

Estimation of COVID-19 dynamics “on a back-of-envelope”: Does the simplest SIR model provide quantitative parameters and predictions?

Eugene B. Postnikov

Department of Theoretical Physics, Kursk State University, Radishcheva st., 33, Kursk 305000, Russia

ARTICLE INFO

Article history:

Received 8 April 2020

Accepted 21 April 2020

Available online 1 May 2020

Keywords:

Covid-19

Compartmental epidemic model

SIR model

Logistic regression

ABSTRACT

Basing on existence of the mathematically sequential reduction of the three-compartmental (Susceptible-Infected-Recovered/Removed) model to the Verhulst (logistic) equation with the parameters determined by the basic characteristic of epidemic process, this model is tested in application to the recent data on COVID-19 outbreak reported by the European Centre for Disease Prevention and Control. It is shown that such a simple model adequately reproduces the epidemic dynamics not only qualitatively but for a number of countries quantitatively with a high degree of correlation that allows to use it for predictive estimations. In addition, some features of SIR model are discussed in the context, how its parameters and conditions reflect measures attempted for the disease growth prevention that is also clearly indicated by deviations from such model solutions.

© 2020 Elsevier Ltd. All rights reserved.

1. Introduction

The outbreak of disease named by the World Health Organisation (WHO) as COVID-19 induced by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is now an officially declared as pandemics, which affected practically all world countries, see [1–3] for a general review.

This situation induces also a demand to the mathematical epidemiologist community [4] for revealing models of outbreak dynamics, which have not only explanatory but also a predictive potential while the outbreak is in an active phase. They are aimed at the fast estimations of the future COVID-2019 impact on the population, measures required from the public health system and effectiveness of different quarantine measures.

Recently, there is a significant number of models proposed to achieve this goal, varying from very simple to complicated ones, which include a lot of variables and parameters. However, it should be pointed out that detailed models, which include a lot of variable parameters require, respectively a detailed statistics for their validation. This is, for example, the case of Chinese outbreak, where the larger specification of affecting factors and decease statistics is available [5–8] but for the rest of the world, the available data variety is quite limited. The last fact argues in favour of simple models, in majority regressions and compartmental ones. Among the last, one can mention variants of SEIR and SIR(D) models. e.g. [9–12]. At the same time, even such models considered directly in

their original differential form contain more parameters than directly available information for their determining that argues that the reduction of variable and parameters number may be preferable, see the discussion in [12]. On the other hand, the pure logistic growth fitting applied *ad hoc* simply as a probable regression for the expected sigmoid curve described the total number of infected persons leads to controversial and uncertain results, e.g. [13–15], that indicates an oversimplification when the fitting model does not have a clear linking to the epidemiological model background.

Thus, the main goal of this work is to fill a gap between two mentioned approaches trying to simplify the model as much as possible but not simpler and to check its reliability with respect to the actual available epidemiological data on COVID-19 outbreak and perspectives of its forecasting for the nearest months.

As a starting simplest model, the SIR model proposed by Kermack and McKendrick [16] is chosen. This approach, which is one of the most popular basic models of mathematical epidemiology [17] operates with the subdivision of the full population into three groups: Susceptible S , Infected I , and Recovered/Removed R , which interact according to the kinetic scheme

$$S + I \xrightarrow{k} 2I, \quad (1)$$

$$I \xrightarrow{\tau^{-1}} R, \quad (2)$$

where the rate coefficients are: k is the contact rate, or, in a general case the probability of being infected (this more abstract definition may be more reasonable when one considers not a localised well-mixed population but the dynamics of data averaged over a

E-mail addresses: postnikov@kursksu.ru, postnicov@gmail.com

wide geographic area, see the respective discussion in [18]); τ is the characteristic duration of illness.

The system of ordinary differential equation, which corresponds to the kinetic scheme (1)–(2) reads as

$$\begin{aligned}\frac{dS}{dt} &= -kSI, \\ \frac{dI}{dt} &= kSI - \tau^{-1}I, \\ \frac{dR}{dt} &= \tau^{-1}I.\end{aligned}$$

Why may this simple system be appropriate for estimating the dynamics of the recent COVID-19 outbreak in different countries? The answer is based on the specificity of available data provided, e.g. by the World Health Organisation [19] and the European Centre for Disease Prevention and Control [20].

The freely available statistics is externally averaged since the daily data are given for each country as a whole, without more precise localisation within country regions and cities; detailed data on intra-country traffic, movements and contacts are absent. Therefore, the consideration can be based on the strong mean field level of approximation only if do not intend to introduce some hypothetical “hidden” parameters but intends to apply these available data only. In addition, the widespread restrictions of the last two months on international flights support the consideration of the country-based data.

