

Gompertz model in COVID-19 spreading simulation

E. Pelinovsky^{a,b}, M. Kokoulina^c, A. Epifanova^c, A. Kurkin^{c,*}, O. Kurkina^c, M. Tang^d,
E. Macau^e, M. Kirillin^b

^a National Research University – Higher School of Economics, Myasnitskaya st., 20, Moscow 101000, Russian Federation

^b Institute of Applied Physics RAS, Ulyanov st., 46, Nizhny Novgorod 603950, Russian Federation

^c Nizhny Novgorod State Technical University n.a. R.E. Alekseev, Minin st., 24, Nizhny Novgorod 603950, Russian Federation

^d East China Normal University, Shanghai 200241, China

^e Universidade Federal de São Paulo, São Paulo 04021-001, Brazil

ARTICLE INFO

Article history:

Received 4 November 2021

Accepted 1 December 2021

Available online 4 December 2021

Keywords:

Gompertz model

Logistic equation

Mathematical modeling

COVID-19

ABSTRACT

The paper reports on application of the Gompertz model to describe the growth dynamics of COVID-19 cases during the first wave of the pandemic in different countries. Modeling has been performed for 23 countries: Australia, Austria, Belgium, Brazil, Great Britain, Germany, Denmark, Ireland, Spain, Italy, Canada, China, the Netherlands, Norway, Serbia, Turkey, France, Czech Republic, Switzerland, South Korea, USA, Mexico, and Japan. The model parameters are determined by regression analysis based on official World Health Organization data available for these countries. The comparison of the predictions given by the Gompertz model and the simple logistic model (i.e., Verhulst model) is performed allowing to conclude on the higher accuracy of the Gompertz model.

© 2021 Elsevier Ltd. All rights reserved.

1. Introduction

The recent coronavirus epidemic, which has affected all countries, arouse a great deal of interest in mathematical models that allow describing the dynamics of the growth of new cases and predicting pandemic development. To the surprise of many, it turned out that even the simplest logistic model based on the first order ordinary differential equation provided reasonable accuracy in describing the number of cases in different countries during the first wave of the epidemic [1–6]. The classical logistic equation was proposed by Verhulst in 1838 [7] and has the following form:

$$\frac{dN}{dt} = rN \left(1 - \frac{N}{N_{\infty}} \right), \quad (1)$$

where N is the current number of infected persons, r is the infection rate, and N_{∞} is the total number of infected persons in the considered wave in the region. This model was further generalized by introducing power laws into various terms in Eq. (1), which followed by the development of various SIR-like models taking into account the number of infected, recovered, etc.; see for instance [8–15].

Meanwhile, historically, the Gompertz equation:

$$\frac{dN}{dt} = rN \left(1 - \frac{\ln N}{\ln N_{\infty}} \right) \quad (2)$$

should be recognized as the first model of this class, which was proposed as early as 1825 [2]. This model is not commonly used in describing the dynamics of the epidemic development, although it is mentioned in several papers and books on mathematical biology [16–21]. We especially note one of the recent papers [20], where this model was generalized in the same way as the classical logistic model.

This paper presents the study of the Gompertz model in its different variants (e. g., continuous and discrete times) applied to the COVID-19 spreading in different countries.

2. Gompertz model

It is worth noting that the parameters of the Gompertz Eq. (2) are the same as the parameters of the classical logistic Eq. (1) and, in this connection, their solutions could be compared quantitatively. An analytical solution for the Eq. (2) could be obtained in the following form:

$$N(t) = N_{\infty} \exp \left[\ln \left(\frac{N_0}{N_{\infty}} \right) \exp \left(-\frac{rt}{N_{\infty}} \right) \right], \quad (3)$$

where N_0 is the initial number of infected persons. This solution will be further employed in Section 4 for analysis of coronavirus

* Corresponding author.

E-mail addresses: aakurkin@nntu.ru (A. Kurkin), oksana.kurkina@mail.ru (O. Kurkina), mtang@ce.ecnu.edu.cn (M. Tang), elbert.macau@unifesp.br (E. Macau), mkirillin@yandex.ru (M. Kirillin).

