Journal of Advances in Mathematics and Computer Science

36(3): 97-117, 2021; Article no.JAMCS.67121
ISSN: 2456-9968
(Past name: British Journal of Mathematics & Computer Science, Past ISSN: 2231-0851)

SIR Model Parameters Estimation with COVID-19 Data

Nilson C. Roberty[1*] and Lucas S. F. de Araujo[2]

[1]*Nuclear Engineering Program, Coppe, Universidade Federal do Rio de Janeiro, Av. Horócio Macedo, 2030, Bloco G, Sala 206, Centro de Tecnologia, Cidade Universitária, Ilha do Fundão, 21941-914, Rio de Janeiro, RJ, Brasil.*
[2]*Nuclear Engineering Department, Poli, Universidade Federal do Rio de Janeiro, Av. Horócio Macedo, 2030, Bloco G, Sala 206, Centro de Tecnologia, Cidade Universitária, Ilha do Fundão, 21941-914, Rio de Janeiro, RJ, Brasil.*

Authors' contributions

This work was carried out in collaboration between both authors. Both authors read and approved the final manuscript.

Article Information

DOI: 10.9734/JAMCS/2021/v36i330349
Editor(s):
(1) Dr. Junjie Chen, University of Texas at Arlington, USA.
(2) Dr. Rodica Luca, Gh. Asachi Technical University, Romania.
(3) Dr. Paul Bracken, The University of Texas RGV, USA.
Reviewers:
(1) Adil Osman Mageet, Ajman University, UAE.
(2) Mojeeb Al-Rahman El-Nor Osman El-Haj, International University of Africa, Sudan.
(3) Li Wang, Ningxia University, China.
(4) Theodore G. Lewis Naval, Postgraduate School Monterey, USA.
(5) Samson Wangila Wanyonyi, University of Eldoret, Kenya.
(6) Vipin Tiwari, Kumaun University, India.
Complete Peer review History: http://www.sdiarticle4.com/review-history/67121

Original Research Article

Received: 14 February 2021
Accepted: 18 April 2021
Published: 24 April 2021

Abstract

Based on the SIR model that divides the population into susceptible, infected and removed individuals, data about the evolution of the pandemic compiled by the Johns Hopkins University Center for Systems Science and Engineering (JHUCSSE) are integrated into the numerical system solution. The system parameters Rate of Contact β, Basic Reproduction Number $\mathcal{R}_0$ and Removal

**Corresponding author: E-mail: nilson@con.ufrj.br;*

Rate γ, also named Rate of Decay, are determined according to a ridge regression approach and a mobile statistical scheme with different averages. Data is automatically downloaded from https://raw.githubusercontent.com/CSSEGISandData/COVID-19. The main Python libraries used are Numpy, Pandas, Skit-Learn, Requests and Urllib.

Keywords: COVID-19; Python3; Inverse Problems; Ridge Regression; ODE; Fixed Points.

2010 Mathematics Subject Classification: 34A55, 62J07, 34XX, 37C25.

1 Introduction

The simplest model capable of describing the three main states of the phenomenology associated with 'COVID-19 infection' is the SIR model ([1],[2]). These three letters designate the three compartments grouping the population affected by the disease. The susceptible group (S) correspond to all people that can be in contact with the virus and contract the disease. By contracting the disease, susceptible people will become part of the temporary group of infected people (I). After the characteristic time of the evolution of the disease, those infected individuals migrate to the removed group (R) that counts both the people who get recovered and those who died due to the severity of clinical conditions. This model is based on two parameters traditionally denoted as β and γ, which contemplate two basic types of fundamental mechanisms found in the phenomena of reaction and decay ([3]).

The emergence of corona virus epidemic at the end of 2019 has led the world to an unprecedented crisis. The last infectious disease of this size had occurred a century ago, exponentially increasing the social consequences that followed the first world war. At this time, the importance of scientific treatment for pests and epidemics emerges in the wake of these consequences and the SIR model appears as a simple and adequate tool for the quantification of this type of problem.Important books on mathematical modeling of populations in biology and epidemiology have been published, we quote ([4]). As a nonlinear system of ordinary differential equations, the SIR model has an important position as a pure and applied mathematics subject. Since in the modelling of real infections the two main parameters depends on social contact and therapeutics procedures, and consequently are strongly determined by public politics, they will change with time and the SIR model will behave as a non autonomous dynamical systems ([5]). Also, very disputed political decisions based on different interests and beliefs of the authorities increase their stochastic nature. Interesting works has been done before the present outbreak, we cite only some([6],[7]). More recently([8]) investigate analytical features of the SIR model and their applications to COVID-19, and ([9]) comments those results. In a more recent work ([10]) he presents an analytical parameter estimation of COVID-19 data with a different methodology from the one used in this work. Other interesting recent paper treat the COVID-19 data forecasting problem ([11]). We will not treat this forecasting problem in this first work.

In section (2) we describe the SIR model and we give a brief description of the parameters. In subsection (2.1) we define the Basic Reproduction Number. In subsection (2.2) we inform more about the data spreadsheets which were used. In Section (3) we derive a fix point equation and define the concept of Local Pandemic Cycle. Section (4) describes how Python3.7 is used to extract the data from JHUCSSE website and produce our results. Section (5) describes the numerical implementation of the Tychonoff regularization for parameters reconstruction. The results are presented in section (6). Some concluding remarks are presented in section (7).

2 The SIR Model

The equations of the SIR model with time-dependent $\beta(t)$ and $\gamma(t)$ are:

$$\begin{cases} \frac{dS(t)}{dt} = -\beta(t)\frac{S(t)}{N}I(t) \\ \frac{dI(t)}{dt} = \beta(t)\frac{S(t)}{N}I(t) - \gamma(t)I(t) \\ \frac{dR(t)}{dt} = \gamma(t)I(t) \end{cases} \tag{2.1}$$

with the initial condition $(S(t_0), I(t_0), R(t_0)) = (S_0, I_0, R_0)$.

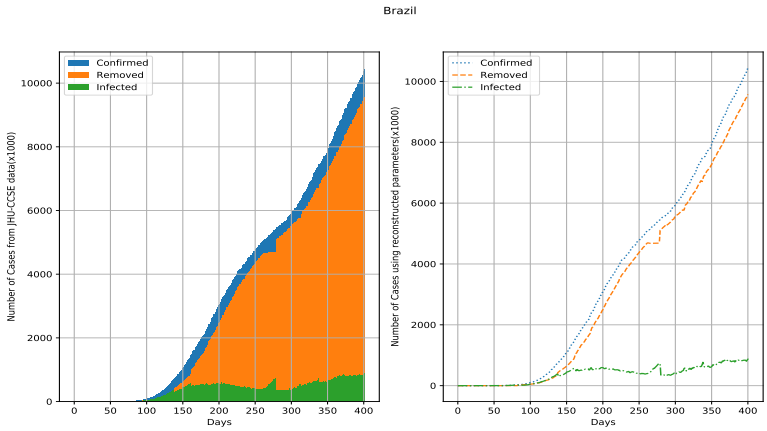

Fig. 1. Comparing SIR model evolution states Data compiled by Johns Hopkins University Center for Systems Science and Engineering (JHUCSSE) with Numerical Simulations of the system (2.1) using reconstructed β and γ parameters - The Brazilian Data Case

The functions $S(t), I(t), R(t)$ respectively represent the susceptible, infected and removed (recovered plus death) number of people at any given time. The shape of these equations allows us to make a parallel with the kinetic nuclear reaction equations which involves reactions and radioactive decay ([3]).

