proceed as before, we get the estimated parameters $\hat{\beta} = 0.89$, $\hat{\gamma} = 0.79$, thus $R_0 \approx 1.13$.

Figure 2 shows the simulation of infected cases with COVID-19 in a worst-case scenario. We get $MAE = 30$ as accuracy of this model. We remark that the predicted number of infected cases with COVID-19 in Algeria increase exponentially fast in this case.

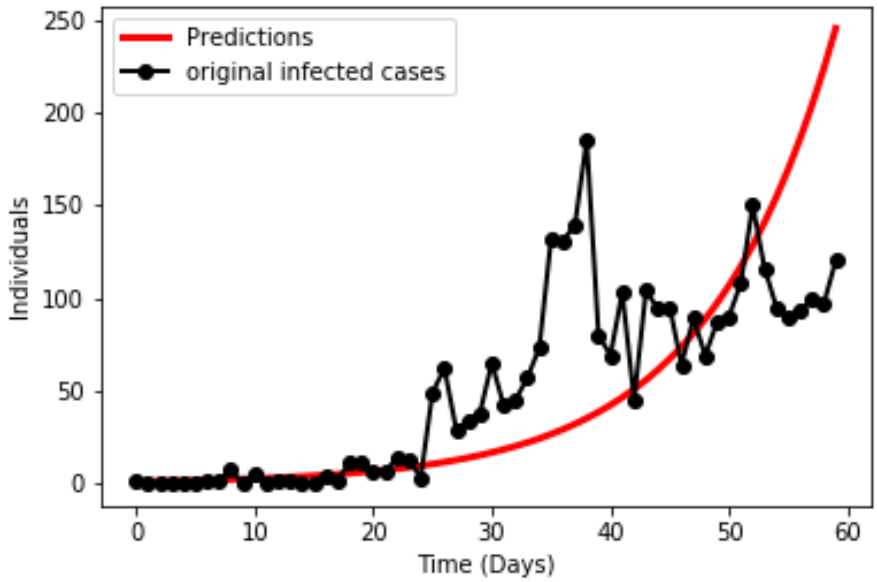

Figure 2: Fit of SIR model for Algeria infected cases in a worst-case scenario covid-19, from 25 February 2020 to 24 April 2020.

The prediction of the epidemic peak of COVID-19 in Algeria using both models is shown in Figure 3. We extract from this Figure that the predicted epidemic peak would be around 25 July 2020 at the latest, with about 2444×10^2 infections, in a worst-case scenario. The predicted epidemic peak based on the optimal SIR model was on 13 April 2020 with 106 infections. The end-date of COVID-19 in Algeria is estimated to be around 16 November 2020 in a worst-case scenario, and around 18 September 2020 based on the optimal SIR model predictions. We can remark easily the large differences in the predicted infections between the two models. These differences represent the potentially preventable infections as a result of containment strategies [11]. This shows the importance of containment strategies and how many COVID-19 infections can be prevented by these strategies. Stand on these results; we conclude that the COVID-19 epidemic will continue to spread in Algeria for about six upcoming months, at the latest.