spreading dynamics in different countries. For a qualitative analysis of the Gompertz equation it is convenient to use new variables. After the substitution:

$$\tau = rt/N_\infty, x = N/N_\infty, \tag{4}$$

the Eq. (2) is reduced to:

$$\frac{dx}{d\tau} = -x \ln x, 0 < x \leq 1, \tag{5}$$

In the paradigm of dynamical systems [22,23], the described system has two equilibrium states: $x = 0$ represents an unstable and non-rough state of equilibrium, while $x = 1$ is a stable rough state of equilibrium. All trajectories monotonically tend to a stable equilibrium state:

$$x(\tau) = \exp[\ln(x_0) \exp(-\tau)]. \tag{6}$$

Note, that after the substitution

$$y = \ln(x), \tag{7}$$

the Eq. (5) is reduced to a trivial linear equation

$$\frac{dy}{d\tau} = -y, -\infty < y \leq 0, \tag{8}$$

Discretization of Eq. (8) leads to the Malthusian discrete equation:

$$y_{n+1} = py_n, 0 < p = 1 - \Delta\tau < 1, \tag{9}$$

where $\Delta\tau$ is the time step which is not expected to be large. This transition has a single fixed point $y = 0$, which is stable. The Eq. (9) can easily be solved based on the initial number of infected persons $y_0 = \ln(N_0/N_\infty)$, we obtain

$$y_n = p^n y_0 \text{ or } \frac{N_n}{N_\infty} = \left(\frac{N_0}{N_\infty}\right)^{p^n}, \tag{10}$$

There is no qualitative difference between solutions of Gompertz equation with discrete and continuous time, and for a small time step they go one into another.

In order to account possible delay in the Gompertz equation let us introduce the delay time. Usually the delay time is introduced into model to take into account the infection period (from 2 up to 14 days for COVID-19). This parameter can be contained in different parts of Eqs. (2) and (5). If the first equation term has a delay, the equation has the following form:

$$\frac{dx}{d\tau} = -x(\tau - T) \ln x(\tau), 0 < x \leq 1, \tag{11}$$

where T is the delay time. The zero state of equilibrium in this case remains unstable, while $x = 1$ is stable for any delay time, and the monotonous variation of the number of infected persons is preserved. If the delay is contained in the second term of the Eq. (5), after the substitution (7) the following equation is obtained:

$$\frac{dy}{d\tau} = -y(\tau - T), -\infty < y \leq 0, \tag{12}$$

This is equation was studied in details earlier [24,25]. The equilibrium state $y = 0$ ($x = 1$) is stable when

$$T < \frac{\pi}{2} \tag{13}$$

The Eq. (12) can be easily solved by the linear multistep method, and in the case of the constant number of new cases in the time interval $\tau < T$ its solution is expressed by

$$y(\tau) = y_0 \sum_{n=0}^{E(\tau/T)-1} \frac{(-1)^n}{n!} [\tau - nT]^n. \tag{14}$$

The function y monotonically tends to zero with time (while x tends to unity), and only the rate of this tendency depends on the delay time. Note, that the presence of a delay time increases the

number of parameters, and, hence, the dimensionality of the problem.

The described above features of the Gompertz equation solutions in different forms will be used below to analyze the first wave of the COVID-19 pandemic in different countries.

3. Analysis of the COVID-19 pandemic data for determination of the model coefficients

Let us compare the two mentioned above models based on logistic (1) and Gompertz (2) equations in description of the COVID-19 epidemic progression using the official data for the first wave of coronavirus infection in 23 countries: Australia, Austria, Belgium, Brazil, Great Britain, Germany, Denmark, Ireland, Spain, Italy, Canada, China, Netherlands, Norway, Serbia, Turkey, France, Czech Republic, Switzerland, South Korea, USA, Mexico, and Japan. Initial data on the number of cases were taken from the World Health Organization website (<https://www.who.int/>). The data is updated daily, thus, the derivative $dN/dt = K$ in the logistic equation can be associated with the number of daily new cases. Since our study is limited by the first wave of the epidemic only, a procedure for separation of the epidemic waves was employed based on calculation of the autocorrelation function. The first wave is determined as the period between the first two zero values for the autocorrelation function.

Fig. 1 shows official statistical data on daily new cases respect to the number of total cases together with two analytical approximations using Eqs. (1) and (2) for 23 countries. The official statistics data features a fast increase in the number of new cases in the beginning of the pandemic followed by a slower decrease by the end of the considered period, which is repeated by the analytical solution (red lines) obtained in the framework of the Gompertz model. Since the logistic equation produces a symmetric solution (blue curves) respect to the center of the half number of the total cases, the Gompertz equation qualitatively provides a more accurate fit of the official data owing to higher flexibility.

The standard approach to determination of the model parameters consists in minimization of the least mean square deviation of the model data from the official statistics.

Table 1 shows the calculated coefficient of determination, R^2 together with coefficients r and N_∞ . The coefficient of determination is computed according to the equation

$$R^2 = 1 - \frac{\sum_{i=1}^n (K_i - \bar{K}_i)^2}{\sum_{i=1}^n (K_i - \bar{K}_i)^2}, \tag{15}$$

where K_i is the number of infected persons on day i , \bar{K}_i is model value, and \bar{K}_i is average value.

