Data analysis on Coronavirus spreading by macroscopic growth laws

P. Castorina[a,b], A. Iorio[b] and D. Lanteri[a,b,c]

[a] INFN, Sezione di Catania, I-95123 Catania, Italy
[b] Faculty of Mathematics and Physics, Charles University
V Holešovičkách 2, 18000 Prague 8, Czech Republic
[c] Dipartimento di Fisica e Astronomia, Università di Catania, Italy

(Dated: March 3, 2020)

To evaluate if the effort to contain the epidemic spreading of the new Coronavirus disease 2019 is really working, we carried out an analysis of the time evolution of the infection in different Countries, by considering two well-known macroscopic growth laws, the Gompertz law and the logistic law. The proposed data analysis permits an evaluation of the maximum number of infected individuals. The daily data must be compared with the obtained fits to verify if the spreading is under control. From our analysis it appears that the spreading is reaching saturation in China and Singapore, due to the strong containment policy of the national governments. For South Korea and Italy, instead, the data of the first week of March on the new infections, will be crucial to understand if the saturation will be reached for lower or higher numbers of infected individuals. This will show whether a strong containment policy is worth keeping.

I. INTRODUCTION

The epidemic spreading of the new Coronavirus disease 2019 (COVID-19) [1] is producing the strongest containment attempt in the history of humankind. In China million of people are forced to live in isolation and in extreme difficult conditions. In other Countries the disease is growing fast enough.

Since the mechanisms of COVID-19 spreading are not completely understood, the number of infected people is large and the containment effects are essentially of empirical nature.

From this point of view a more quantitative analysis of the epidemic spreading can be interesting, and in the literature there is a large number of mathematical models (see for example [2, 3]).

However, in our opinion, this stage of the disease does not permit detailed analyses, since the available data consist of the number of infected patients in different geographic areas, as shown in Fig. 1 for China and in Fig. 2 for South Korea and Italy [4].

In other words, one has "coarse grain" information, and detailed "microscopic" studies that are, at the moment, of limited use.

Moreover, there is an impressive number of experimental verifications, in many different scientific sectors, that coarse-grain properties of systems, with simple laws with respect to the fundamental microscopic algorithms, emerge at different levels of magnification providing important tools for explaining and predicting new phenomena.

Therefore, an analysis based on macroscopic laws can be useful to understand the behavior of growth rate of the infection and to verify if its containment is indeed working.

A general classification of macroscopic growth laws is reported in Refs. [5, 6]. In the present study we focus on two well-known laws: the Gompertz law (GL) [7], and the logistic law (LL) [8].
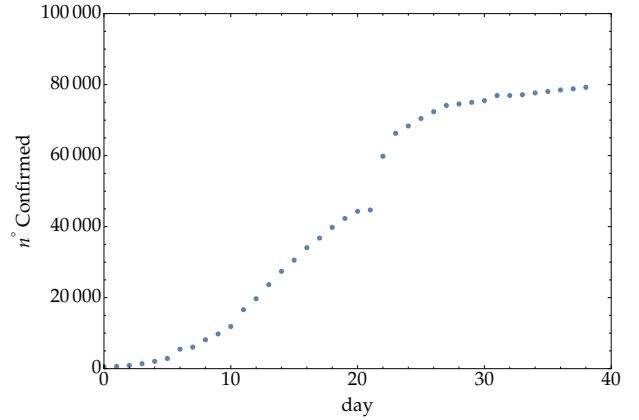

Figure 1: Number of infected individuals in China [4]. The jump corresponds to a different counting rule of infected people. Day zero is January the 22th.

The GL [7], initially applied to human mortality tables (i.e. aging) and tumor growth [9, 10], also describes kinetics of enzymatic reactions, oxygenation of hemoglobin, intensity of photosynthesis as a function of CO_2 concentration, drug dose-response curve, dynamics of growth, (e.g., bacteria, normal eukaryotic organisms).

The LL [8] has been applied in population dynamics, in economics, in material science and in many other sectors.

For a discussion of the COVID-19 data, one has to know that each of the considered macroscopic laws is characterized by two parameters, α_g, N_∞^g, for the GL, and α_l, N_∞^l, for the LL (the mathematical details are reported in appendix A). The meaning of the parameters is crucial to understand the evolution of the epidemic spreading.

