PREDICTING THE EVOLUTION OF COVID-19 IN PORTUGAL USING AN ADAPTED SIR MODEL PREVIOUSLY USED IN SOUTH KOREA FOR THE MERS OUTBREAK

Pedro Teles
Department of Physics and Astronomy
Faculty of Sciences of the University of Porto, Portugal
Rua do Campo Alegre s/n, 4169-007 Porto

Abstract

Since original reports in Wuhan, China, the new coronavirus covid-19 has spread very quickly worldwide, leading the World Health Organization (WHO) to declare a state of pandemic. Moreover, as of the 13th of March 2020, the WHO has announced that the European continent is now the main centre of the pandemic. Many European governments have already implemented harsh measures to attempt to contain the spread of the virus. In Portugal, there are, as of the 18th of March 2020, 642 confirmed cases. In this study, I applied an adapted SIR model previously used in South Korea to model the MERS outbreak, which is also caused by a coronavirus, to estimate the evolution of the curve of active cases in the case of the Portuguese situation. As some of the parameters were unknown, and the data for Portugal is still scarce, given that the outbreak started later (first case on the 2nd of March) I used Italian data (first reported case in Italy on the 31st of January) to predict them. I then construct two or three different scenarios for the evolution of covid-19 in Portugal, considering both the effectiveness of the mitigation measurements implemented by the government, and the self-protective measures taken by the population, as explained in the South Korean model. In the out of control scenario, the number of active cases could reach as much as ~40,000 people by the beginning of April. If measures are appropriately taken this number can be reduced to about 700 cases. In case that some measures are taken, but their implementation isn't done appropriately, the active cases could reach circa 7,000 people. The actual figures probably lie between the interval (~700-7,000) and the peak will be reached between 9th and the 20nd of April 2020.

1. INTRODUCTION.

There is already abundant information on the new coronavirus and its spread worldwide [1-3], as well as the risks it poses to the population. The person infected with the severe acute respiratory syndrome coronavirus 2 (SARS-cov-2)[4] may develop the disease called COVID-19, which manifests with mild to severe symptoms of fever, cough, shortness of breath and other signs [5]. In the most severe cases, the infection can lead to the development of acute respiratory distress syndrome (ARDS) causing respiratory failure, septic shock, or multi-organ failure, and even death [6].

There is yet a lot about this new coronavirus that is still unknown. Studies suggest that the case fatality rate of the virus is of about 3.5% in mainland China, but the authors suggest considering the range 0.2-3.0% [7]. However, this value seems to be much higher in Italy [8], suggesting its strong dependence on demographics. In fact, the Chinese Center for Disease Control released an
analysis that showed that the case fatality rate ranged from 0% in children aged 0-9 years old to 14.8% in those over 80 [9], which implies that countries with a high number of elderly citizens, as is the case of Italy and Portugal, are more prone to having higher death rates.

The quick spread of the virus in Europe has led the WHO to declare Europe as the new epicentre of the disease on the 13th of March of 2020[10]. The rapid growth of the number of active cases presenting severe symptoms has saturated the health services in most countries in the continent, especially in Italy [11]. It is yet unknown when the epidemic will reach an end, but governments throughout Europe have implemented harsh measures to prevent and mitigate the spread of the virus. Yet, as of today there are as many as 77,394 confirmed cases on the European continent, in 46 territories, resulting in 3,457 deaths, and rising[2-3]. In Portugal, the two first confirmed cases in the Portuguese territory were announced on the 2nd of March, which puts the country slightly behind the curve in the outbreak in the European continent. As of today, there are 642 confirmed cases, 2 deaths, and 3 recovered [12]. Also, the country is on lockdown since the 13th of March 2020, and the Portuguese Parliament declared a State of Emergency today (18th of March 2020).

One of the questions many policy makers, and governments are asking themselves is how the spread is going to evolve in time. A timely idea of the amount of cases that will exist in a near future can allow governments and policy makers to act accordingly.