First of all, it is interesting to analyze the ratio of number of the total infected people to number of population in each country (Fig. 2a). Such values confirm the dangerous character of the COVID-19 epidemic. It is worth mentioning that Asian countries (except Turkey) demonstrate the smallest values of this ratio, while, on the other hand, American countries (except Canada) demonstrate the highest value. All the European countries demonstrate moderate values exceeding those for Asian one, however, smaller than that for American ones. Australia demonstrates the value below that for European countries, however, higher than that for Asian ones. Although, the set of countries is not fully representative, one could expect typical regional values for different parts of the world. It is worth noting, that despite their geographical location, Turkey and Canada demonstrate values close to those typical for European countries.

Correlation of the parameters derived from the both considered models (i.e., Gompertz and logistic) is shown in Figs. 3–5. First of

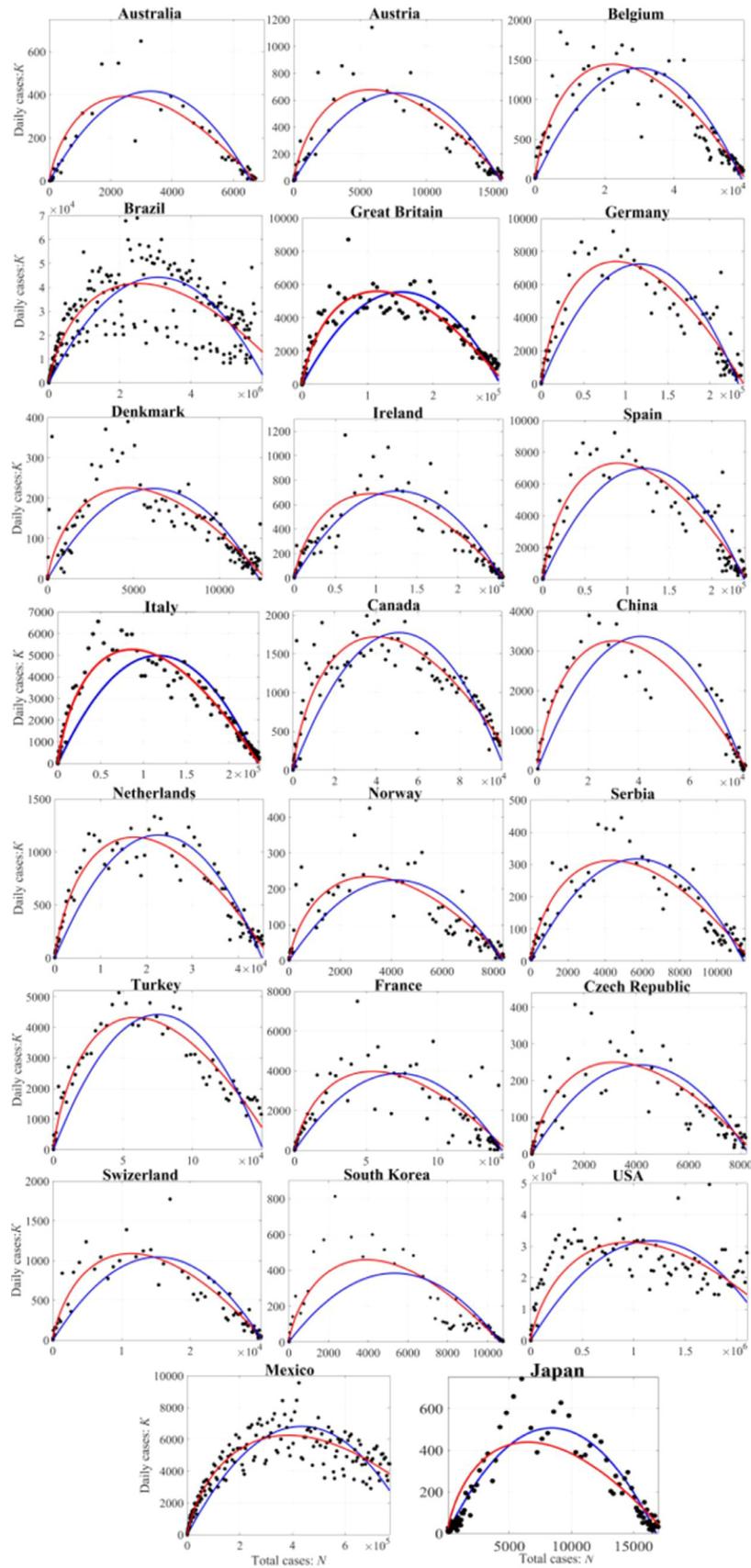


Fig. 1. Number of daily new cases K versus total number of cases N and corresponding analytical approximations with logistic model (blue line) and Gompertz model (red line) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Model and determination coefficients for two considered approximations (N_p is population in each country).

