Chinese and Italian COVID-19 outbreaks can be correctly described by a modified SIRD model

Diego CACCAVO[1,2,3]

[1] Department of Industrial Engineering, University of Salerno, 84084 Fisciano (SA), Italy
[2] Department of Pharmacy, University of Salerno, 84084 Fisciano (SA), Italy
[3] Eng4Life Srl, Academic spin-off, Via Fiorentino, 32, 83100 Avellino (AV), Italy

dcaccavo@unisa.it

Abstract

The COVID-19 disease is rapidly spreading in whole globe, affecting millions of people and pushing governments to take drastic measures to contain the outbreaks. The understanding of the dynamics of the epidemic is of great interest for the governments and health authorities that are facing COVID-19 outbreaks. The scarce presence of epidemiologic data, due to the still ongoing outbreaks, makes prediction difficult and mainly based on heuristic (fitting) models. However, these models with non-physical based parameters, can only give limited insight in the evolution of the outbreaks. In this work a SIRD compartmental model was developed to describe and predict the evolution of the Chinese and Italian outbreaks. Exploiting the similarities of the measures taken by the governments to contain the virus and of the total population number of Hubei province and Italy, the model was tuned on the Chinese outbreak (almost extinguished) and by perturbation the Italian outbreak was describe and predicted. With this approach the lack of data due to the still ongoing Italian epidemic was compensated by the Chinese experience, allowing to obtain a SIRD model for this relatively new COVID-19 outbreak. According to the modeling prediction the peak of infection should be on the 28th of March, with almost 42,000 infected people.

Keywords

Coronavirus; Epidemic; Modeling; Compartmental models; China; Italy.

1. Introduction

The COVID-19 disease, caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has spread rapidly in the first months of the 2020 and the World Health Organization (WHO) on the 11th of March declared a global pandemic [1]. COVID-19 symptoms are often (80%) mild [2], may appear 2-14 days after exposure (mean incubation period 5-6 days) [2], and include fever, cough and shortness of breath [3]. However the 13.8% of infected have severe disease and the 6.1% are critical (respiratory failure, septic shock, and/or multiple organ dysfunction/failure) [2]. Mild cases typically recover within two weeks, while those with severe or critical disease may take three to six weeks to recover. Among those who have died, the time from symptom onset to death has ranged from two to eight weeks [2]. In the severe and critical disease, the patient admission to intensive care unit (ICU) is necessary and the mechanical ventilation (artificial assistance to support breathing), may be required [4]. This special treatment is the bottle neck of most of the health care systems, which are not prepared to deal with the numbers of patients that the COVID-19 disease could produce. Therefore, measures of containment of the SARS-CoV-2 outbreaks are essential to not overstress the health care systems and to manage the pandemic.

The SARS-CoV-2 was firstly reported at the end of 2019 in the Wuhan city (Hubei province), China, reaching epidemic proportion at the end of January 2020. On the 23rd of January the Chinese government imposed the lockdown in Wuhan and other cities in Hubei province in an effort to quarantine the center of a COVID-19 outbreak. These drastic measures, with almost 60 million of people in quarantine, were defined “unprecedented in public health history” by the WHO. Unfortunately, these measures were not able to stop the spreading of SARS-CoV-2, which caused several outbreaks around the globe.

Italy was one of the countries most affected by the SARS-CoV-2, reporting the first positive patient on the 20th of February 2020, followed by an extensive (and unexpected) epidemic in the north of Italy, especially in the Lombardy region. The Italian government, similarly to the Chinese government, progressively from the 8th of March to the 11th of March 2020 imposed the lockdown in the whole nation, introducing the so called “social distancing”.

