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
\(^\dagger\) V Holešovičkách 2, 18000 Prague 8, Czech Republic
\(^{(c)}\) Dipartimento di Fisica e Astronomia, Università di Catania, Italy

\(\text{(Dated: March 17, 2020)}\)

To evaluate the effectiveness of the containments on the epidemic spreading of the new Coronavirus disease 2019, we carry on an analysis of the time evolution of the infection in different Countries, by considering well-known macroscopic growth laws, the Gompertz law, and the logistic law. We also propose here a generalization of Gompertz law. Our 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, due to the strong containment policy of the national governments. In Singapore a large growth rate, recently observed, suggests the start of a new strong spreading. For South Korea and Italy, instead, the next data on new infections will be crucial to understand if the saturation will be reached for lower or higher numbers of infected individuals.

I. INTRODUCTION

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

Since the mechanisms of COVID-19 spreading are not completely understood, the number of infected people is large and the effects of containment are evaluated essentially on an empirical basis. Therefore, a more quantitative analysis of the epidemic spreading can be interesting. In the literature there is a large number of mathematical models (see for example \(^{[2–5]}\)).

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 \(^{[6]}\).

In other words, one has “coarse grained” 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. \(^{[7, 8]}\). In the present study we focus on well-known laws: the Gompertz law (GL) \(^{[9]}\), a new proposed generalized GL (GGL) and the logistic law (LL) \(^{[10]}\), which will be compared with the exponential

![Figure 1: Number of infected individuals in China \(^{[3]}\). The jump corresponds to a different counting rule of infected people. Day zero is January the 22th.](image1)

![Figure 2: Number of infected individuals in South Korea (orange point) and Italy (blue point) \(^{[6]}\). Day zero is February the 20th for South Korea, and February the 22th for Italy.](image2)
spreading, which means that the containment efforts have no effect.

The GL, initially applied to human mortality tables (i.e. aging) and tumor growth, also describes kinetics of enzymatic reactions, oxygenation of hemoglobin, intensity of photosynthesis as a function of CO2 concentration, drug dose-response curve, dynamics of growth, (e.g., bacteria, normal eukaryotic organisms). The GGL is the generalization of the GL.

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

The previous laws differ in the description of the virus containment effects, which in the LL is stronger (power law behavior) than in the GL and GGL, which have a logarithmic decrease of the specific growth rate (see appendix A).

For a discussion of the COVID-19 data, one has to know that each of the considered macroscopic laws is characterized by two important parameters, \(\alpha_g, N^g_\infty\), for the GL, \(\alpha_l, N^l_\infty\) for the LL, and by three parameters, \(\alpha_{gg}, N^g_\infty\) and \(\beta\) for the GGL (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 \(\alpha_g, \alpha_l, \alpha_l\) describe the specific rate of the initial exponential growth, after which there is a slowdown of the disease, due to contrast mechanisms. In particular, \(N^g_\infty, N^l_\infty\), called carrying capacities, fix the maximum number of infected people in the models.

The contrast effect, mathematically represented by the second term in Eqs. (A1), (A2) and (A3), 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 \(N^g_\infty, N^l_\infty\) and \(N^l_\infty\) 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, GGL 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. DATA ANALYSIS

The cumulative number of infected people, in the different Countries, is used to describe the evolution of the infection spreading. However, the reliability of the data could depend on the status of the spreading also: in China, the counting of the infected people has been going on for a long time, the data are stable, and one does not expect any systematic error due to external limiting factors.

In Italy and South Korea, where the spreading is in a critical phase, and the containment effects is at an early stage, some limiting external factors could reduce the effective number of infected people. If, for example, the number of available kits (swabs) to detected the disease has a maximum number per day, one cannot detected, in a single day, a larger number of infected individuals. Moreover the infected, but asymptomatic, people introduces some degree of uncertainty (although, the number of the “truly” asymptomatic infections appears to be relatively rare).

On the other, if the swab number is large enough and if one considers that the usual chain is from asymptomatic to symptomatic (and than detectable) individuals, the previous effects could be strongly reduced.

With the previous warnings, in the next sections the global data about the cumulative number of infected people is discussed, for the different Countries, and compared with the macroscopic and exponential growth laws.