In this study, I adapted a model of the SIR type used in South-Korea to model the evolution of the active cases of the MERS outbreak in the country back in 2015 [13]. Although there are other models being used [14-15], this model seemed most appropriate given that the MERS and the SARS-cov-2 are related illnesses, both provoked by a coronavirus strand, and with similar symptoms, although MERS was much deadlier. In order to fit the parameters and given that there is still very little data in the case of Portugal, I used the data of Italy instead. This allowed me to fit the curve of current active cases in Portugal with a model, which I then use, by implementing the control measure parameters predicted in the model to predict the future number of cases, in three different scenarios (the out-of-control scenario, a scenario in which the government is successful in applying control measures and citizens mildly adhere to self-protection measures, and a scenario in which the government fails to completely apply control measures, yet citizens still adhere to some measure of self-protection).

2. METHODOLOGY.

The model used is the model described in Xia et al[13], which can be understood by the flow diagram shown in figure 1 (which was taken from the paper, page 4). In this model, S corresponds to the number of susceptible individuals; E, the number of exposed individuals; A, the number of asymptomatic infected cases; I, the total number of mild-to-severe infected patients. H, the number of hospitalized cases; R, the number of removed cases, and finally N, the total population of Portugal.

\( \beta_1 \) is the transmission coefficient of the asymptomatic infected cases, \( \beta_2 \) is the transmission coefficient of the symptomatic infected cases (mild infected person and severe patients) to the susceptible, \( \beta_3 \) is the transmission coefficient of the
hospitalized cases to the susceptible, $1/\sigma$ is the mean incubation period, $1/\lambda$ is the mean time between symptom onset to hospitalization, $1/k_1$ is the mean infectious period of asymptomatic infected person for survivors, $1/k_2$ is the mean duration for hospitalized cases for survivors, $1/\delta$ is the mean time from hospitalization to death, $\gamma$ is the clinical outbreak rate in all the infected cases. The time unit is 1 day.

Figure 1. Flow chart of the SIR-type model used in this work, taken from [13]

The set of differential equations can then be written as:

\[
\begin{align*}
\frac{dS(t)}{dt} &= -\beta_1 \frac{S(t)A(t)}{N} - \beta_2 \frac{S(t)I(t)}{N} - \beta_3 \frac{S(t)H(t)}{N}, \\
\frac{dE(t)}{dt} &= \beta_1 \frac{S(t)A(t)}{N} + \beta_2 \frac{S(t)I(t)}{N} + \beta_3 \frac{S(t)H(t)}{N} - \sigma E(t), \\
\frac{dA(t)}{dt} &= (1 - \gamma)\sigma E(t) - k_1 A(t), \\
\frac{dI(t)}{dt} &= \gamma \sigma E(t) - \lambda I(t), \\
\frac{dH(t)}{dt} &= \lambda I(t) - k_2 H(t) - \delta H(t), \\
\frac{dR(t)}{dt} &= k_1 A(t) + k_2 H(t) + \delta H(t).
\end{align*}
\]

Given that the available data for covid-19 is still very preliminary, the model was further adapted. The only parameters that were considered “known” were the mean incubation time, which was taken as $1/\sigma = 5.1$ days from the literature (here using the median as equal to the mean) [16], and the mean infectious period of asymptomatic people, where a conservative estimate of $1/k_1 = 14$ days was used. All other parameters were taken to be “unknown”. Furthermore, and because the fitting of the model was difficult to achieve, I took $\beta_1 = \beta_2 = \beta_3$ as a first approximation, meaning that the mean transmission coefficient is the same independently of type and setting of the transmission. Finally, unlike Xia et al[13], I consider that the epidemic starts with only one person infected ($I_0 = 1$), being the rest of the values 0, except for $S_0$ which is simply $N-1$. See table 1 for a breakdown of the different parameters of this model.