On the other hand, it is not reflected in the available data, at what stage of illness each infected person enters the statistics as a registered infected one. The fact can be revealed from the latent phase during some testing of population groups suspicious to the risk to be infected up to the late phases of active illness when a person is taken to hospital under serious acute conditions. Whence, the division into latent (exposed) and active forms, i.e. any kind of SEIR-models will be over-complexification with respect to the actual data uncertainty.

Thus, this model taken as a background will be a subject of study with respect to the further reducing to the form, which should provide a simple and easily realisable procedure as well as for testing of its relevance for modelling COVID-19 in different countries of the world.

2. Methods

2.1. Analytic treatment

The main fact, which allows the further consideration, is based on the proof given in [21] that the system of SIR Eqs. (3)–(3) is already one-dimensional due to the presence of two conservation laws and, moreover, can be sequentially reduced to the Verhulst (logistic) equation under the very natural assumptions. For the self-consistency of explanations, this line of reasoning is given below as well as its adjustment to the considered situation.

For simplicity, it is convenient to make this derivation for S , I , R considered as fractions of the full initially susceptible population taken as unity: $S + I + R = 1$. This gives the first conservation law reducing the number of equations from three to two.

Further, Eqs. (3) and (3) give

$$\frac{dS}{dR} = -k\tau S$$

with the solution $\log(S) + k\tau R = 1$ written under the condition $S(0) = 1$ (this means that one practically neglect by the fraction of initially infected persons in the case of a large population). This gives the second conservation law that reduces the SIR system to one differential equation, which is convenient to write with respect

to R :

$$\frac{dR}{dt} = \tau^{-1} [1 - R - e^{-k\tau R}]. \quad (3)$$

Under conditions of “weak outbreak” that means $k\tau \max(R) \ll 1$, or, if to consider the dimensional population $\mathbf{R} = \mathbf{R}\mathbf{N}$, where $\mathbf{N}$ is the total population, $k\tau \max(R) \ll \mathbf{N}$ (this, fortunately, fulfils for all countries, the cumulative number of infected is much less (e.g. for China, where the outbreak is almost finalized) and expected to be much less for the rest of world that the total country population), the exponential term in Eq. (3) can be expanded into the Taylor series that results in the classic Verhulst (logistic) equation

$$\frac{dR}{dt} = \frac{\Re_0 - 1}{\tau} R \left[1 - \frac{\Re_0^2}{2(\Re_0 - 1)} R \right], \quad (4)$$

where $\Re_0 = k\tau$ is the basic reproductive number.

Note that the transition to the dimensional quantities does not change its functional form and the initial growth rate $((\Re_0 - 1)\tau^{-1})$ but simply rescales the saturation values from $\Re_0^2(2(\Re_0 - 1))^{-1}$ to $\Re_0^2(2(\Re_0 - 1)\mathbf{N})^{-1}$.

Note also that the number of recovered/removed persons modelled by the Verhulst equation can be deduced from the number of infected persons $\mathbf{I}(t)$ (in the original dimensional form of data), due to Eq. (3) as the integral with varying upper limit

$$\mathbf{R}(t) = \frac{1}{\tau} \int_0^t \mathbf{I}(t) dt. \quad (5)$$

Note that Eq. (5) contains the characteristic disease duration τ , which is, in principle, known from medical practice but is not required from the pure data-driven approach to the problem. Actually, since τ is simple scaling constant, the replacement of $\mathbf{R}(t)$ by the cumulative number of infected persons

$$\Sigma I = \int_0^t \mathbf{I}(t) dt \quad (6)$$

does not change the form of the Verhulst equations written now as

$$\frac{d\Sigma I}{dt} = r\Sigma I \left(1 - \frac{\mathbf{R}}{K} \right), \quad (7)$$

where r is the growth rate and K is the saturation value.

Its solution depends on a unique initial condition $\Sigma I(0) = I(0)$. However, the practical impossibility to determine the actual zero time of the outbreak, when the very first patient was infected, leads to the more convenient for data analysis purposes form of the solution of Eq. (7),

$$\Sigma I(t) = \frac{K}{1 + e^{-r(t-t_m)}}, \quad (8)$$

where t_m defines the time moment when the cumulative number of infected persons reaches half of the saturated values. For the Verhulst equation, this moment of time also exactly corresponds to the maximum of the outbreak that is also practically important information from the prognostic point of view.

