

disCOVIDer19

A path-guide inside the COVID-19 pandemia

Fabio Caironi, Marzio De Corato,
Andrea Ierardi, Federico Matteucci,
Gregorio Saporito

"The laws of history are as absolute as the laws of physics, and if the probabilities of error are greater, it is only because history does not deal with as many humans as physics does atoms, so that individual variations count for more" (I. Asimov, Foundation and Empire)

CONTENTS

1	Introduction	5
2	Theoretical Background	7
2.1	Population growth dynamics	7
2.2	Application to the epidemiology	8
2.3	Time series approach	9
2.4	An ab-initio approach: the SEIR model	9
3	Data Origin	9
4	Panels	10
4.1	Home	10
4.2	Data Inspection	11
4.3	Data Analysis	13
5	Conclusions	18
6	Packages used	18

LIST OF FIGURES

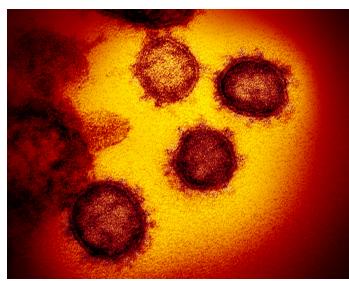
Figure 1	Different pictures of the virus as reported by the NIAID's Rocky Mountain Laboratories (RML) in Hamilton, Montana [1]	5
Figure 2	The Malthusian growth (red line) compared to the logistic growth (blue line). The carrying capacity is marked with a black dashed line. Image taken from [2]	8
Figure 3	Choropleth map by province showing absolute number of Covid-19 cases	10
Figure 4	Input selector where the user can choose the three different breakouts absolute, percentage, and density	11
Figure 5	Boxes with syncronised barometer level statistic about the Coronavirus	11
Figure 6	•	12
Figure 7		13
Figure 8		16
Figure 9		17

LIST OF TABLES

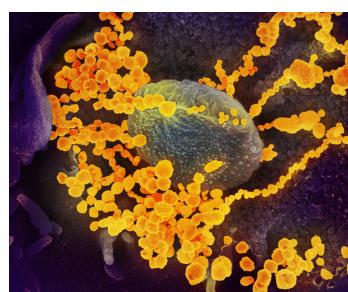
ABSTRACT

1 INTRODUCTION

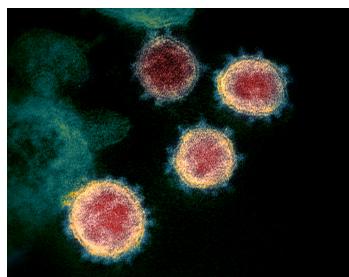
BACKGROUND In December 2019 different cases of pneumonia were reported in Wuhan (China) [3]. Their origin was later ascribed to a new virus classified as *Severe acute respiratory syndrome coronavirus 2* (SARS-CoV-2) whose TEM and SEM pictures are reported in the Fig. 1. The origin of this virus is still subject of scientific debate between the scientific community, however one of the most common opinion is that this virus comes from bats, in particular the genus *Rhinolophus* [4]. The most compelling feature of this virus is that is its ability to spread also via coughing and sneezing [5], and also by touching infected surfaces [6]. Differently with respect to the SARS-CoV the virus seems to have a lower mortality rate [7, 8].



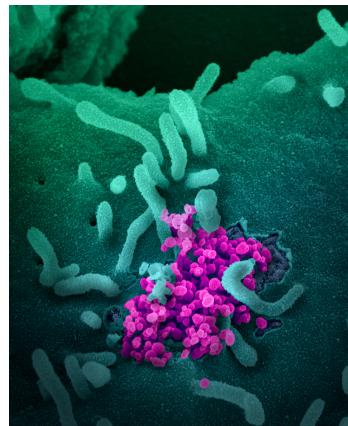
(a) Transmission electron microscope (TEM) image of SARS-CoV-2—also known as 2019-nCoV, emerging from the surface of cells cultured



