Modelling the spread of Covid19 in Italy using a revised version of the SIR model

Andrea Palladino[1,2], Vincenzo Nardelli[1,3], Luigi Giuseppe Atzeni[1], Nane Cantatore[1,5], Maddalena Cataldo[1,6], Fabrizio Croccolo[1,4], Nicolas Estrada[1,7], and Antonio Tombolini[1,5]

[1] CoVstat (Italy)
[2] DESY, Platanenallee 6, Zeuthen (Germany)
[3] Universit degli Studi di Milano - Bicocca, Milano (Italy)
[4] Universite de Pau et des Pays de lAdour, E2S UPPA, CNRS, TOTAL, LFCR UMR5150, Anglet, (France)
[5] Tombolini & Associati srl, Loreto (Italy)
[6] Universit degli Studi dell'Aquila, L'Aquila, (Italy)
[7] Universit degli studi di Padova, Viccolo dell'Osservatorio 3, Padova (Italy)

May 2020

Abstract

In this paper, we present a model to predict the spread of the Covid-19 epidemic and apply it to the specific case of Italy. We started from a simple Susceptible, Infected, Recovered (SIR) model and we added the condition that, after a certain time, the basic reproduction number R_0 exponentially decays in time, as empirically suggested by world data. Using this model, we were able to reproduce the real behavior of the epidemic with an average error of 5%. Moreover, we illustrate possible future scenarios, associated to different intervals of R_0. This model has been used since the beginning of March 2020, predicting the Italian peak of the epidemic in April 2020 with about 100.000 detected active cases. The real peak of the epidemic happened on the 20th of April 2020, with 108.000 active cases. This result shows that the model had predictive power for the italian case.

1   Introduction

At the beginning of 2020, a previously-unknown respiratory tract disease was reported in China [1]. This event is having a huge negative impact worldwide, not only under an healthcare perspective, but also under the economic, social and cultural ones. SARS-CoV-2 has been identified as the causative agent of the pandemic outbreak. It is a newly encountered member of the coronavirus family belonging to the RNA-viruses. Its behaviour

is comparable to influenza viruses or SARS-CoV the causative agent of the pandemic outbreak 2002/03 of [1][2]. As soon as virus particles get into a host (human), they start invading cells (in this case, predominantly the ones in the respiratory tract) and these replicate the genome of the virus. Virus particles get into the hosts saliva and humans infect each other by talking to infected individuals, touching hands and close face-to-face interaction [2][3]. A number that is a good landmark for the transmission rate, or the infectiousness of any infectious disease, is the basic reproduction number (R_0): this defines the average number of people that are infected by a single carrier over a defined period of time. R_0 is an indicator of the transmissibility of the epidemic and it has been defined as the average number of secondary cases that a single case can generate in a completely susceptible population. R_0 is of course dependent on the characteristics of the epidemic itself; however, it also depends on the population sample we are considering. The higher the human interaction in a population is, the higher the value R_0 will be. Among the factors that can influence R_0 in a given population there are therefore social habits and social organization. The basic reproduction number is therefore mutable depending on these aspects, and the analysis of its variation in time can be crucial to monitor the trend in the transmissibility among a single population. For our specific case of study, i.e. the spread of the COVID-19 epidemic in Italy, it is important to study the variation of R_0 in time before and after the implementation of lockdown measures, as well as after its removal. In our modified SIR model, we let R_0 vary in time as a consequence of quarantine. As we will show later, this will also allows us better reproduce real data, as well as have a predictive view on the whole trend of the epidemic.

Symptoms of coronavirus disease (COVID-19) are widespread: from asymptomatic patients to patients with flu-like symptoms up to a severe pneumonia leading to a severe acute respiratory distress symptom (ARDS) [1][2], making the ventilation of patients unavoidable.

Given the lack of reliable and long-term data regarding incubation period, virulence, contagiousness, and other transmission parameters [1] for the novel coronavirus SARS-CoV-2 and the lack of reliable drugs and vaccines [3], containment measurements, the tracking of infected people and the treatment of patients in the early stage of the illness, remain the only feasible option to face the ongoing outbreak of the virus that is leading to a collapsing health system with thousands of deaths, as seen in hotspots.