2.2. Numerical treatment

The data for analysis were downloaded from the official information page of the European Centre for Disease Prevention and Control [20] as a table in .csv format. They cover today's data on the geographic distribution of COVID-19 cases worldwide starting from December 31, 2019 (in some cases, from the date of the first reporting to the ECDC database) up to the day of downloading,

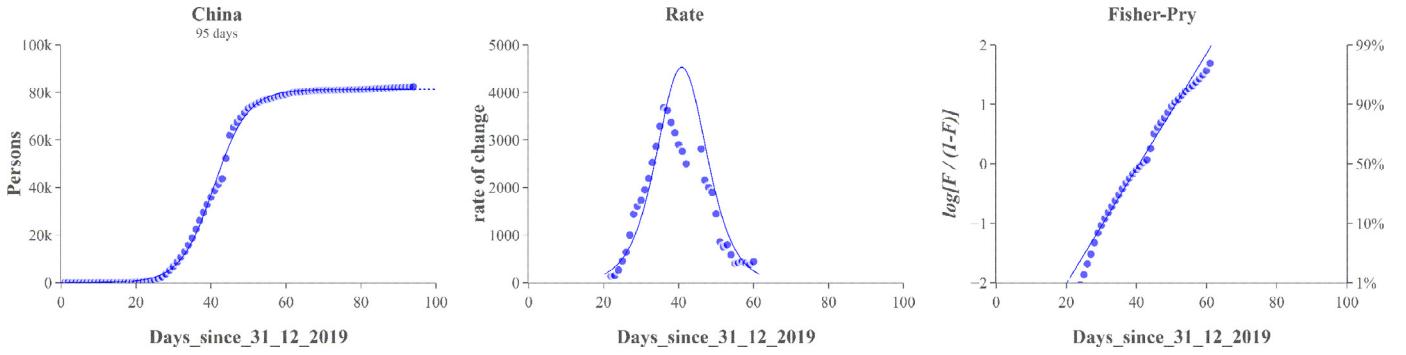


Fig. 1. The comparison of COVID-19 outbreak in China (markers) and its fitting with the SIR model (solid line) reduced to the logistic equation with respect to ΣI (left subpanel) and other characteristics: daily I (rate, middle panel), and the Fisher-Pry representation of the logistic growth, where $F = \Sigma I/K$.

daily. The last day in the database used in the section “Results” is April 3, 2022.

The downloaded table was processed: for each country used for analysis identified by its geoID the column of daily infected cases was extracted and sorted in the descended order of dates (in the ECDC database, new data are added, conversely, to the top of the column). The first day of the records was identified and the time variable stated from this day as $t = 0$ with daily step was formed. Such a two-columns matrix was saved as a file in .csv format for further parameter determination. The MATLAB script, which evaluated these procedures, is given in the Appendix.

The obtained data dale was processed using Loglet Lab 4 software (<https://logletlab.com/>) especially developed for high-quality fitting logistic and related models. A genetic algorithm (GA) based on Monte-Carlo Annealing was applied with the root mean square objective function; the uncertainty in parameter values was estimated using the bootstrap method. The Monte Carlo ensemble consisted of 10^4 iterations for both procedures, the confidence interval is equal to 95 %.

As one of the specific features applied in Loglet Lab for both fitting procedure and output representation, is the Fisher-Pry transform [22], which introduces the variable alteration (in the notation used in this work and in figures below) $F = \Sigma I/K$ and transformation of the logistic function (8) transforms to the linear dependence

$$\frac{F}{1-F} = -r(t - t_m).$$

In addition to computational issue, this representation provides visual guidance for estimating time intervals, when the data already satisfies (if satisfies) the solution of the SIR model in the logistic representation that is one of the targets of this work.

3. Results

The first test evaluated is naturally related to China, where the active outbreak is finished, and the situation with the cumulative number of infected persons reached the saturation, see Fig. 1(left panel). For this quantity, the fitting with the solution (8) not only visually gives reproducing the actual dynamics but also has a relatively high value for the coefficient of determination $R^2 = 0.982$. However, ΣI is an integral value, which smooths instant dynamics for the daily number of infected persons.

This takes place actually, see Fig. 1(middle panel), where the real data markers correspond to the solid line curve representing the rate of change of the logistic curve

$$\frac{d\Sigma I}{dt} \equiv I = \frac{K r t e^{-r(t-t_m)}}{(1 + e^{-r(t-t_m)})^2} \quad (9)$$

more qualitatively than quantitatively.

In more details, one can note that they go close to each other during the first stage of the outbreak but start deviate drastically not reaching the maximum, which is completely “cut-off” the peak and lead to the fast decay of the number of infected persons. This effect is definitely connected with the quarantine measures introduced by the Chinese government that is supported by several studies building more complex models aligned with the timeline of these measures, e.g. [6,7].