These drastic measures, also adopted by other EU countries, apart from the health sector, have a dramatic impact on the global economy as well as on the lifestyle habits of millions of people. Epidemiologic descriptions and predictions could help in the understanding of the dynamic of the COVID-19 disease, supporting governments and health authorities in the resources allocation [5]. Most of the models proposed until now to describe the COVID-19 outbreaks, due to the scarcely data available of a still ongoing epidemic, are heuristic models. These models are “fitting” equation, exponential [5] or logistic [6], used to describe the infection data, whose parameters have scarce (or none) physical meaning and their use for prediction, especially at long term, should be avoided. On the contrary compartmental models could constitute the right trad-off between the need of having physical based parameters, the easiness of solution and the scarce presence of data in the evolving epidemic. Typically, the population of interest is subdivided into a small number of compartments based on infection status (e.g. susceptible, infectious or recovered) and the flows between these compartments are described by ordinary differential equations [7]. Aim of this work is to develop a compartmental model (SIRD) to describe first the Chinese COVID-19 epidemic and then, by perturbation of the model parameters, to describe and predict the Italian COVID-19 epidemic.

2. SIRD model

In compartmental models the population under study is divided into compartments and, with assumption that every individual in the same compartment has the same characteristics, the movements of people within the compartments can be described. In particular with "SIRD" model a four compartments model is indicated, where "S" indicates the "Susceptible" compartment, "I" specifies the "Infected" compartment, "R" denotes the "Recovered" compartment and "D" identifies the "Death" compartment, as schematized in Figure 1.

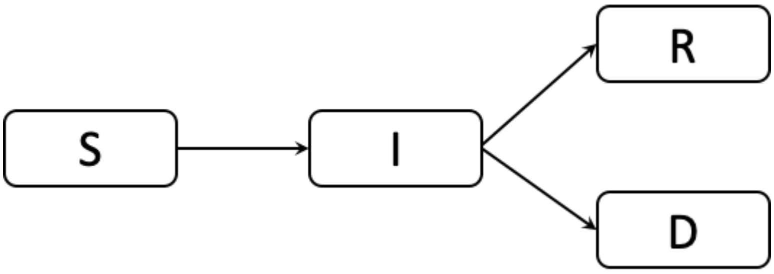

Figure 1. Schematization of the SIRD model. "S" indicates the "Susceptible" compartment, "I" specifies the "Infected" compartment, "R" denotes the "Recovered" compartment and "D" identifies the "Death" compartment.

It is assumed that the total population of reference is constant ($N = 60\,million$), meaning that in the considered time period the rate of birth and death (including the one related to the COVID-19) are equal [8]. This is particularly true for low lethality diseases with very large outbreaks (like a whole nation). It is also assumed that recovered people cannot be re-infected [9]. The ordinary differential equations, with their initial conditions, describing the movements of people from one compartment to the other are reported in the eq. (1-4).

$$\frac{dS}{dt} = -\frac{\beta I}{N} S \qquad (1)$$
$$\mathrm{S}(\mathrm{t}=0) = \mathrm{S}_0 = N - I_0 - R_0 - D_0$$

$$\frac{dI}{dt} = \frac{\beta I}{N} S - \gamma I - \delta I \qquad (2)$$
$$I(t=0) = I_0 = I_{data}(1)$$

$$\frac{dR}{dt} = \gamma I \qquad (3)$$
$$R(t=0) = R_0 = R_{data}(1)$$

$$\frac{dD}{dt} = \delta I \qquad (4)$$
$$D(t=0) = D_0 = D_{data}(1)$$

In eq. 1, S represents the number of people susceptible to the disease (total minus the already interested ones), β is the average number of contacts per person per time [1/day] (multiplied by the probability of disease transmission in a contact between a susceptible and an infectious subject). The probability to get infected is proportional to the product of the infected and the susceptible individuals, or better to the fraction of infected individuals (I/N) and the susceptible.

In eq. 2, I represents the number of infected people, whose initial value is taken from the literature data. Susceptible people get infected by the contact with other infected people (first term RHS), whereas infected people can recover (γI) with the kinetic γ, which is inversely proportional to the duration of the infection and directly proportional to the probability of recovery. Moreover, critical patients can die (δI) with the kinetic δ, which is inversely proportional to the average time between the infection and the death and directly proportional to the lethality of the disease.

