Fitting SIR model to COVID-19 pandemic data and comparative forecasting with machine learning

Mouhamadou A.M.T. Baldé
University of Cheikh Anta Diop.
Laboratory of Mathematics of Decision and Numerical Analysis.
BP 45087, 10700. Dakar, Senegal.
Department of Mathematics of Decision(DMD)-FASEG.
mouhamadouamt.balde@ucad.edu.sn

Abstract

In this work, we use a classical SIR model to study COVID-19 pandemic. We aim to deal with the SIR model fitting to COVID-19 data by using different techniques and tools. We particularly use two ways: the first one start by fitting the total number of the confirmed cases and the second use a parametric solver tool. Finally a comparative forecasting, machine learning tools, is given.

Keywords— fitting, COVID-19, SIR model scaling, SIR model fitting, predict, machine learning

I. INTRODUCTION

The Coronavirus disease 2019 (COVID-19) epidemic has to get the attention of all scientists around the world. The COVID-19 appears in November 2019 in Wuhan, central China. In March 2020, the epidemic was reclassified as a pandemic by the World Health Organization (WHO). The COVID-19 pandemic is spreading rapidly in many other countries. Measures, limiting human-to-human contact (social distancing, barrier measures, confinement), are taken to stem the spread, causing a sudden slowdown in the world economy and a stock market crash on March 12, 2020. The COVID-19 is contagious with human-to-human transmission. The incubation period is generally between two and fourteen days, with an average of five days.

Many researchers have used new or classical models of infectious diseases to describe, study or predict the evolution of COVID-19 pandemic.

In this work, we aim to use the well known classical SIR model [Anderson and May, 1991], [Hethcote, 2000], to fit it to world data of COVID-19. We discuss and apply different methods and show numerical results. The methods consist in first to use the work in [Liu, Magal, Seydi and Webb, 2020] which start by fitting the total number of the confirmed cases and in second to use parametric solver. SIR stands for “Susceptible-Infected-Removed”. If we consider that all removed individuals are the recovered individuals, then we can say “Susceptible-Infected-Recovered”. In this paper, we first introduce the classical mathematical SIR models we use to fit COVID-19 data. In the second we present different mechanisms we use to fit the data. Then we show numerical results. And finally, we show some prediction for Senegal of the SIR models we study compared with forecasting using machine learning.
II. CLASSICAL SIR MODEL

The classical SIR model is given as follow:

\[
\begin{align*}
\frac{dS}{dt} &= -\frac{\beta(t)S(t)I(t)}{N} \\
\frac{dI}{dt} &= \frac{\beta(t)S(t)I(t)}{N} - \gamma I(t) \\
\frac{dR}{dt} &= \gamma I(t)
\end{align*}
\]

(1)

with the dependent variables \(S(t), I(t), R(t)\) being respectively the number of susceptible individuals at time \(t\), the number of infectious individuals at time \(t\) and the number of recovered individuals at time \(t\). In addition, the unknowns \(S(t), I(t), R(t)\) satisfy \(N = S(t) + I(t) + R(t)\). \(N\) is the total number of individuals.

The initial conditions are \(S(0) = S_0 \geq 0, I(0) = I_0 \geq 0\) and \(R(0) = R_0 \geq 0\).

Some parameters must be presented: \(\beta\) is the contact rate, \(1/\gamma\) is the average infectious period, \(R_0 = \beta/\gamma\) is the basic reproduction number.

The SIR model satisfies some properties:

- The susceptible function \(S(t)\) decreases to \(S_\infty = S_0 \exp(-\frac{R_0}{N} (R_\infty - R_0))\) as \(t \to \infty\).
- The recovered function \(R(t)\) increases up to \(R_\infty = N - S_0 \exp(-\frac{R_0}{N} (R_\infty - R_0))\) as \(t \to \infty\).
- If \(R_0 > N/S_0\), then \(I(t)\) increases up to a maximum value \(I_{\max} = I_0 + S_0 - \left(1 + \ln\left(\frac{R_0S_0}{N}\right)\right) \frac{N}{R_0}\).
- If \(R_0 < N/S_0\), then \(I(t)\) decreases to 0.

