



Article

# Gaussian Parameters Correlate with the Spread of COVID-19 Pandemic: The Italian Case

Carmelo Corsaro , Alessandro Sturniolo  and Enza Fazio 

Department of Mathematical and Computational Sciences, Physics Science and Earth Science, University of Messina, Viale F. Stagno D'Alcontres 31, I-98166 Messina, Italy; alessandro.sturniolo1@studenti.unime.it (A.S.); enfazio@unime.it (E.F.)

\* Correspondence: ccorsaro@unime.it

**Abstract:** Until today, numerous models have been formulated to predict the spreading of Covid-19. Among them, the actively discussed susceptible-infected-removed (SIR) model is one of the most reliable. Unfortunately, many factors (i.e., social behaviors) can influence the outcomes as well as the occurrence of multiple contributions corresponding to multiple waves. Therefore, for a reliable evaluation of the conversion rates, data need to be continuously updated and analyzed. In this work, we propose a model using Gaussian functions, coming from the solution of an ordinary differential equation representing a logistic model, able to describe the growth rate of infected, deceased and recovered people in Italy. We correlate the Gaussian parameters with the number of people affected by COVID-19 as a function of the large-scale anti-contagion control measures strength, and also of vaccines effects adopted to reach herd immunity. The superposition of gaussian curves allow modeling the growth rate of the total cases, deceased and recovered people and reproducing the corresponding cumulative distribution and probability density functions. Moreover, we try to predict a time interval in which all people will be infected or vaccinated (with at least one dose) and/or the time end of pandemic in Italy when all people have been infected or vaccinated with two doses.

**Keywords:** coronavirus; growth rate model; statistical modelling; gaussian fitting



**Citation:** Corsaro, C.; Sturniolo, A.; Fazio, E. Gaussian Parameters Correlate with the Spread of COVID-19 Pandemic: The Italian Case. *Appl. Sci.* **2021**, *11*, 6119. <https://doi.org/10.3390/app11136119>

Academic Editors: Johann Michael Köhler and Stefan Wölfl

Received: 26 May 2021  
Accepted: 28 June 2021  
Published: 30 June 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The novel Coronavirus disease (COVID-19) was first reported on 31 December 2019 in Wuhan, Hubei Province, China. Then, it started spreading rapidly across the world and still today, more than one year later, we have to face it. The cumulative incidence of the causative virus (Severe Acute Respiratory Syndrome- CoronaVirus, SARS-CoV-2) is even increasing and has already involved 219 countries and territories of which the most affected in terms of total deaths are USA, Brazil, Mexico, India, UK, Spain, Italy, Russia and France [1]. On 30 January 2020, World Health Organization (WHO) Director-General declared the COVID-19 outbreak a Public Health Emergency of International Concern [2]. Later, on 11 February 2020, WHO named the disease “COVID-19”, indicating explicitly the year of occurrence. As of 24 May 2021, there have been 166,860,081 confirmed cases of COVID-19, including 3,459,996 deaths. Fortunately, vaccines have been already tested and administrated especially to the eldest because they are mostly exposed to COVID-19 issues. As of 24 May 2021, a total of 1,489,727,128 vaccine doses have been administered [3]. The major difference between the pandemic caused by CoV-2 and related coronaviruses, like SARS-CoV and MERS-CoV (Middle East Respiratory Syndrome), is the ability of CoV-2 to spread rapidly through human contact and leave nearly 20% infected subjects as symptom-less carriers [4–6].

Because of the global spreading of COVID-19, numerous mathematical models were developed with the aim to reproduce and predict its potential diffusion [7–9]. Of course, each model tells a different story about the loss of life to come, making it hard to know which could be the most reliable. However, robust models guessing the prognosis of

COVID-19 are urgently needed to support decisions about hospital admission and population level interventions [10,11]. Today, all the proposed models emphasize the fact that it is not possible to separately predict either the probability of getting severe COVID-19 disease or of dying from, owing primarily to incomplete knowledge of who actually has the disease. The risk of getting COVID-19 depends on the individual's behavior and the local dynamics of the disease which evolve continuously. It is also worth mentioning how some individuals' emotions can be influenced by random factors (web searches, social media interactions and the corresponding spread of fake news) [12]. Hence, an adaptive tuning of the model is expected to be repeated periodically. Furthermore, the discovery of new virus variants as well as a shift in the age distribution of cases towards younger people, strongly limit the accuracy of possible predictions [13]. This occurrence is provoked by the depletion of older susceptible individuals, either because they have already been infected or have been (preferentially) vaccinated. A further challenge is given by local interventions that governments adopt to reduce the virus spreading at the time the model was developed. Indeed, a very reliable model should include too many uncontrolled variables to predict the outcomes of a specific decisions (e.g., restaurants opening).

Predictive mathematical models often make use of statistical approaches. Data fitting with an analytically defined function allows extrapolating a possible future trend without any knowledge of the process driving the data trend. For example, a thorough forecasting analysis of epidemics data from several countries of the world has been performed by Valvo with a bimodal lognormal distribution [14]. Therefore, only short-term forecasts and without correlations can be generally performed. On the other hand, complex and innovative strategies, making use of machine learning or regression, have been implemented with the aim to project SARS-CoV-2 cases into the future [15,16]. Mechanistic models based on differential equations, such as the well-known SIR model (Susceptible–Infectious–Recovered) and its numerous modifications, are the most used to mimic and predict the spread of COVID-19 by including specific assumptions on the chosen parameters which control transmission, disease, testing capacity and immunity [17–20]. These models, on the contrary of crude statistical approaches, are able to consider nonlinear dependences such as the increasing diffusion speed with the number of infected people. On the other hand, our poor knowledge of this novel virus, as well as the different variants met in different countries, limits model accuracy. Note that the highest source of uncertainty influencing all models lies in the number of really infected people. In fact, available data depend on the number of performed virologic tests; so a certain number of infected (but not tested) people is not considered (essentially asymptomatic people). This issue is called *under-reporting* and it has been proposed to model its process in order to avoid an underestimation of the infection rate [21,22] and therefore to reduce the model uncertainty also considering the heterogeneous (in space and time) distribution of the acquired data [11]. Hence, hospitalizations and deaths could be considered more reliable data although, from one side, they may underestimate disease burden and, from the other side, correlations with the cases in the community should be unveiled [10]. Thus, until today, COVID-19 pandemic evolution is very difficult to predict because many factors (i.e., social behaviors) are involved as well as the occurrence of multiple contributions corresponding to multiple “waves”. Summarizing, two main modelling approaches are being used today: (i) time series regression analysis furnishes reliable estimate only on a short time window. (ii) SIR model or its modifications that, by using coupled differential equations, justify inter-conversions between the different fraction of populations considered within the model. One of the most extended SIR models is called SIDARTHE and includes inter-conversions between susceptible (S), infected (I), diagnosed (D), ailing (A), recognized (R), threatened (T), healed (H) and extinct (E) [18]. However, for a reliable evaluation of the conversion rates, data need to be continuously analyzed over varying time windows for improved predictive power during virus spreading and evolution [23]. Unfortunately, predicting the next wave is really a hard task, even for mechanistic models, that by employing constant “conversion rates” are indeed unimodal. This would require identifying a change (likely increase) in the relevant rate.