At the same time, the last stage of the decaying outbreak again follows the bell-shaped curve given by Eq. (9) that is more-or-less natural from the point of view of the theory of dynamical systems since the solution goes fast to the stationary point via the universal exponential function.

As well, the total duration of the outbreak is also well reproduced by the simple logistic model. It is confirmed by Fig. 1(right panel), where the Fisher-Pry plot is shown. Although the real data markers visibly fluctuate around the fitting straight line, they are most close to the latter in the first growing stage of the outbreak, i.e. the prognosis of the total duration made before the expected maximum gave the correct total duration. Thus, the main effect of quarantine measures, which changed the functional form of the plot was mainly in a drastic reduction of the number of affected people during these predicted time interval that is definitely the most important form the point of view of public health.

However, the current situation in several other countries argues in favour of better applicability of the proposed simplified model. Italy can be considered as the most demonstrable example since it is a country the most heavily affected among European ones now and the time course of the outbreak seems to leave the maximum behind. This provides an opportunity to explore a predictive capacity of the simplified model taking into consideration different time intervals from the outbreak’s beginning as shown in Fig. 2. The respective of the parameters and their uncertainty found from the data fitting with different lengths of time series are given in Table 1.

From the set of plots in Fig. 2 one can see that in the case of epidemic outbreak in Italy, in contrast to the case of China, the correspondence between the solutions of the SIR model reduced to the logistic equation and the real recordings is quantitatively accurate. It is seen not only from the plots of the cumulative number ΣI ($R^2 = 0.995$ for the first plot based on the shortest sample, and $R^2 = 0.997$ for the rest three) but also in the subpanels illustrating the daily number of infected persons, i.e. the rate of change of the logistic function, where markers are slightly and symmetrically scattered around the bell-shaped functions (9) as well as in the Fisher-Pry plots, where the data accurately follow the straight line after some short setting interval affected by small numbers of the daily data counts.