The impact of Covid-19 in Italy has been and is still very severe, with a death toll notably high. Up to now, more than 200.000 people have been tested positive in Italy and more than 30.000 died due to the Coronavirus. In this paper, we present the model used by the CoVstat group [7] to model the spread of the epidemic. Using this model, we had been able to predict with good accuracy the peak of the active infected, both related to its location in time (the date of the peak) and its amplitude (the maximum number of active infected).

2 Methods

In this paragraph we present the SIR model, that is used as base for the development of further models from the CoVstat group. We point out that the SIR model is not predictive. In Sec.2.2 we present the ingredients that represent a novelty here and have been included to improve the performance of the model and to describe the real data in a reasonable manner.
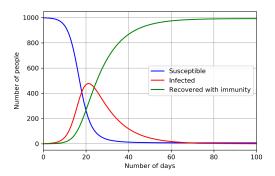
2.1 The standard SIR model

SIR is one of the simplest models to describe the spread of an epidemic [8]. The SIR model is based on the assumption of a totally susceptible population at time t_0, i.e. the beginning of the spreading. In the SIR model, the overall population of **N** individuals is divided into 3 categories: susceptibles (**S**), infected (**I**) and removed (**R**). Hence, at a given time t from the beginning of the spreading of the epidemic, $I(t)$ and $S(t)$ are the number of infected people present in the population and the number of vulnerable people that have not contracted the virus yet, respectively, while $R(t)$ is the sum of the ones that have developed immunity (recovered) or deceased and are therefore removed from the susceptible count. It is straightforward to notice that, at any time t, $S(t) + I(t) + R(t) = N$. The SIR model purpose is to describe the variation in time of S(t), I(t), and R(t) meaning the migration in time of individuals among these 3 categories. The model consists of 3 categories: **S** for the susceptible people, **I** for the infected people and **R** for the sum of recovered and deceased people. The classic SIR model is described by 3 ordinary differential equations:

$$
\begin{aligned}
\frac{dS}{dt} &= -\beta \frac{I\,S}{N} \\
\frac{dI}{dt} &= \beta \frac{I\,S}{N} - \gamma I \\
\frac{dR}{dt} &= \gamma I
\end{aligned}
$$

where β is related to the velocity of diffusion of the virus and γ is related to the time required to infected people to become removed (recovered or deaths). Both parameters have the dimension of time^{-1}. The three equations written above can be interpreted in the following manner:

- at the beginning of the epidemic, the entire population is susceptible to the infection ($S(0) = N$). If there is a single infected person, other people can get the infection, going from the category **S** to the category **I**. The strength (speed) of the spread of the virus is determined by the parameter β;

- the number of infected people increases when susceptible people get infected. After a typical timescale equal to $1/\gamma$ infected people **I** go in the third category **R**;
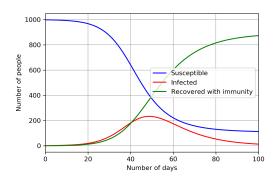
3

Figure 1: Evolution of an epidemic using the SIR model, with $\gamma = 0.1$ and $\beta = 0.5$ on the left panel and $\beta = 0.25$ on the right panel. In the classical SIR model, R_0 is proportional to the fraction of susceptible people. Particularly it is equal to S/200 in the left panel and S/400 in the right panel. The condition $R_0 = 1$ occurs when the red curves reach their maximum.

- the category **R** includes the sum of people that recovered or died after infection.

Therefore, in the classic SIR model there are only 2 free parameters that can be used to fit real data: β and γ. Both β and γ have dimension of time^{-1}. From here on, we will use days as unit of time.