(b) Scanning electron microscope of the SARS-CoV-2 emerging from the surface of cells cultured



(c) TEM image of SARS-CoV-2. Note the spikes that gives the name coronavirus to the virus



(d) SEM image of the virus

Figure 1: Different pictures of the virus as reported by the NIAID's Rocky Mountain Laboratories (RML) in Hamilton, Montana [1]

In January the Chinese government imposed the quarantine for the city of Wuhan (almost 11M people); the quarantine was later expanded to the full province of Hubei (60M people) and then to the neighbour provinces Huanggang, Ezhou and Xianning. The virus then spread in Canada, Germany, Thailand and Japan and then in other different countries included Italy [9]. In Italy the first cases,

two Chinese tourist from Hubei, were reported in Rome [10]; then other cases were reported in Codogno (Lodi,Lombardy). Later the virus spread almost in all regions of Italy with a higher density in Lombardy, Veneto, Emilia-Romagna, Piemonte and Marche. Starting from 22 of February, the Italian government started to impose the quarantine (red-zone) for eleven different municipalities, in particular Codogno, Casalpusterlengo, Lodi (included the neighbour municipalities) and Vo'. Starting from this date different restrictive measures were imposed starting from the regions with the highest number of cases: public entrainment were almost suspended as well as schools and universities. Workers (in public and private sectors) were allowed, when possibile, to work at home (smart-working). Such measures culminated with a decree approved by the Italian government that divided Italy in three areas: the red zones in which the municipalities with highest number of cases in which all the population was subjected to quarantine, the yellow one (Lombardia, Veneto, Emilia Romagna) in which schools, universities and public events ,sports as well cinemas were suspended and the rest of Italy in which no restrictive measures were adopted [11]; however 3 days later all schools and universities were suspended [12]. The restrictive measures were also extend up to all Italy with a further decree [13].

MOTIVATION In order to have a better understanding about the spread and the effects on the population of the COVID19 disease, we propose a ShinyApp, called disCOVIDer19 , that gathers the data , mostly provided by the Protezione Civile [14], about the number of infections(current as well cumulative), hospitalized people as well in intensive care, deaths and recovered for Italy, its regions and provinces; more details will be discussed in sections 3 and 4. Furthermore we fitted the cumulative cases for Italy and all its regions and provinces with a logistic curve, in order to get a bird-of-eye view on the trend followed as well on the effectiveness of the restrictive measures imposed by the Italian government. The theoretical foundations of this fit ,as well how to manage and use it, will be discussed in the section 2. Beside the logistic curve, we also considered a statistical approach, largely diffused among economists, that considers the number of cases in each day as a time series: on this basis we were able to make use of a particular tool, which will be briefly introduced in the section 2, that allowed us to make a further forecast about future cases. This latter becomes useful in case of large deviations from logistic distribution.

2 THEORETICAL BACKGROUND

Different models were proposed in the literature for modelling the spread of a disease [15], here we are going to consider a relative simple model taken from the growth dynamics of populations.

2.1 Population growth dynamics

The dynamics of population was founded by Thomas Malthus in 1817 [16] with his well known equation:

$$x_{n+1} = x_n(1 + r) \quad (1)$$

Where x_n stands for the population at time and r the growth rate. Thus the following discrete progression will be obtained:

$$P_{n+1} = P_0(1 + r)^{n+1} \quad (2)$$

in which P_0 stands for the initial population. This discrete model can be reshaped in to a continuous one in the following way:

$$\dot{P}(t) = rP(t) \quad (3)$$

this is a relative simple differential equation with separable variable. Its integration gives:

$$P(t) = P_0(t) \quad (4)$$

Malthus, of course, did not believed that the population could grow ad infinitum with an exponential growth, but since he estimated that the resources growth follows a linear path, he argued that, at the intersection of the two curves, a consistent part of the population will not have access to the resources. Thus he expected that the exponential growth realizes only in the first part of the growth. Basing on these arguments, a Belgian mathematician Pierre-François Verhulst in 1838, proposed a model [17] in which the key point is the maximum number of people that the resources allows to live, usually called the carrying capacity here indicated as K (P and r have the same meaning of the previous model):

$$\dot{P} = rP \left(1 - \frac{P}{K}\right) \quad (5)$$

that have the following analytic solution

$$P = \frac{KP_0e^{rt}}{K + P_0(e^{rt} - 1)} \quad (6)$$

In Fig. 2 this solution is compared to the Malthusian one: as one can point out the logistic growth has a saturation effect near the

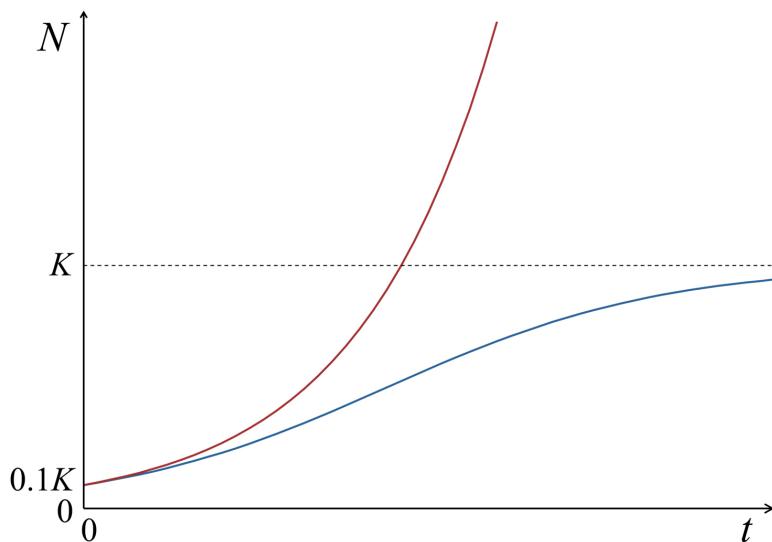


Figure 2: The Malthusian growth (red line) compared to the logistic growth (blue line). The carrying capacity is marked with a black dashed line. Image taken from [2]

asymptote of carrying capacity. This represent the limit capacity that Malthus proposed to indicate as the intersection between the linear and exponential growth. It is worth nothing that near the origin the two growth are identical, then the logistic curve becomes linear up to the saturation region.

2.2 Application to the epidemiology

The dynamics of the two models described before can be used to shape the spread of an infection [18, 19] in this case the population P is replaced with the total number of infected people, P_0 the number of first infected people, r the spreading rate and the carrying capacity K with the maximum number of people that can be infected. The latter is the key point for the modelling: in principle this parameter would be equal to the population number, in practice due to the restrictive measure a large part of population can be removed from this computation: this lowers the predicted number of infected people. Thus we can argue that the effectiveness of the restrictive measures can be inferred from the linear behaviour of the cumulative curve (except, of course, if the numer of infected people is close to the population size). It is worth nothing that if the quarantine fails for also one/few infected people, the carrying capacity is increased up to the people that can be infected by these subjects. If this contribution is not negligible a new logistic growth may be found, with of course an exponential starting behaviour. Thus the prediction with the logistic curve should be taken as an evaluation of the effectiveness of restrictive measures and as the best scenario that can may realize. For this reason it is

advisable to see the logistic curve as a local (in time) estimator. In principle one can model a differential set of equation were the carrying capacity is time dependent. This possibility may be considered by the authors in a future version of this App. Furthermore it is worth nothing that other similar models such as Gopertz, Richards or Bertalanffy can be considered [19] : the authors may consider to include them in a future version of this App.