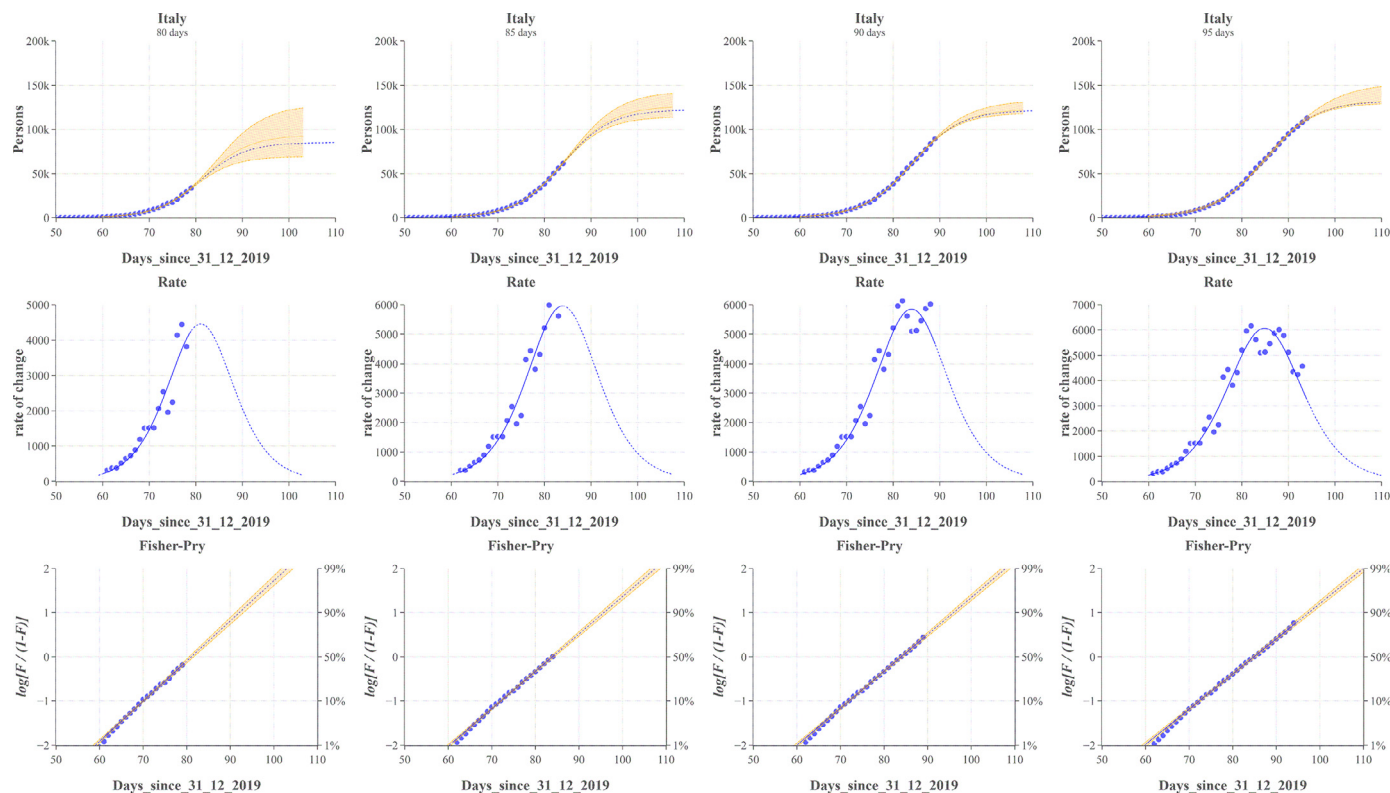


Fig. 2. Series of plots, which demonstrate logistic curve fittings (solid parts of curves) and predictions(dashed parts) using the data (markers) registered for different time intervals from the outbreak's beginning (80, 85, 90, and 95 days since 31/12/2019, day 0). The yellow filled areas denote uncertainty ranges obtained via the Monte Carlo simulations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Parameters of the data fitting for Italy corresponding to Fig. 2. Here *Fitted* means the direct fitting of the data, *Min*, *Max*, and *Median* are the boundary values and the median for the distribution obtained via the Monte Carlo simulations (note that the order of magnitudes for *r* values is reverse to the order for *K* and *t_m* but they are kept in this order to keep the set of parameters for each curve in the same column. Dimensionalities for *K*, *r*, and *t_m* are Persons, day⁻¹ and day, respectively. The notation in parenthesis denotes the number of days since 31/12/2019, for which the data were used for fitting.

Parameter	<i>Fitted</i>	<i>Min</i>	<i>Max</i>	<i>Median</i>
<i>K</i> (80 days)	84784	70917	124125	92812
<i>r</i> (80 days)	0.209	0.219	0.190	0.203
<i>t_m</i> (80 days)	81	79	84	81
<i>K</i> (85 days)	123148	113506	136838	123024
<i>r</i> (85 days)	0.194	0.201	0.186	0.194
<i>t_m</i> (85 days)	83	83	84	83
<i>K</i> (90 days)	121848	118000	135206	122401
<i>r</i> (90 days)	0.193	0.197	0.179	0.191
<i>t_m</i> (90 days)	84	83	85	84
<i>K</i> (95 days)	132455	130766	152018	133371
<i>r</i> (95 days)	0.183	0.187	0.161	0.182
<i>t_m</i> (95 days)	85	85	87	85

Thus, it is possible to conclude that the SIR-model in its logistic growth representation allows the accurate fitting the epidemic process in the case of Italy from the accelerated growth stage to the stage of the epidemic's decay (decelerated growth). However, another important question arises: what is the capacity of this model for predicting further epidemic course. The sequence of fittings limited by different final dates separated by ten-days intervals gives information for such analysis. First of all, it should be pointed out that the growth of sample leads to the di-

minishing value of the coefficient *r* and the subsequent growth of the coefficient *K* (the maximal estimated number of infected person that is on of the key demanded characteristics expected from the prognosis). Such behaviour is not only typical for the logistic regression [23] but also has an explanation based on the model's constructions: comparing Eqs. (4) and (7) one can see that both *r* and *K* depend on one parameter $\mathfrak{R}_0$ reciprocally. Whence, the decay of *r* from 0.209 down to 0.183 is not connected with the actual reducing the averaged contact rate (this case would result in the changing shape of the rate's bell curve as in the case of China that is not observed in the case of Italy) but with the sequential better and better adjustment of the logistic curve to the longer and longer time series of data. This conclusion is also supported by the fact that all Fisher–Pry plots are accurately linear. At the same time, an application of the Monte Carlo method for estimating parameters uncertainty provides some opportunity for forecasting and estimation its reliability even when the outbreak still does not reach its peak. This method not only gives a range of predicted values, which is wider for less warrant early forecast and squeezes to the purely fitted value when the processed sample gets longer, but also propose some criteria for estimating relevance of the extrapolation. In particular, there is visible trend in the displacement of the predicting extrapolation of the simple fitting to the lower boundary of the prediction uncertainty defined by the Monte Carlo method. An explanation of this trend is based on the better relevance of the fitting for the data, which are in the regime of decelerating growth, i.e. the rate maximum is behind. Thus, the uncertainty in the data, which will be added, lead to the possible growth of the cumulative variable considered. On the contrary, before the rate maximum both uncertain results are possible, over-fitting as well as under-fitting. From the practical point of view, for predictions on earlier stages, it is better to take into

