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 29 April 2020 for the prediction. We estimate the parameters of our model by minimizing the negative log likelihood function using the Nelder- Mead method. We found that the epidemic peak reach July month and the COVID-19 disease will disappear in the month of October. We suggest that Algerian authorities need to implement a strict containment strategy over a long period to effectively reduce the epidemic size.

Keywords: Prediction, COVID-19, Algeria, the 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 (World Health Organization. a, 2020). The Chinese authorities identifies on 7 January the coronavirus disease (COVID-2019) as the causative virus (World Health Organization.a , 2020.) This ongoing epidemic continues to cause infections in other Chinese cities and in multiple countries around the world (Roosa et al. a, 2020). 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 (World Health Organization.b, 2020).

Algeria has reported the first case of COVID-19 on 25 February 2020, in Ouargla region and he was an Italian national (HAMIDOUCHE, 2020; Algerian Ministry of Health 2020). Two other cases were reported on 01 march 2020 in Blida region in the North of Algeria (HAMIDOUCHE, 2020; Algerian Ministry of Health 2020).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 (Algerian Ministry of Health, 2020).

Various earlier studies have predicted the confirmed cases of COVID-19 infection in different countries. (Kuniya, 2020) 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. b, 2020) 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, 2020) 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 coming months. (Hamidouche, 2020). According to his results, the number of infected cases in Algeria will exceed 5000 on April 7th, 2020 and it will double to 10000 on April 11th, 2020. (Hamidouche, 2020).

This study applies the Susceptible-Infected-Recovered (SIR) model without demographics (no births, deaths, or migration) (Kermack and McKendrick, 1927; Dietz 1967; Keeling and Rohani, 2008) to predict the daily confirmed cases of COVID-19 infection in Algeria. We use data of reported cases between February 25, 2020 and April 17, 2020. We predict the epidemic peak and the end-date of this disease. These results could be helpful for decision makers in Algeria to implement their strategies.

Method:

1) Model:

The SIR model without demographics is given by the following non linear system of ordinary differential equation (Kermack and McKendrick 1927, Keeling and Rohani, 2008, Harko, 2014):

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

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 recovered cases from the disease. The parameters β and γ are called the contact rate and the recovery 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 (Harko, 2014).

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 $\mathrm{S}(0)>0, \mathrm{I}(0)>0, R(0)=0$ (Keeling and Rohani, 2008).

The basic reproductive ratio $R_0=0$ 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'' (Keeling and Rohani, 2008).

We integrate numerically our non-linear system using the Livermore Solver for Ordinary Differential Equations (LSODA) algorithm (Hindmarsh, 1983; Petzold, ,1983).This algorithm handle 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 (Hindmarsh, 1983). 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 true numbers of daily infected cases of COVID-19 in Algeria), we suppose that y_t follows a Poisson distribution with parameter $\lambda = p\tilde{I}(\mathrm{t})$ (King, 2017) :

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

Where the parameter p is a constant reflects a combination of sampling efficiency and the detectability of infections (King, 2017) .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 (Nelder and Mead,1965) 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 (Gao and Han, 2010; baudin, 2010)

For more details about the Nelder-Mead algorithm, see ((Nelder and Mead, 1965; baudin, 2010; Gao and Han, 2010)

The procedure of estimation can be resumed in the following three steps (Nsoesie et al, 2013):

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 and discussion:

The data were obtained from Johns Hopkins University, Center for systems science and engineering (Johns Hopkins University, 2020). We used the number of daily reported confirmed cases for the COVID-19 epidemic in Algeria, from 25 February 2020 to 17 April 2020.

We used equation (1.2) with initial parameters $\beta = 0.95$, $\gamma = 0.75$ 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 $N = 43851044$ is the population size of Algeria (Worldometer, 2020).