In this model, the basic reproduction number R_0 is a dimensionless number obtained using a combination of the previous parameters:

$$R_0(t) = \frac{\beta S(t)}{N\gamma}.$$

$R_0(t)$ provides direct and quantitative information on the spread of the epidemic. If $R_0 \leq 1$ the epidemic will stop spontaneously, while with $R_0 > 1$ it will continue spreading. In Fig. 1 we show two different evolution patterns of the pandemic, to demonstrate the effect of different values of β. Both simulations start with the initial condition of 1 infected person, for a population of N =1000 people in total, and fixing $\gamma = 0.1$ (i.e. typical duration of the illness of 10 days). In both panels the blue line represents the number **S** of susceptible people which at time $t = 0$ days coincides with almost the entire population (for our simulation: **S**(0) = 1000 − 1) and then decreases as both the recovered and infected numbers (green and red lines, respectively) increase. After an interval of time, in both the simulations, the infected number **I** reaches a peak, the value of which is dependent on the parameter β. After that time, the number of infected decreases as the recovered increase and ultimately reaches the total number of individuals. In the simulation on the left we set $\beta = 0.5$ while on the right we assume $\beta = 0.25$, to simulate the implementation of social distancing actions. We notice that in the left panel the peak of infected people comes after

20 days, with roughly half the population infected. On the right panel the peak is shifted and comes after 50 days and the number of infected people at the peak is less than the half of the previous case. This shows the importance of social distancing, since social distancing and quarantines reduce the parameter β, helping in reducing the number of infected people at the peak of the epidemic. This is indispensable to avoid the collapse of the healthcare system, and especially the intensive care units. Although a simulation with the standard SIR appears to be adequate to describe an epidemic spreading in a sample where all the initial conditions remain constant throughout the period of time, it is not sufficient when it comes to a more complex and realistic situation such as the population of a given country, where the parameters of the model are influenced by other external factors. From this the necessity to modify the model for our case of study. In the classic SIR model the parameter β is constant, therefore it cannot account for the effect produced by quarantine actions, that would have also a dynamical impact on the parameter $R_0(t)$. For this reason, it is useful to go beyond the classic SIR model, as explained in the next section.

2.2   Beyond the classic SIR model

In the classic SIR model, the parameter β is constant in time. This means that it cannot account for the slowdown of the spread due to the quarantine. To simulate a more realistic scenario, it is necessary to go beyond the classic SIR model. We denote this model as SIR 2.0. The open source code is available in the CoVstat repository [11]. Compared to the classic SIR it contains the following new features:

- the parameter $\beta(t)$ changes in time, to account for the effects due to quarantine and social distancing. Particularly $\beta(t) = \beta_0$ before a time t_{th}, while it exponentially decays for $t > t_{th}$:
$$\beta(t) = \beta_0 \ e^{-(t-t_{th})/\tau} \quad \text{for } t \geq t_{th},$$
where t_{th} and τ are two additional parameters of the model. The time t_{th} represents the starting time of the quarantine actions, while τ refers to the decaying period and it has the dimension of time. The previous assumption is driven by empirical observations, since this behavior of $R_0(t)$ has been observed in several countries during the Covid19 pandemic;

- we take into consideration the possible presence of asymptomatic patients. The virologist Ilaria Capua has suggested that 2/3 of patients in Italy might be asymptomatic [9]. The study conducted on the population of Vo' Euganeo [10] reaches similar conclusions. Therefore, it is reasonable to assume that the total number of infected people is roughly 3 times higher that the number of people that were tested positive.

Compared to the classic SIR model, in this revised version there are 2 more parameters, t_{th} and τ. Therefore the SIR 2.0 is characterized by 4 parameters in total: β_0, γ, t_{th} and τ.
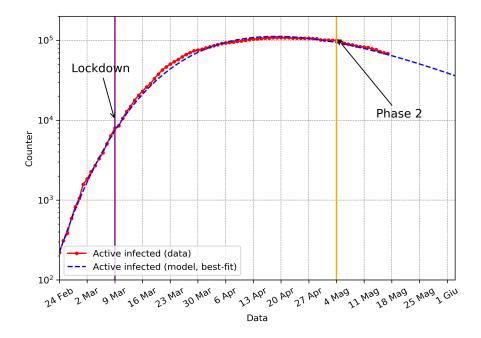
Figure 2: Simulation of the Italian case using the SIR 2.0 model. The red points represent the real number of active cases, while the blue curve represent the best-fit curve of the model. The magenta vertical line indicates the beginning of the quarantine, while the orange vertical line indicates the end of it.

3 Results

In this paper we present and discuss varios results obtained using the model described above. We first focus on the curve of the active infected in Italy. Then we discuss possible future scenarios and how to compute the parameter R_0 for Italy and italian regions. From here on, we will always use data from the 24th of February to the 17th of May.