Remark II.1. At the outset of an epidemic, nearly everyone (except the infected case) is susceptible. So we can say that \(S_0 \approx N\) and then we can replace \(N/S_0\) by 1.

It is also possible to couple the SIR model with other differential equations like death equation to study the number of death due the epidemic disease or economic model equations to study the impact of the epidemic disease on the economy. To couple the SIR model with a death equation, we consider the death rate \(\mu\) due to the infection. So the system is given by:

\[
\begin{align*}
\frac{dS}{dt} &= -\frac{\beta(t)S(t)I(t)}{N} \\
\frac{dI}{dt} &= \frac{\beta(t)S(t)I(t)}{N} - \gamma I(t) - \mu I(t) \\
\frac{dR}{dt} &= \gamma I(t) \\
\frac{dD}{dt} &= \mu I(t)
\end{align*}
\]

(2)

with \(D\) being the number of death due to the infection. Then the total removed is \(R(t) + D(t)\) and \(N = S(t) + I(t) + R(t) + D(t)\).
III. Methods

i. Fitting function to data

The starting point of the work is to find a function that fits the data of total confirmed cases. As in [Liu, Magal, Seydi and Webb, 2020], we set the total number of infectious cases by:

\[ TNI(t) = \int_{t_0}^{t} I(s) ds. \]  

(3)

The total number of infectious cases function, that fit the data, can be written as \( TNI(t) = b \exp(ct) - a \). Where \( a, b, c \) are parameters to estimate by using the least square method. Then with \( TNI(t) \) we can obtain information on the unknown functional variables and parameters of the SIR model.

After calculation as in [Liu, Magal, Seydi and Webb, 2020], we obtain:

\begin{itemize}
  \item \( t_0 = \ln(a) - \ln(b) \).
  \item \( I(t) = I(t_0) \exp(c(t - t_0)) \), with \( I(t_0) = I_0 = ac \).
  \item \( R(t) = \gamma(I(t) - I_0) \).
  \item \( \beta = \frac{Nc + \gamma c}{S_0} \).
\end{itemize}

The simulation results are given in subsection i, figure 1a, 1b, 1c and 1d. Also the values of the parameters \( \beta \) and \( \gamma \) are given.

We consider now that at a time \( T \) some measures are taken like social distancing, half or full confinement. It can be interpreted as the contact rate is reduced by some factor or at a time point, the contact rate is close to 0. Indeed in [Liu, Magal, Seydi and Webb, 2020], the authors considered that the transmission of COVID-19 from infectious to susceptible individuals stopped after strong measures has been taken in China. Then they have fixed the contact rate to 0. In [Lauro, Kiss and Miller, 2020], the authors considered that the contact rate is reduced by some factor when a social distancing intervention is introduced with a certain duration. Then they have replaced \( \beta \) by \((1 - c)\beta\). It is also possible to consider that \( \beta \) will decrease depending on time like in [Liu, Magal, Seydi and Webb, 2020], where authors has considered an exponential decreasing function of time.

In this work, we consider that the contact rate decreases progressively to 0. Then we choose continuous function with a slow decrease to describe the contact rate starting at the date-time of the measures.

We propose two types of function for \( \beta \) noted \( \tilde{\beta}(t) \).