Eq. 3 and eq. 4 describe the time evolution of the number of recovered patients R and the death patients D. Actually, the four equations are not independent and the fourth one can be derived from the total balance on the population:

$$D(t) = N - S(t) - I(t) - R(t) \tag{5}$$

Despite very simple SIR (o SIRD) models are based on constant kinetics (β, γ, δ) to describe the epidemic evolution until the depletion of the susceptible people, the measures taken by the governments have had huge impact on the epidemic and therefore on its evolution kinetics.

2.1 β the average number of contacts per person per time

The kinetic β is most of the time used as fitting parameter to describe the epidemiologic data. In this case, with the "quarantine" protocol taken by the governments, the number of contacts per person per unit time can dramatically decrease during the epidemic. In this work the kinetic of infection was described with an exponential function:

$$\beta(t) = \beta_0 \exp\left(-\frac{t}{\tau_\beta}\right) + \beta_1 \tag{6}$$

β_0 (actually $\beta_0 + \beta_1$) is the initial (high) infection kinetics that decreases exponentially thanks to the lockdown measures. β_1 is the infection kinetic at infinite time, which can be set to be zero due to a long enough period of people isolation. τ_β is the characteristic time of decrease, at $t = 3\,\tau_\beta$ the kinetic of infection would be decreased of the 90%.

2.2 γ the kinetic of recovery

The kinetic of recovery γ, in case of a new disease like the COVID-19, is not a constant. The health care system, the medical doctors with all their staff have to learn new therapeutic procedures, get used to treat patients with new symptoms etc. This means that initially the time of recovery will be longer, therefore the kinetic of recovery slower, until a new constant value is reached that should be inversely proportional to the average duration of the infection. In this work γ was modeled with a logistic function:

$$\gamma\,(t) = \gamma_{\,0} + \frac{\gamma_{\,1}}{1 + \exp\left(-t + \tau_\gamma\right)} \tag{7}$$

Where $\gamma_{\,0}$ is the recovery kinetic at time zero, $\gamma_{\,0} + \gamma_{\,1}$ is the recovery kinetic at regime, which is reached after τ_γ days of adaptation.

2.3 δ the kinetic of death

The kinetic of death δ, especially for new diseases, it is not constant and decreases with time, luckily. This is due to both the adaptation of the pathogen, whose less lethal strains tend to be the most spread, as well as due to the development of new clinical treatments. In this work an exponential decay function was used to describe this evolution:

$$\delta\,(t) = \delta_0 \exp\left(-\frac{t}{\tau_\delta}\right) + \delta_1 \tag{8}$$

Where (δ_0+δ_1) is the initial death rate, which decreases substantially (90%) at $t = 3\,\tau_\delta$, and reaches a long-term lethality rate δ_1.

2.4 Implementation and optimization

The model was implemented and solved in MATLAB 2019b, with the ode45 solver, based on an explicit Runge-Kutta (4,5) formula. The optimization of the model parameters to describe the experimental results was performed via the minimization of the Sum of Squared Errors (SSE) between the modeling and epidemiologic data. The SSE function was minimized with application of gradient-based method (fmincon) first, followed by a step of minimization with a gradient-free algorithm (pattern-search), followed by a third step of gradient-based method. The use of these three steps was necessary to reach the local minimum, whereas the reaching of a global minimum was proved with the multi-starting technique.

3. The epidemiologic data

The epidemiologic data were taken from an open source repository operated by the Johns Hopkins University Center for Systems Science and Engineering (JHU CSSE) [10]. The data regard the total number of cases, the recovered cases and the death cases. The number of infected can be estimated by the difference of the total number of cases minus the recovered and the death cases. As it can be seen the Chinese outbreak has been almost suppressed, with the peak of infection around the 18th of February. The Italian outbreak is instead still in the fast-growing phase but with the number of deaths already close to the Chinese one.