The parameter N can be interpreted as the total number of people involved in the epidemic outbreak. The sum of the three equations (2.1) establishes N as a time invariant

$$S(t) + I(t) + R(t) = N \tag{2.2}$$

The first parameter is the Contact Rate, β. It establishes the rate of contact per person in a population with N individuals. Since the susceptible fraction, $\frac{S}{N}$, represents the probability of finding a susceptible individual in the population, an infected person have in average $\beta\frac{S}{N}$ contacts with susceptible people.

The second parameter is the removal rate, γ, which controls the decay rate from the infected state to the removed state. These type of parameters are also found in nuclear reaction phenomenology. The probability of an individual leave the infected condition in an infinitesimal increment of time $\delta\tau$ is $\gamma\delta\tau$. Therefore, the probability of remaining in the infected condition is $1 - \gamma\delta\tau$. Thus, in the classical limit, we have the quantification of the probability that in a total time τ an individual remains in the condition of infected

$$\lim_{\delta\tau\to 0}(1-\gamma\delta\tau)^{\frac{\tau}{\delta\tau}} = e^{-\gamma\tau} \tag{2.3}$$

This behavior is typical in the nuclear radioactive decay phenomena. Meanwhile in the context of the epidemic outbreak it is observed that individuals are more likely to recover or die in a few days after contracting the infection, this simplification is systematically adopted and useful.

The population number N is normally very large when compared with the confirmed number of cases directly involved in the epidemic, at least in its beginning. Thus, the hypothesis $\frac{S}{N} \simeq 1$ makes the model linear. It is appropriate to define the Effective Value of the Rate of Contact β_{ef}.

$$\beta_{ef} = \beta\frac{S}{N} \tag{2.4}$$

The average number of contacts of an infected individual in an infinitesimal time $\delta\tau$ is $\beta_{ef}\delta\tau$. Thus $1-\beta_{ef}\delta\tau$ represents the probability that the infection will not be transmitted in the time interval $\delta\tau$. Since this interval has $\frac{\tau}{\delta\tau}$ sub-intervals, an infected individual remains unlikely to infect others with probability:

$$\lim_{\delta\tau\to 0}(1-\beta_{ef}\delta\tau)^{\frac{\tau}{\delta\tau}} = e^{-\beta_{ef}\tau}$$

Consequently the probability of the transmission is $1-e^{-\beta_{ef}\tau}$ ([1]).

2.1 The basic reproduction number $\mathcal{R}_0$

An individual who remains infected for a time τ will have in average $\beta\tau$ contacts and will infect susceptible individuals who have had contact with him ([1]). The probability that this individual remain infected in this interval of time is given by the equation(2.3), and the probability of finishing the period of infection at the moment τ will be $\gamma e^{-\gamma\tau}$ for all time τ. In this circumstance, the average Basic Reproduction Number of the infection will be:

$$\mathcal{R}_0 = \int_0^\infty \beta\tau\gamma e^{-\gamma\tau}d\tau = \frac{\beta}{\gamma}\ . \tag{2.5}$$

The Basic Reproduction Number $\mathcal{R}_0 := \frac{\beta}{\gamma}$ has a direct influence on local evolution of the pandemic. As it will be seen in the next sections, it controls the evolution of the epidemic cycle, as shown in the figures 1 and 2. The control of the Contact Rate between individuals can reduce the basic reproduction number to levels low enough to prevent an epidemic outbreak of great proportions. All this depends of the society's ability to reduce its values to below 1. Different countries are at different stages in the development of the pandemic.

Effective health authorities efforts to contain the outbreak will introduce a trend towards a reduction in the basic number of reproduction by decreasing the Contact Rate β. Further study can be conducted also with network science ([1]).

2.2 The COVID-19 data

On the website https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases, ([12]), it is available 3 time series with information used for analysis on the cumulative number of cases of Covid19 registered as confirmed, recovered and deaths for several countries. For the automatic download there is an alternative GitHub link in the website https://raw.githubusercontent.com which is synchronized with the JHUCSSE data. For some countries, there is discrimination against provinces and regions. The combination of the numerical solution of the dynamic SIR equations solved as a deterministic system of ordinary differential equations with parameters β , γ and N allows a temporal characterization of the system states.

Those results can be compared with the observational data in the JHUCSSE worksheets:

$$\{confirmed.csv, recovered.csv, death.csv\} \ . \tag{2.6}$$

In this way, we establish an Observational-Evolution model that can be seen both as a Statistical filtering problem for discrete times and also as a problem of deterministic extrapolation with constant parameters in a long-range time. In this first work we will not explore the extrapolation ([13]),([14]).

3 The Existence of the Outbreak in the SIR Model

The first system equation (2.1) with time constant parameters β and γ can be solve explicitly giving $S(t) = S(0)exp(-\frac{\beta \int_0^t I(\tau)d\tau}{N})$. By integrating the third equation $R(t) - R(0) = \gamma \int_0^t I(\tau)d\tau)$. The combination of these equations gives

$$S(t) = S(0)exp(-\mathcal{R}_0 \frac{R(t) - R(0)}{N}) \tag{3.1}$$

The replacement of (3.1) in the equation (2.2) provides the following equation with normalized variables $r(t) := \frac{R(t)}{N}$,$i(t) := \frac{I(t)}{N}$ and $S(t) := \frac{S(t)}{N}$ for the removal rate

$$\frac{dr(t)}{dt} = \gamma(1 - r(t) - (1 - i(0)))e^{-\mathcal{R}_0 r(t)}). \tag{3.2}$$

This equation (3.2) is equivalent to the SIR model in the case of constant parameters.

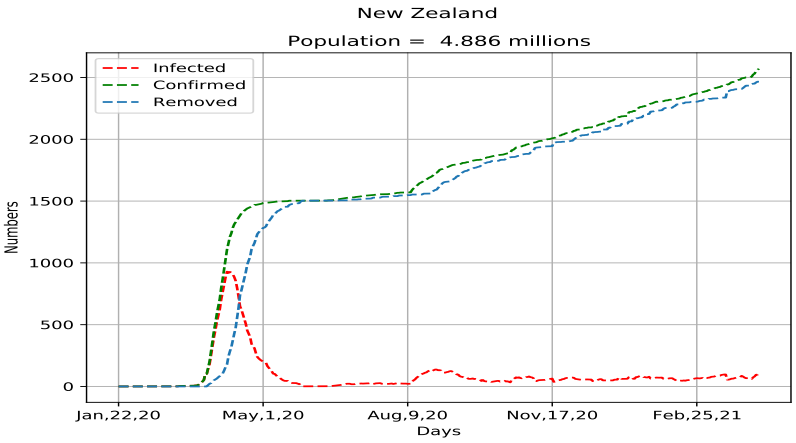

Fig. 2. Local pandemic cycle in New Zealand. The Local Pandemic Cycle is defined between March and June 2020 by the area delimited by the curves confirmed and removed, respectively

We can see that the rates $\frac{dr}{dt}$ and $\frac{di}{dt}$ are both null at the beginning and at the end of the outbreak. The evolution of the curves that represents the number of confirmed cases $c(t) = r(t) + i(t)$ and that represents the number of removed cases $r(t)$ delimits a cycle of the state evolution process. The area between the curves can be seen as a measure of the size of the outbreak. This region of the process will be called as Local Pandemic Cycle. And, of course, after a complete cycle, nothing

prevents a new outbreak from starting after the final state of the first outbreak, starting a new local pandemic cycle. In the Fig. 2 we can see the formation of a complete first cycle and the adequate control of the disease. The rates $\frac{dr}{dt}$ and $\frac{dc}{dt}$ are both null at the beginning and at the end of the outbreak between March and June 2020. At this period of time, only the residual number of infected people is observed.