The first one is:

\[
\tilde{\beta}(t) = \begin{cases} 
\beta & \text{if } t \in [t_0, T] \\
\beta \left( \frac{T}{t} \right)^{\delta/p} & \text{if } t > T,
\end{cases}
\]  

(4)

where \( \delta \) and \( p \) are parameters to choose. The second one is:

\[
\tilde{\beta}(t) = \begin{cases} 
\beta & \text{if } t \in [t_0, T] \\
\beta - \frac{(\ln(T/t))^\kappa}{q} & \text{if } t > T,
\end{cases}
\]  

(5)

where \( \kappa \) and \( q \) are parameters to choose.
Then replacing $\beta$ by $\tilde{\beta}(t)$ in the SIR model, we obtain:

$$
\begin{align*}
\frac{dS}{dt} &= -\frac{\tilde{\beta}(t)S(t)I(t)}{N} \\
\frac{dI}{dt} &= \frac{\tilde{\beta}(t)S(t)I(t)}{N} - \gamma I(t) \\
\frac{dR}{dt} &= \gamma I(t)
\end{align*}
$$

(6)

By solving (6), we obtain the new expressions of $TN I(t)$, $I(t)$ and $R(t)$. We show the results in subsection III.

Remark III.1.
1. The values of the parameters $\delta$, $p$, $\kappa$ and $q$ can be fixed such that $\tilde{\beta}(t)$ decreases slowly.
2. Since $\tilde{\beta}$ depends on time, we must always consider values of $t$ such that $\tilde{\beta}$ remains positive.
3. $\tilde{\beta}$ given by (2) is positive for all time while $\tilde{\beta}$ given by (3) is positive only if $t \in [0, T \exp((\beta q)^{1/\kappa})]$.

ii. Fit to data with scaling

When there is not enough data the fit is generally difficult since the maximal values of the dependent variables can be very huge. For example, $S_0$ depends on the total population and generally $S_0 \approx N$, with $N$ the size of a chosen sample of the population. In SIR models, it is generally possible to choose different population sizes with similar characteristics. If we have characteristic measures we can make the model dimensionless so that the quality of the results is always good. Then we scale the SIR model to the data by using scaling [Langtangen and Pedersen, 2016].

Let us fix the following constant characteristics size: $t_c$, $S_c$, $I_c$, $R_c$ respectively of the time $t$, the susceptible $S(t)$, the infected $I(t)$ and the recovered $R(t)$. Then the dimensionless variables $\bar{t}$, $\bar{S}$, $\bar{I}$, $R$ are given by $t = t_c$, $S = \frac{S}{S_c}$, $I = \frac{I}{I_c}$, $R = \frac{R}{R_c}$.

Replacing in (1) and calculating, we obtain the scaled system:

$$
\begin{align*}
\frac{d\bar{S}}{dt} &= -\frac{\beta t_c}{N} \bar{S}(t)\bar{I}(t) \\
\frac{d\bar{I}}{dt} &= \frac{\beta S_c t_c}{N} \bar{S}(t)\bar{I}(t) - \gamma t_c \bar{I}(t) \\
\frac{d\bar{R}}{dt} &= \gamma \frac{I_c t_c}{R_c} \bar{I}(t)
\end{align*}
$$

(7)

Remark III.2. For the adjusting of the data, we choose $t_c = t_d^{\text{max}}$, with $t_d^{\text{max}}$ being the last time of the data.
The scaled system can be written, by dropping the “-”, as follows:

\[
\begin{align*}
\frac{dS}{dt} &= -\tilde{\beta}_1 S(t) I(t) \\
\frac{dI}{dt} &= \tilde{\beta}_2 S(t) I(t) - \tilde{\gamma}_1 I(t) \\
\frac{dR}{dt} &= \tilde{\gamma}_2 I(t)
\end{align*}
\]

with \( \tilde{\beta}_1 = \frac{\beta I_c t_c}{N}, \tilde{\beta}_2 = \frac{\beta S_c t_c}{N}, \tilde{\gamma}_1 = \gamma t_c \) and \( \tilde{\gamma}_2 = \gamma \frac{I_c t_c}{R_c} \).

**Remark III.3.** By choosing the characteristics such that \( S_c = I_c = R_c \), we obtain \( \tilde{\beta}_1 = \tilde{\beta}_2 \) and \( \tilde{\gamma}_1 = \tilde{\gamma}_2 \). And then we get a new SIR model.