The parameters α_g, α_l describe the specific rate of the initial exponential growth, after which there is a slowdown of the disease, due to contrast mechanisms. In particular, N_∞^g, N_∞^l, called *carrying capacities*, fix the maximum number of infected people in the two models.
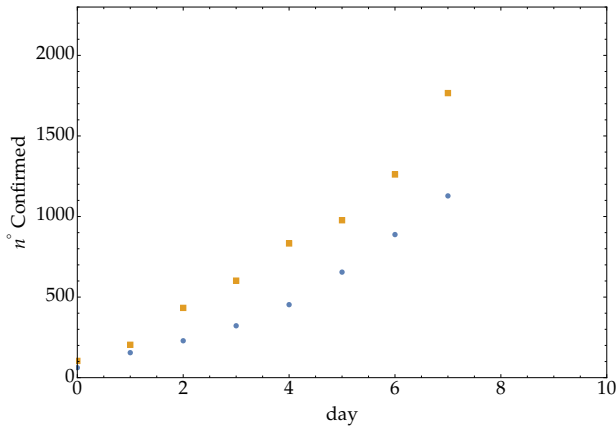
Figure 2: Number of infected individuals in South Korea (orange point) and Italy (blue point) [4]. Day zero is February the 20th for South Korea, and February the 22th for Italy.
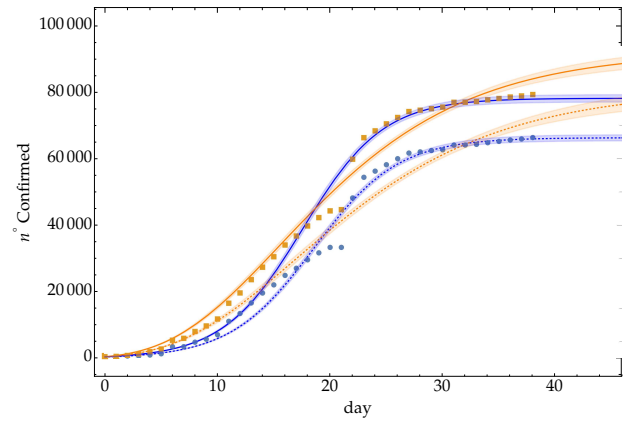

Figure 3: Number of infected individuals, fit of the chinese data (orange points). The Gompertz law prediction (orange) and the logistic curve (blue) are depicted with a band representing the 68% of confidence level. The blue points and dotted curves describe to the same fit for the Hubei region. Time zero corresponds to the initial day - 22/01.
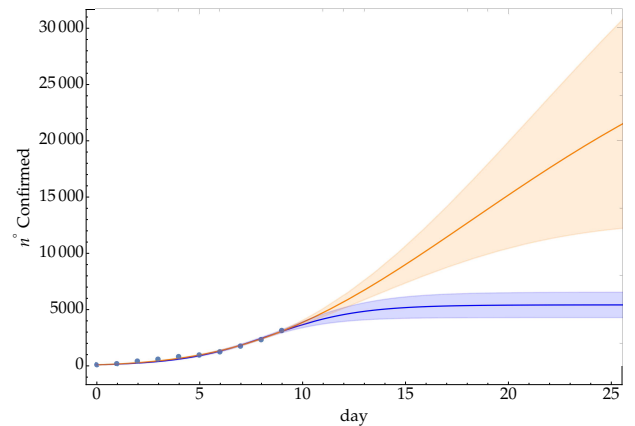

Figure 4: Number of infected individuals, fit of the South Korea data. GL (orange) and LL (blue) with a band representing the 68% of confidence level. Time zero corresponds to the initial day - 20/02.

The *contrast effect*, mathematically represented by the second term in Eqs. (A1) and (A2), depends on many possible mechanisms of pathological and political origin (medical cure, biological conditions, isolation, information, *et cetera*).

It should be clear that the present analysis does not give any specific indication in this respect, however the fitted value of N_∞^g, N_∞^l tell us how far is the disease evolution from the saturation point where the restraint effort is such that the spreading is practically over.

Indeed, a fit of the available data by GL and LL determines the values of the corresponding parameters, giving information on the possible behavior of the spreading.

We apply the analysis to China, South Korea, Italy and Singapore since one needs the number on infected people in a large enough time interval for a reliable fitting procedure.

II. CHINA

The available data cover the long period from January the 22th to February the 29th and, therefore, the numerical fit is more reliable. The results are depicted in Fig. 3. In the final period, when Chinese Government decided a different counting rule, the available data are well fitted by the logistic curve with $\alpha_l = 0.282 \pm 0.004$ (per day) and $N_\infty^l = 78121 \pm 1106$. Gompertz law predicts a larger saturation value $N_\infty^g = 92213 \pm 2339$, with $\alpha_g = 0.105 \pm 0.003$ (per day). Notice that the error is small due to the large number of available data. The number of infected chinese is today about 79251, which means that the effort to contrast the disease has been successful and almost completed.