In this work, we propose a model using a Gaussian mixture to describe the growth rate of infected, deceased and recovered people. Generally, even if it is a simple approach, Gaussian fitting procedure has been largely debated [24–27] to analyze epidemic trend given the low accuracy of Gaussian process approximations to non-linear stochastic individual-based models of epidemics. Nevertheless, in light of some measures currently in use to judge the lockdowns or to prevent spreading of the virus, Gaussian modelling provides exact and simple approximate relationships between the two relevant parameters of the function (peak time and width).

Specifically, this manuscript aims to discuss in detail the mutual relations between the growth rate of infected, deceased and recovered people as a function of the large-scale anti-contagion control measures strength, and also of vaccines effects adopted to reach herd immunity. We take into account some relevant aspects such as the slowdown and trend inversion in infected growth for days close to the observed infection peaks, and/or societal system resilience. Indeed, it is found that the infection rate data show a characteristic rapid growth to a peak value followed by a slower decline if countermeasures are employed/deployed effectively. Moreover, our Gaussian model allows reproducing the probability density function (PDF) and cumulative distribution function (CDF) of infected, deceased and recovered people. Finally, the comparison between infections and vaccines data provides an estimation of the time interval to achieve the herd immunity and remove restrictions. The paper is organized as follows. In Section 2 the proposed Gaussian model is explained and analyzed with specific attention to its sensitivity with respect to the herd immunity. In Section 3 the achieved results are shown in three separated subsections. The first deals with the growth rate analysis of infected, deceased and recovered people performed with the adopted model. The second subsection concerns the correlation analysis between raw variables and the obtained fitting parameters. Finally, the third subsection reports the statistical approach in terms of cumulative and density functions that our model also allows to obtain. At the end, in Section 4 conclusions and future perspectives are drawn also by considering vaccinations data.

## 2. The Gaussian Growth Rate Model

The classical logistic model corresponds to the second-order approximation of the epidemic curve growth described by the original Kermack–McKendrick SIR model [28–30], which is indeed a mechanistic compartmental model aimed at describing the dynamics of an epidemics in a population. In classical SIR model, the population is assumed to be constant over time, so that the following equality holds for each time  $t$ :

$$S(t) + I(t) + R(t) = N \quad (1)$$

although deaths caused by the disease itself should be considered as well. However, we note that in the Italian case we are interested in, there have been 125,335 deaths as of 24 May 2021 [31] over an initial population of 59,641,488 reported in 2019, just before the pandemic emergency was declared. Therefore, we will assume the Italian population to be approximately constant, as the decrease caused by COVID-19 pandemic was less than 1% (0.2%) of the initial population. We can translate these words into a system of non-linear ordinary differential equations (ODEs) [32]:

$$\frac{dS}{dt}(t) = -\alpha(t)S(t)I(t) \quad (2)$$

$$\frac{dI}{dt}(t) = \alpha(t)S(t)I(t) - \beta(t)I(t) \quad (3)$$

$$\frac{dR}{dt}(t) = \beta(t)I(t) \quad (4)$$

where  $\alpha(t)$  and  $\beta(t)$  are defined as transmission ( $\alpha(t)$ ) and removing (i.e., death or recovery,  $\beta(t)$ ) rates [33]. This system should be accompanied by suitable initial conditions  $S(0) > 0$ ,

$I(0) > 0$  and  $R(0) \geq 0$  [32]. The first condition requires the susceptible number  $S$  to be non-zero at initial time  $t = 0$ , because if it were, the disease could not find a vulnerable population to spread through. The second condition states that the initial number of infected should be non-zero too, or else there would not be any epidemics outbreaks.

Despite its shortcomings in terms of data availability, SIR model provides a simple, intuitive and effective description of epidemics in a population. Herein we propose a model based only on one equation of the SIR model (Equation (2)).

We propose a Gaussian mixture able to describe the growth rate of infected, deceased and recovered people in Italy. We correlate the Gaussian parameters with the number of people affected by COVID-19 as a function of the large-scale anti-contagion control measures strength, and also of vaccines effects adopted to reach herd immunity. In detail, we consider a Gaussian mixture to model the growth rate  $r_{Est}(t)$  of the total cases  $I_{tot}(t)$ , deceased  $E(t)$  and recovered  $R(t)$  (estimated from National Health Institute of Italy (it. *Istituto Superiore di Sanità*, or ISS)) daily data reports in json format [31]. Data processing and analysis have been performed by using a custom build algorithm with MATLAB (2018a, MathWorks, Natick, MA, USA). As we will detail in the next, the growth rate  $r_{Est}(t)$  can be expressed by the following relation, according to a generalization of classical logistic models [28,34]

$$r_{Est}(t) \equiv \frac{dy(t)/dt}{y(t) \left[1 - \frac{y(t)}{N}\right]} \quad (5)$$

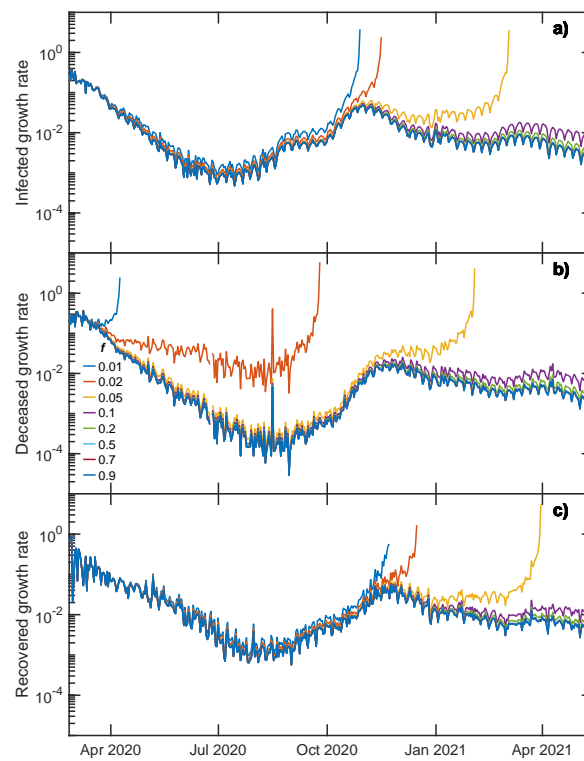
where  $y$  in our case indicates infected, deceased or recovered people,  $N$  is the number of considered people and  $t$  is time. We will show that the estimated growth rate  $r_{Est}(t)$  can be described, and fitted to, by a Gaussian mixture, from which the name of our model: Gaussian Growth Rate (GGR) model.

First, we assume that only a sub-population  $N_H$  is considered as susceptible rather than the total population  $N$ . It is worth to mention that  $N_H$  in SIR models coincides with the so-called attack size, i.e., the total number of infected individuals. Since the logistic model is governed by two parameters, the growth rate and the saturation (asymptote),  $N_H$  is related with the asymptotic value through:

$$S(t) + I_{tot}(t) = N_H \quad (6)$$

where  $I_{tot}$  corresponds to the cumulative total cases (i.e., the total number of infected people). Then we assume that the ratio  $f$  between  $N$  and  $N_H$  may be connected with the long-debated herd immunity ( $N_H = fN$ ). In fact,  $f$  corresponds to the portion of total population which must be immunized for the epidemic to die out. As reported by WHO, both a good estimate of the numerical value of  $f$  and the achievability of herd immunity has been questioned [35]. So, in Figure 1 the time trends of infected, deceased and recovered growth rate are shown for different values of the ratio  $f$ . A clear divergence emerges for  $f < 0.1$  while trends remain almost unchanged for  $f > 0.1$ . Moreover, the obtained result is related with the considered fraction of the population being infected over the total (around 7% [36]), which further justifies the no relevant changes found for  $f > 0.1$ .