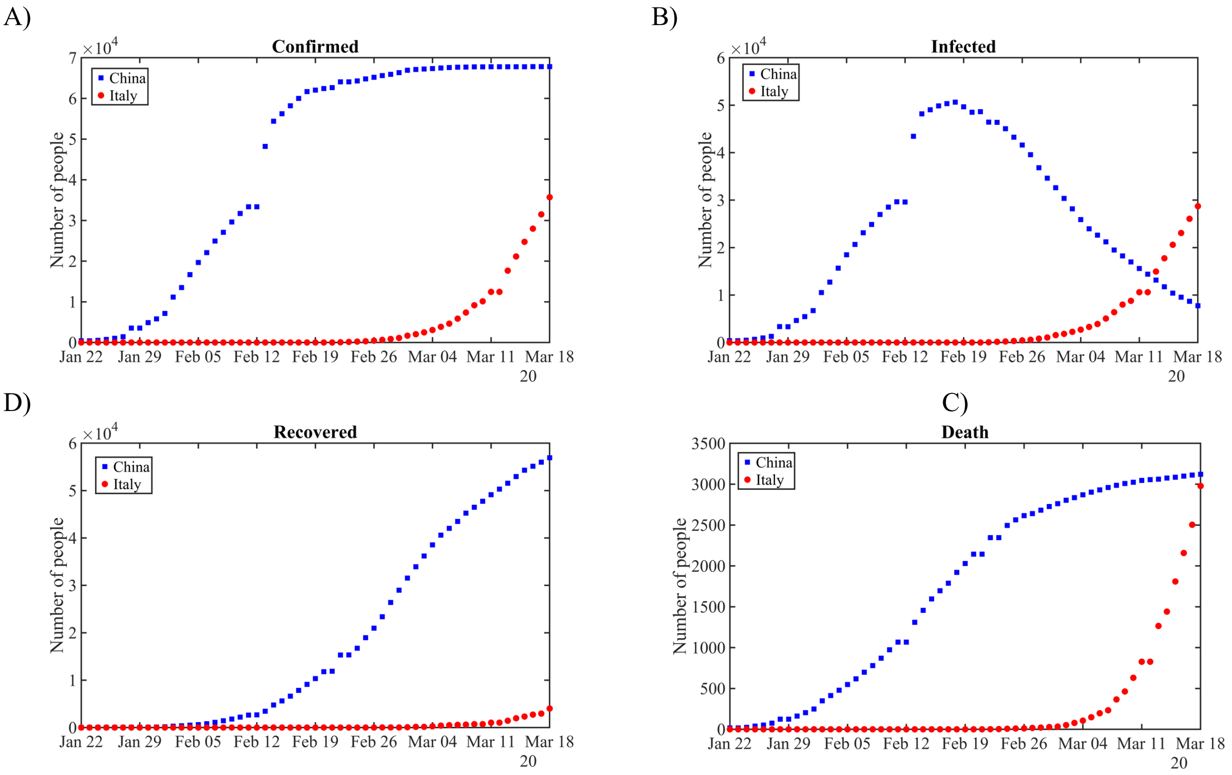

Figure 2. A) Confirmed cases. B) Infected people. As it can be seen the Chinese outbreak has reached the peak of the infected people around the 18th of February, whereas the Italian outbreak it is still in the exponential phase. C) The number of recovered patients. D) The number of death patients.

4. Results and discussion

The Chinese outbreak of COVID-19, almost at the end of its development, can be used to understand the dynamic of COVID-19 disease. The model parameters obtained for the description of the Chinese outbreak were perturbated to describe the initial Italian outbreak and to predict its evolution.

4.1 Chinese COVID-19 epidemic

The SIRD model results for the Chinese COVID-19 outbreak are reported in Figure 3. The model parameters (8 parameters) were fitted on the experimental data, obtaining a satisfactorily description of the epidemiologic data published so far.