So, the Fig. 2 shows the first New Zealand epidemic cycle in which the result of controlling the $\mathcal{R}_0$ parameter to values under 1 led to a low value of $r_\epsilon \simeq \frac{310}{4886000} \simeq 6.3410^{-5}$ which measures the infected fraction of the population.

Since at the extremes of a cycle the derivatives are zero, we get the following fixed point equation, which is also known as giant component size equation,

$$r = 1 - (1 - \epsilon)e^{-\frac{\beta}{\gamma}r} \tag{3.3}$$

The numerical solution of this fix point equation is done with the Newton method.

When the influence of the initial disturbance $\epsilon = \frac{I_0}{N} \simeq 0$ is neglected, this equation will always have a zero solution, $r(t) = 0$, and can, depending of the value $\mathcal{R}_0 = \frac{\beta}{\gamma}$ have another non-null solution([1]). In the case of $\mathcal{R}_0 = 2$, the non-null solution is $r \simeq 0.7968121301668857$, showing that almost 80% of the initially susceptible population will be infected during the cycle formation. This is the size of this outbreak for this $\mathcal{R}_0$ parameter. In network science theory, this number quantifies the so-called "giant component" that indicates a phase change in the system.

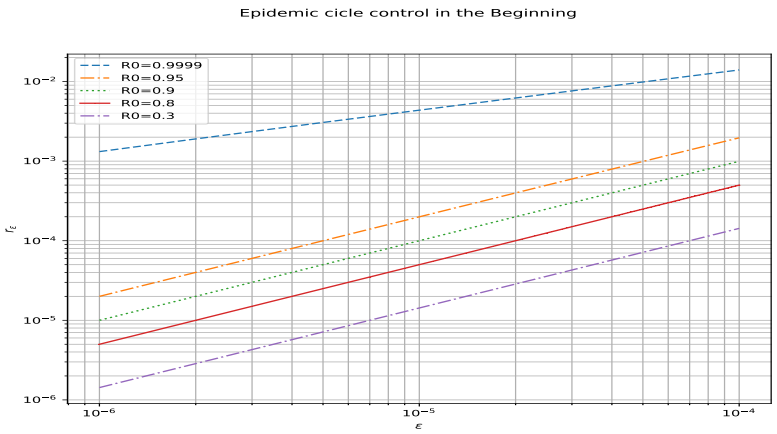

Fig. 3. Variation of the giant component size with the initial infinitesimal fraction of infected people in an outbreak controlled at the beginning

On the other hand, the same calculation with $\mathcal{R}_0 = 0.9999999$ does not have a second non-null solution. This indicates the absence of outbreak with no formation of a large giant component. In this case, the number of infected individuals, which is initially small, decreases instead of growing as in the first case, and the infection is extinguished instead of propagates forming the outbreak. To adequate the fix point model giving by equation (3.3) to predict the outbreak it is necessary to introduce a non null $\epsilon \neq 0$ disturbance term at the initial value. Doing this we take into account the nucleation of the infection which must be initiated by some infected person. This will bring up a new fixed point solution to the fixed point equation (3.3), $r_\epsilon = \frac{C_\epsilon}{N} = \frac{R_\epsilon}{N}$. If the infected population is very small at beginning and quantitatively negligible when compared to the total of

susceptible individuals, it is then important the non-null ϵ in the fix point equation, since it gives a quantification of the total fraction of recovered or deaths individuals even when $\mathcal{R}_0 < 1$. This is the case of the absence of an outbreak of higher proportions.

Calculations for formation of the "small giant component" are presented in Fig. 2. The log-log graphicshows the final fraction of infected people r_ϵ for an initial fraction ϵ of infected individuals during a controlled epidemic cycle.

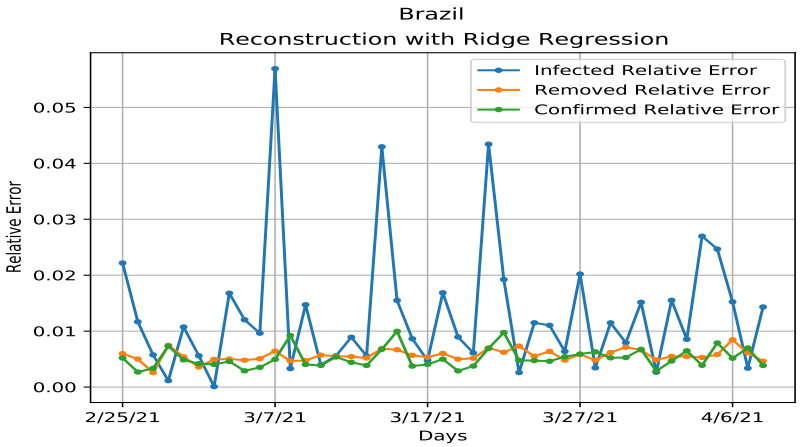

Fig. 4. Relative errors for infected, removed and confirmed population in SIR model solved with reconstructed parameters β and γ. The Brazilian case

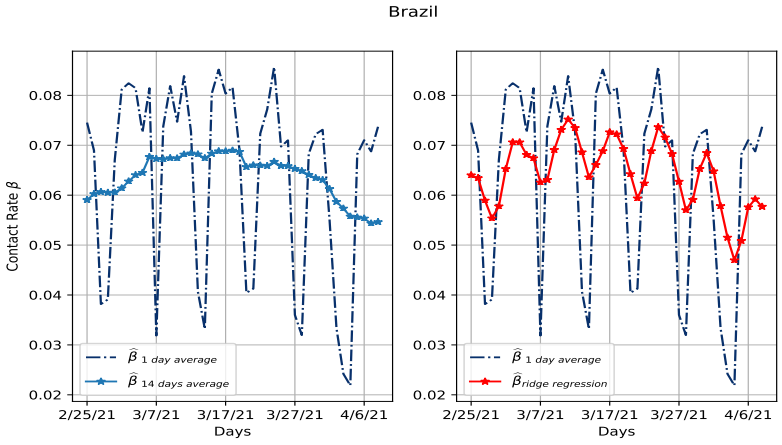

Fig. 5. In the first figure we compare β values calculated by 14-days average method with the 1-day average estimation. In the second one the ridge regression estimation is compared with the 1-day average estimation

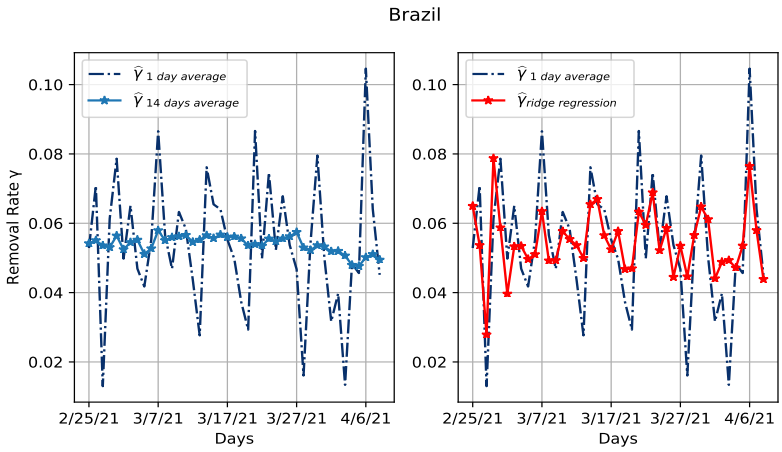

Fig. 6. In the first figure we compare γ values calculated by 14-days average method with the 1-day average estimation. In the second one the ridge regression estimation is compared with the 1-day average estimation