III. SOUTH KOREA

Fig. 4 shows the result of the fits using the South Korea data, from February the 20th to February the 29th. The reduced number of data increases the error in the fitted parameter: $\alpha_g = 0.10 \pm 0.02$, $\alpha_l = 0.47 \pm 0.02$, $N_\infty^g = 35056 \pm 23775$ and $N_\infty^l = 5419 \pm 1060$. The two growth laws differ in the saturation values, although they are compatible within the 68% of confidence level (see the band in Fig. 3). Therefore one has to carefully follow if the next data are in agreement with the Gompertz evolution or with the logistic one.
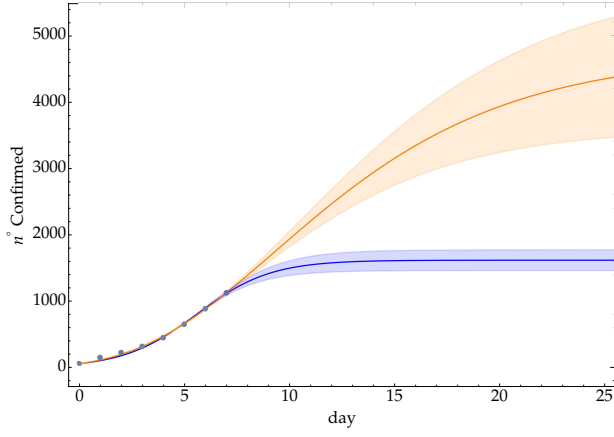
Figure 5: Number of infected individuals, fit of Italy data. GL (orange) and LL (blue) with a band representing the 68 % of confidence level. Time zero corresponds to the initial day - 22/02.
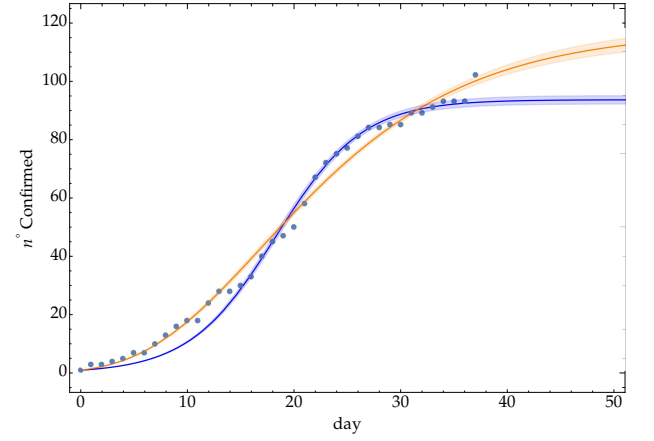

Figure 6: Number of infected individuals, fit of Singapore data. GL (orange) and LL (blue) with a band representing the 68 % of confidence level. Time zero corresponds to the initial day - 23/01.

	α^g	α^l
China	0.105 ± 0.003	0.282 ± 0.004
South Korea	0.10 ± 0.02	0.47 ± 0.02
Italy	0.16 ± 0.01	0.57 ± 0.02
Singapore	0.092 ± 0.002	0.247 ± 0.003

Table I: The value of the parameters α_g and α_l for different Nations.

	N_∞^g	N_∞^l
China	92213 ± 2339	78121 ± 1106
South Korea	35056 ± 23775	5419 ± 1060
Italy	4724 ± 1049	1618 ± 143
Singapore	117 ± 3	93 ± 1

Table II: The maximum number of infected individuals evaluated by the fitting procedure in different Countries.

IV. ITALY

The Italian data cover the short time from February the 22th to February the 29th. The results are depicted in Fig. 5, where the band represent the 68% of confidence level. Again the parameters have large errors: $\alpha_g = 0.16 \pm 0.01$, $\alpha_l = 0.57 \pm 0.02$, $N_\infty^g = 4724 \pm 1049$ and $N_\infty^l = 1618 \pm 143$. In the next days it will be clear if the data follows the GL or the LL, which implies that the saturation value could be very different. If the contrast effort is working, the final number of infected individuals should be 1618 ± 143 or less.

V. SINGAPORE

For Singapore the number of infected people is much smaller and the resulting fit is depicted in Fig. 6.

Finally, the values of the fitted parameters are summarized in tables I and II for a comparison between different nations.

VI. HOW TO USE THE FITS

To avoid possible misunderstandings, it is useful to comment on how to use of the previous fits in the next days.