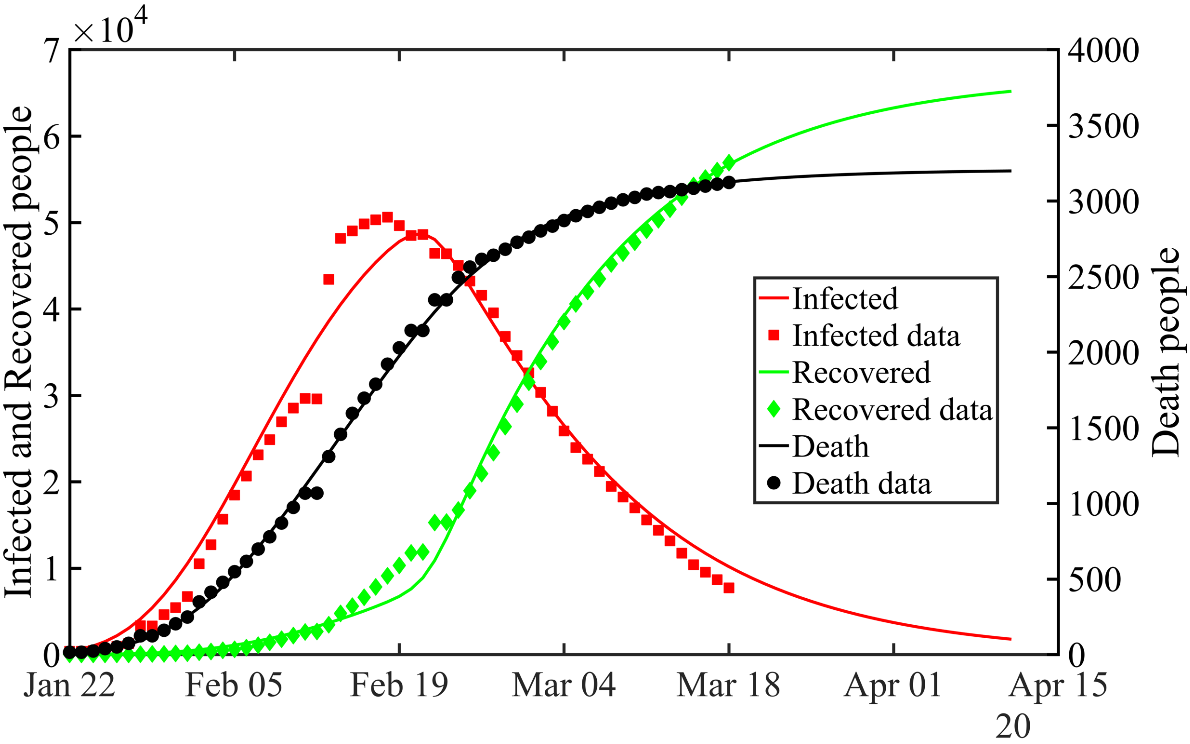

Figure 3. Chinese COVID-19 epidemiologic data descripted and predicted (at time longer than the data points) by the results of the SIRD model

The modeling infection peak is around the 18th of February, with almost 50,000 people infected, in accordance with the data, predicting the end of the outbreak at the beginning of May.

It is interesting to analyzed the value of the obtained modeling parameters, reported in Table 1, to make some physical considerations. The average number of contacts per person per time decreased with a characteristic time τ_β=10.46 [day]. It is worth to note that the epidemiologic data are referred to the lockdown day -1. Since than the average number of contacts per person decreased exponentially, due to the quarantine, and diminished of 90% after almost 30 days ($3\tau_\beta$). This demonstrate, along with the experimental date itself, that it required 30 days after the lockdown to substantially slow down the spreading of the virus.

The parameters referred to the kinetic of recovery γ_0, γ_1, τ_γ can give information on time of recovery ($d_{recovery}$). Actually $\gamma \propto P_{recovery}/d_{recovery}$, therefore considering the probability of recovery ($P_{recovery}$) of 90% (R/(R+I+D)), the recovery time at regime (after 30 days (τ_γ)) was $d_{recovery} = 13$ days.