**iii. Parametric solve**

We estimate the parameters \( \beta \) and \( \gamma \) by using parametric solver in Mathematica. With the parametric solver, we solve the SIR model (2) with solutions depending on the parameters we want to estimate such that the model fits the data. Then plotting the parametric solution for different values of \( \beta \) and \( \gamma \), finally, give a fit. Parametric solver typically solve differential equations by going through several different stages, depending on the type of equations.

**iv. Machine learning**

Machine learning is programming computers to optimize a performance criterion using example data or past experience. We have a model defined up to some parameters, and learning is the execution of a computer program to optimize the parameters of the model using the training data or past experience. The model may be predictive to make predictions in the future, or descriptive to gain knowledge from data or both.

Machine learning uses the theory of statistics in building mathematical models because the core task is making inferences from a sample. (See [Alpaydín, 2010].)

Then machine learning can be used for the estimation of parameters of an epidemic model. We can go deep by doing data mining since we have massive world data. We can use machine learning to learn about the worldwide data of pandemic COVID-19 and then to predict the future evolution of the disease.

We can also try to learn on the spatial distribution and progression of the disease and predict the location susceptible to be a high level of risk.

In this work, we just use a machine learning tool to learn on small data of Senegal’s COVID-19 cases and then to predict the evolution in future days. We use it as a comparison with the other forecasts based on the work we do in this paper.

**IV. Numerical simulations**

**i. The data**

In this subsection we show table of data we use for simulation. The tables [1] [2] and [8] give data respectively of Senegal, China and France. Data for Senegal is obtained from daily press releases on
the COVID-19 from the Ministry of Health and Social Action (http://www.sante.gouv.sn/) and the data for China and France come from Wolfram Research. Wolfram Research provide estimated confirmed COVID-19 infection trend by country or region, based on WHO, U.S. CDC, ECDC, China CDC (CCDC), NHC and DXY. The source dataset is compiled daily by Johns Hopkins CSSE. For details see: https://github.com/CSSEGISandData/COVID-19/issues.

Table 1: COVID-19 data for Senegal: 2020 March 02-31.

<table>
<thead>
<tr>
<th>Date</th>
<th>2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31</th>
</tr>
</thead>
<tbody>
<tr>
<td>New cases</td>
<td>1 1 1 1 0 0 0 0 0 1 5 11 3 2 1 4 5 2 9 9 11 12 7 13 6 14 11 20 13</td>
</tr>
<tr>
<td>Total cases</td>
<td>1 2 3 4 4 4 4 4 5 10 21 24 26 31 36 38 47 56 67 79 86 99 119 130 142 162 175</td>
</tr>
<tr>
<td>Recovered cases</td>
<td>0 0 0 0 1 1 1 1 1 2 2 2 2 2 2 2 2 2 8 8 9 9 11 18 27 28 40</td>
</tr>
<tr>
<td>Death cases</td>
<td>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1</td>
</tr>
</tbody>
</table>

Table 2: COVID-19 data for Hubei, China: 2020 January 22- March 31.

<table>
<thead>
<tr>
<th>Date</th>
<th>1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infected cases</td>
<td>399 399 494 689 964 1302 3349 3341 4651 5461 6736 10532 12722 15677 18483 20677 23139 24881</td>
</tr>
<tr>
<td>Recovered cases</td>
<td>28 28 31 32 42 45 80 88 90 141 168 295 386 522 633 817 1115 1439</td>
</tr>
<tr>
<td>Death cases</td>
<td>17 17 24 40 80 88 89 104 114 162 204 213 276 329 392 479 618 780</td>
</tr>
</tbody>
</table>

ii. The function fit

Here we show simulations related to the subsection of section III. The formula of \( TNI(t) \) is given by \( TNI(t) = b \exp(ct) - a \), with \( a = 13.9324 \), \( b = 9.61779 \) and \( c = 0.100095 \) (figure 1a). We use \( \gamma = \frac{1}{7} \) and then we obtain: \( t_0 = 3.7051 \), \( I_0 = 1,39456 \). The total population of Senegal is \( N = 16743927 \) from Senegal Population (2020) - Worldometer (www.worldometers.info) and then we obtain \( S_0 = N - I_0 \). Since \( S_0 \approx N \) we simplify the calculation and obtain \( \beta = 0, 242957 \).