In the following, we are going to set  $f = 0.7$  (70% herd immunity), a proportion which has been frequently cited as a likely immunity threshold [37,38], especially in the early days of vaccination campaign. We should keep in mind, however, that there are several variables to be considered, such as social interactions in reopening scenarios and variations of transmission rate for different virus strains, which may alter this threshold value [39].



**Figure 1.** (a) Infected, (b) deceased and (c) recovered growth rate (see Equation (5) for definition)  $r(t)$  vs.  $t$  (days) in a log-lin plot for different values of herd immunity ratio:  $f = 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 0.7, 0.9$ . While  $r(t)$  shows a clear divergence for  $f < 0.1$ , it does not change significantly for  $f > 0.1$ .

As for infected people, we may set their number  $I(t)$  as equal to a time-dependent fraction of total cases:

$$I(t) = \eta(t)I_{tot}(t) \tag{7}$$

where  $0 \leq \eta(t) \leq 1$  for each  $t$ . Therefore, we obtain from previous equations:

$$\frac{dS}{dt}(t) = -\alpha(t)S(t)I(t) = -\alpha(t)\eta(t)S(t)I_{tot}(t) \tag{8}$$

$$\frac{dI_{tot}}{dt}(t) = N_H\alpha(t)\eta(t)I_{tot}(t) \left[ 1 - \frac{I_{tot}(t)}{N_H} \right] \tag{9}$$

Furthermore, we use the Italy case-fatality ratio (CFR), which is the ratio (that we call  $m$ ) between the number of deceased  $E(t)$  with respect to the total cases  $I_{tot}$ :

$$E(t) = mI_{tot}(t) \tag{10}$$

$$\frac{dE}{dt}(t) = N_H\alpha(t)\eta(t)E(t) \left[ 1 - \frac{E(t)}{mN_H} \right] \tag{11}$$

Note that, the CFR value is estimated to be of 3.0% by the Johns Hopkins University of Medicine [36] from available data on COVID-19 death toll. We have chosen to use CFR and not IFR value (COVID-19 infection fatality ratio) since estimating IFR is challenging. In fact, it would require to know the real number of total cases and deceased, which may have been underestimated in times when Italy’s testing capacity was limited and the pressure on the healthcare system was high [40]. Data currently available provide information only about confirmed cases, possibly underestimating real data, as De Natale et al. pointed out already in May 2020 [41]. Indeed, given the relatively high number of tests performed so far, the use of CFR instead of IFR for  $m$  seems a correct assumption.

Likewise, we can set the recovered number  $R(t)$  as proportional to  $I_{tot}(t)$  by a ratio  $h$ :

$$R(t) = hI_{tot}(t) \tag{12}$$

$$\frac{dR}{dt}(t) = N_H\alpha(t)\eta(t)R(t) \left[ 1 - \frac{R(t)}{hN_H} \right] \tag{13}$$

The ratio  $h$  is in general a time-dependent value. However, for our analysis it is sufficient to consider its limit for large times, that is:

$$h = 1 - m \tag{14}$$

because for  $t \rightarrow +\infty$  all infected people will have either died or recovered, whereas the active cases will vanish. So, the problem can be described by a logistic ODE. We may further define the function  $r(t)$ , referred to the growth rate:

$$r(t) \equiv N_H\alpha(t)\eta(t) \tag{15}$$

As one can easily verify by separation of variables, the solution of Equations (9), (11) and (13) can be written in the following general form:

$$y(t) = \frac{L}{1 + \left( \frac{L-y(0)}{y(0)} \right) e^{-\int_0^t r(t')dt'}} \tag{16}$$

where  $y(t) \equiv I_{tot}(t), E(t), R(t)$  and correspondingly  $L \equiv N_H, mN_H, hN_H$  is to be interpreted as the asymptotic value of  $y(t)$ . As for the functional form of the growth rate  $r(t)$ , it is well known that it is constant in classical logistic problems [42]. This makes the mathematical difficulties of integrating  $r(t)$  over time  $t$  less daunting, but as we will see, real data suggest us to consider non-negligible time variations of growth rate. Hence, we will assume  $r(t)$  to be a function of time, whose form is to be derived from data analysis.

To get an idea of the mathematical form of growth rate  $r(t)$ , we observe from Equations (7) and (15) that it increases and decreases as the number of active cases (i.e., infectious people). On this theoretical basis, firmly rooted in classical SIR model mathematics, we assume for  $r(t)$  daily variations the same feedback mechanism characterising daily variations of infected number  $I(t)$ , that is  $dr(t)/dt \sim r(t)$ . Then, the simplest analytical form to model  $r(t)$  is a Gaussian function:

$$Ae^{-\frac{1}{2}\left(\frac{t-\tau}{\sigma}\right)^2} \tag{17}$$

which is centered at a time  $\tau$  (centroid).  $\sigma$  represents, as well known, the standard deviation related to the full width at half maximum of each Gaussian, while  $A$  is the maximum height of the curve. As we would expect from a growth rate, a single curve does not blow up to infinity, and its derivative is indeed proportional to the Gaussian curve itself by a factor  $-(t - \tau)/\sigma^2$ . In practice, the constantly evolving dynamics of the pandemic has alternated phases of slower diffusion with periods of rapid growth, corresponding to several observable peaks in growth rate data. Therefore, a Gaussian mixture, plus a constant background ( $Bkg$ ) accounting for all random events, is needed in order to fit all available data:

$$r(t) = Bkg + \sum_{i=1}^n A_i e^{-\frac{1}{2}\left(\frac{t-\tau_i}{\sigma_i}\right)^2} \tag{18}$$