The parameters related to the kinetic of death, analogously can be used to obtain information about the average time between the symptoms and the death. Considering the death probability of 4.5% (D/(R+I+D)), the average death time was progressively increased from 5 days in the first week (22-28th January) to 9 days considering the first 3 week, until reaching two months (meaning a lethality close to zero) at the end of the epidemic. This unexpected result can be explained considering that the treatments were improved as well as considering a parameter not directly accounted in this model: the relation between lethality and patience age [11]. Indeed the "quarantine" measures may be particular effective for old people (>60 years) and for people with more comorbidities, who form the group of patients in which the COVID-19 lethality is the highest.

Table 1. Model parameters obtained from a best fit optimization, with the * the parameters used in the Italian outbreak and imported from the Chinese SIRD model.

Parameter	China	Italy
β_0	0.5364 [1/day]	0.2584 [1/day]
τ_β	10.460 [day]	10.000* [day]
γ_0	0.011 [1/day]	0.020 [1/day]
γ_1	0.061 [1/day]	0.027 [1/day]
τ_γ	30.4 [day]	22.1 [day]
δ_0	0.010 [1/day]	0.019 [1/day]
δ_1	7.2×10^{-4} [1/day]	7.2×10^{-4}*[1/day]
τ_δ	14.0 [day]	14.0* [day]

4.2 Italian COVID-19 epidemic

The Italian modeling of the outbreak was performed by perturbation of the Chinese parameters (starting points of the optimization) and adding some constraints on the range of variability of the parameters. The modeling starts from the 26/02/2020, day in which the total cases in Italy were close to the cases in Hubei on the 22/01/2020 (around 400). The population of the Hubei province and Italy is also roughly the same: 60 million.

The main Italian lockdown was performed on the 11th of March, when the cases were already 10590 (against the 438 on the 23rd of January in China). To account for this delay, the equations for the description of the Italian outbreak were slightly adjusted for $\beta(t)$ and $\delta(t)$.

$$\beta(t) = \begin{cases} \beta_0 \,, t < t_{lockdown} \\ \beta_0 \exp\left(-\dfrac{(t - t_{lockdown})}{\tau_\beta}\right) \end{cases} \tag{9}$$

$$\delta(t) = \begin{cases} \delta_0 \,, t < t_{lockdown} \\ \delta_0 \exp\left(-\dfrac{(t - t_{lockdown})}{\tau_\delta}\right) + \delta_1 \end{cases} \tag{10}$$

Considering a constant rate for the number of contacts per person and also a constant kinetic of death before the date of lockdown (between the 26th of February and the 11th of March). After the lockdown both the function decrease exponentially due to the isolation of people (reduction of β) and especially for categories that are at greatest risk of death (reduction of δ). From the Chinese model, the characteristic time of infection τ_β, the characteristic time of death τ_δ as well as the lethality rate at long time δ_1 were imported and not used as optimizable parameters, considering comparable the measures taken by the Chinese and the Italian government and therefore their impact on the outbreaks.

The modeling results obtained are reported in Figure 4. The epidemiologic data are well described with the optimization of the five adjustable parameters. According to the model prediction the peak is expected between the end of March and the beginning of April (28th of March precisely) with almost 42,000 infected people. According to the data reported in [2], if the 6% of patients will require an admission to the intensive care unit, the number of beds in those units should be around 2500 (in accordance with [5]) which is almost half of the total availability [5].