The solving of the system of differential equations was performed using the Mathematica code [17], which also allows for the use of fittable parameters, using the function “NonLinearModelFit”[18]. Given that there is still very little data for Portugal, the data for Italy was used as proxy. This seems like the most reasonable approach (instead of using data from China or another Asian country), given the cultural and health system proximity between the two
countries. In order to determine the parameters $\lambda$, $\delta$, I used the number of deaths, as reported by the Italian ministry of Health and available online [19]. For the estimation of the value of $k_2$, the number of recovered was used.

Table 1. The different parameters of the model described by Eq. 1

<table>
<thead>
<tr>
<th>parameter</th>
<th>Xia et al value (days)</th>
<th>this work (days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1/\sigma$</td>
<td>5.2 (3~8)</td>
<td>5.1</td>
</tr>
<tr>
<td>$1/\lambda$</td>
<td>5 (3~7)</td>
<td>To be fitted</td>
</tr>
<tr>
<td>$1/\delta$</td>
<td>15.16(0~42)</td>
<td>To be fitted</td>
</tr>
<tr>
<td>$1/k_1$</td>
<td>5</td>
<td>14</td>
</tr>
<tr>
<td>$1/k_2$</td>
<td>7</td>
<td>To be fitted</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.8756(0.853~0.9324)</td>
<td>To be fitted</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.7833(0.5925~0.8592)</td>
<td>To be fitted</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>0.4568(0.3839~0.6751)</td>
<td>To be fitted</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.0348(0.0285~0.353)</td>
<td>To be fitted</td>
</tr>
<tr>
<td>$N$</td>
<td>49520000</td>
<td>10290000(Portugal);60480000(Italy)</td>
</tr>
<tr>
<td>$S_0$</td>
<td>49519960</td>
<td>10289999(Portugal);60479999(Italy)</td>
</tr>
<tr>
<td>$E_0$</td>
<td>16~32</td>
<td>0</td>
</tr>
<tr>
<td>$A_0$</td>
<td>16~36</td>
<td>0</td>
</tr>
<tr>
<td>$I_0$</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>$H_0$</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$R_0$</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

After determining the values for these parameters, they were used to fit the curve of infected cases for Portugal, where the values of $\beta_1=\beta_2=\beta_3$ and $\gamma$ were fitted to obtain a model. This model was then used to estimate the future evolution of future cases in Portugal, according to three different scenarios:

- Out-of-control scenario (nothing done, the virus is free to spread);
- Scenario in which the government takes severe mitigating measures and the population mildly adheres to self-protecting measures;
- Scenario in which the government fails to implement mitigating measures, yet still implements some, and the population mildly adheres to self-protecting measures.

3. RESULTS.

3.1. Fitting of parameters $\lambda$ and $\delta$ (Italy).

The number of deaths in Italy was taken from [18], and the set of differential equations described in eq. 1 fitted to obtain the best possible fit for parameters $\lambda$ and $\delta$. An $R^2$ value of 0.99255 was obtained with the fit. The rest of the parameters are presented in table 3. Results are shown in figure 1.
Table 2. Table of fitted parameters $\lambda$ and $\delta$ as obtained using “FindFit” in Mathematica

<table>
<thead>
<tr>
<th>parameter</th>
<th>estimate</th>
<th>standard error</th>
<th>t-statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda$</td>
<td>0.2548</td>
<td>0.0002</td>
<td>879</td>
<td>9.83x10^-70</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.0168</td>
<td>0.0002</td>
<td>73</td>
<td>3.02x10^-36</td>
</tr>
</tbody>
</table>

Figure 1. Graphical representation of the fitted model to the Italian government official death toll.

3.2. Fitting of parameter $k_2$ (Italy).

Now, using the values for $\lambda$ and $\delta$ found in 3.1, parameter $k_2$ was obtained again with a fit of Eq. 1, but now to the number of recovered cases in Italy, which again was taken from [18]. The best possible fit for $k_2$ gave an $R^2$ value of 0.86178 was obtain with the fit. The rest of the parameters are presented in table 3.