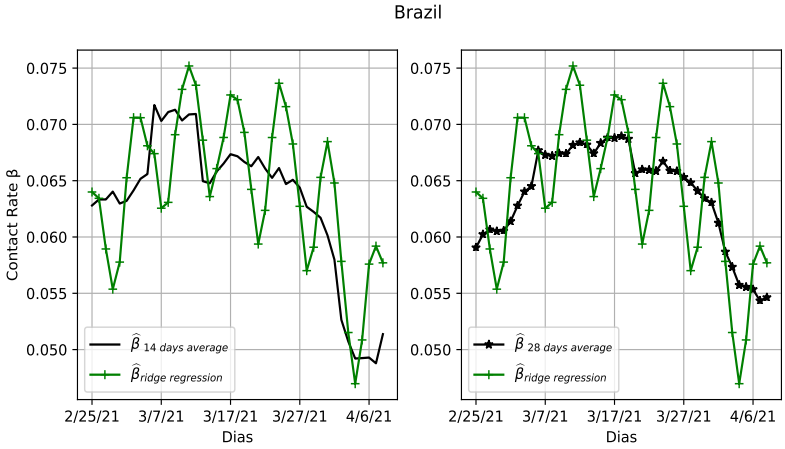

Fig. 7. In the first figure we compare β values calculated by 14-days average method are compared with ridge regression. In the second one the 28-days average estimation is compared with ridge regression

4 Python 3.7 Implementation of COVID-19 Data Processing

In order to automate the extraction of information, a library was developed to be capable of extracting and processing the time series data (2.6), as well as combining these series with deterministic solutions of the SIR equations. The result is being shown in the graphical that makes up the selection shown in this work. It is focus on the graphs presented mainly in the data referring to Brazil.

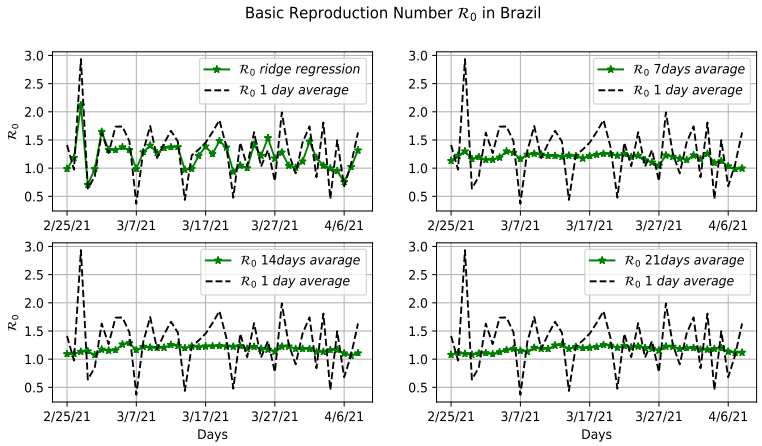

Fig. 8. Comparison of estimation of the Basic Number of Reproduction $\mathcal{R}_0$ for ridge regression, 7 , 14 and 21 days average with 1 day average

The system, however, works for any country in that database and can be easily adapted to other databases. The data made available by JHUCSSE were processed in the Python 3.7 language. A reference on Python we quote ([15]). The Python libraries such as Pandas for extracting information and Skit-Learning ([16]) for parameters prevision are used. In addition, other libraries such as SciPy, Math and Matplotlib were extensively used.

5 Estimates for the Contact Rate Parameter β and the Removal Rate Parameter γ

The main statistical tool used for estimating the parameters β and γ was the ridge regression, which implements the Tychonoff regularization for the linear system formed with the regression of the time series data.

The SIR equations (2.1) can be rewritten in terms of the number of confirmed cases up to a time t, $C(t) := I(t) + R(t)$. If we notice that

$$\frac{S(t)}{N} = 1 - \frac{I(t) + R(t)}{N} \tag{5.1}$$

The second equation of the SIR model can be rewritten as:

$$-N\frac{d\ln(1 - \frac{(I(t)+R(t)}{N})}{dt} = \beta(t)I(t) \tag{5.2}$$

while the equation for the removed people rate remains unchanged. Thus, by integrating these equations between $[t - \tau, t]$, we obtain

$$-N\ln(1 - \frac{(I(t) + R(t))}{N}) + N\ln(1 - \frac{(I(t-\tau) + R(t-\tau))}{N}) = \int_{t-\tau}^{t} \beta(t_1)I(t_1)dt_1$$

and

$$R(t) - R(t-\tau) = \int_{t-\tau}^{t} \gamma(t_1) I(t_1) dt \ .$$

Defining the $\overline{\beta}^{\tau}$ and $\overline{\gamma}^{\tau}$ means as:

$$\overline{\beta}^{\tau} = \frac{\int_{t-\tau}^{t} \beta(t_1) I(t_1) dt_1}{\int_{t-\tau}^{t} I(t_1) dt_1} \tag{5.3}$$

$$\overline{\gamma}^{\tau} = \frac{\int_{t-\tau}^{t} \gamma(t_1) I(t_1) dt_1}{\int_{t-\tau}^{t} I(t_1) dt_1} \tag{5.4}$$

we have the average values for τ days with respect the total number of infected cases in these days

$$\overline{\beta}^{\tau} = \frac{-N \ln(1 - \frac{(I+R)}{N}(t)) + N \ln(1 - \frac{(I+R)}{N}(t-\tau))}{\int_{t-\tau}^{t} I(t_1) dt_1}$$

$$\overline{\gamma}^{\tau} = \frac{R(t) - R(t-\tau)}{\int_{t-\tau}^{t} I(t_1) dt_1} \tag{5.5}$$

The expression for $\overline{\beta}^{\tau}$ can be written in terms of a linear factor and a non-linear correction factor:

$$\overline{\beta}^{\tau} = \frac{(I+R)(t) - (I+R)(t-\tau)}{\int_{t-\tau}^{t} I(t_1) dt_1} \frac{-N \ln(1 - \frac{(I+R)}{N}(t)) + N \ln(1 - \frac{(I+R)}{N}(t-\tau))}{(I+R)(t) - (I+R)(t-\tau)}$$

The linear factor can be identified as β_{ef} given by equation (5.6)

$$\overline{\beta}^{\tau}_{ef} = \frac{(I+R)(t) - (I+R)(t-\tau)}{\int_{t-\tau}^{t} I(t_1) dt_1} \tag{5.6}$$

and the non-linear factor:

$$\overline{F}^{\tau}_{\beta,nL} = \frac{-N \ln(1 - \frac{(I+R)}{N}(t)) + N \ln(1 - \frac{(I+R)}{N}(t-\tau))}{(I+R)(t) - (I+R)(t-\tau)}$$

Therefore we have:

$$\overline{\beta}^{\tau} = \overline{\beta}^{\tau}_{ef} \overline{F}^{\tau}_{\beta,nL} \tag{5.7}$$

Note that in the limit $\tau \to 0$ we have:

$$\overline{F}^{\tau}_{\beta,nL} \to \frac{1}{1 - \frac{(I+R)}{N}} = \frac{N}{S}$$

This is the inverse of the term that contributes for the system non-linearity. Therefore, the non-linear correction factor will only be significant for a very high percentage of confirmed cases, what is not observed in the time series data. In this way, for these first pandemic months, the contribution of the non-linear factor is negligible and $\beta \simeq \beta_{ef}$.

The main statistical tool used for estimating the parameters β and γ was the ridge regression, which implements the Tychonoff regularization for the linear system formed with the regression of the time series data. When the regularization considers daily updating, the equations (5.6) and (5.5) become:

$$\beta_L = \frac{C(t+1) - C(t)}{C(t) - R(t)} \simeq \overline{\beta}^{1}_{L} \tag{5.8}$$

$$\gamma = \frac{R(t+1) - R(t)}{C(t) - R(t)} \simeq \overline{\gamma}^{1} \tag{5.9}$$

By the n-days average method we mean the calculation of γ and β with equations (5.5) and (5.6) with $\tau = n$.