2.3 Time series approach

2.4 An ab-initio approach: the SEIR model

3 DATA ORIGIN

4 PANELS

4.1 Home

The home panel provides an overview of the COVID19 spread in Italy. The data shown are synchronised with the civil protection database. Whenever the App is started, a check for the updates is performed and their occurrence is indicated in the *most recent updates* section. The home panel consists of two main sections: the choropleth map and the summary statistics. The choropleth map (Fig 3) is an interactive heat map with breakouts by region and by province tracking the number of Covid-19 cases.

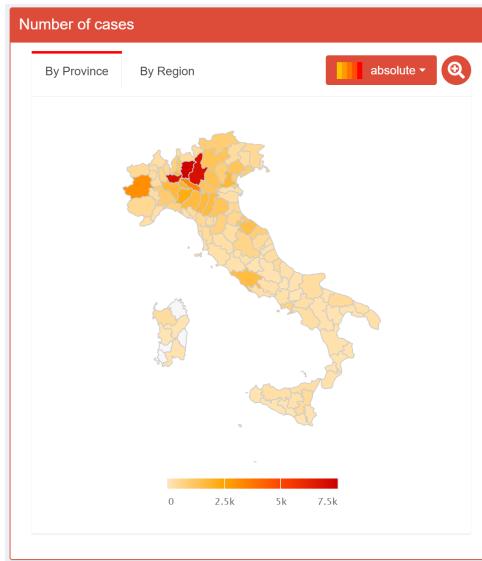


Figure 3: Choropleth map by province showing absolute number of Covid-19 cases

Beyond the raw absolute number of total cases by province/region we propose two other indicators which introduce some form of normalisation to improve the comparability of the different geographical areas. These indicators are percentage and density (Fig. 4).

The percentage indicator is calculated dividing the total number of COVID19 cases by the population of the respective region/province (the latter retrieved from Istat [20, 21]).

$$\text{percentage cases} = \frac{\text{total cases}}{\text{local population}} \times 100 \quad (7)$$

This indicator ensures that the number of cases in more populated areas are not over-indexed compared to less populated ones. Similarly, it ensures less populated areas are not under-indexed. The density indicator accounts for the territorial extension (in km²) of re-

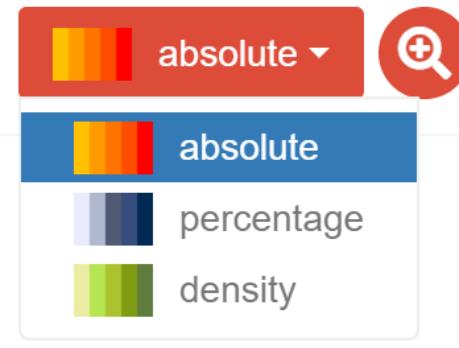


Figure 4: Input selector where the user can choose the three different breakouts absolute, percentage, and density

gions/provinces expressed in parts per thousand (‰) and is calculated as follows:

$$\text{density} = \frac{\text{total cases}}{\text{territorial extension}} \times 1000 \quad (8)$$

This indicator ensures that we are not underestimating small regions/provinces with fewer cases than larger ones but with more concentrated cases per Km². Finally, the summary statistics section (Fig. 5) provides some barometer level statistics regarding the current total number of COVID19 cases in Italy with three further breakouts: intensive care, hospitalised, and home isolation.



Figure 5: Boxes with synchronised barometer level statistic about the Coronavirus

4.2 Data Inspection

The *Data Inspection* panel provides a data visualisation of the general information and a deeper overview about the hospital occupancy, growth monitoring and test tracking for the COVID19 in Italy. It is divided in main two section: introduction and deeper inspection. In first chart (Fig. 7a) of the introduction are presented the general information about the total cases, total recovered, total hospitalised and intensive care occupancy. It is possible to select the entire Italian country or a particular region or province. Moreover, it is possible to visualise and filter the raw data for country, region and province in

the panel Raw Data (Fig. 6b). In the deeper inspection panel there are four charts: the first two charts are in two different panels of a box and visualise information about intensive care occupancy in the Italian hospital in different regions, while the others represent the growth monitoring of total cases and test tracking. It is worth noting that the occupation may be higher than their total number, due to the fact that the places in intensive therapy that we considered may be upgraded (see the section 3). In the first chart 6a is represented the percentage hospital occupancy in the selected day divided by capacity with respect to the initial intensive care capacity at the start of the pandemic.