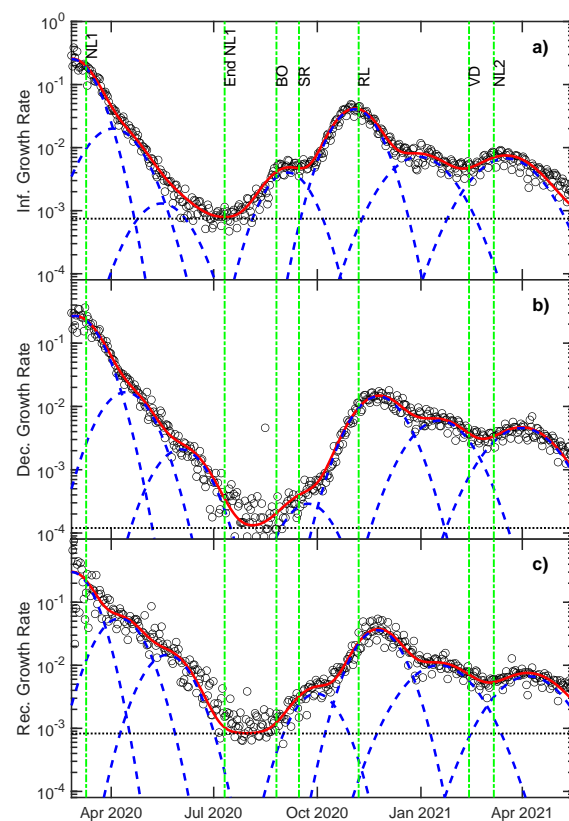
### 3. Results and Discussion

#### 3.1. Growth Rate Analysis

To extract significant information from the epidemics data, we should perform a careful analysis of growth patterns in daily and total numbers of confirmed cases and deceases, looking for details about growth maxima positions and widths for both infected



and deceased, and how they correlate with each other. For this purpose, we have fitted growth rate  $r(t)$  data (Figure 2) of infected (panel (a)), deceased (panel (b)) and recovered (panel (c)) people in Italy since the beginning of the pandemic (26 February 2020 for first official case in Italy) up to 24 May 2021, by using seven Gaussian bands (dashed blue lines) plus a constant background (Equation (18)), according to the GGR model. The constant background assumption means that random contagion events are uniformly distributed in time regardless of the general epidemics data trend, in a consistent way with respect to the uncertainties and incompleteness of available data. This leaves a certain margin of error to undetected cases and the resulting, seemingly random, contagion events [11,43]. The lag in the growth rate of deceased and recovered with respect to infected people is intrinsically included in the theory by means of the Gaussians centroid. In fact, we considered that only the first Gaussian band for infected people is centered at zero (days). We will enter more in details in the next section.



**Figure 2.** (a) Infected, (b) deceased and (c) recovered growth rate fitted by a Gaussian mixture (Equation (18)) with seven components (blue dashed lines) plus a constant background (black dotted line). The red solid line corresponds to the total fit and the vertical green dashed lines mark important dates as described in the text.

For the chosen modelling, we have selected 7 Gaussian components after an optimization procedure that took into account the use of the same standard deviation for all Gaussian bands (within an interval of 15–25 days in order to maintain a comparable spread for the considered events) and the lowest  $\chi^2$  corresponding to a statistical significance of the fitting parameters ( $p < 0.05$ ). Then we have focused our attention on the time and intensity of the events (respectively, the centroid and amplitude of the Gaussian functions). The width of the last Gaussian band reaches the highest bound value because the end of data is still not definitive. Our results indicate that deceased growth rate peaks are shifted from 3 (only the first peak) up to 21 days ahead of corresponding infected peaks. This is overall consistent with the hospitalization time intervals for severe Covid-19 cases [44,45] and also holds for the time shifts between recovered and infected contributions, ranging

from less than 1 day for the first peak up to 25 for the fourth one. The uncertainties obtained after the fitting procedures are less than 10 days for all centroids, less than 5 days for all standard deviations and of few percent for all Gaussian amplitudes.

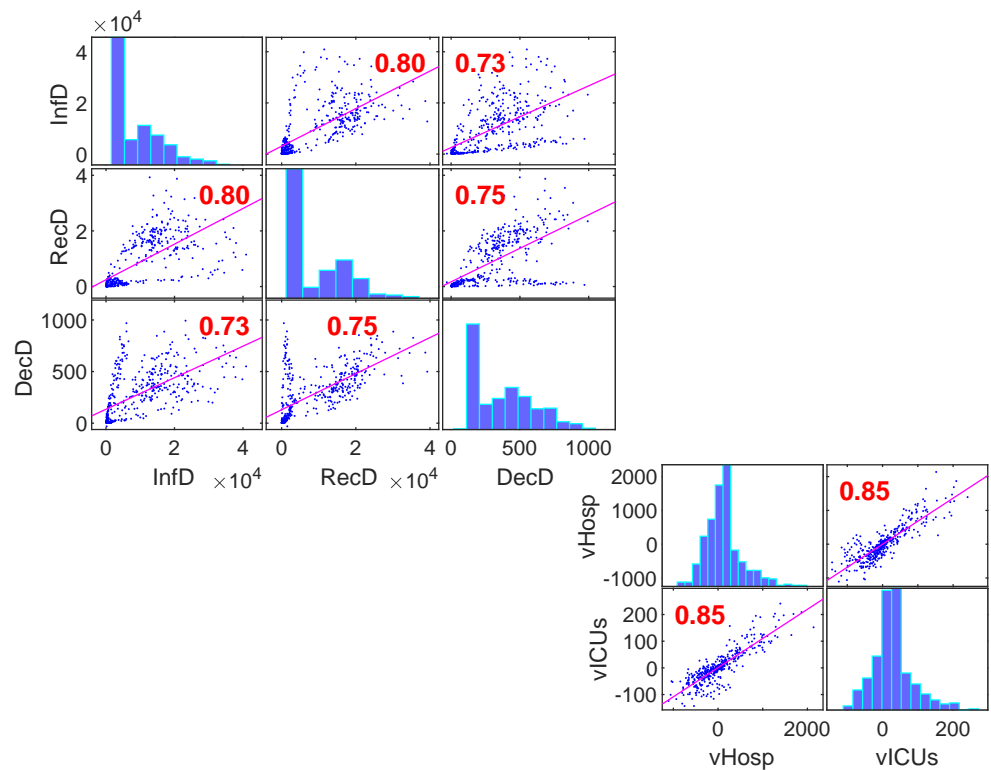
The vertical green dashed lines reported in Figure 2 remark the date of important events in the timeline of Covid-19 pandemic evolution in Italy [46], useful to understand the implication of relevant episodes (e.g., the application of tight measures for anti-contagion control and the discovery of new dangerous virus variants) [47]. First of all, the national-scale lockdown (*NL1*) declared on 9 March 2020 to contain the early spread of Covid-19 and its official ending (*End NL1*) on 10 July 2020. In these days, the infected growth rate reached a minimum in Italy; however, we would like to point out that, in the same period, pandemic was far from slowing down on a global scale, as the number of daily cases was growing up to a plateau of ~250,000 daily cases in August 2020 [1]. On 25 August 2020 a local, yet significant, outbreak started in a few touristic sites in Sardinia island, in particular at Billionaire Club in Porto Cervo (*BO*). A peak of the infected growth rate curve has been detected a few days after this event, as can be seen in Figure 2. Then, another steep increase in infected, deceased and recovered growth rate followed the schools reopening on 14 September 2020 (*SR*). This caused a new regional-scale lockdown on 6 November 2020 (*RL*), after that the growth rate started to decrease steadily. However, the detection of new variants (most importantly B.1.1.7, *VD*), in February 2021 marked a new growth rate increase in our dataset which, once again, ended a few weeks after another national lockdown was declared on 6 March 2021 (*NL2*). Finally, we outline that total fits, represented by the solid red lines in Figure 2, have been obtained with an  $R^2$  of 0.97 for infected and deceased data (panels a and b, respectively) and 0.77 for recovered data (panel c). This relatively small value is due to the data acquired in the first days that are very noisy.