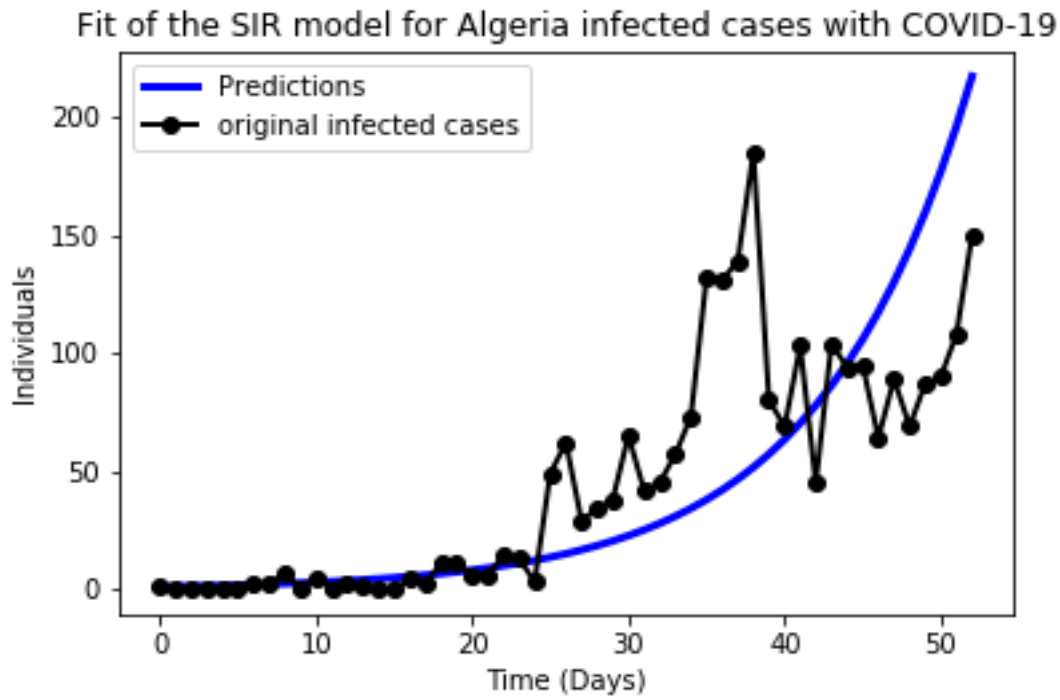
Following the procedure of estimation cited previously we get the new parameters: $\beta = 0.925$ and $\gamma = 0.82$.

Thus the basic reproduction number R_0 is given by:

$$R_0 = \frac{\beta}{\gamma} = \frac{0.925}{0.82} \approx 1.13$$

We substitute the new parameters in our SIR 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, from 25 February 2020 to 17 April 2020.

Figure 1



The mean absolute error (MAE) is used to evaluate the accuracy of our 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 (Hyndman and Athanasopoulos , 2018 ; Boudrioua, 2019). We get $MAE = 25$ for the prediction of new infected cases.

Figure 2 below shows the prediction of epidemic peak of COVID-19 in Algeria. We extract from this Figure that the predicted epidemic peak will be in 10 July 2020 and the estimated end-date of COVID-19 epidemic in Algeria is 23 October 2020.

Figure 2

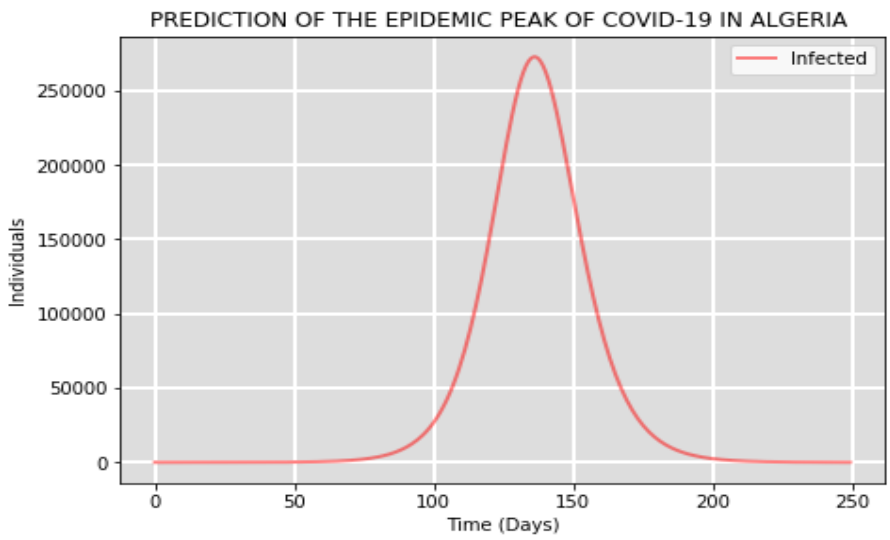
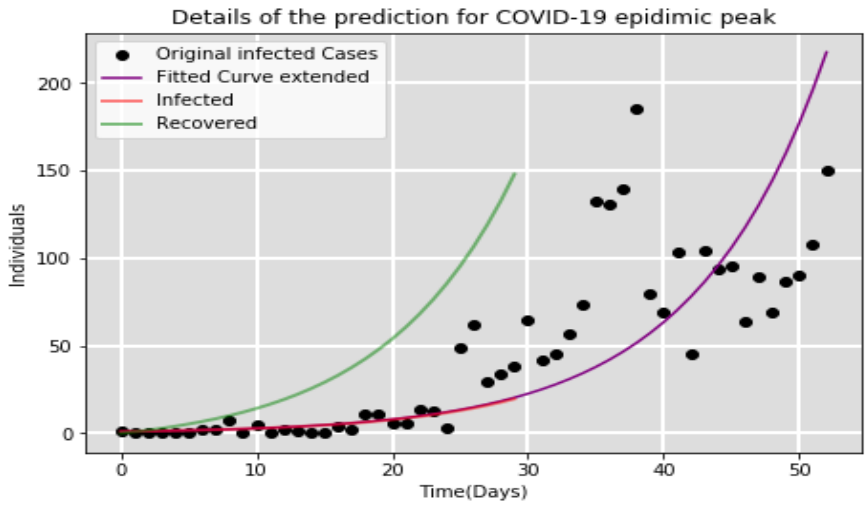