III. HOW TO USE THE FITS

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

With the caveats discussed in the previous section, the parameters \(N^g_\infty\)s give information on the maximum number of infected individuals. For each nation, one has to follow the day by day data, refitting the parameters until they stabilize. The key point is to check whether the data are in agreement either with the exponential, GL, GGL and LL, or else are in between. A typical example is given in table IIII where is reported for Italy the predictions obtained by using the available data until a specific day (March the 8th, for table IIII). The day after, one has to repeat the numerical analysis, which implies a redefinition of the parameters, i.e. of the specific growth rate, until they stabilize.

This is highly relevant, because the GL, GGL, describing a less effective containment effort, predict a much larger maximum number of infected. Hence, in this case, the contrast effect has to be improved and, probably, diversified. On the other hand, one gets a very 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, as in China.

IV. CHINA

The available data cover the long period from January the 22th to March the 10th and, therefore, the numerical fit is more reliable. The results are depicted in Fig. 3.
the final period, when Chinese Government decided a different counting rule, the available data are well fitted by the logistic curve with \( \alpha_l = 0.279 \pm 0.003 \) (per day) and \( N_{l\infty} = 79304 \pm 734 \). Gompertz law predicts a larger saturation value \( N_{g\infty} = 86376 \pm 1282 \), with \( \alpha_g = 0.111 \pm 0.002 \) (per day). Notice that the error is small due to the large number of available data. The number of infected Chinese is today about 80757, which means that the effort to contrast the disease has been successful and almost completed. Previous analysis have been done by considering the growth of mortality \[11\].

V. SINGAPORE

For Singapore the number of infected people is much smaller and the previous considered external limiting factor does not, presumably, apply. The resulting fit is depicted in Fig. 4 and, as shown by data, there is a recent strong increase in the growth rate: a clear signal that there is some new uncontrolled outbreak of the infection.

VI. SOUTH KOREA

Fig. 5 shows the result of the fits using the South Korea data, from February the 20th to March the 10th. The reduced number of data increases the error in the fitted parameter: \( \alpha_g = 0.152 \pm 0.005 \), \( \alpha_l = 0.422 \pm 0.005 \), \( N_{g\infty} = 10232 \pm 379 \) and \( N_{l\infty} = 7579 \pm 103 \). The GGL parameters are \( \alpha_{gg} = 0.224 \pm 0.004 \), \( N_{gg\infty} = 7589 \pm 106 \) and \( \beta = 0.37 \pm 0.03 \)

The Gl and LL differ in the saturation values, although they are compatible within the 68% of confidence level (see the band in Fig. 4). Therefore one has to carefully follow if the next data are in agreement with the Gompertz evolution or with the logistic one. The exponential behavior is strongly disfavored by the data.

The mortality growth follows the same trend has shown in Fig. 6.

VII. ITALY

The Italian data cover the time range going from February the 22th to March the 8th. The results are depicted in Fig. 7 where the band represent the 68% of confidence level. Previous analysis has been done in ref. \[13\], looking at the mortality table and at the number of patients in the Italian hospitals. Recent data shows an agreement with the exponential law and the last two data are shown more clearly in Fig. 7. The exponential trend forces the GL to reproduce the data with an artificially large value of \( N_{gl\infty} \), due to the logarithmic behavior. The
Figure 6: Mortality growth, fit of the South Korea data. GL (orange), GGL (red), LL (blue). Time zero corresponds to the initial day - 20/02.

Figure 7: Number of infected individuals, fit of Italy data. Exponential law (purple), GL (orange) and LL (blue) with a band representing the 68% of confidence level. Time zero corresponds to the initial day - 22/02. The parameters are: \( \alpha_g = 0.072 \pm 0.004 \), \( \alpha_l = 0.388 \pm 0.008 \), \( N_{g\infty} = 84577 \pm 22488 \) and \( N_{l\infty} = 10566 \pm 1009 \). The GGL gives results very similar to the GL. Red data have not been included in the fit.

Figure 8: Mortality growth, fit of the Italian data by exponential law. Time zero corresponds to the initial day - 20/02.

In Italy the mortality clearly follows an exponential growth, see Fig. 8, confirming that the data on the cumulative number of infected people could still have strong fluctuation.