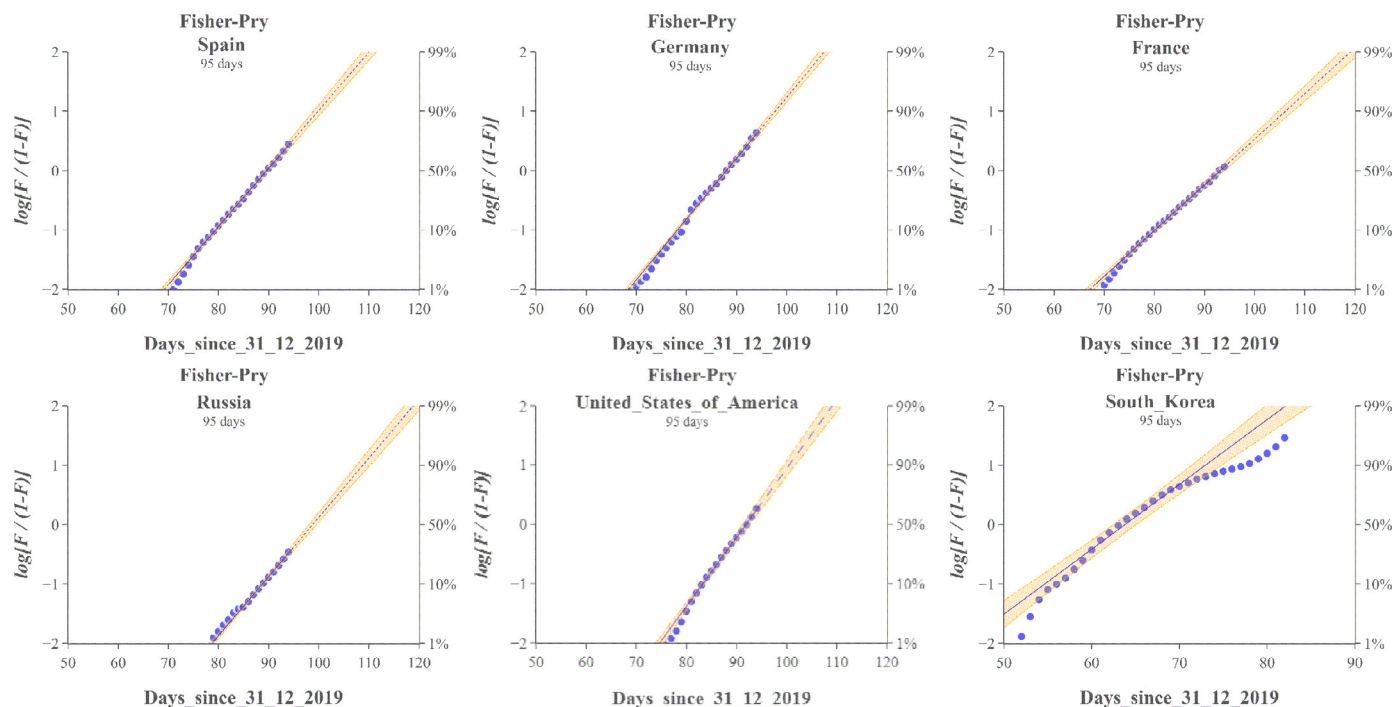


Fig. 3. The Fisher-Pry plots for different countries based on the data available to April 3, 2020 (markers). The solid and dashed line show their fitting and forecast, respectively; the yellow area covers the uncertainty range. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

account the Monte Carlo-based upper boundary, which as it seen in Fig. 2 and Table 1 gives reasonable estimations even with such a simple model is used.

To determine, either the case of Italy is an exception or the reduced SIR model is already wide applicable, such tests were evaluated for a number of other countries and argued in the positive answer for the majority of them. Some typical examples for different stages of outbreak are shown in Fig. 3, where the Fisher-Pry representations on the date of April 3rd of 2020 are given.