Table 3. Table of fitted parameter $k_2$ as obtained using “FindFit” in Mathematica

<table>
<thead>
<tr>
<th>parameter</th>
<th>estimate</th>
<th>standard error</th>
<th>t-statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_2$</td>
<td>0.0875</td>
<td>0.012</td>
<td>7.32</td>
<td>3.09x10^-9</td>
</tr>
</tbody>
</table>
As can be seen, in the case of recovered cases, the fit has not the same quality as the fit in 3.1. This is further evidenced by figure 2, where a graphical representation of the fit is shown.

![Graphical representation of the fitted model to the Italian government official recoveries.](image)

Figure 2. Graphical representation of the fitted model to the Italian government official recoveries.

### 3.3. Fitting of parameters $\beta$ and $\gamma$ (Portugal).

After the two previous procedures, we now have all the parameters we need to fit Eq. 1 to the number of Portuguese active cases, in order to determine the value of $\beta$ and $\gamma$. The number of Portuguese active cases was taken from [20], which is updated daily.

The values previously obtained of parameters $\lambda$, $\delta$, and $k_2$ were used. As explained previously, because the fitting was difficult to obtain, as a first approximation, we considered $\beta_1 = \beta_2 = \beta_3$. The obtained values for $\beta$ and $\gamma$ are given in table 4.
Table 4. Table of fitted parameters $\beta$ and $\gamma$ as obtained using “NonLinearModelFit” in Mathematica

<table>
<thead>
<tr>
<th>parameter</th>
<th>estimate</th>
<th>standard error</th>
<th>t-statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>1.16</td>
<td>0.033</td>
<td>34.997</td>
<td>$1.51 \times 10^{-16}$</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.0158</td>
<td>0.0034</td>
<td>4.634</td>
<td>$2.74 \times 10^{-5}$</td>
</tr>
</tbody>
</table>

A graphical depiction of the fitted model is given in figure 3.

![Graphical representation of the fitted model to the Portuguese government official active cases.](image)

Figure 3. Graphical representation of the fitted model to the Portuguese government official active cases.

In this scenario, the out-of-control scenario, in which the corona virus would be “free” to spread within the country would lead to a pic of ~40,000 infected around the 6th of April 2020. A scenario that will very likely be avoided given the strong containment measures implemented by the government.
3.4. Discussion of the obtained values for $\lambda$, $\delta$, $k_2$, $\beta$, and $\gamma$.

The fitting of these parameters was difficult to obtain. There is enough systematic error and lack of statistics to make the present results very unreliable. In order to obtain the parameters for $\lambda$, $\delta$, and $k_2$, an ersatz value for $\gamma$ had to be given. In this model, $\gamma$ is originally defined as "the clinical outbreak rate in all the infected cases", which is in itself difficult to understand what it means. In the cases of the fits of $\lambda$, $\delta$, and $k_2$, the ersatz value used for $\gamma$ had a very strong impact in the results of the other parameters. In fact, as it comes multiplied with $\sigma$ in the transfer from compartment $E(t)$ to compartments $A(t)$ and $S(t)$, I think the high variation in this value could very well be due to the huge uncertainty there is at the moment for the appropriate incubation time of this virus, which some authors claim can be of up to 27 days [21], although this number is probably an outlier. Finally, a case of an incubation period of 18 days seems to have been confirmed by a recent study [22]. The other parameters are somewhat less elusive. The obtained value for $\gamma=0.0168$ is 50% the value obtained by Xia et al [13] of 0.0348, which shows the difference in case fatality rates between the SARS-covid-2 (7.2% currently for Italy), and MERS (estimated at 34%), but then again the value should be 20% of this value and not 50%, further implying that the case fatality rate of the SARS-covid-2 is still difficult to determine, and could be higher than previously thought.