### 3.2. Statistical Analysis—Correlations

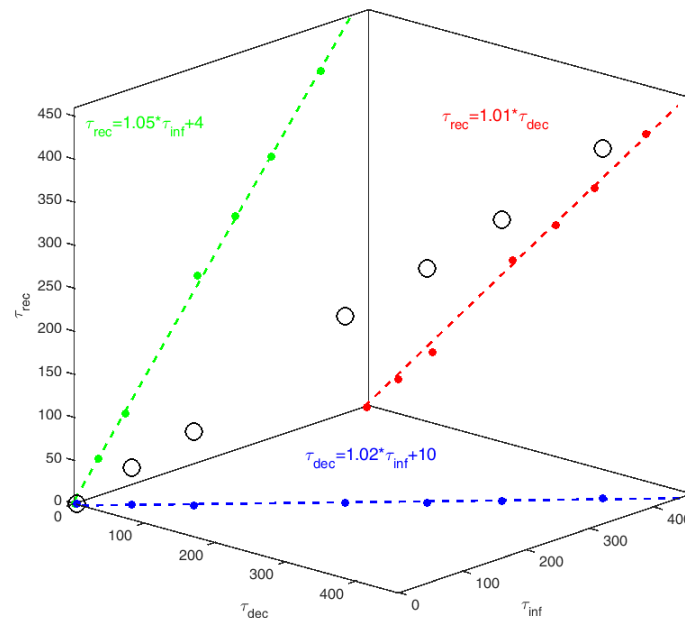
A statistical analysis was carried out with the aim to find specific correlations between raw data corresponding to daily new counts for infected, deceased and recovered people and daily variations of hospitalized and admitted to intensive care (ICUs). Figure 3 reports the correlation matrix of these variables.

Spearman's rank correlation coefficient was used to ascertain the strength and direction of the relationship between two variables (off-diagonal plots in Figure 3). This statistical analysis was again performed using MATLAB and red values correspond to a statistically significant level with  $p < 0.05$ . It emerges that the daily number of deceased and recovered people shows a strong positive correlation with that of the infected people of 0.80 and 0.73, respectively. This trivially confirms that both deceased and recovered people increase with infected. Moreover, we observe a corresponding strong positive correlation of 0.75 between the number of daily deceased and recovered people and, an even stronger positive correlation of 0.85 (easily understandable) between daily variations of hospitalized and ICUs. Unfortunately, the exact mechanisms underlying this association are uncertain since several possible reasons may determine these trends. However, in order to at least find a correlation between the time of the events considered crucial and the change in the growth rate of infected, deceased and recovered people, we analyze the centroids values with respect to each other. The tridimensional graph, plotted in Figure 4, shows how peak centroids for all three growth rates are linearly dependent with an average offset of 10 days between infected and deceased peaks and 4 days between recovered and infected peaks. These values correspond to the above mentioned average lag between the growth rate of recovered and deceased people with respect to that of infected people. Unsurprisingly, the obtained angular coefficients of the straight lines vary in a range of values very close to 1. This confirms our previous observation of a correspondence between infected growth rate peaks and deceased or recovered growth rate peaks shifted forward in time.





**Figure 3.** Scatter plot matrix between daily data for infected, deceased and recovered people (top left), and for daily variations of hospitalized and admitted to intensive care (ICUs) (bottom right). The results of Spearman test are shown in the off-diagonal plots whereas the diagonal panels report the single variables histograms.



**Figure 4.** Tridimensional representation of the correlation between the timing of the events corresponding to the centroids of the Gaussian mixture used to fit infected, deceased and recovered people. Dashed lines correspond to linear fits between the corresponding variables.

### 3.3. Statistical Analysis—Cumulative and Density Functions

In order to numerically compute the number of total cases, deceased and recovered as a function of time, we substituted the obtained analytical fit curves  $r(t)$  (red lines plotted in Figure 2) in Equation (16) for each data:

$$I_{tot}(t) = \frac{I_{\infty}}{1 + \left(\frac{I_{\infty}-I_{tot}(0)}{I_{tot}(0)}\right)e^{-\int_0^t r(t')dt'}} \tag{19}$$

$$E(t) = \frac{E_{\infty}}{1 + \left(\frac{E_{\infty}-E(0)}{E(0)}\right)e^{-\int_0^t r(t')dt'}} \tag{20}$$

$$R(t) = \frac{R_{\infty}}{1 + \left(\frac{R_{\infty}-R(0)}{R(0)}\right)e^{-\int_0^t r(t')dt'}} \tag{21}$$

Herein  $I_{\infty}$ ,  $E_{\infty}$  and  $R_{\infty}$  represent the asymptotic limits of total cases, deceased and recovered for very large times. According to our model, these values are not independent. Instead, we have assumed a simple proportionality relation between deceased, recovered and infected (see Equations (10), (12) and (14)). We noticed that the normalized curves  $I_{tot}(t)/I_{\infty}$ ,  $E(t)/E_{\infty}$  and  $R(t)/R_{\infty}$  can be considered, mathematically speaking, as cumulative distribution functions (CDFs)  $F(t)$ , since they both are monotonic increasing and their output value for large times equals 1. Therefore, we interpret such functions as describing the probability of an entire subpopulation  $I_{\infty}(E_{\infty},R_{\infty})$  having been infected (being deceased or recovered) by day  $t$  so they can be directly compared with the total cases (total infected, deceased and recovered people), normalized at the value of the last considered day (24 May 2021 in our case).

$$F_I(t) \equiv \frac{I_{tot}(t)}{I_{\infty}} = \frac{1}{1 + \left(\frac{I_{\infty}-I_{tot}(0)}{I_{tot}(0)}\right)e^{-\int_0^t r(t')dt'}} \tag{22}$$

$$F_E(t) \equiv \frac{E(t)}{E_{\infty}} = \frac{1}{1 + \left(\frac{E_{\infty}-E(0)}{E(0)}\right)e^{-\int_0^t r(t')dt'}} \tag{23}$$

$$F_R(t) \equiv \frac{R(t)}{R_{\infty}} = \frac{1}{1 + \left(\frac{R_{\infty}-R(0)}{R(0)}\right)e^{-\int_0^t r(t')dt'}} \tag{24}$$

where:

$$F_{I,E,R}(t) \equiv P[y(t') = y_{\infty} | t' \leq t] \tag{25}$$

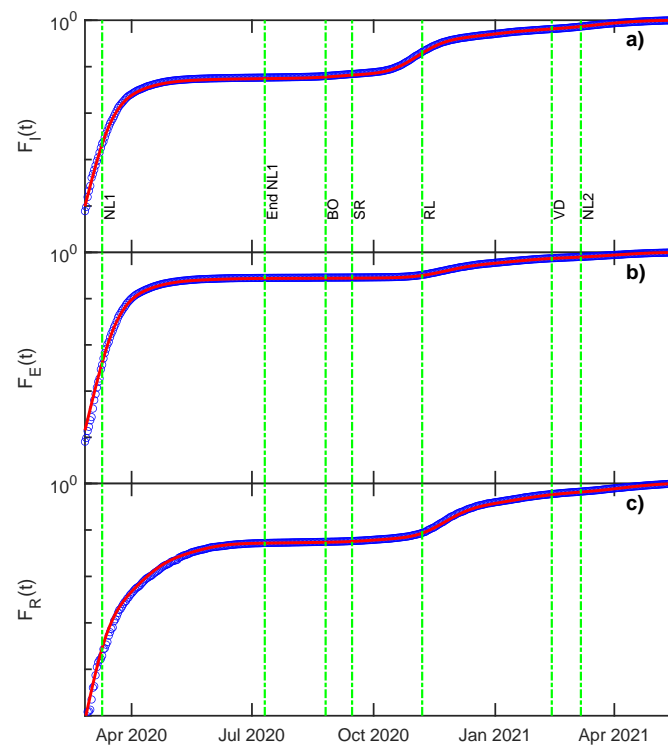
and for each case we can set  $y(t) \equiv I_{tot}(t), E(t), R(t)$ . The corresponding probability density functions (PDFs) are then defined:

$$p_I(t) \equiv \frac{dF_I}{dt}(t) = \frac{1}{I_{\infty}} \frac{dI_{tot}}{dt}(t) \tag{26}$$