One can see that five out of six given examples follow the linearity of the Fisher-Pry plot; moreover, this behaviour practically does not depend on the current stage of the epidemic outbreak and is observed for Germany and Spain, where the raw data and fitting indicate the decelerating phase, for France, which seems to be close to the peak as well as for Russia where the data argue the first, accelerating, phase. Except for France, where the $r = 0.17 \div 0.18$, three other countries indicate similar $r \approx 0.215 \pm 0.015$ by the line's slope with uncertainties diminishes as the data approaching the latter phase. As for the USA, the Fisher-Pry plot is also linear but with the drastically larger $r = 0.26 \pm 0.01$. However, it is clearly seen that the SIR is not universal for the COVID-19 pandemic. In addition to the partial deviation for China visible in Fig. 1, Fig. 3 indicates that South Korea is a complete exception: the plot in Fisher-Pry's variables is not linear almost everywhere; the only short linear interval is located around the middle point (50%), where the maximum of I occurs, i.e. is simply connected with the linearity of ΣI around this point, therefore it is a pure mathematical artefact.

4. Discussion and conclusions

The main conclusion is that the evaluated analysis shows that the SIR being sequentially reduced to the Verhulst equation can provide an accurate description of the COVID-19 epidemic on the level of a country-wide averaged statistics. Moreover, the time course of outbreaks follows such functional dependence practically

as a rule. The clearly visible exceptions are exhibited by the countries, where very strong and wide-ranged measures were undertaken. First of all it is China, and the change of epidemic dynamics is visible in Fig. 1 (middle panel) that allows to assign and localise in time an effect of this measures be a deviation from the "natural logistic course" typical for the majority of countries. The second demonstrable example is South Korea, where completely different but also effective kind of measures was undertaken [24]. This is also reflected in the respective subpanel of Fig. 3, where the Fisher-Pry representation of the data differs from both China and other countries represented there. Thus, it is possible to conclude that the proposed SIR-Verhulst representation can play the role of a criterion for quantification such public health measures.

For other countries, for which illustrations are given in Figs. 2 and 3 as well as many others the strictly logistic growth of the cumulative number of infected persons is demonstrated as typical. Note also a recent study [15] the logistic regression for a wide range of countries and their territories was evaluated. However, the solution of the Verhulst equation was applied just as a kind of sigmoidal function without its linking to any epidemiological model, and, as a result, reliability of the obtained parameters were discussed on the base of pure computational issues only.

On the contrary, in this work, the parameters of the logistic function are proven as connected with the parameters of the well-established Kermack-McKendrick model that reveal their meaning in the context of the epidemic process. In particular, the proven interrelation between r and K parameters indicates that the diminishing of r during the outbreak can originate not from reducing contact rate but principally from the better regression accuracy when the time series get longer and take into account the normal effect of deceleration characterising the full interval of the outbreak. Simultaneously, this provides an opportunity to estimate the quality of the logistic curve-based forecast and its uncertainty more warrantably.

Form this point of view, the Fisher-Pry plot is significant since it indicates the logistic character of the cumulative number of in-

fectured persons when this plot is linear even if its slope determined for time series of different length is different. This is a reason why only such plots are given for different countries in Fig. 3. In addition, the length of data series grows daily; therefore more and more accurate estimations will be available but their possibility is assured namely by this form of data and model representation.

Declaration of Competing Interest

The author declare that he has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A

The script for MATLAB, which reads the file downloaded from ECDC site with the standard name 'COVID-19-geographic-disbtribution-worldwide.csv' and transforms it into the text file readable by Loglet Lab 4 software. The last day number since 31/12/2019 considered during a fitting in the time series is assigned to the variable `t_end`. The script also supplies the output file with columns names and the file name, which coincides with the country name and the used duration of the time period (in particular, they will automatically be displayed in Loglet Lab's output pictures, see the figures above).

```
clear;close all;clc;
TBL=readtable('COVID-19-geographic-
disbtribution-worldwide.csv');
countries=TBL.geoId;
geoIdcode='CN';
idx=find(strcmp(countries,geoIdcode));
len=length(idx);
I=TBL.cases(idx);
I=flipud(I);
t=0:len-1;
t=t';
D=TBL.day(idx);
D=flipud(D);
dD=abs(diff(D));
dmissd=find((dD>1)&(dD<28));
tu=flipud(t);
for j=1:length(dmissd);
t(dmissd(j)+1:end)=t(dmissd(j)+1:end)
+dD(dmissd(j))-1;
end
csI=cumtrapz(t,I);
t_end=98;
ind_end=find(t==t_end);
u(:,1)=t(1:ind_end);
u(:,2)=csI(1:ind_end);
indnon100=find(u(:,2)>=100);
A=['Days_since_',cell2mat(TBL.dateRep
(idx(end)))];
A=strrep(A,'/','_');
out=array2table(u,'VariableNames',A,'Persons');
outname=[cell2mat(TBL.countriesAndTerritories
(idx(1))),'.
(num2str(u(end,1)+1),'_days').csv'];
writetable(out,outname);
```