Finally, from the obtained values it is also possible to determine a value for the basic reproduction number $R_0$, which here I take as being simply the value $R_0=\beta/\sigma=5.916$ which demonstrates how strong the capacity for spreading of this particular virus is. In table 4, the final values of the obtained parameters are given.

<table>
<thead>
<tr>
<th>parameter</th>
<th>this work (days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1/\sigma$</td>
<td>5.1 (fixed)</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>0.2548±0.0002</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.0168±0.0002</td>
</tr>
<tr>
<td>$k_1$</td>
<td>0.0714 (fixed)</td>
</tr>
<tr>
<td>$k_2$</td>
<td>0.0875±0.012</td>
</tr>
<tr>
<td>$\beta$</td>
<td>1.16±0.033</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.0158±0.034 (Portugal)*</td>
</tr>
</tbody>
</table>

*This value could go as high as 35 as an ersatz for the determination of the other values for the Italian data. Please see discussion.

4. DIFFERENT SCENARIOS FOR THE EVOLUTION OF SARS-COVID-2 IN PORTUGAL.

The paper by Xia et al [13] further presents an extra 7 parameters to take into account control measures in time evolution of the epidemic. These parameters are meant to take into account two things:

- Isolation and monitoring measures taken by the government (parameters $d_1$, $d_2$, $d_3$, and $d_4$);
- Self-protection measures taken by the population (parameters $l_1$, $l_2$, and $l_3$).
The way these parameters are introduced in the model are as follows:

\[
\begin{align*}
\frac{dS(t)}{dt} &= -l_1 \beta_1 \frac{S(t)A(t)}{N} - l_2 \beta_2 \frac{S(t)I(t)}{N} - l_3 \beta_3 \frac{S(t)H(t)}{N}, \\
\frac{dE(t)}{dt} &= l_1 \beta_1 \frac{S(t)A(t)}{N} + l_2 \beta_2 \frac{S(t)I(t)}{N} + l_3 \beta_3 \frac{S(t)H(t)}{N} - (\sigma + d_1)E(t), \\
\frac{dA(t)}{dt} &= (1 - \gamma)\sigma E(t) - (k_1 + d_2)A(t), \\
\frac{dl(t)}{dt} &= \gamma \sigma E(t) - (\lambda + d_3)I(t), \\
\frac{dH(t)}{dt} &= \lambda I(t) - k_2 H(t) - (\delta + d_4)H(t), \\
\frac{dR(t)}{dt} &= k_1 A(t) + k_2 H(t) + \delta H(t) + d_1 E(t) + d_2 A(t) + d_3 I(t) + d_4 I(t).
\end{align*}
\]

The \(l_1\), \(l_2\), and \(l_3\) parameters are basically multiplication factors for the transmission parameter \(\beta\), so in our model we only consider one value for it as we did for \(\beta\). The \(d_1\), \(d_2\), \(d_3\), and \(d_4\) parameters are summed to the parameters of spread between the infected, asymptomatic and hospitalized cases, further stressing their connection to implemented government measures. The values provided by Xia \textit{et al.}[13], when applied to the Portuguese case, seem to provide a very unrealistic scenario, in which the values would immediately start to diminish. Perhaps this made sense in the case of the MERS epidemic, given that it was much more localised that the one provoked by this virus. As such, and for the lack of a better model, I devised two different scenarios in addition to the “out-of-control” scenario described in the previous section:

- Scenario in which the government takes severe mitigating measures and the population mildly adheres to self-protecting measures (scenario 1);
- Scenario in which the government fails to implement mitigating measures, yet still implements some, and the population mildly adheres to self-protecting measures (scenario 2).