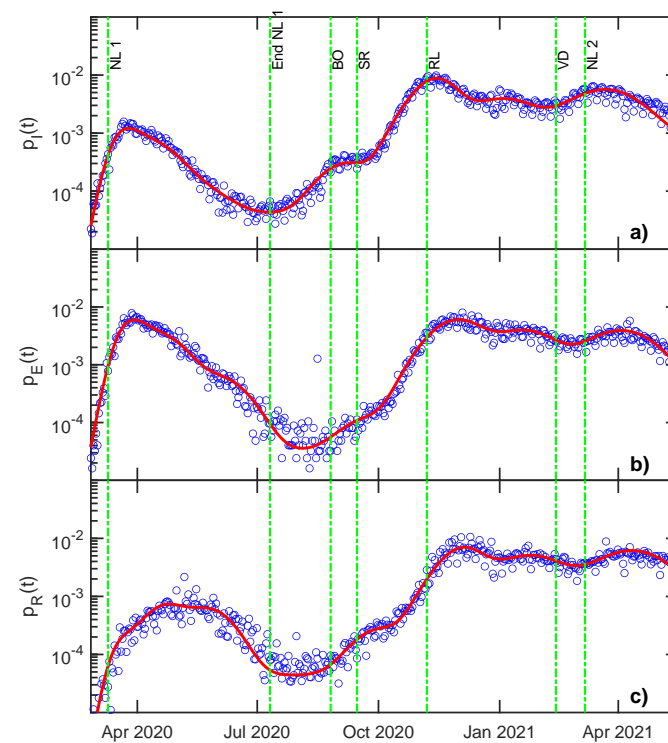
$$p_E(t) \equiv \frac{dF_E}{dt}(t) = \frac{1}{E_{\infty}} \frac{dE}{dt}(t) \tag{27}$$

$$p_R(t) \equiv \frac{dF_R}{dt}(t) = \frac{1}{R_{\infty}} \frac{dR}{dt}(t) \tag{28}$$

and can be directly compared with the new daily counts provided for these variables. Having interpreted  $F(t)$  as the probability of total cases, deceased and recovered number reaching their maxima by day  $t$ , the corresponding PDFs  $p(t)$  are the probabilities of reaching these maxima on a day  $t$ . PDFs could provide useful information about future massive outbreaks in the short-term, since they are usually anticipated by a significant increase of growth rate  $r(t)$  and of PDF  $p(t)$ . After computing  $F_{I,E,R}(t)$  and  $p_{I,E,R}(t)$  by fitting  $r_{Est}(t)$  with our GGR model, we compared raw and modeled data. The corresponding behaviors are shown in Figures 5 and 6.



**Figure 5.** (a) Infected, (b) deceased and (c) recovered cumulative distribution functions. Blue open circles correspond to raw data, while the red solid lines represent model curves (Equations (22)–(24)).



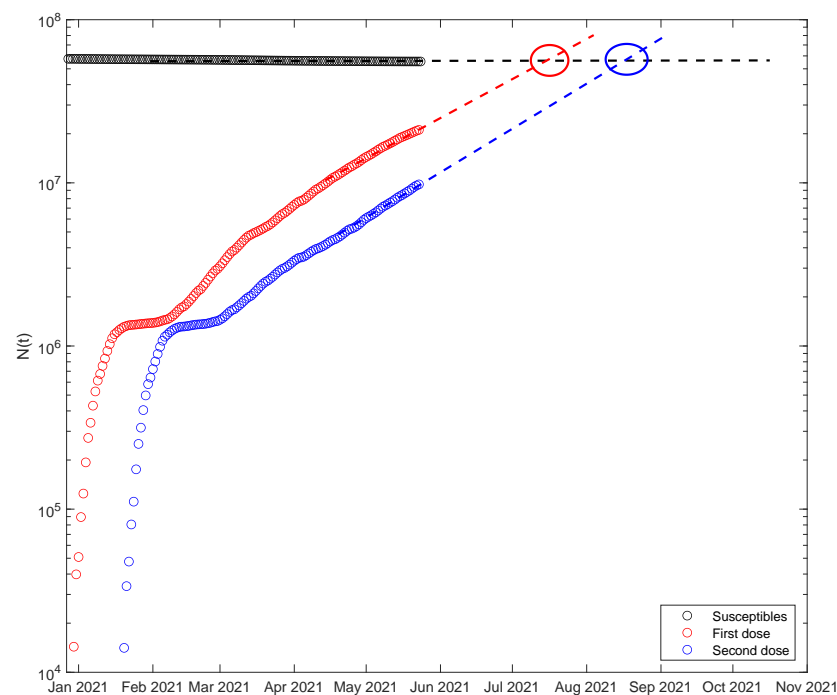
**Figure 6.** (a) Infected, (b) deceased and (c) recovered probability distribution functions. Blue open circles correspond to raw data, while the red solid lines represent model curves (Equations (26)–(28)).

We underline that raw data series have been estimated by normalizing cumulative infected, deceased and recovered by their latest values reported in Italian Ministry of Health data, while daily variations of the same quantities have been normalized by the sum of all daily increments. It is noteworthy that our GGR model is able to perfectly describe

both cumulative and distribution functions of infected, deceased and recovered people. Although we use the sum of relatively simple functions (Gaussians) for fitting the growth rate of those people characterizing the pandemic spread, this allows to replicate also daily and cumulative data with a very good precision. This endows the GGR model with a short-term predictive power especially in terms of PDFs.

#### 4. Conclusions and Future Perspectives

A simple model constituted by a Gaussian mixture, defined taking into account anti-contagion control measures, has been used to fit the growth rate of COVID-19 affected, deceased and recovered people in Italy. The statistical correlation between considered pair of variables allowed us inferring the variation of the virus diffusion in correspondence of the effect of well identified events during the pandemic spread in Italy. Furthermore, our GGR model allowed to reproduce the cumulative distribution functions of total cases, deaths and recovers as well as their daily distribution in terms of probability density functions, giving information about the statistical distribution of events and conferring a short-term predictive power to the model that is unfortunately limited to few days. However, in terms of survival analysis, an interesting prediction has been obtained. To this purpose we compared the trend of susceptible and vaccinated people in order to establish a time interval within which all people will be infected or vaccinated (with at least one dose) and/or the definitive time of the end of pandemic in Italy when all people will be infected or vaccinated after the two doses. Figure 7 shows on a log-lin scale the number of susceptible people (black squares, obtained by subtracting the cumulative cases from the total of the population) and those of vaccinated people (red and blue symbols for people vaccinated with a single or double doses, respectively) as a function of time. By a simple data extrapolation, we can locate the first mentioned time interval around the mid of July and the definitive end of pandemic in Italy one month later, around the mid of August, if the vaccination rate remains the same.