Now we consider the time of the measures as 2020, March 23. Then \( T = 23 \). For \( \tilde{\beta} \) given by (4), results are shown by figures 2a, 2c and 2e. For \( \beta \) given by (5), results are shown by figures 2b, 2d.

<table>
<thead>
<tr>
<th>Date</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infected cases</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Recovered cases</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Death cases</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Date</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>26</td>
<td>27</td>
<td>28</td>
<td>29</td>
<td>30</td>
<td>31</td>
<td>32</td>
<td>33</td>
<td>34</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>Infected cases</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Recovered cases</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Death cases</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Date</td>
<td>36</td>
<td>37</td>
<td>38</td>
<td>39</td>
<td>40</td>
<td>41</td>
<td>42</td>
<td>43</td>
<td>44</td>
<td>45</td>
<td>46</td>
<td>47</td>
<td>48</td>
<td>49</td>
<td>50</td>
<td>51</td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>Infected cases</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Recovered cases</td>
<td>11</td>
<td>11</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Death cases</td>
<td>12</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Date</td>
<td>53</td>
<td>54</td>
<td>55</td>
<td>56</td>
<td>57</td>
<td>58</td>
<td>59</td>
<td>60</td>
<td>61</td>
<td>62</td>
<td>63</td>
<td>64</td>
<td>65</td>
<td>66</td>
<td>67</td>
<td>68</td>
<td>69</td>
<td>70</td>
</tr>
<tr>
<td>Infected cases</td>
<td>1366</td>
<td>4396</td>
<td>6413</td>
<td>7492</td>
<td>8883</td>
<td>10616</td>
<td>12150</td>
<td>13708</td>
<td>13144</td>
<td>16796</td>
<td>17923</td>
<td>20002</td>
<td>22511</td>
<td>25269</td>
<td>29561</td>
<td>33599</td>
<td>39161</td>
<td></td>
</tr>
<tr>
<td>Recovered cases</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>Death cases</td>
<td>91</td>
<td>91</td>
<td>148</td>
<td>148</td>
<td>148</td>
<td>243</td>
<td>450</td>
<td>562</td>
<td>674</td>
<td>800</td>
<td>1100</td>
<td>1331</td>
<td>1696</td>
<td>1905</td>
<td>2314</td>
<td>2606</td>
<td>3024</td>
<td>3573</td>
</tr>
</tbody>
</table>

and 24

We do the same as in the paragraphs above for France. $TNI(t) = b \exp(ct) - a$, with $a = 1$, $b = 0.88$ and $c = 0.158$ (figure 3a). We use $\gamma = 1/7$ and then we obtain: $t_0 = 0.809$, $I_0 = 0.158$. The total population of France is $N = 65241903$ from France Population 2020 (https://worldpopulationreview.com) and then we obtain $S_0 = N - I_0$. Since $S_0 \approx N$ we simplify the calculation and obtain $\beta = 0.300857$.

For France, we consider the time of the measures as 2020, March 17. Then $T = 56$. For $\tilde{\beta}$ given by (4), results are shown by figures 4a and 4b. For $\hat{\beta}$ given by (5), results are shown by figures 5a and 5b.
Figure 1: Plot of the total number of cases using the SIR model and compared with plot of Senegal’s data. On the abscissa axis, the graduation 30 represents 2020, March 31.
Figure 2: On the top: plot of the contact rate $\tilde{\beta}(t)$ with $\delta = 15$ and $p = 10$ and $\tilde{\beta}(t)$ with $\kappa = 1/3$ and $q = 10$. In the left for figures related to $\tilde{\beta}(t)$ and in the right for figures related to $\tilde{\beta}(t)$, we plot the two functions $TNI(t)$ compared with plot of Senegal’s data in table 1 completed until April 06. On the abscissa axis, the graduation 30 represents 2020, March 31. Then 2020, April 06 is the graduation 36.
Figure 3: Plot of the total number of case data using the SIR model\(^1\) and compared with plot of France's data. On the abscissa axis, the graduation 70 represents 2020, March 31.