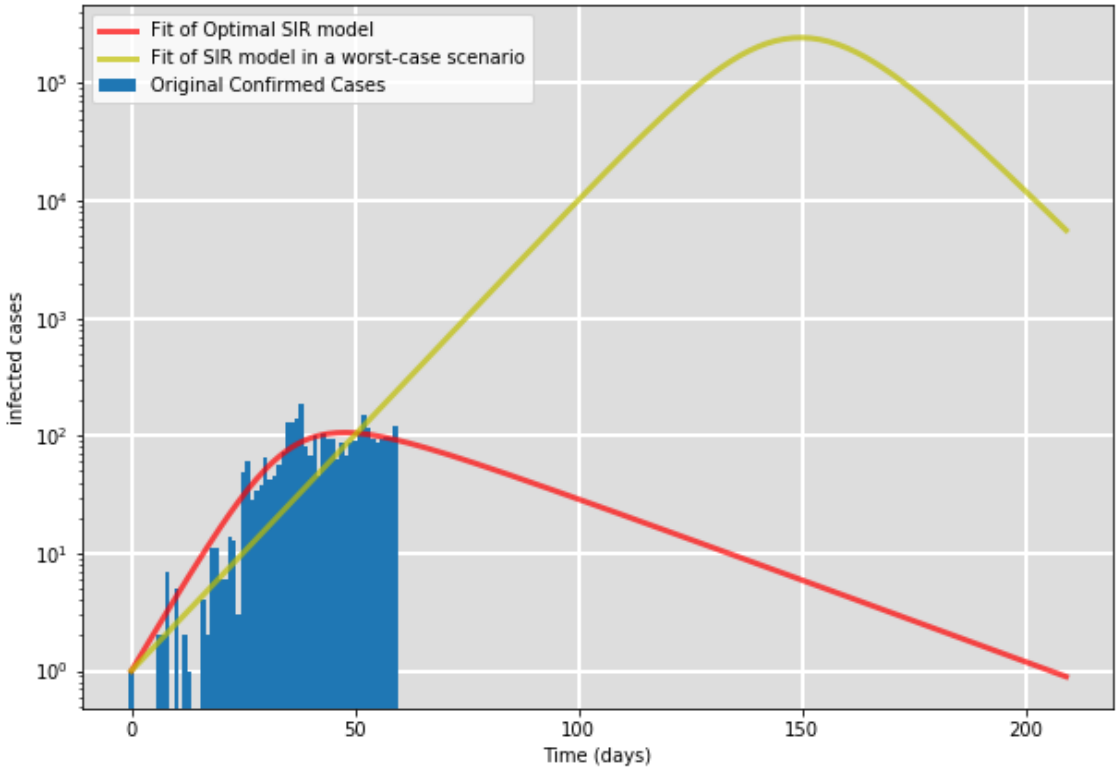

Figure 3: Prediction of the epidemic peak of COVID-19 in Algeria by fitting an optimal SIR model and an SIR model in a worst-case scenario.

Conclusion:

In this study we predict the COVID-19 epidemic outbreak in Algeria using the number of daily infected cases, from 25 February 2020 to 24 April 2020. We have developed two SIR compartmental models without demographics, an optimal model and a model in a worst-case scenario. The estimated value of the basic reproduction number R_0 using the optimal SIR model is 5.45, while R_0 is estimated to be equal to 1.13 in the worst-case scenario COVID-19. R_0 has a various determinants such as social behavior, population structure, viral evolution, environment, immunity as well as immigration policies, healthcare and Infrastructure of the health sector [21, 22, 4]. Comparing the prediction results of both models, we find that the COVID-19 epidemic peak in Algeria might attain 24 July 2020 with about 2444×10^2 infections in a worst-case scenario. The end of this epidemic in Algeria is expected to be between 18 September 2020 and 16 November 2020, at the latest. Our

findings agree with the results of Kuniya [16] in the case of COVID-19 outbreak in Japan, and with the WHO's statement that COVID-19 will maybe disappear in the summer [16].

Following Kuniya [16] and based on the large differences in the predicted infections between the two models that represent the potentially preventable infections, we suggest that a strict containment strategy is needed in Algeria as soon as possible, over a relatively long period, in order to decrease the epidemic size.

In perspective, we will try to capture the effect of quarantine measures on COVID-19 outbreak in Algeria. For this, we would like to apply the SIQR model (Q stands for quarantined) [19].

References:

1. Algerian Ministry of Health 2020. Available at http://covid19.sante.gov.dz/carte/.
2. Baudin .M. Nelder mead user's manual 2010. Available at: https://wiki.scilab.org/The%20Nelder-Mead%20Component?action=AttachFile&do=view&target=neldermead.pdf
3. Boudrioua, M. S. Modeling and Forecasting the Algiers Stock Exchange Returns using the Box-Jenkins Methodology. Preprints 2019, 2019090134 (doi: 10.20944/preprints201909.0134.v1).
4. Delamater, P. L., Street, E. J., Leslie, T. F., Yang, Y. T., Jacobsen, K. H. Complexity of the Basic Reproduction Number (R_0). Emerg. Infect. Dis. 2019; 25(1), 1–4. https://doi.org/10.3201/eid2501.171901
5. Dietz. K. Epidemics and Rumours: A Survey. J. Roy. Stat. Soc. 1967; A **130** 505–528.
6. Gao, F , Han, L. Implementing the Nelder-Mead algorithm with adaptive parameters. Comput Optim Appl. 2012; 51.259-277.
7. Hamidouche M. COVID-19 outbreak in Algeria: A mathematical Model to predict cumulative cases. [Preprint]. Bull World Health Organ. E-pub: 25 March 2020. doi: http://dx.doi.org/10.2471/BLT.20.256065
8. Harko.T, Lobo. F. S. N, Mak. M. K.. Exact analytical solutions of the Susceptible-Infected- Recovered (SIR) epidemic model and of the SIR model with equal deaths and births. APPL MATH COMPUT. 2014; 236 (2014), 184–194.
9. Hindmarsh, A. C. ODEPACK, A Systematized Collection of ODE Solvers; in p.55--64 of Stepleman, R.W. et al.[ed.] 1983 Scientific Computing, North-Holland, Amsterdam.