**Figure 7.** Semi-logarithmic plot of the intersection points between trends of Susceptible and vaccinated people with first (red data) and second (blue data) doses (data available from [48]). Dotted lines correspond to data extrapolation for evaluating the intersection that may mark the end of the pandemic in Italy.

Finally, for a complete description of the whole pandemic until the foreseen end (mid of August), our model predicts the occurrence of one, two at most, additional soft events during which we must cohabit with the virus aiming to achieve the herd immunity. Our estimation for new soft events refers to the time needed to reach the herd immunity if the vaccination rate remains unchanged. If so, only a low-intensity additional Gaussian band would be needed to fit the data up to the mid of August. This prediction cannot consider possible effects due to new virus variants or other uncontrolled events. Hence, in the near future, we will continue monitoring the virus spreading and running our model to confirm its actual prediction and to further test its ability to reproduce COVID-19 data including any other unexpected circumstances.

**Author Contributions:** Conceptualization, C.C., A.S. and E.F.; methodology, C.C. and E.F.; data curation, A.S.; writing—review and editing, C.C., A.S. and E.F. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Publicly available datasets were analyzed in this study. Data have been lastly accessed on 24 May 2021 and can be found here: <https://raw.githubusercontent.com/pcm-dpc/COVID-19/master/dati-json/dpc-covid19-ita-andamento-nazionale.json> and <https://raw.githubusercontent.com/italia/covid19-opendata-vaccini/master/dati/somministrazioni-vaccini-summary-latest.json>.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

COVID-19	2019 Coronavirus Disease
SIR	Susceptible-Infected-Recovered
ODE	Ordinary differential equation
ISS	National Health Institute of Italy
GGR	Gaussian Growth Rate
WHO	World Health Organization
CFR	Case-Fatality Ratio
IFR	Infection Fatality Ratio
ICU	admitted to Intensive Care
CDF	Cumulative Distribution Function
PDF	Probability Density Function

## References

1. COVID Live Update. Reported Cases and Deaths by Country or Territory. Available online: <https://www.worldometers.info/coronavirus/#countries> (accessed on 24 May 2021).
2. WHO. A Year without Precedent: WHO's COVID-19 Response. Available online: <https://www.who.int/news-room/spotlight/a-year-without-precedent-who-s-covid-19-response> (accessed on 24 May 2021).
3. WHO. WHO Coronavirus (COVID-19) Dashboard. Available online: <https://covid19.who.int/> (accessed on 24 May 2021).
4. Zhu, Z.; Lian, X.; Su, X.; Wu, W.; Marraro, G.; Zeng, Y. From SARS and MERS to COVID-19: A brief summary and comparison of severe acute respiratory infections caused by three highly pathogenic human coronaviruses. *Respir. Res.* **2020**, *21*, 224. [[CrossRef](#)]
5. Abdelrahman, Z.; Li, M.; Wang, X. Comparative Review of SARS-CoV-2, SARS-CoV, MERS-CoV, and Influenza A Respiratory Viruses. *Front. Immunol.* **2020**, *11*, 2309. [[CrossRef](#)]
6. Soufi, G.J.; Hekmatnia, A.; Nasrollahzadeh, M.; Shafiei, N.; Sajjadi, M.; Irvani, P.; Fallah, S.; Irvani, S.; Varma, R.S. SARS-CoV-2 (COVID-19): New Discoveries and Current Challenges. *Appl. Sci.* **2020**, *10*, 3641. [[CrossRef](#)]
7. Vytla, V.; Ramakuri, S.K.; Peddi, A.; Srinivas, K.K.; Ragav, N.N. Mathematical Models for Predicting Covid-19 Pandemic: A Review. *J. Phys. Conf. Ser.* **2021**, *1797*, 012009. [[CrossRef](#)]
8. Anirudh, A. Mathematical modeling and the transmission dynamics in predicting the Covid-19—What next in combating the pandemic. *Infect. Dis. Model.* **2020**, *5*, 366–374. [[CrossRef](#)]