Figure 4: Plot of the contact rate \(\tilde{\beta}(t)\)\(^4\) with \(\delta = 1\) and \(p = 1/2\). Plot of the two functions \(TNI(t)\) compared with plot of France's data in table 3 completed until April 06. On the abscissa axis, the graduation 70 represents 2020, March 31.

Figure 5: Plot of the contact rate \(\tilde{\beta}(t)\)\(^5\) with \(\kappa = 1/4\) and \(q = 13\). Plot the two functions \(TNI(t)\) compared with plot of France's data in table 4 completed until April 06. On the abscissa axis, the graduation 70 represents 2020, March 31. Then 2020, April 06 is the graduation 76.
iii. The fit by scaling

The scaled SIR model (8) is plotted in figure 6a. And in figure 6b we fit with the data of Senegal in table 1. We fix the characteristic parameters as: $t_c = 30$ which is the number of days of the data of Senegal in table 1, $I_c = S_c = R_c = 1000$. We use $\beta = 5.0753$, $\gamma = 1/45$ to obtain $\beta_1 = \beta_2 = 0.00939349$ and $\gamma_1 = \gamma_2 = 0.535714$.

![Scaled SIR model](image)

(a) Plot of the scaled SIR model (8) with parameters $t_c = 30$, $I_c = S_c = R_c = 1000$, The susceptible $S(t)$ in blue line, The population in red line, the infected $I(t)$ in orange and the recovered in green line.

![Data fit](image)

(b) Fit of the infected $I(t)$ in orange to the effective case data in black dotted. And fit of the recovered $R(t)$ in green to the recovered case data in blue dotted. The plotted data is for Senegal. The data is padded 1.8 to the right.

**Figure 6:** Plot of the scaled SIR model (8) compared with plot of Senegal’s data in table 1. The date 2020 March 31 corresponds to the graduation 2.8.

iv. The fit by parametric resolution

Here we use a parametric solver in Wolfram Mathematica to solve the SIR model (2) with respect to parameters $\beta$ and $\gamma$. We start by using the values $\beta = 5.0753$ and $\gamma = 1/45$. Then after we get the new values for the fit. The results are shown in figures 7a, 7b and 7c for Senegal, in figures 8a, 8b and 8c for China and in figures 9a, 9b and 9c for France.
Figure 7: Plot of the SIR model with the parameters $\beta = 5.0753$ and $\gamma = 1/45$ and the proportion of death $\mu = 0.01\gamma$. The susceptible $S(t)$ in red line, the population in blue line, the infected $I(t)$ in green line, the recovered in orange line and the death $D(t)$ in purple.

(a) Plot of the SIR model with the parameters $\beta = 5.0753$ and $\gamma = 1/45$ and the proportion of death $\mu = 0.01\gamma$. The susceptible $S(t)$ in red line, the population in blue line, the infected $I(t)$ in green line, the recovered in orange line and the death $D(t)$ in purple.

(b) Plot of the SIR model with the parameters $\beta = 5.0753$, $\gamma = 1/45$, the proportion of death $\mu = 0.01\gamma$ and reduced population. And plot without fit to the infected data in red dotted, the recovered data in blue dotted, the death data in black dotted.

(c) Plot of the SIR model with the parameters $\beta = 0.15$, $\gamma = 1/45$ and the proportion of death $\mu = 0.01\gamma$. And plot with fit to data.