10. Hopkins University, J. 2019 novel coronavirus (nCoV) data repository. Center for systems science and engineering. 2020 https://github.com/CSSEGISandData/2019-nCoV.

11. Huang Y, Yang L, Dai H, Tian F , Chen K. Epidemic situation and forecasting of COVID-19 in and outside China. [Submitted]. Bull World Health Organ. E-pub: 16 March 2020. doi:http://dx.doi.org/10.2471/BLT.20.255158

12. Hyndman, R.J , Athanasopoulos, G. Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2. 2018. Accessed on Mars 2020.

13. Keeling, M. , Rohani, P. Modeling infectious diseases in humans and animals. Princeton University Press. 2008.

14. Kermack, W.O.& McKendrick, A.G. A Contribution to the Mathematical Theory of Epidemics. Proc. Roy. Soc. Lond. 1927; A **115** 700–721.

15. King**.** A.A. Introduction to inference: parameter estimation. ICTP Workshop on Mathematical Models of Climate Variability, Environmental Change and Infectious Diseases 2017**.** Available at: https://kingaa.github.io/clim-dis/parest/parest.htm

16. Kuniya, T. Prediction of the Epidemic Peak of Coronavirus Disease in Japan,. J. Clin. Med. **2020**, 9, 789.

17. Nelder, J.A. & Mead, R. A simplex method for function minimization. Comput. J. 1965; **7**, 308–313 .

18. Nsoesie, E.O, Beckman R.J, Shashaani S, Nagaraj K.S, Marathe M.V. A Simulation Optimization Approach to Epidemic Forecasting. PLoS ONE 2013; 8(6): e67164. https://doi.org/10.1371/journal.pone.0067164

19. Nuño M., Castillo-Chavez C., Feng Z., Martcheva M. Mathematical Models of Influenza: The Role of Cross-Immunity, Quarantine and Age-Structure. In: Brauer F., van den Driessche P., Wu J. (eds) Mathematical Epidemiology. Lecture Notes in Mathematics, vol 1945. Springer, Berlin, Heidelberg. 2008.

20. Petzold, L.R. Automatic Selection of Methods for Solving Stiff and Nonstiff Systems of Ordinary Differential Equations. Siam J. Sci. Stat. Comput 1983; **4**, 136--148.

21. Ridenhour, B., Kowalik, J. M., & Shay, D. K. . Unraveling R_0: Considerations for Public Health Applications. Am. J. Public Health. 2018; *108*(Suppl 6), S445–S454. https://doi.org/10.2105/AJPH.2013.301704r.

22. Ranjan. R. Predictions for COVID-19 outbreak in India using Epidemiological models medRxiv 2020.04.02.20051466; doi: https://doi.org/10.1101/2020.04.02.20051466.

23. Roosa, K.; Lee, Y.; Luo, R.; Kirpich, A.; Rothenberg, R.; Hyman, J.M.; Yan, P.; Chowell, G. a. Short-term Forecasts of the COVID-19 Epidemic in Guangdong and Zhejiang, China: February 13–23, 2020. J. Clin. Med. **2020**, 9, 596.

24. Roosa, K.; Lee, Y.; Luo, R.; Kirpich, A.; Rothenberg, R.; Hyman, J.M.; Yan, P.; Chowell, G. b. Real-time forecasts of the COVID-19 epidemic in China from Feb 5 to Feb 24, 2020. Infect Dis Modell, 5 (2020), pp. 256-263

25. World Health Organization. a. 2020. Available at: https://www.who.int/csr/don/05-january-2020-pneumonia-of-unkown-cause-china/en/

26. World Health Organization. b. 2020. Available at: https://covid19.who.int/

27. Worldometer. 2020. Available at https://www.worldometers.info/world-population/population-by-country/