$$\text{percentage occupancy} = \frac{\text{occupancy}}{\text{capacity}} \times 100 \quad (9)$$

The second chart (Fig. 7c) visualises a bar chart of the hospital occupancy and capacity of intensive care in different region of Italy at the selected day. The third chart (Fig. 7b) represents the percentage growth and growth change of total cases day by day. The fifth chart (Fig. 6c) visualises the daily cases with respect to the daily tests.

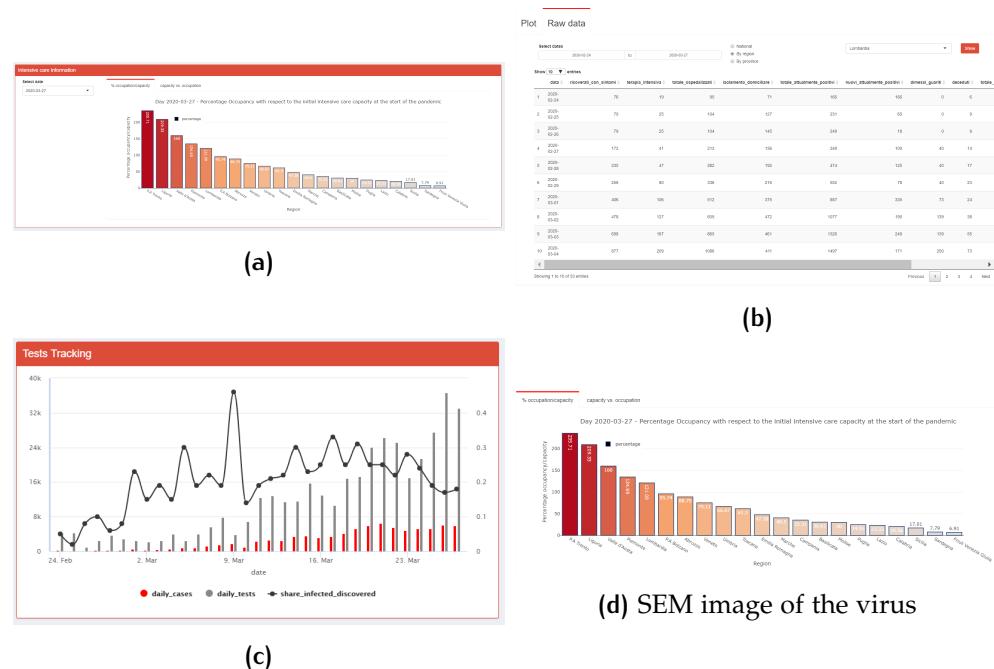


Figure 6: •

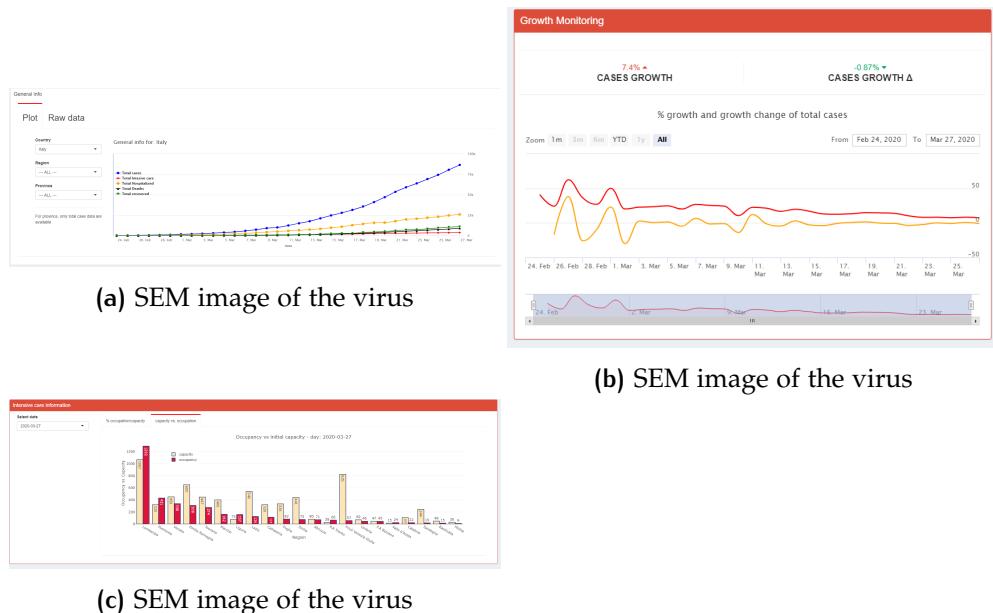


Figure 7

4.3 Data Analysis

Panel *Analysis* explores different mathematical models for shaping the infection's time series, namely the ones introduced in Section 2. As previously argued, each of these model's predictive power hinges on uncertain actual scenarios, such as the effectiveness of restrictive measures, the presence of several independent and spread outbreaks, and so on.

Package [22] has been used for growth curve fitting. This package provides function *SummarizeGrowth*, which inputs the time series of sample data and outputs the model represented by the growth metrics. Such parameters include those of the logistic equation that best fit the data, namely K , P_0 and r , as in Eq.6. The reasons why this package was chosen are chiefly two. At the one hand, it performs non-linear curve fitting with the Levenberg-Marquardt algorithm [23], which happens to be one of the most robust nls algorithms. For this step, function [24] is used. On the other hand, the package is endowed with methods for background correction.

At the top of analysis panel the user is allowed to pick one dataset among national, by region and by province. Hence, any following analysis will use the selected dataset as input (the user can browse the raw dataset in panel inspection, see paragraph 4.2).

The first section is devoted to the logistic model. An input box (Fig.8a) lists several dynamic inputs:

1. A date range that the calculator will use for curve fitting. The option to choose initial and final dates different from those at,

respectively, the beginning and the end of the dataset, is useful in two ways:

- It allows to cut initial dates which possibly correspond to zero or constant infections data, as a consequence of a delay in the outbreak in that territory. Actually, an automatic process will do so after the choice of territory, adjusting the initial date to the first day whose new infections exceed a given threshold (currently set to 1).
 - It allows to compute the available prediction n days ahead, by cutting the last n dates, and compare it to the real data.
2. A check-box *standardise positive cases by total swabs*. If selected, the share of infected among the tested, instead of the infected themselves, will be used for computations and rendering. This may be useful to capture the selection bias, but only in the scenario in which tests are randomly assigned to the population, or at least are assigned with a logic common to all dates and territories. This option is not currently available for provinces.
 3. Plot type selection boxes. The available plot types are cumulative cases and new cases. (The user can always show and hide plot's components by clicking on their label in the legend).
 4. Residuals plot type. Four options for graphically rendering the residuals: Residuals vs. fitted values; Standardized residuals vs. fitted values; Autocorrelation; Square root of absolute residuals vs. fitted values.