Prediction of β and γ mean values using Tychonoff regularization.

The representation for the estimated values $\widehat{\beta}(t)$ and $\widehat{\gamma}(t)$ using multi-linear regression is:

$$\widehat{\beta}_J(t) = a_0 + \sum_{j=1}^{J} a_j \beta(t-j) \tag{5.10}$$

$$\widehat{\gamma}_K(t) = b_0 + \sum_{j=1}^{K} b_j \gamma(t-j) \tag{5.11}$$

where J and K are the orders of the two filters and a_j , $j = 1, ..., J$ and b_j , $j = 1, ..., K$ are the coefficients of the time series values that participate in the regression.

The coefficients are calculated minimizing the functional of Tychonoff

$$F[a_1, ..., a_J] := \sum_{t=J}^{T-2} (\beta(t) - \widehat{\beta(t)})^2 + \alpha_1 \sum_{j=0}^{J} a_j^2 \tag{5.12}$$

$$F[b_1, ..., b_J] := \sum_{t=K}^{T-2} (\gamma(t) - \widehat{\gamma(t)})^2 + \alpha_2 \sum_{j=0}^{K} b_j^2 \tag{5.13}$$

where α_1 and α_2 are regularization parameters.

The implementation of this optimization problem is already done in the Skit-Learn library ([16],[17]).

6 Results

The main result of this work is the interpretation of the temporal data evolution in the current COVID-19 pandemic with data based graphs in terms of the SIR model with time dependent parameters.

A second important result is the investigation about the Local Pandemic Cycle concept, with the basic observation that there is a formation of a giant component even when the basic reproduction number is less than one, $\mathcal{R}_0 < 1$.

Another important result is the derivation of the giant component size fix point equation (3.3) which allow us to calculate the fraction of number of removed people at the end of an epidemic cycle for a negligible initial disturbance. It is shown by this equation, how the parameter $\mathcal{R}_0$ can control the dynamics of the epidemic cycle.

Lastly, it is shown in the work that $\frac{S(t)}{N}$ is a good simplification to the model which enable us to compute the systems parameters by ridge regression or n-days average. It is show that these parameters can be used to reconstruct the systems variables and arrive in values with controlled relative error as seen in Fig. 4.

A brief summary of the graphics.

Fig. 1 shows the temporal evolution of the incomplete local pandemic cycle in Brazil. Those data compare values from JHUCSSE data with numerical values solution of the SIR differential equation system done with parameters estimated from ridge regression of the same data.

Fig. 2 shows the formation of a Local Pandemic Cycle in New Zealand between march and July 2020. Since the Basic Reproduction Number has been controlled to be less than one, $\mathcal{R}_0 < 1$, the initial infinitesimal fraction of infect people start locally the pandemic but the pandemic die out

locally. This shown one strict correlation between the control measures adopted by local authorities and the local control of the pandemic.

Fig. 3 shows the fact that every epidemic outbreak has the form of a giant component, even if that component is infinitesimal when compared to the quantities of nations populations. This type of structure appears when actions are made to reduce the Contact Rate reducing the Basic Reproduction Number to values $R_0 < 1$ at the beginning of the epidemic outbreak. An example of this is shown in Fig. 2 and in Fig. 10.

Fig. 4 shows typical relative errors occurring when the dynamical systems SIR is solved with the reconstructed parameters. We observe a good agreement between the compiled observational data evolution and our numerical solution of the SIR model with the reconstructed parameters.

Figs. 5 and 6 present estimations for β and γ, respectively, using 14 days average and the ridge regression method and compare it with the 1 day average estimation. The ridge regression preserves oscillations from the data with some smoothing and this permits a more realist adequacy of the SIR model as observed in Fig. 1.

Fig. 7 shows the estimated average Contact Rate β for 14 and 28 days, respectively, and compare it with the reconstruction by ridge regression. We see that the ridge regression preserves the oscillations in the observational data. An excessive smoothing is observed in these 14 and 28 averages.

Fig. 8 makes use of several averages that point to values that can be associated with the Basic Number of Reproduction $\mathcal{R}_0$ using the average values for 2, 7, 14 and 28 days. Although the ridge regressions oscilations shows the difficulty in the control of $\mathcal{R}_0$, with a increase and decrease of the parameter value, the averages shown the the parameter is in fact in a level between one and one and half, perhaps $\mathcal{R}_0 \approx 1,2$. An important alert to authorities.

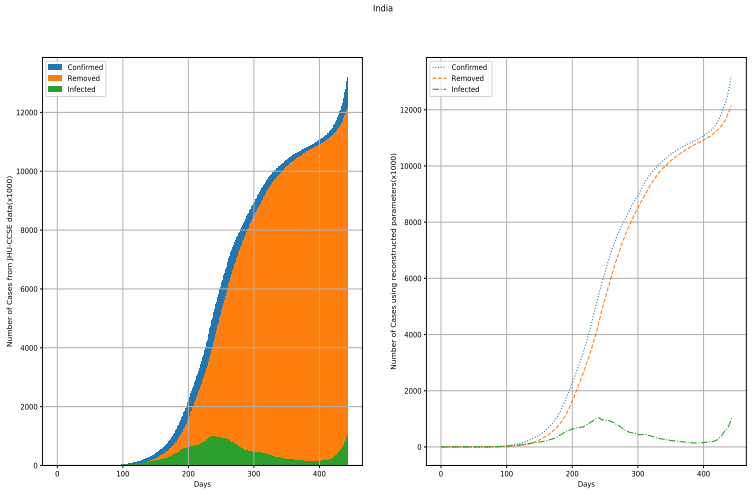

Fig. 9. Comparing SIR model evolution states Data compiled by Johns Hopkins University Center for Systems Science and Engineering (JHUCSSE) with Numerical Simulations of the system (2.1) using reconstructed β and γ parameters - The India Data Case

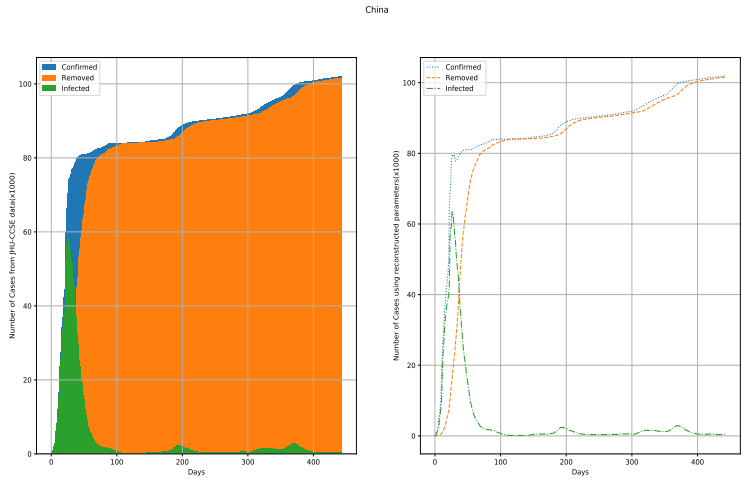

Fig. 10. Comparing SIR model evolution states Data compiled by Johns Hopkins University Center for Systems Science and Engineering (JHUCSSE) with Numerical Simulations of the system (2.1) using reconstructed β and γ parameters - The Chinese Data Case

Fig. 9 shows the temporal evolution of the incomplete local pandemic cycle in India. Those data compare values from JHUCSSE data with numerical values solution of the SIR differential equation system done with parameters estimated from ridge regression of the same data.

Fig. 10 shows the temporal evolution of the incomplete local pandemic cycle in China. Those data compare values from JHUCSSE data with numerical values solution of the SIR differential equation system done with parameters estimated from ridge regression of the same data.