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

Table I: The value of the parameters \( \alpha_g \) and \( \alpha_l \) for different Nations.

| Country   | \( \alpha_g \)       | \( \alpha_l \)       |
|-----------|-----------------------|-----------------------|
| China     | 0.111 \( \pm \) 0.002 | 0.279 \( \pm \) 0.003 |
| South Korea | 0.152 \( \pm \) 0.005 | 0.422 \( \pm \) 0.005 |
| Italy     | 0.072 \( \pm \) 0.004 | 0.388 \( \pm \) 0.008 |
| Singapore | 0.073 \( \pm \) 0.003 | 0.213 \( \pm \) 0.007 |

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

| Country     | \( N_{g\infty} \)         | \( N_{l\infty} \)         |
|-------------|--------------------------|--------------------------|
| China       | 86376 \( \pm \) 1282     | 79304 \( \pm \) 734      |
| South Korea | 10232 \( \pm \) 379       | 7579 \( \pm \) 103        |
| Italy       | 84577 \( \pm \) 22488     | 10566 \( \pm \) 1009      |
| Singapore   | 158 \( \pm \) 7           | 119 \( \pm \) 4           |

VIII. COMMENTS AND CONCLUSIONS

Let us state clearly that, the take-home message of our analysis is that, beyond any doubts, a strong containment policy should be kept.

As for countries with a longer (known) exposure to COVID-19, our analysis clearly shows that the spreading is: a) reaching saturation in China, b) but in Singapore, after a period of important slow down, a new increase is clearly visible. As for countries with a shorter (known) exposure, keeping in mind the limitations recalled in Sec. II our analysis, depicted in Figs. 5 and 7 shows that South Korea and Italy are in a different situation (see also [16, 17]. In those cases, the observed data in the near future will be crucial to understand if the evolution will either follow an exponential growth, or the GL, or the GGL or else the LL. This will allow to understand if the saturation will be reached for lower or higher numbers of infected individuals. 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 genomics methods \[18\].

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.

**Acknowledgments**

The authors thank Giorgio Parisi for useful discussions and comments. A.I. is partially supported by UNCE/SCI/013.

---

### 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

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

(A1)

the Generalized Gompertz law is solution of

$$
\frac{1}{N(t)} \frac{dN(t)}{dt} = \alpha_{gg} \ln^{(1-\beta)} \left( \frac{N_\infty}{N(t)} \right),
$$

(A2)

while the logistic law equation is

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

(A3)

The exponential behaviour (i.e. no reduction of the spreading) is

$$
\frac{1}{N(t)} \frac{dN(t)}{dt} = \text{constant},
$$

(A4)

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

The general solution of the Gompertz equation is

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

(A5)

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

$$
N^{gg}(t) = N_\infty^{gg} \exp \left\{ - \left[ \ln^\beta \left( \frac{N_\infty^{gg}}{N(0)} \right) - \alpha_{gg} \beta \left( t - t_0 \right) \right] \right\}.
$$

(A6)

For the logistic equation one has

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

(A7)

---

| day (March) | N   | Exp | GL  | LL  |
|-------------|-----|-----|-----|-----|
| 1 th        | 1694| 1512| 1439| 1226|
| 2 th        | 2036| 1898| 1906| 1713|
| 3 th        | 2502| 2383| 2476| 2345|
| 4 th        | 3089| 2991| 3159| 3127|
| 5 th        | 3858| 3755| 3964| 4041|
| 6 th        | 4636| 4713| 4896| 5042|
| 7 th        | 5883| 5916| 5960| 6060|
| 8 th        | 7375| 7425| 7157| 7023|
| 9 th        | 9172| 9320| 8487| 7871|
| 10 th       | 10419| 11699| 9946| 8574|
| 11 th       | 14685| 15303| 9127|
| 12 th       | 18433| 13229| 9545|
| 13 th       | 23137| 15036| 9851|
| 14 th       | 29042| 16940| 10070|
| 15 th       | 36455| 18928| 10224|
| 16 th       | 45759| 20888| 10332|
| 17 th       | 57437| 23107| 10406|
| 18 th       | 72096| 25271| 10457|
| 19 th       | 90496| 27468| 10492|
| 20 th       | 113593| 29685| 10515|

Table III: Number of confirmed sick in Italy predicted by exponential, Gompertz and logistic fits, compared with data (column N). Fits are made by using the available data until March the 8th. Bold data have not been included in the fit.
[7] 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.

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

[9] 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.

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

[11] G. Parisi, private communication.

[12] World Health Organization, Coronavirus disease (COVID-19) report, https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf

[13] E. Bucci, E. Marinari, L’evoluzione dell’epidemia da coronavirus in Italia, Scienza in Rete (in Italian) https://www.scienzainrete.it/

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

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

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

[17] A. Arianna and P. Giudici, A Poisson Autoregressive Model to Understand COVID-19 Contagion Dynamics (March 9, 2020). Available at SSRN: https://ssrn.com/abstract=3551626 or http://dx.doi.org/10.2139/ssrn.3551626

[18] 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.