(d) Plot of the SIR model with the parameters $\beta = 0.15$, $\gamma = 1/45$ and the proportion of death $\mu = 0.01\gamma$. And plot of the fit without data.
Figure 8: Plot of the SIR model\(^2\) compared with plot of China’s data in Table 2. The total population of Hubei we use is \(N=59172000\) (https://en.wikipedia.org/wiki/2020_Hubei_lockdowns). On the abscissa axis, the graduation 63 represents 2020, March 08.
v. Discussion

The SIR model is fitted to the data. This allows us to note that the maximum number of infected can go up to 900000 (figure 7d) in Senegal and 2700000 (figure 9d) in France. While in the scaling the figures 6a shows that the maximum number of infected can go up to 320000. However, this analysis does not take into account the nationwide anti-pandemic measures. Using the results shown in the figures 2e, 2f, 4b, and 5b we can see that the data deviate from their first exponential evolution to follow a new, slower trajectory which can still be exponential. To understand the nature of this new trajectory, the way to manage the contact rate is decisive and can lead to different analyzes. However, having considered a slow decrease over time of the contact rate, we can see in the figures 2e, 2f, 4b, and 5b that the updated data pass under the new trajectory for Senegal while in France the data follow the new trajectory. This could, therefore, be due to the effects of the measures taken by these countries.
V. Prediction

In this section, we plot some predictions for Senegal. The prediction does not concern the total confirmed cases, but only the effective cases obtained by reduced from the total confirmed cases, the recovered cases and the death cases.

We show the curve of effective infected cases given by $TNI(t) - R(t) - D(t)$ with both $\hat{\beta}$ (4) and (5), the curve of the fitted infected $I(t)$ in figure 7c, and the curve obtained by using machine learning. We plot these curves with additional dates until 2020, April 21.

We use machine learning, based only on data, to do forecasting. “Predict” is a function of “Automated Machine Learning” in Wolfram Mathematica. It allows for automatic training and data prediction. We can choose different method of regression algorithm: “RandomForest”, “LinearRegression”, “NeuralNetwork”, “GaussianProcess”, “NearestNeighbors”, etc. We use the “NeuralNetwork” regression algorithm which predicts using an artificial neural network.

Let’s recall that our forecasting use in a first part the results of the works in subsection ii and iii where we use a SIR model to fit data. And in a second part, the forecasting using only data. The forecasting show in figures 10a, 10c an optimistic situation for Senegal. Also the machine learning figure 10e shows an optimistic forecasting. But in all cases, we see that the additional data (March 31-06) goes down the predicting path. It may be caused by the nationwide anti-pandemic measures in Senegal. In addition, we have considered in figures 10a and 10b a contact rate gradually reduced since the measures were taken and we see that the data come below the path of the predicted curve.
SIR model fit to COVID-19 data • 2020, April • Vol., No.

(a) Plot of the number of infected cases using the scaling SIR model (6) with $\tilde{\beta}$ given by (4). On the abscissa axis, the graduation 30 represents 2020, March 31. Then 2020, April 06 is the graduation 36.

(b) Plot of the number of infected cases using the scaling SIR model (6) with $\tilde{\beta}$ given by (5). On the abscissa axis, the graduation 30 represents 2020, March 31. Then 2020, April 06 is the graduation 36.

(c) Plot of the number of infected cases using the scaling SIR model (8). On the abscissa axis, the graduation 2.8 represents 2020, March 31. Then 2020, April 06 is the graduation 3.0.

(d) Plot of the number of infected cases using SIR model (2) with parametric solve. On the abscissa axis, the graduation 40 represents 2020, March 31. Then 2020, April 06 is the graduation 46.

(e) Plot of the number of infected cases by training a Predictor Function to predict the average next values of infected. On the abscissa axis, the graduation 30 represents 2020, March 31. Then 2020, April 06 is the graduation 36.