3.1 The Italian case

Using the model SIR 2.0 it is possible to reasonably describe the spread of COVID-19 in Italy. We used the data of Protezione Civile [12], starting on February 24th, 2020 (therefore, this date corresponds to time zero of our model). We use $N = 60.36 \times 10^6$ people as population and, as initial conditions, we choose $R_0 = 0$ and $I_0 = 221$, corresponding to the number of infected people on the 24th of February. In order to find the best model, we minimize the mean squared error between predictions and true data, allowing the 4 free parameters to vary. We consider as a last update the 17th of May. The best fit is given by
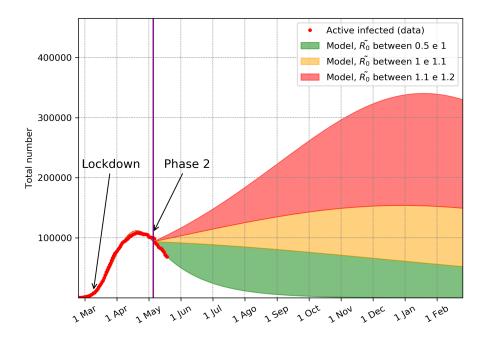
6

Figure 3: Simulation of future scenarios, fixing a certain value of R_0 at the 4th of May. We identify 3 regions, corresponding to low risk (green), mid risk (orange) and high risk (red) situations.

the following set of parameters:

$$\beta_0 = 0.384, \quad \gamma = 0.048, \quad t_{th} = 0, \quad \tau = 26.33 \ .$$

The best model has an average error of 4.9% compared to the data and is shown in Fig.2 as the dashed blue line. On the home page of [7] the best fit model is updated daily. However it is important to recall that the predictions during the pandemic were stable. Since the beginning of March the model has predicted the peak of the infected people in April, with 100.000 of infected people (see Palladino's talk [13]).

3.2 Future predictions

The phase 2 started in Italy on the 4th of May with progressive removal of quarantine. The conditions are different from phase 1, however, since the lockdown has not been completely released we cannot assume the initial conditions are fully restored. Even when the country's lockdown will formally be over, the local legislation about social distancing and the civic consciousness of population will be dramatically different than before. On the other side we still don't have enough information about the immunity for those who have contracted

7

the virus. Therefore, it is hard, or even impossible, to make accurate predictions without a reasonable model for the evolution of $R_0(t)$. However, it is interesting to understand which future scenarios are associated to different intervals of the parameter $R_0(t)$.

In order to do that, we fit the past data with the SIR 2.0, as explained in the previous section. Then, we use the basic SIR model for new future predictions, fixing a certain ratio $\beta(t)/\gamma$ on the last day of data. Therefore, the value of $\tilde{R}_0$ at the t_d=4th May is given by:

$$\tilde{R}_0 = \frac{\beta(t_d)S(t_d)}{\gamma N}$$

where $S(t_d)$ is the number of susceptible people on t_d. Then $R_0(t)$ evolves as:

$$R_0(t) = \tilde{R}_0 \frac{S(t)}{N}$$

Since in our model we assume that the total number of infected people is 3 times higher that the number of tested positives $I(t)$, the previous expressions becomes equal to:

$$R_0(t) = \tilde{R}_0 \left(1 - \frac{3I(t)}{N}\right)$$

The epidemic starts decreasing when $R_0(t) < 1$. This condition is always satisfied when $\tilde{R}_0 < 1$. In Fig.3 we report 3 different regions corresponding to 3 different intervals of $\tilde{R}_0$, as explained in the legend of the figure.