7 Conclusions

The results found are conveniently presented in a graphic form to facilitate understanding and use it in decision making, as well as to warn about the seriousness of the consequences of the permissiveness in acceptance of inadequate parameters values, which here means a high level of Contact Rates.

It can be concluded that the ridge regression is an appropriated tool for estimation of parameters for the SIR model and numerical reconstruction of observational data in short range. Besides that Python 3.7 is an adequate computational ambient for treat this kind of data.

By inspection of the graphics, it can be concluded that the n days average can facilitate data inspection due to its smoothing properties.

The immediate lockdown during the pandemic outbreak and social distancing are the main tools for local pandemic control by controlling the Contact Rate β and consequently controlling the Basic Reproduction Number $\mathcal{R}_0$.

Future works involving forecast of investigated data the consequences of parameters in different levels, adequate or not, will be developed.

Acknowledgements

To Brazilian agencies CNPq and Capes for financial support.

Competing Interests

Authors have declared that no competing interests exist.

References

[1] Newman M. Networks. Oxford University Press; 2018.

[2] Kermack William Ogilvy, McKendrick Anderson G. A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character. 1927;115(772):700-721.

[3] Kaplan I. Nuclear Physics. Addison Wesley; 1963.

[4] Castillo-Chavez Carlos. Mathematical models in population biology and epidemiology. New York: Springer; 2012.

[5] Kloeden Peter E, Rasmussen Martin. Nonautonomous dynamical systems. American Mathematical Soc; 2011.

[6] Allen Linda JS. A primer on stochastic epidemic models: Formulation, numerical simulation, and analysis. Infectious Disease Modelling Journal; 2017.

[7] Cai Y, Kang Y, Banerjee M, Wang W. A stochastic SIRS epidemic model with infectious force under intervention strategies. Journal of Differential Equations. 2015;259(12):7463-7502.

[8] Kudryashov Nikolay A, Chmykhov Mikhail A, Vigdorowitschr Michael. Analytical features of the SIR model and their applications to COVID-19. Applied Mathematical Modelling. 2021;90:466-473.

[9] Prodanov Dimiter. Comments on some analytical and numerical aspects of the SIR model. Applied Mathematical Modelling. 2021;95:236-243.

[10] Prodanov Dimiter. Analytical parameter estimation of the SIR epidemic model. Applications to the COVID-19 pandemic. Entropy. 2021;23(1):59.

[11] Perc Matja, et al. Forecasting Covid-19. Frontiers in Physics. 2020;8:127.

[12] Novel coronavirus (COVID-19) cases data. Available:https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases
(Accessed from 2020 May to 2021 March)

[13] Kaipio J, Somersalo E. Statistical and computational inverse problems. Applied Mathematical Science. Springer Science & Business Media. 2006;160.

[14] Brunton Steven L, Nathan Kutz J. Data-driven science and engineering: Machine learning, dynamical systems, and control. Cambridge University Press; 2019.

[15] Hill C. Learning scientific programming with Python. Cambridge University Press; 2016.

[16] Raschka Sebastian. Python machine learning. Packt publishing ltd; 2015.

[17] Chen Yi-Cheng, Lu Ping-En, Chang Cheng-Shang, Liu, Tzu-Hsuan. A Time-dependent SIR model for COVID-19 with undetectable infected people; 2020.
arXiv preprint arXiv:2003.00122

Appendix

We select some countries to extract its data from the JHUCSSE spreadsheet. Results are similar to those presented for the Brazilian Case.

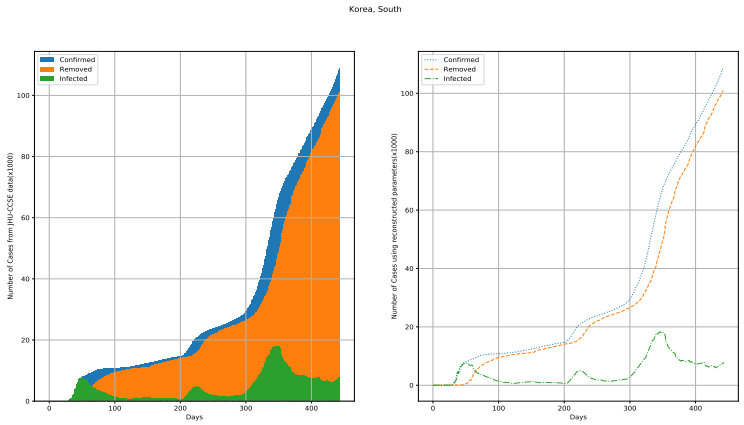

Fig. 11. Comparing SIR model evolution states Data compiled by Johns Hopkins University Center for Systems Science and Engineering (JHUCSSE) with Numerical Simulations of the system (2.1) using reconstructed β and γ parameters - The South Korea Data Case

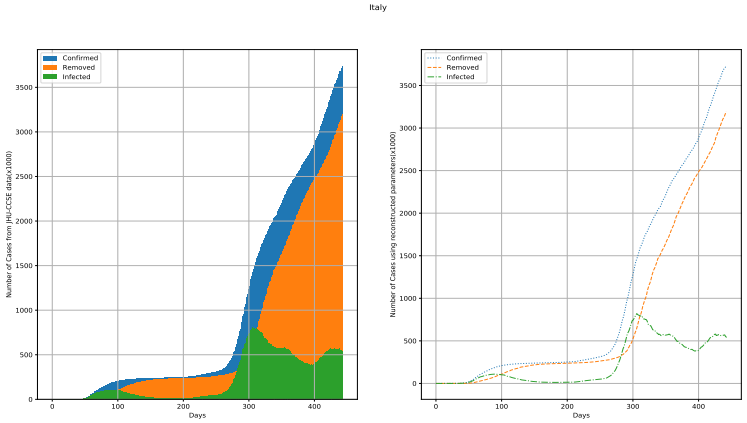

Fig. 12. Comparing SIR model evolution states Data compiled by Johns Hopkins University Center for Systems Science and Engineering (JHUCSSE) with Numerical Simulations of the system (2.1) using reconstructed β and γ parameters - The Italy Data Case

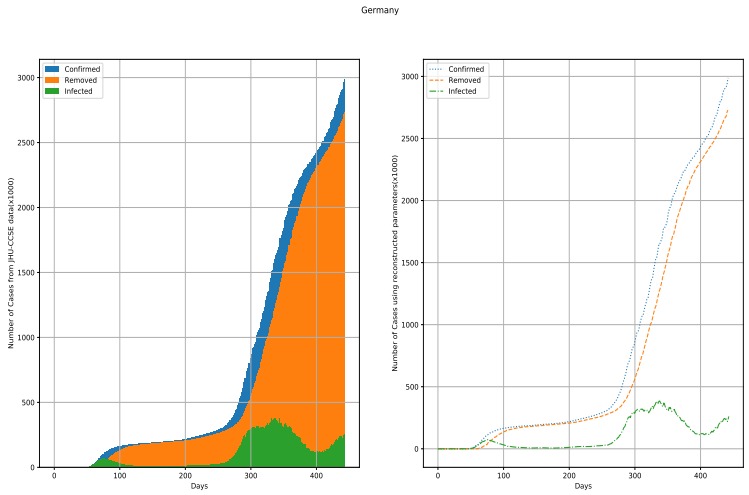

Fig. 13. Comparing SIR model evolution states Data compiled by Johns Hopkins University Center for Systems Science and Engineering (JHUCSSE) with Numerical Simulations of the system (2.1) using reconstructed β and γ parameters - The Germany Data Case