Figure 10: Forecasting for Senegal, using the results of the works in subsections and machine learning. The data plotted (red dotted, green dotted), is given in table 1, but completed until 2020, April 06.
Figure 11: Forecasting for France, using the results of the works in subsection ii and iii and machine learning. The data plotted (red dotted, green dotted), is given in table 5 but completed until 2020, April 06.
VI. CONCLUSION AND PERSPECTIVES

In this paper, we have used the classical SIR model for fitting data and then we have done forecasting. We have also estimated the contact rate, and the average infectious period parameters of the SIR model to obtain fit. Machine learning can help in epidemiology to understand the disease, but also to study the impact or effectiveness of the anti-pandemic measures taken. We can also study the ideal date for stopping the containment measures. Ideal in the sense that even without the measures the disease can no longer spread. We can also study the possibility of making periodic confinements to reduce the economic impact of the measures. The economy may be impacted by the evolution of the disease and measures. It would, therefore, be interesting to couple economic models with epidemiological models. Since the impact can also be long-term, scaling can be used to dimension the epidemiological-economic coupling. The impact of the environment is also to be taken into account in the models. This means studying the contamination due to the environment (air, objects, etc).

VII. APPENDIX

i. Using function fit in SIR model

From $TNI(t) = b \exp(ct) - a$, we have $TNI(t_0) = b \exp(ct_0) - a$ and using (9), we obtain

$$b \exp(ct_0) - a = 0.$$ (9)

We deduce $a = b \exp(ct_0)$, then $t_0 = \frac{\ln(a) - \ln(b)}{c}$.

Again from (3) we have

$$I(t) = TNI(t) = bc \exp(ct).$$ (10)

Then $I(t_0) = bc \exp(ct_0) = ac = I_0$ and $\frac{I(t)}{I(t_0)} = \exp(c(t - t_0))$. Hence, we obtain

$$I(t) = I(t_0) \exp(c(t - t_0)),$$ (11)

then $\dot{I}(t_0) = cI(t_0)$.

From the second equation in the SIR model (1), we obtain at $t_0$, $\dot{I}(t_0) = \frac{\beta I_0 S_0}{N} - \gamma I_0 = cI(t_0)$. Hence, we get

$$c = \frac{\beta S_0}{N} - \gamma.$$ (12)

Using the third equation in the SIR model (1) yields

$$\dot{R}(t) = \gamma I(t_0) \exp(c(t - t_0)),$$ (13)

then integrating, we obtain

$$R(t) = \frac{\gamma}{c}(I(t) - I(t_0)).$$ (14)
ii. Determination of the scaled SIR

$S(t) = S(t_c, \tilde{t})$, $I(t) = I(t_c, \tilde{t})$ and $R(t) = R(t_c, \tilde{t})$, then replacing in left members of the SIR model (1), we obtain:

$$\dot{S} = \frac{S_c}{t_c} \tilde{S}$$
$$\dot{I} = \frac{I_c}{t_c} \tilde{I}$$
$$\dot{R} = \frac{R_c}{t_c} \tilde{R}$$

Replacing again in right members of the SIR model (1), we get:

$$\dot{S} = -\frac{\beta S_c I_c}{N} \tilde{S}$$
$$\dot{I} = \frac{\beta S_c I_c}{N} \tilde{S} - \gamma I_c \tilde{I}$$
$$\dot{R} = \gamma I_c \tilde{I}$$

Then we have:

$$\frac{S_c}{t_c} \ddot{S} = -\frac{\beta S_c I_c}{N} \tilde{S}$$
$$\frac{I_c}{t_c} \ddot{I} = \frac{\beta S_c I_c}{N} \tilde{I}S - \gamma I_c \tilde{I}$$
$$\frac{R_c}{t_c} \ddot{R} = \gamma I_c \tilde{I}$$

Finally, we obtain:

$$\begin{cases} 
\dot{S} = -\frac{\beta I_c t_c}{N} \tilde{S} \\
\dot{I} = \frac{\beta S_c t_c}{N} \tilde{I}S - \gamma t_c \tilde{I} \\
\dot{R} = \gamma \frac{I_c t_c}{R_c} \tilde{I}
\end{cases} 
$$

(15)

REFERENCES


SIR model fit to COVID-19 data