9. Adiga, A.; Dubhashi, D.; Lewis, B.; Marathe, M.; Venkatramanan, S.; Vullikanti, A. Mathematical Models for COVID-19 Pandemic: A Comparative Analysis. *J. Indian Inst. Sci.* **2020**, *100*, 793–807. [CrossRef] [PubMed]
10. Holmdahl, I.; Buckee, C. Wrong but Useful—What Covid-19 Epidemiologic Models Can and Cannot Tell Us. *N. Engl. J. Med.* **2020**, *383*, 303–305. [CrossRef]
11. Bertuzzo, E.; Mari, L.; Pasetto, D.; Miccoli, S.; Casagrandi, R.; Gatto, M.; Rinaldo, A. The geography of COVID-19 spread in Italy and implications for the relaxation of confinement measures. *Nat. Commun.* **2020**, *11*, 4264. [CrossRef]
12. Ortega, J.M.; Bernabé-Moreno, J. Modelling the Degree of Emotional Concern: COVID-19 Response in Social Media. *Appl. Sci.* **2021**, *11*, 3872. [CrossRef]
13. Sperrin, M.; McMillan, B. Prediction models for covid-19 outcomes. *BMJ* **2020**, *371*. [CrossRef]
14. Valvo, P.S. A Bimodal Lognormal Distribution Model for the Prediction of COVID-19 Deaths. *Appl. Sci.* **2020**, *10*, 8500. [CrossRef]
15. Syeda, H.B.; Syed, M.; Sexton, K.W.; Syed, S.; Begum, S.; Syed, F.; Prior, F.; Yu, F., Jr. Role of Machine Learning Techniques to Tackle the COVID-19 Crisis: Systematic Review. *JMIR Med. Inform.* **2021**, *9*, e23811. [CrossRef] [PubMed]
16. Alvarez, M.; Gonzalez, E.; Trujillo de Santiago, G. Modeling COVID-19 epidemics in an Excel spreadsheet to enable first-hand accurate predictions of the pandemic evolution in urban areas. *Sci. Rep.* **2021**, *11*, 4327. [CrossRef] [PubMed]
17. Bertozzi, A.L.; Franco, E.; Mohler, G.; Short, M.B.; Sledge, D. The challenges of modeling and forecasting the spread of COVID-19. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 16732–16738. [CrossRef]
18. Giordano, G.; Blanchini, F.; Bruno, R.; Colaneri, P.; Filippo, A.; Matteo, A.; Colaneri, M. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nat. Med.* **2020**, *26*, 1–6. [CrossRef]
19. Wang, H.; Xu, K.; Li, Z.; Pang, K.; He, H. Improved Epidemic Dynamics Model and Its Prediction for COVID-19 in Italy. *Appl. Sci.* **2020**, *10*, 4930. [CrossRef]
20. Capuano, F. Modeling Growth, Containment and Decay of the COVID-19 Epidemic in Italy. *Front. Phys.* **2020**, *8*, 554. [CrossRef]
21. Jersakova, R.; Lomax, J.; Hetherington, J.; Lehmann, B.; Nicholson, G.; Briers, M.; Holmes, C. Bayesian imputation of COVID-19 positive test counts for nowcasting under reporting lag. *arXiv* **2021**, arXiv:2103.12661.
22. Moriña, D.; Fernández-Fontelo, A.; Cabaña, A.; Arratia, A.; Puig, P. Bayesian Synthetic Likelihood Estimation for Underreported Non-Stationary Time Series: Covid-19 Incidence in Spain. *arXiv* **2021**, arXiv:2104.07575.
23. Mukherjee, S.; Mondal, S.; Bagchi, B. Persistence of a pandemic in the presence of susceptibility and infectivity distributions in a population: Mathematical model. *medRxiv* **2021**. [CrossRef]
24. Kröger, M.; Schlickeiser, R. Gaussian Doubling Times and Reproduction Factors of the COVID-19 Pandemic Disease. *Front. Phys.* **2020**, *8*, 276. [CrossRef]
25. Wu, D.; Mac Aonghusa, P.; O’Shea, D.F. Correlation of national and healthcare workers COVID-19 infection data; implications for large-scale viral testing programs. *PLoS ONE* **2021**, *16*, e0250699. [CrossRef]
26. Singhal, A.; Singh, P.; Lall, B.; Joshi, S.D. Modeling and prediction of COVID-19 pandemic using Gaussian mixture model. *Chaos Solitons Fractals* **2020**, *138*, 110023. [CrossRef]
27. Schüttler, J.; Schlickeiser, R.; Schlickeiser, F.; Kröger, M. Covid-19 Predictions Using a Gauss Model, Based on Data from April 2. *Physics* **2020**, *2*, 197–212. [CrossRef]
28. Wu, K.; Darcet, D.; Wang, Q.; Sornette, D. Generalized logistic growth modeling of the COVID-19 outbreak: Comparing the dynamics in the 29 provinces in China and in the rest of the world. *Nonlinear Dyn.* **2020**, *101*, 1561–1581. [CrossRef] [PubMed]
29. Ma, J. Estimating epidemic exponential growth rate and basic reproduction number. *Infect. Dis. Model.* **2020**, *5*, 129–141. [CrossRef]
30. Kermack, W.O.; McKendrick, A.G.; Walker, G.T. A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond. Ser. Contain. Pap. Math. Phys. Character* **1927**, *115*, 700–721. [CrossRef]
31. Protezione Civile Italiana. COVID-19 Epidemiological Data Monitoring. Available online: <https://github.com/pcm-dpc/COVID-19> (accessed on 24 May 2021).
32. Rodrigues, H. Application of SIR epidemiological model: New trends. *arXiv* **2016**, arXiv:1611.02565.
33. Wacker, B.; Schlueter, J. Time-Continuous and Time-Discrete SIR Models Revisited: Theory and Applications. *Adv. Differ. Equ.* **2020**, *2020*, 556. [CrossRef]
34. Chowell, G.; Hincapie-Palacio, D.; Ospina, J.; Pell, B.; Tariq, A.; Dahal, S.; Moghadas, S.; Smirnova, A.; Simonsen, L.; Viboud, C. Using Phenomenological Models to Characterize Transmissibility and Forecast Patterns and Final Burden of Zika Epidemics. *PLoS Curr.* **2016**, *8*. [CrossRef] [PubMed]
35. Ashwanden, C. Why herd immunity for COVID is probably impossible. *Nature* **2021**, *591*, 520–522. [CrossRef]
36. Johns Hopkins University of Medicine. Mortality Analyses. Available online: <https://coronavirus.jhu.edu/data/mortality> (accessed on 24 May 2021).
37. European Commission. EU Vaccines Strategy. Available online: [https://ec.europa.eu/info/live-work-travel-eu/coronavirus-response/public-health/eu-vaccines-strategy\\_en](https://ec.europa.eu/info/live-work-travel-eu/coronavirus-response/public-health/eu-vaccines-strategy_en) (accessed on 24 May 2021).
38. Hopkins Tanne, J. Covid-19: Biden aims for 70 vaccination by 4 July. *BMJ* **2021**, *373*, n1155. [CrossRef] [PubMed]
39. Johns Hopkins University of Medicine. What Is Herd Immunity and How Can We Achieve It with COVID-19? Available online: <https://www.jhsph.edu/covid-19/articles/achieving-herd-immunity-with-covid19.html> (accessed on 6 April 2021).
40. Modi, C.; Böhm, V.; Ferraro, S.; Stein, G.; Seljak, U. Estimating COVID-19 mortality in Italy early in the COVID-19 pandemic. *Nat. Commun.* **2021**, *12*, 2729. [CrossRef]

41. De Natale, G.; Ricciardi, V.; De Luca, G.; De Natale, D.; Di Meglio, G.; Ferragamo, A.; Marchitelli, V.; Piccolo, A.; Scala, A.; Somma, R.; et al. The COVID-19 Infection in Italy: A Statistical Study of an Abnormally Severe Disease. *J. Clin. Med.* **2020**, *9*, 1564. [[CrossRef](#)]
42. Tsoularis, A.; Wallace, J. Analysis of Logistic Growth Models. *Math. Biosci.* **2002**, *179*, 21–55. [[CrossRef](#)]
43. Phipps, S.J.; Grafton, R.Q.; Kompas, T. Robust estimates of the true (population) infection rate for COVID-19: A backcasting approach. *R. Soc. Open Sci.* **2020**, *7*, 200909. [[CrossRef](#)] [[PubMed](#)]
44. Rees, E.; Nightingale, E.; Jafari, Y.; Waterlow, N.; Clifford, S.; Jombert, T.; Procter, S.; Knight, G. COVID-19 length of hospital stay: A systematic review and data synthesis. *BMC Med.* **2020**, *18*, 270. [[CrossRef](#)]
45. Bellan, M.; Patti, G.; Hayden, E.; Azzolina, D.; Pirisi, M.; Acquaviva, A.; Aimaretti, G.; Valletti, P.; Angilletta, R.; Arioli, R.; et al. Fatality rate and predictors of mortality in an Italian cohort of hospitalized COVID-19 patients. *Sci. Rep.* **2020**, *10*, 20731, [[CrossRef](#)]
46. Il Sole 24 Ore. History of Covid-19 from the Beginning. Available online: <https://lab24.ilssole24ore.com/storia-coronavirus/> (accessed on 24 May 2021).
47. Loewenthal, G.; Abadi, S.; Avram, O.; Halabi, K.; Ecker, N.; Nagar, N.; Mayrose, I.; Pupko, T. COVID-19 pandemic-related lockdown: Response time is more important than its strictness. *EMBO Mol. Med.* **2020**, *12*, e13171. [[CrossRef](#)]
48. COVID-19 Health Emergency Committee (Italy). Anti-COVID-19 Vaccine Administrations Open Data. Available online: <https://github.com/italia/covid19-opendata-vaccini> (accessed on 24 May 2021).