The values for these scenarios are given in Table 5.

| Table 4. Control parameters of the model described by Eq. 2 |
|---------------------------------|-----------------|-----------------|
| parameter | Xia \textit{et al.}[13] | this work (scenario 1) | this work (scenario 2) |
| \(d_1\) | 0.1922 | 0.1922 | 0.0961 |
| \(d_2\) | 0.1922 | 0.1922 | 0.0961 |
| \(d_3\) | 0.1922 | 0.1922 | 0.0961 |
| \(d_4\) | 0.1922 | 0.1922 | 0.0961 |
| \(l_1\) | 0.3239 | 0.5 | 0.5 |
| \(l_2\) | 0.9175 | 0.5 | 0.5 |
| \(l_3\) | 0.5316 | 0.5 | 0.5 |

The day considered for the initiation of these measures is day 33 corresponding to the 18th of March 2020 which is 5 days after the government implemented the measures, so giving the virus another cycle of infection before the measures take effect. Obviously other values could have been used, but there was no clear...
indication of what these values could be, therefore I just assumed, as a first approximation, that the protective measures would manage to diminish the transmission of the contamination by 50%. For the government related measures I used the same value as the ones from Xia et al[13]. In the second scenario I keep the 50% reduction in the transmission rate but diminish the efforts of the government by half. Results are shown in figure 4.

![Graphical representation of the different scenarios considered in this study.](image-url)

Figure 4. Graphical representation of the different scenarios considered in this study.

Results show that measures can be effective in “flattening the curve”, as it has come to be called this effect in recent times. In the first scenario the effectiveness of the control measures seems to be at a maximum, and a maximum of just about 700 cases would be reached. This seems unlikely as of today, there are 641 active cases in the country. In scenario 2, the maximum is shifted to the 20th of April, reaching a maximum number of ~7,000 active cases in the country. In this scenario, the transmission rate has been cut to a half, but the government has not been effective in implementing control measures. As said previously, in the out-of-control scenario, the peak would be reached at about ~40,000 people.

5. LIMITATIONS OF THIS STUDY AND SCOPE OF APPLICATION

Given the huge uncertainties concerning these numbers, and the current lack of knowledge of the ongoing epidemic, it is very hard to make sense of its evolution, even using well-established SIR methods. The current work makes a series of
assumptions which may or may not be proven correct in the future, including the incubation period of the virus, which is still not well-known. This proved to be a major source of uncertainty to fit the results, because it had a great influence in the other parameters, especially parameter $\gamma$, which seems to describe a probability of the person being exposed to the virus being symptomatic or not. Instead this parameter seems to behave in this study as an indicator of the huge uncertainty of the value for the incubation period. Finally, epidemiologists and pathologists will tell you that these parameters, unlike what is considered in this study, are not only not constant, but dependent on time, location, culture, and other factors. It is therefore particularly hard to model the current epidemic of covid-19.

Also, this study takes into account the particular situation for one country (Portugal), and we know that in globalised times, the likelihood that new foci of infections arise from outside visitors is also very high. Although currently most government have implemented serious travel restrictions, it is now well-known exactly how this parameter could affect this model, which does not take that into account. In fact, one of the major issues with the SARS-covid-2 virus is that, unlike the other SARS, it seems to be much less contained locally, as is evident by the amount of countries afflicted.

The amount of actual active cases is another clear source of uncertainty. This model takes the values of active cases and compares them with official confirmed numbers, which not only may be smaller than the actual cases, as are released at an administratively defined time, which probably does not correspond to something realistic.

The author believes that the values obtained in this study should be taken qualitatively rather than quantitatively. One can think of the scenarios as defining two limits of a likelihood interval, and as such, the actual figure of maximum number of active cases in Portugal will likely be in the interval $\sim700$-$7,000$ between the 9th-20th of April. This can perhaps serve to advise the competent authorities in taking the appropriate measures. It also serves to show the importance of control and self-protecting measures. It is possible to bring down the number of affected people by following the recommendations of the WHO and health authorities. Without them, this model predicts that the figures of active cases would reach a staggering $\sim$40,000 people. It also shows that with the appropriate measures this can be indeed brought down. Hopefully that will be the case not just in Portugal, but in the rest of the World.

REFERENCES


[21] Coronavirus incubation could be as long as 27 days, Chinese provincial government says - Reuters, Feb. 22, 2020