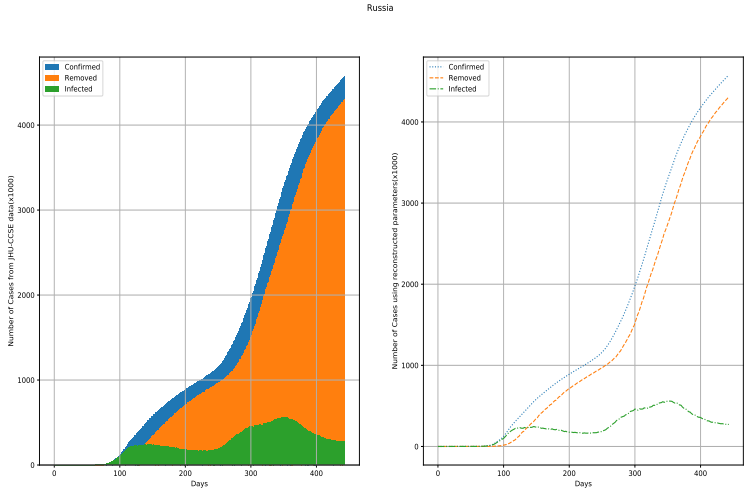

Fig. 14. Comparing SIR model evolution states Data compiled by Johns Hopkins University Center for Systems Science and Engineering (JHUCSSE) with Numerical Simulations of the system (2.1) using reconstructed β and γ parameters - The Russia Data Case

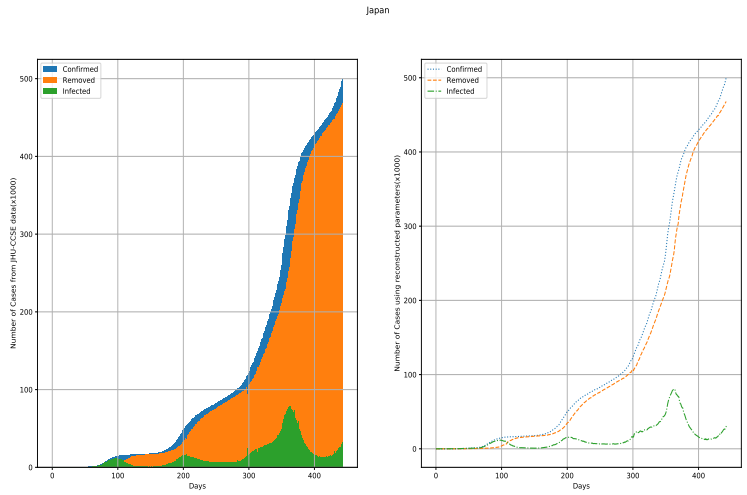

Fig. 15. Comparing SIR model evolution states Data compiled by Johns Hopkins University Center for Systems Science and Engineering (JHUCSSE) with Numerical Simulations of the system (2.1) using reconstructed β and γ parameters - The Japanese Data Case

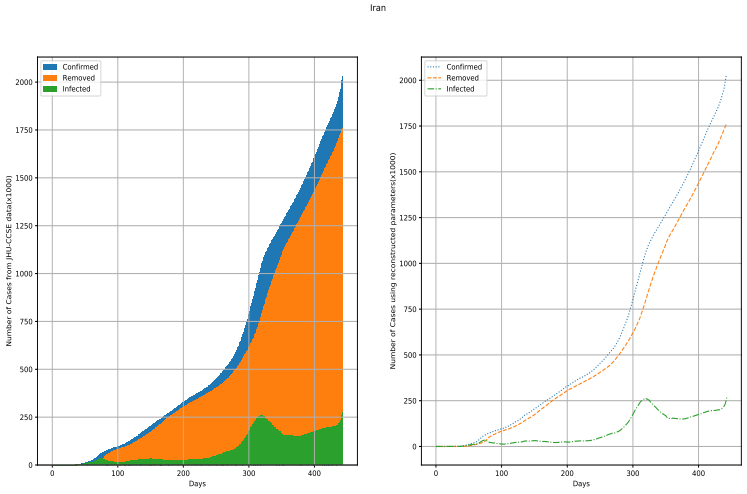

Fig. 16. Comparing SIR model evolution states Data compiled by Johns Hopkins University Center for Systems Science and Engineering (JHUCSSE) with Numerical Simulations of the system (2.1) using reconstructed β and γ parameters - The Iran Data Case

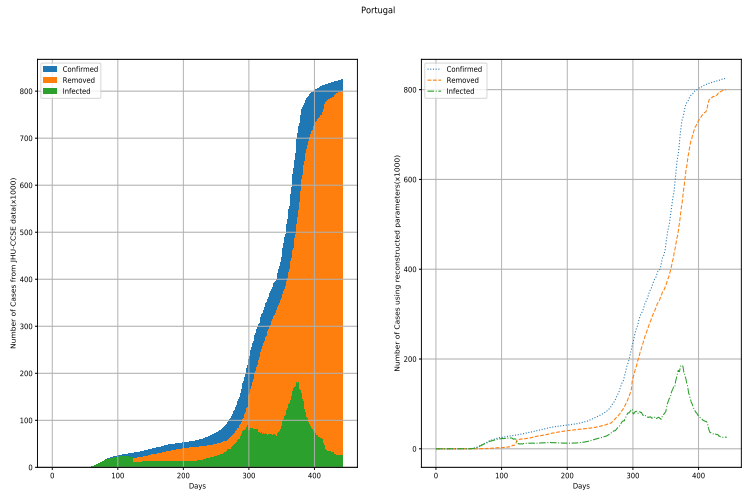

Fig. 17. Comparing SIR model evolution states Data compiled by Johns Hopkins University Center for Systems Science and Engineering (JHUCSSE) with Numerical Simulations of the system (2.1) using reconstructed β and γ parameters - The Portugal Data Case

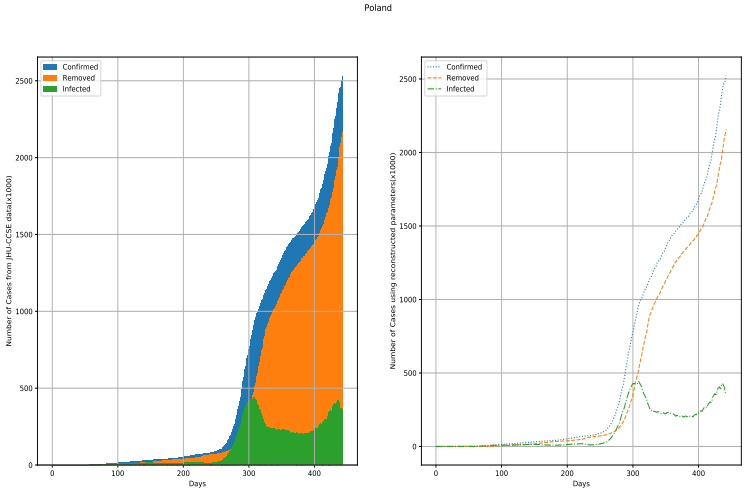

Fig. 18. Comparing SIR model evolution states Data compiled by Johns Hopkins University Center for Systems Science and Engineering (JHUCSSE) with Numerical Simulations of the system (2.1) using reconstructed β and γ parameters - The Poland Data Case