Figure 3 shows more details on the prediction for COVID-19 epidemic in Algeria.

Figure 3



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 17 April 2020. We have applied the SIR compartmental model without demographics in this task. We find that the COVID-19 epidemic peak reach the month of July and the end of this epidemic will be in the month of October.

Our findings agree with the results of (Kuniya, 2020) in the case of COVID-19 outbreak in Japan, and with the WHO's statement that COVID-19 will maybe disappear in the summer (Kuniya, 2020). Following (Kuniya, 2020) we suggest that a strict containment strategy is needed in Algeria, over a relatively long period, to reduce the epidemic size.

References:

[1] Algerian Ministry of Health.. (2020). http://covid19.sante.gov.dz/carte/.

[2] Baudin .M. (2010). Nelder mead user's manual. Available at: http://wiki.scilab.org/The_Nelder-Mead_Component.

[3] Boudrioua, M. S. (2019). Modeling and Forecasting the Algiers Stock Exchange Returns using the Box-Jenkins Methodology. Preprints 2019, 2019090134 (doi: 10.20944/preprints201909.0134.v1).

[4] Dietz. K. (1967). Epidemics and Rumours: A Survey. J. Roy. Stat. Soc. A **130** 505–528.

[5] Gao, F & Han, L. (2012). Implementing the Nelder-Mead algorithm with adaptive parameters. Comput Optim Appl 51.259-277.

[6] Hamidouche M. (2020) COVID-19 outbreak in Algeria: A mathematical Model to predict cumulative cases. medRxiv 2020.03.20.20039891; doi: https://doi.org/10.1101/2020.03.20.20039891

[7] Harko.T, Lobo. F. S. N and Mak. M. K. (2014). Exact analytical solutions of the Susceptible-Infected-Recovered (SIR) epidemic model and of the SIR model with equal deaths and births, Applied Mathematics and Computation, 236 (2014), 184–194.

[8] Hindmarsh, A. C. (1983) ODEPACK, A Systematized Collection of ODE Solvers; in p.55--64 of Stepleman, R.W. et al.[ed.] (1983) Scientific Computing, North-Holland, Amsterdam.

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

[10] Hyndman, R.J & Athanasopoulos, G. (2018). Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2. Accessed on Mars 2020.

[11] Keeling, M. & Rohani, P. (2008). Modeling infectious diseases in humans and animals. Princeton University Press.

[12] Kermack, W.O.& McKendrick, A.G. (1927). A Contribution to the Mathematical Theory of Epidemics. Proc. Roy. Soc. Lond. A **115** 700–721.

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

[14] Kuniya, T. (2020). Prediction of the Epidemic Peak of Coronavirus Disease in Japan,. J. Clin. Med. **2020**, 9, 789.

[15] Nelder, J.A. & Mead, R. (1965). A simplex method for function minimization. Comput. J. **7**, 308–313 .

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

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

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

[19] Roosa, K.; Lee, Y.; Luo, R.; Kirpich, A.; Rothenberg, R.; Hyman, J.M.; Yan, P.; Chowell, G. b . (2020). 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

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

[21] World Health Organization. b. (2020). Available at: https://covid19.who.int/

[22] Worldometer. (2020). https://www.worldometers.info/world-population/population-by-country/