The parameters N_∞ give information on the maximum number of infected individuals. For each nation, one has to follow the data in the next days. The key point is to check whether the data are in agreement either with the GL, or with the LL, or else are in between.

This is highly relevant, because the GL predicts a much larger number of maximu number of infected. Hence, in this case, the contrast effort has to be improved and, probably, diversified. On the other hand, one gets a good signal that the disease is slowing down to a smaller saturation values, if the data agree, or are less than, the values predicted by the LL.

VII. COMMENTS AND CONCLUSIONS

From this analysis it appears that the spreading is reaching saturation in China and in Singapore, due to the strong containment policy of the national governments. On the other hand, Figs. 4 and 5 show that South Korea and Italy are in a different situation. In the latter cases, the observed data in the first week of March will be crucial to understand if the evolution will follow the GL or the LL, hence to understand if the saturation will be reached for lower or higher numbers of infected individuals. Therefore, this analysis suggests that a strong containment policy should be kept, at least until the fits clearly shows which growth law the evolution follows.

The proposed approach for monitoring the evolution of the epidemic spreading of COVID-19 has to be consider as a complementary tool to more fundamental genomic methods [11, 12].

Of course, this analysis needs to be updated on a daily basis. The daily data must be compared with the fits, to verify if the spreading is under control or not (out of control being the exponential growth). This will help to understand quantitatively the status of the COVID-19 spreading.

Appendix A:

Let us call $N(t)$ the number of infected individuals at time t. The Gompertz evolution law is the solution of the differential equation

$$\frac{1}{N(t)} \frac{dN(t)}{dt} = \alpha_g \ln \frac{N_\infty^g}{N(t)}, \qquad (A1)$$

while the logistic law equation is

$$\frac{1}{N(t)} \frac{dN(t)}{dt} = \alpha_l \left(1 - \frac{N(t)}{N_\infty^l}\right). \qquad (A2)$$

The two laws differ in the description of the contrast term in the second member.

The general solution of the Gompertz equation is

$$N^g(t) = N_\infty^g \, \exp\left\{\ln\left(\frac{N(0)}{N_\infty^g}\right) \, e^{-\alpha_g(t-t_0)}\right\} \qquad (A3)$$

where t_0 is the initial time, $N(0)$ the initial value of the infected individuals coming from the available data.

For the logistic equation one has

$$N^l(t) = \frac{N(0) \, e^{\alpha_l(t-t_0)}}{1 - \frac{N(0)}{N_\infty^l} \left[1 - e^{\alpha_l(t-t_0)}\right]} . \qquad (A4)$$

[1] World Health Organization, Coronavirus disease (COVID-19) outbreak, https://www.who.int/emergencies/diseases/novel-coronavirus-2019.

[2] S. A. Herzog, S. Blaizot and Niel Hens, Mathematical models used to inform study design or surveillance systems in infectious diseases: a systematic review, BMC Infectious Diseases **17** (2017) 775.

[3] N. C. Grassly and C. Fraser, Mathematical models of infectious disease transmission, Nature Reviews Microbiology **6** (2008) 477.

[4] Novel Coronavirus (COVID-19) Cases, provided by JHU CSSE, https://github.com/CSSEGISandData/COVID-19.

[5] P. Castorina, P. P. Delsanto, C. Guiot, Classification Scheme for Phenomenological Universalities in Growth Problems in Physics and Other Sciences, Phys. Rev. Lett. **96** (2006) 188701.

[6] P. Castorina and P. Blanchard, Unified approach to growth and aging in biological, technical and biotechnical systems, SpringerPlus **1** (2012) 7.

[7] B. Gompertz, On the nature of the function expressive of the law of human mortality and a new mode of determining life contingencies, Phil. Trans. R. Soc. **115** (1825) 513.

[8] P. F. Verhulst, Notice sur la loi que la population poursuit dans son accroissement, Correspondance Mathématique et Physique, **10** (1838) 113.

[9] G.G: Steel, Growth kinetics of tumours, Clarendon Press, Oxford, 1977.

[10] L. A. Norton, Gompertzian model of human breast cancer growth, Cancer. Res. **48** (1988) 7067.

[11] A. Lai, A. Bergna, C. Acciarri, M. Galli, G. Zehender, Early Phylogenetic Estimate of the Effective Reproduction Number Of Sars-CoV-2, J. Med. Virol. 2020 Feb 25, doi: 10.1002/jmv.25723.

[12] Y. Chen, Q. Liu, D. Guo, Emerging coronaviruses: Genome structure, replication, and pathogenesis, J. Med. Virol. **92** (2020) 418.