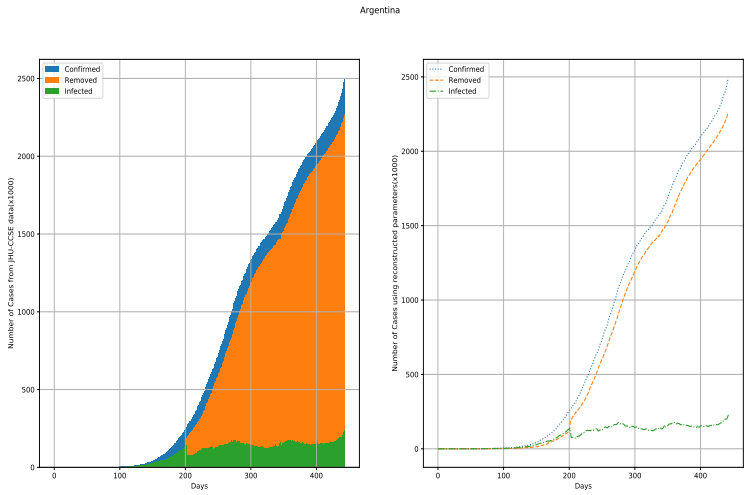

Fig. 19. Comparing SIR model evolution states Data compiled by Johns Hopkins University Center for Systems Science and Engineering (JHUCSSE) with Numerical Simulations of the system (2.1) using reconstructed β and γ parameters - The Argentina Data Case

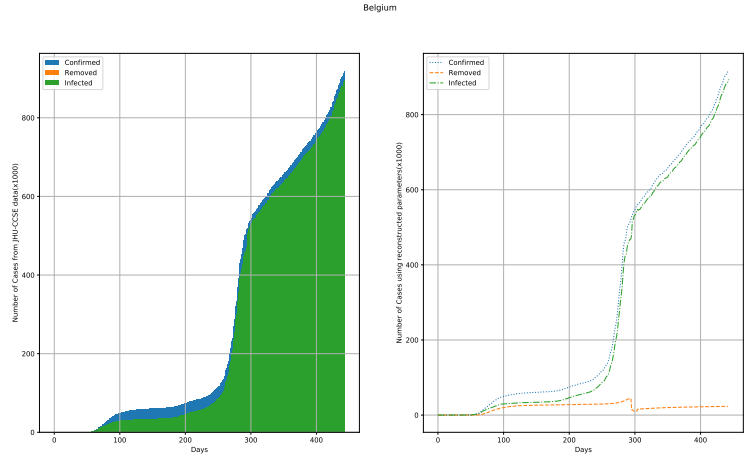

Fig. 20. Comparing SIR model evolution states Data compiled by Johns Hopkins University Center for Systems Science and Engineering (JHUCSSE) with Numerical Simulations of the system (2.1) using reconstructed β and γ parameters - The Belgium Data Case. The recovery sheet for this country was no longer updated since 11/11/2020

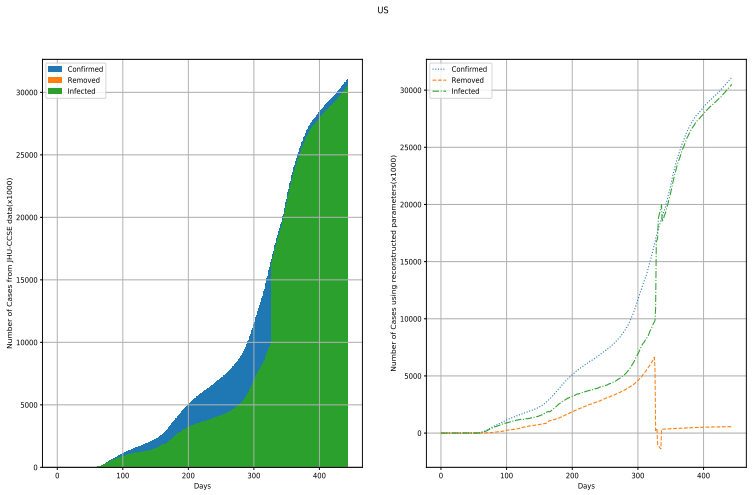

Fig. 21. Comparing SIR model evolution states Data compiled by Johns Hopkins University Center for Systems Science and Engineering (JHUCSSE) with Numerical Simulations of the system (2.1) using reconstructed β and γ parameters - The United State of America Data Case. The recovery sheet for this country was no longer updated since 12/14/2020

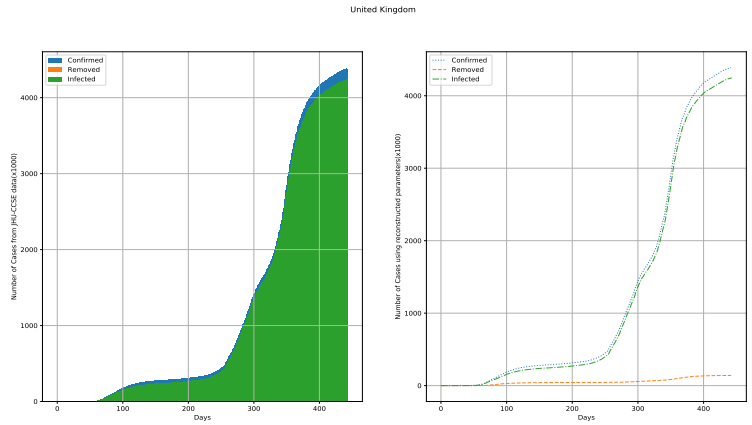

Fig. 22. Comparing SIR model evolution states Data compiled by Johns Hopkins University Center for Systems Science and Engineering (JHUCSSE) with Numerical Simulations of the system (2.1) using reconstructed β and γ parameters - The United Kingdom Data Case.The recovery sheet for this country is not being updated in the same way as other countries

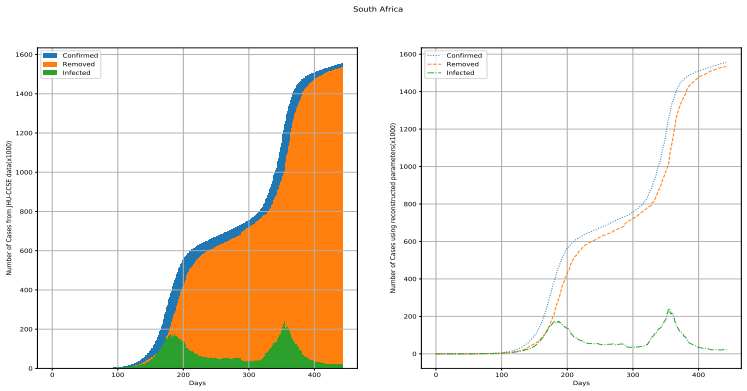

Fig. 23. Comparing SIR model evolution states Data compiled by Johns Hopkins University Center for Systems Science and Engineering (JHUCSSE) with Numerical Simulations of the system (2.1) using reconstructed β and γ parameters - The South Africa Data Case

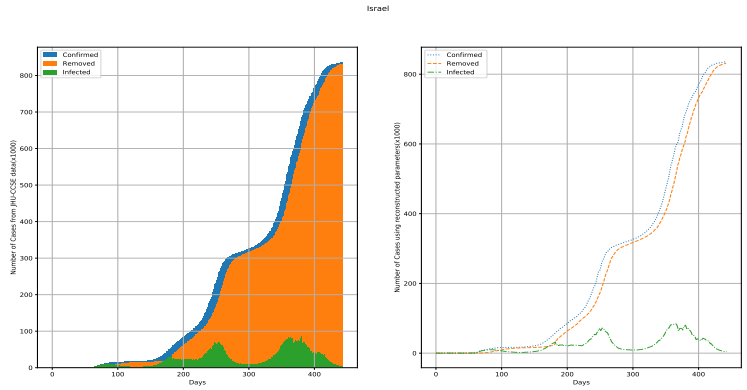

Fig. 24. Comparing SIR model evolution states Data compiled by Johns Hopkins University Center for Systems Science and Engineering (JHUCSSE) with Numerical Simulations of the system (2.1) using reconstructed β and γ parameters - The Israel Data Case

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
http://www.sdiarticle4.com/review-history/67121