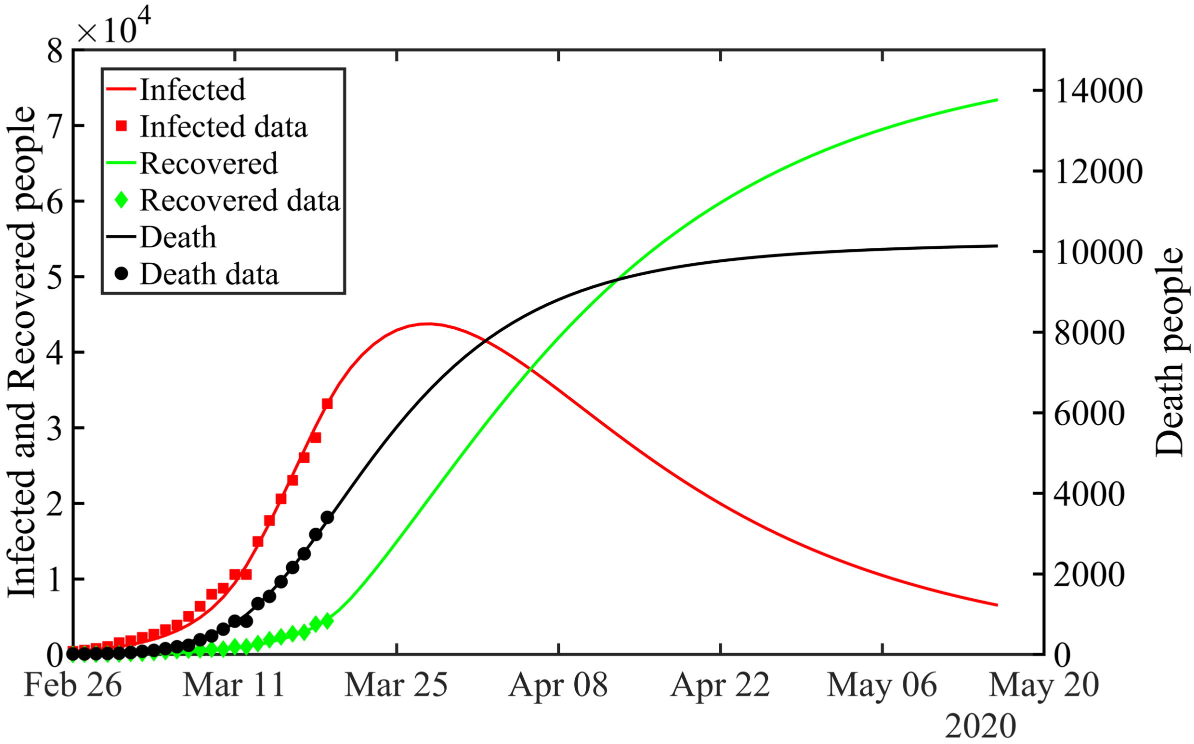

Figure 4. Italian COVID-19 epidemiologic data descripted and predicted (at time longer than the data points) by the results of the SIRD model

It is interesting to note that despite the same characteristic time of infection and the delay in the government restrictive measures, the peak is expected before April 11th (one month after the lockdown). This is because, as it can be seen from the Chinese and Italian infection kinetics, the “force of infection” is lower in Italy, which could be explained considering that the mains outbreaks in Italy have been developed in small cities, whereas in the Hubei province the principal outbreak was in the metropolitan city of Wuhan (11 million of people). It has also to be said that the peak is anticipated because of a greater lethality, that removes at higher rate (almost the double with the respect to the Chinese outbreak) patients from the infected compartment. Higher lethality, which is estimated by the model to be around the 11%, reflects also on the total number of death people at the end of the outbreak that, sadly, should be close to 10,000. The D-curve in Figure 4 is of course a sigma-shaped one, and it can be of interest to fit the experimental data by a proper equation, in order to provide easy estimate of proximal behavior. It is disputed if the D-curve has to obey the Logistic behavior (Logistic is the analytic solution for constant-parameter SIS model), or something different (e.g. Gompertz curve). The fit of the model calculation given in Figure 4 is in favor of the use of Gompertz equation as a descriptive tool. The higher lethality could be due to the higher median age of the Italian population (44.3 years) with respect to the Chinese population (37.4 years) [12]. However further studies should be performed to investigate such a high lethality, i.e. in [13] it is suggested that the reasons could be found in the Italian extensive intergenerational contacts, which are supported by a high degree of residential proximity between adult children and their parents. Regarding the recovered people, it has to be said that Italy had an initial higher recovery kinetic, probably due to the experience and treatments inherited from the Chinese case. At the end of the Italian epidemic the model estimates that the 0.136% of the Italian population will be interested by the SARS-CoV-2 infection.