The output is displayed within three boxes:

1. Inside the [8b](#) box the sample points, the fitted logistic curve and a confidence band at 95 % level for the logistic curve are overlayed. The user can view the sample and fitted values by hovering over or touching (for touch-screen devices) them [8b](#). In addition, if *new cases* is selected, a plot of sample cases differences (which correspond to new cases, in fact) is shown over the logistic estimated distribution [8c](#). The latter is simply obtained by deriving the right hand side of formula (4.3.1):

$$N'(t) = \frac{\left(\frac{K-N_0}{N_0}\right) K r e^{-rt}}{[1 + \left(\frac{K-N_0}{N_0}\right) e^{-rt}]^2} \quad (10)$$

The user is suggested to take full advantage of plotly features. Besides showing points labels when passing over it and enabling plot selection by clicking on label items, plotly charts are endowed with a sequence of tools shown at the top-right of the box [9a](#).

2. Summary output box [9c](#) contains principal information about the selected model and a list of goodness-of-fit tests on the residuals.
3. Residuals box contains the plots of residuals, rendered in the user-chosen fashion. (see Fig.[9b](#)).

Input

Choose fitting interval

24 Feb

Standardise positive cases by total swabs

Plot type

Cumulative cases
 New cases

Residuals

Residuals plot type

Residuals

(a) Set of useful tools for navigating into plots, by plotly.

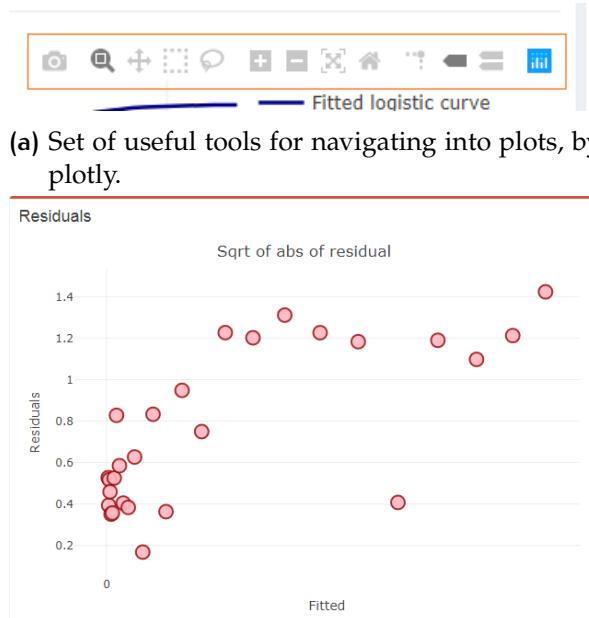


(b) Summary output box of logistic section.

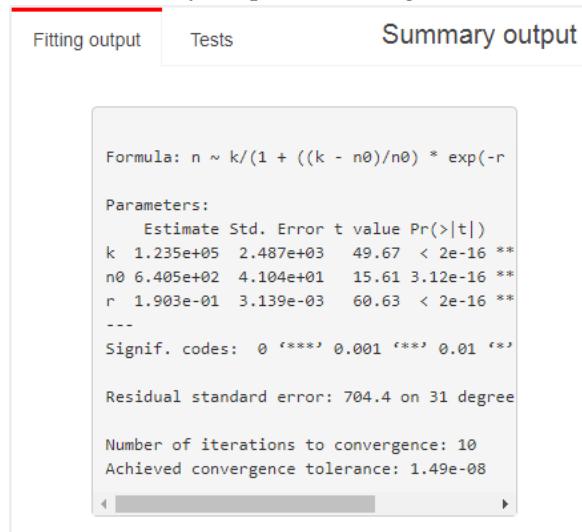


(c) One plot of nls residuals: square root of absolute residuals vs. fitted values.

Figure 8



(b) Summary output box of logistic section.



(c) One plot of nls residuals: square root of absolute residuals vs. fitted values.