References

- [1] Sohrabi C, Alsafi Z, O'Neill N, Khan M, Kerwan A, Al-Jabir A, Iosifidis C, Agha R. World health organization declares global emergency: a review of the 2019 novel coronavirus (COVID-19). *Int J Surg* 2020;76:71–6. doi:10.1016/j.ijssu.2020.02.034.
- [2] He F, Deng Y, Li W. Coronavirus disease 2019 (COVID-19): what we know? *J Med Virol* 2020. doi:10.1002/jmv.25766.
- [3] Wang L-s, Wang Y-r, Ye D-w, Liu Q-q. A review of the 2019 novel coronavirus (COVID-19) based on current evidence. *Int J Antimicrob Agents* 2020:105948. doi:10.1016/j.ijantimicag.2020.105948.
- [4] Rhodes T, Lancaster K, Rosengarten M. A model society: maths, models and expertise in viral outbreaks. *Crit Public Health* 2020. doi:10.1080/09581596.2020.1748310.
- [5] Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, Eggo RM, Sun F, Jit M, Munday JD, et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infectious Dis* 2020. doi:10.1016/S1473-3099(20)30144-4.
- [6] Maier B.F., Brockmann D.. Effective containment explains sub-exponential growth in confirmed cases of recent COVID-19 outbreak in mainland china. *ArXiv:2002.07572*.
- [7] Fang Y, Nie Y, Penny M. Transmission dynamics of the COVID-19 outbreak and effectiveness of government interventions: a data-driven analysis. *J Med Virol* 2020. doi:10.1002/jmv.25750.
- [8] Lin Q, Zhao S, Gao D, Lou Y, Yang S, Musa SS, Wang MH, Cai Y, Wang W, Yang L, et al. A conceptual model for the coronavirus disease 2019 (COVID-19) outbreak in Wuhan, China with individual reaction and governmental action. *Int J Infectious Dis* 2020;93:211–16. doi:10.1016/j.ijid.2020.02.058.
- [9] Fanelli D, Piazza F. Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos Solitons Fractals* 2020;134:109761. doi:10.1016/j.chaos.2020.109761.
- [10] Liu Q, Liu Z, Zhu J., Zhu Y., Li D., Gao Z., Zhou L., Yang J., Wang Q.. Assessing the global tendency of COVID-19 outbreak. *MedRxiv*. 10.1101/2020.03.18.20038224v2
- [11] Calafiore G.C., Novara C., Possieri C.. A modified sir model for the covid-19 contagion in Italy. *ArXiv:2003.14391*.
- [12] Roda WC, Varughese MB, Han D, Li MY. Why is it difficult to accurately predict the COVID-19 epidemic? *Infect Dis Model* 2020;5:271–81. doi:10.1016/j.idm.2020.03.001.
- [13] Zhou X, Hong N, Ma Y, He J, Jiang H, Liu C, Shan G, Su L, Zhu W, Long Y. Forecasting the worldwide spread of COVID-19 based on logistic model and SEIR model. *medRxiv* 2020. doi:10.1101/2020.03.26.20044289.
- [14] Dattoli G., Palma E.D., Licciardi S., Sabia E.. A note on the evolution of covid-19 in Italy. *ArXiv:2003.08684*.
- [15] Tátrai D., Várallyay Z.. COVID-19 epidemic outcome predictions based on logistic fitting and estimation of its reliability. *ArXiv:2003.14160*.
- [16] Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. *Proc R Soc Lond A* 1927;115:700–21. doi:10.1098/rspa.1927.0118.
- [17] Murray JD. *Mathematical biology: i. an introduction*. Springer Science & Business Media; 2007.
- [18] Postnikov EB. Dynamical prediction of flu seasonality driven by ambient temperature: influenza vs. common cold. *Eur Phys J B* 2016;89:13. doi:10.1140/epjb/e2015-50845-7.
- [19] WHO. 2020. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>.
- [20] ECDC. 2020. <https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide>.
- [21] Lavrova AI, Postnikov EB, Manicheva OA, Vishnevsky BI. Bi-logistic model for disease dynamics caused by mycobacterium tuberculosis in Russia. *R Soc Open Sci* 2017;4:171033. doi:10.1098/rsos.171033.
- [22] Fisher JC, Pry RH. A simple substitution model of technological change. *Technol Forecast Soc Change* 1971;3:75–88. doi:10.1016/S0040-1625(71)80005-7.
- [23] Motulsky H, Christopoulos A. *Fitting models to biological data using linear and nonlinear regression: a practical guide to curve fitting*. Oxford University Press; 2004.
- [24] Normile D. Coronavirus cases have dropped sharply in South Korea. what's the secret to its success. *Science* 2020. doi:10.1126/science.abb7566.