3.3   The computation of R0(t) for Italy and regions

In this section we focus on the computation of $R_0(t)$ for Italy and Italian regions. In order to do that, we focus on a subset of data. Particularly we use that last 5 days of data for Italy and last 7 days of data for Italian regions (due to the smaller statistics). This is done to avoid fast oscillations of R_0 and to better understand the general trend of the epidemic during the last week. In order to do that, we use the standard version of the SIR model. We assume an average duration of the illness of 14 days, i.e. $\gamma = 1/14$. Then we minimize the mean squared errors between real data and model, varying $R_0(t)$, with the usual definition given for the SIR model. Let us notice that, even assuming that the total number of infected people is 3 times larger than that the tested positive ones, the number of susceptible people is still very high. Therefore the value of $R_0(t)$ corresponds, in very good approximation, to:

$$R_0(t) \simeq \frac{\beta(t)}{\gamma}$$

With the present numbers, this remains true even if the number of infected people were 10 times larger. Indeed up to now order of 200 K cases were tested positive. So, if the true number of infected people were 2.2 millions, the ratio $S(t_d)/N = 0.97$, very close to 1. The

Figure 4: Left panel: computation of $R_0(t)$ in Italy using the last 5 days of data. Right panel: evolution of $R_0(t)$ in Italy as function of time, using time window of 5 days. Mid panel: $R_0(t)$ computed for italian regions (updated to the 17th of May).

evolution of $R_0(t)$ in Italy is shown on the right panel of Fig.4. On the left panel of the same figure, we represent $R_0(t)$ computed on the last 5 days.

In the central panel of the same figure we report the last value of $R_0(t)$ for italian regions, updated to the 17th of May. In this case we use a time window of 7 days, to compensate

the smaller statistics and to reduce the oscillations of $R_0(t)$. Using this procedure, $R_0(t)$ becomes a good indicator of the behavior of the pandemic during the last week.

4 Conclusion

We have built a model to predict the spread of the Covid-19 epidemic in Italy. We started from a simple SIR model and we added the condition that, after a threshold time, the basic reproduction number $R_0(t)$ exponentially decays in time, as empirically suggested by the spread of the epidemic in different countries. Using this model we were able to predict the peak of the epidemic 1.5 months before it happened, with an error of 1 week on the period and an error smaller than 10% on the absolute numbers. We have also presented possible future scenarios, assuming different intervals of the parameter $R_0(t)$ after the 4th of May, i.e. when the lockdown in Italy has been released. We conclude explaining our procedure to compute $R_0(t)$, as a function of time, for Italy and Italian regions. This paper shows that the model has a good predictive power, when a period of quarantine is observed with fixed conditions.

References

[1] Guo, Y., Cao, Q., Hong, Z. et al. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak an update on the status. *Military Med Res* 7, 11 (2020). https://doi.org/10.1186/s40779-020-00240-0

[2] Singhal, T. A Review of Coronavirus Disease-2019 (COVID-19). *Indian J Pediatr* 87, 281286 (2020). https://doi.org/10.1007/s12098-020-03263-6

[3] Lai, C. C., Shih, T. P., Ko, W. C., Tang, H. J., & Hsueh, P. R. (2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. *Int J Antimicrob Agents*, 55(3), 105924. doi:10.1016/j.ijantimicag.2020.105924

[4] Wu, J.T., Leung, K., Bushman, M. et al. Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nat Med* (2020). https://doi.org/10.1038/s41591-020-0822-7

[5] Velavan, T. P., & Meyer, C. G. (2020). The COVID-19 epidemic. *Trop Med Int Health*, 25(3), 278-280. https://doi.org/10.1111/tmi.13383

[6] Li, Q., et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. New England Journal of Medicine 382, 1199-1207 (2020).

[7] www.covstat.it

[8] Kermack, W.O., McKendrick, A.G. Contributions to the mathematical theory of epidemics I-II-II. *Bltn Mathcal Biology* 53, 3355, 5787, 89118 (1991). https://doi.org/10.1007/BF02464423, https://doi.org/10.1007/BF02464424, https://doi.org/10.1007/BF02464425,

[9] https://www.la7.it/aggiornamenti-sul-coronavirus/video/la-virologa-ilaria-capua-23-dei-contagi-potrebbero-essere-asintomatici-25-02-2020-309481

[10] Lavezzo, Enrico et al., Suppression of COVID-19 outbreak in the municipality of Vo, Italy medRxiv,10.1101/2020.04.17.20053157

[11] https://github.com/vincnardelli/covid19-italia

[12] https://github.com/pcm-dpc/COVID-19/raw/master/dati-andamento-nazionale/dpc-covid19-ita-andamento-nazionale.csv

[13] Link to the DESY seminar