Figure 9

5 CONCLUSIONS

6 PACKAGES USED

REFERENCES

- [1] <https://www.niaid.nih.gov/news-events/novel-coronavirus-sarscov2-images>.
- [2] https://commons.wikimedia.org/wiki/File:Malthusian_growth_vs_logistic_growth.png.
- [3] Chaolin Huang, Yeming Wang, Xingwang Li, Lili Ren, Jianping Zhao, Yi Hu, Li Zhang, Guohui Fan, Jiuyang Xu, Xiaoying Gu, et al. Clinical features of patients infected with 2019 novel coronavirus in wuhan, china. *The Lancet*, 395(10223):497–506, 2020.
- [4] Peng Zhou, Xing-Lou Yang, Xian-Guang Wang, Ben Hu, Lei Zhang, Wei Zhang, Hao-Rui Si, Yan Zhu, Bei Li, Chao-Lin Huang, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, pages 1–4, 2020.
- [5] Isaac Ghinai, Tristan D McPherson, Jennifer C Hunter, Hannah L Kirking, Demian Christiansen, Kiran Joshi, Rachel Rubin, Shirley Morales-Estrada, Stephanie R Black, Massimo Pacilli, et al. First known person-to-person transmission of severe acute respiratory syndrome coronavirus 2 (sars-cov-2) in the usa. *The Lancet*, 2020.
- [6] De Chang, Huiwen Xu, Andre Rebaza, Lokesh Sharma, and Charles S Dela Cruz. Protecting health-care workers from sub-clinical coronavirus infection. *The Lancet Respiratory Medicine*, 8(3):e13, 2020.
- [7] Morten Dræby Sørensen, Brian Sørensen, REGINA GONZALEZ-DOSAL, Connie Jenning Melchjorsen, Jens Weibel, Jing Wang, Chen Wie Jun, Yang Huanming, and Peter Kristensen. Severe acute respiratory syndrome (sars) development of diagnostics and antivirals. *Annals of the New York Academy of Sciences*, 1067(1):500–505, 2006.
- [8] Paul Weiss and David R Murdoch. Clinical course and mortality risk of severe covid-19. *The Lancet*, 2020.
- [9] <https://hgis.uw.edu/virus/>.
- [10] <https://www.wired.co.uk/article/coronavirus-italy>.
- [11] https://www.repubblica.it/politica/2020/03/01/news/coronavirus_misure_governo-249980561/.

- [12] https://www.repubblica.it/politica/2020/03/04/news/coronavirus_governo_chiuse_le_scuole_in_tutta_italia_-250209485/.
- [13] http://www.salute.gov.it/portale/news/p3_2_1_1_1.jsp?lingua=italiano&menu=notizie&p=dalministero&id=4186.
- [14] <https://github.com/pcm-dpc/COVID-19>.
- [15] Matt J Keeling and Pejman Rohani. *Modeling infectious diseases in humans and animals*. Princeton University Press, 2011.
- [16] Thomas Robert Malthus. *An essay on the principle of population, as it affects the future improvement of society. With remarks on the speculations of mr. Godwin, m. Condorcet, and other writers. By TR Malthus*. 1817.
- [17] PF Verhulst. Notice sur la loi que la population poursuit dans son accroissement in: Correspondance mathématique et physique, vol. 10. 1838.
- [18] Robert E Serfling. Historical review of epidemic theory. *Human biology*, 24(3):145–166, 1952.
- [19] Junling Ma, Jonathan Dushoff, Benjamin M Bolker, and David JD Earn. Estimating initial epidemic growth rates. *Bulletin of mathematical biology*, 76(1):245–260, 2014.
- [20] <https://www.istat.it/it/files//2015/04/Superfici-delle-unit%C3%A0-amministrative-Dati-comunali-e-provinciali.zip>.
- [21] http://dati.istat.it/Index.aspx?DataSetCode=DCIS_POPRES1.
- [22] <https://cran.r-project.org/web/packages/growthcurver/index.html>.
- [23] Jorge J Moré. The levenberg-marquardt algorithm: implementation and theory. In *Numerical analysis*, pages 105–116. Springer, 1978.
- [24] <https://www.rdocumentation.org/packages/minpack.lm/versions/1.2-1/topics/nlsLM>.