5. Conclusions

COVID-19 disease, caused by SARS-CoV-2, has spread rapidly and the World Health Organization (WHO) on the 11th of March declared a global pandemic. Despite in most of the cases (80%) COVID-19 presents mild symptoms, in some cases the admission to intensive care units (ICU) and the use of mechanical ventilation are necessary. These types of “special treatments” are limited and an uncontrolled COVID-19 epidemic could easily overwhelm any health care system. In order to contain the outbreaks that are being developing in the whole globe, the governments are taking drastic measures, locking down whole nations. The understanding of the dynamic of the epidemic could be useful to take decisions in time, preparing hospitals to deal with the infection peak. The ongoing outbreaks are often described with simple fitting models, i.e. exponential functions functions (and/or sigma-shaped curves such as Logistic or Gompertz, this last being the one to be preferably used to fit the data), however in this work, exploiting the similarity between the almost extinguished Chinese epidemic and the relatively new Italian epidemic, a SIRD compartmental model was developed and used to describe and predict the evolution of both the outbreaks. The advantages of SIRD models with respect to heuristic models reside mainly in the physical-based SIRD parameters, which can give useful insight and/or can be properly chosen on the basis of physical-based assumptions. The present SIRD model was able to describe the Chinese outbreak, obtaining the kinetic of infection, the kinetic of recovery as well as the kinetic of death. Several modeling parameters obtained from the fitting of the Chinese epidemic were used without further modification for the Italian outbreak, considering the similarity of actions taken by the governments. Instead the best fit kinetic of infection in Italy was lower than the Chinese infection, probably because the Italian outbreaks are mainly based in small hospitals/cities, whereas the main Chinese outbreak was in the metropolitan city of Wuhan. The Italian epidemic showed instead a higher death kinetic, causing at the end of infection three times the Chinese death (modeling results). In conclusion the developed SIRD model predicts for the Italian epidemic the infection peak between the end of March and the beginning of April 2020, with almost 42,000 people infected at the same time and a number of death people at the end of the epidemic of about 10,000.

References

1. *WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020*. 2020; Available from: https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020.
2. *Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19)*. 2020; Available from: https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf.
3. *Centers for Disease Control and Prevention. Coronavirus symptoms and diagnosis.* 2020; Available from: https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html.
4. Guan, W.-j., et al., *Clinical Characteristics of Coronavirus Disease 2019 in China.* New England Journal of Medicine, 2020.
5. Remuzzi, A. and G. Remuzzi, *COVID-19 and Italy: what next?* The Lancet.
6. Batista, M., *Estimation of the final size of the second phase of the coronavirus epidemic by the logistic model.* medRxiv, 2020: p. 2020.03.11.20024901.
7. Lloyd, A.L. and S. Valeika, *Network models in epidemiology: an overview*, in *Complex population dynamics: nonlinear modeling in ecology, epidemiology and genetics*. 2007, World Scientific. p. 189-214.
8. Brauer, F., P.d. DRIESSCHE, and J. Wu, *Lecture notes in mathematical epidemiology*. 2008: Berlin, Germany: Springer.
9. Bao, L., et al., *Reinfection could not occur in SARS-CoV-2 infected rhesus macaques.* bioRxiv, 2020: p. 2020.03.13.990226.
10. *2019 Novel Coronavirus COVID-19 (2019-nCoV) Data Repository by Johns Hopkins CSSE*. Available from: https://github.com/CSSEGISandData/COVID-19.
11. *The Novel Coronavirus Pneumonia Emergency Response Epidemiology Team. The Epidemiological Characteristics of an Outbreak of 2019 Novel Coronavirus Diseases (COVID-19).* China CDC Weekly, 2020, 2(8): 113-122, 2020.
12. *Population Data by country (Recent years)* Available from: https://apps.who.int/gho/data/view.main.POP2040.
13. Dowd, J.B., et al., *Demographic science aids in understanding the spread and fatality rates of COVID-19.* medRxiv, 2020: p. 2020.03.15.20036293.