Model		Gompertz equation				Simple logistic equation			
№	Country	r	N_∞	R^2	$N_\infty / N_p \times 10$	r	N_∞	R^2	$N_\infty / N_p \times 10$
1	Australia	1.297	6974	0.81	0.003	0.21	6901	0.73	0.003
2	Austria	1.079	16,155	0.82	0.019	0.16	15,950	0.67	0.018
3	Belgium	0.743	59,278	0.79	0.052	0.1	57,350	0.66	0.050
4	Brazil	0.26	6,834,000	0.67	0.329	0.03	6,120,000	0.58	0.295
5	Great Britain	0.6	318,100	0.89	0.048	0.074	302,300	0.75	0.046
6	Germany	0.9	183,500	0.83	0.022	0.1	181,300	0.67	0.022
7	Denmark	0.5	12,580	0.74	0.022	0.075	12,110	0.64	0.021
8	Ireland	0.75	25,620	0.8	0.054	0.12	24,990	0.79	0.053
9	Spain	1.035	240,385	0.86	0.052	0.12	234,800	0.73	0.051
10	Italy	0.749	253,625	0.92	0.043	0.12	234,800	0.73	0.040
11	Canada	0.5	108,700	0.78	0.030	0.069	101,900	0.67	0.028
12	China	1.21	82,700	0.91	0.001	0.16	81,860	0.8	0.001
13	Netherlands	0.705	47,099	0.90	0.028	0.1	45,560	0.79	0.027
14	Norway	0.701	8350	0.73	0.016	0.114	8182	0.6	0.016
15	Serbia	0.69	11,650	0.83	0.013	0.12	11,140	0.77	0.013
16	Turkey	0.826	166,042	0.89	0.021	0.1	161,100	0.62	0.020
17	France	0.84	150,700	0.7	0.023	0.1	148,100	0.57	0.023
18	Czech Republic	0.775	8191	0.71	0.008	0.134	7846	0.58	0.007
19	Switzerland	1.005	30,638	0.87	0.036	0.14	30,230	0.76	0.036
20	South Korea	1.026	10,829	0.72	0.002	0.13	10,850	0.49	0.002
21	USA	0.56	2,247,000	0.78	0.068	0.05	2,351,000	0.52	0.071
22	Mexico	0.23	1,044,000	0.89	0.08	0.031	869,200	0.88	0.067
23	Japan	0.66	17,670	0.73	0.001	0.12	16,640	0.89	0.001

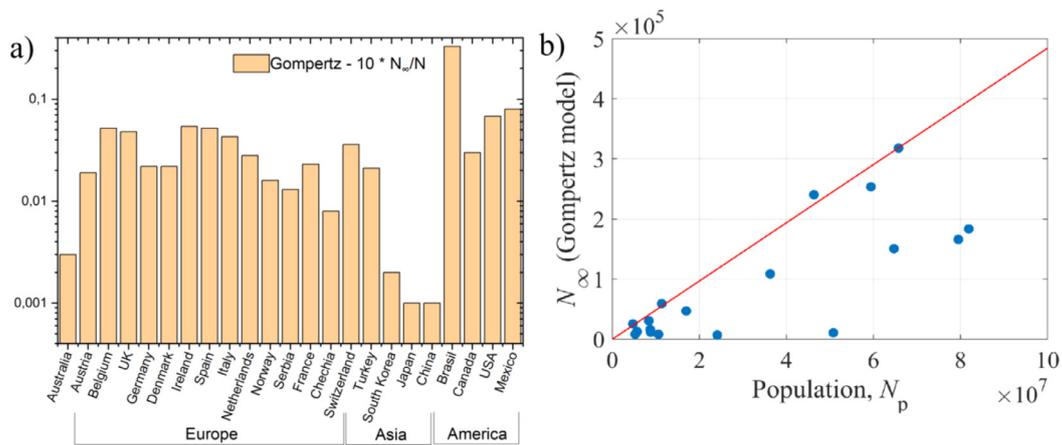


Fig. 2. Maximal number of infected persons N_∞ normalized for country population N_p (a) and maximal number of infected persons N_∞ versus N_p (b) in Gompertz model.

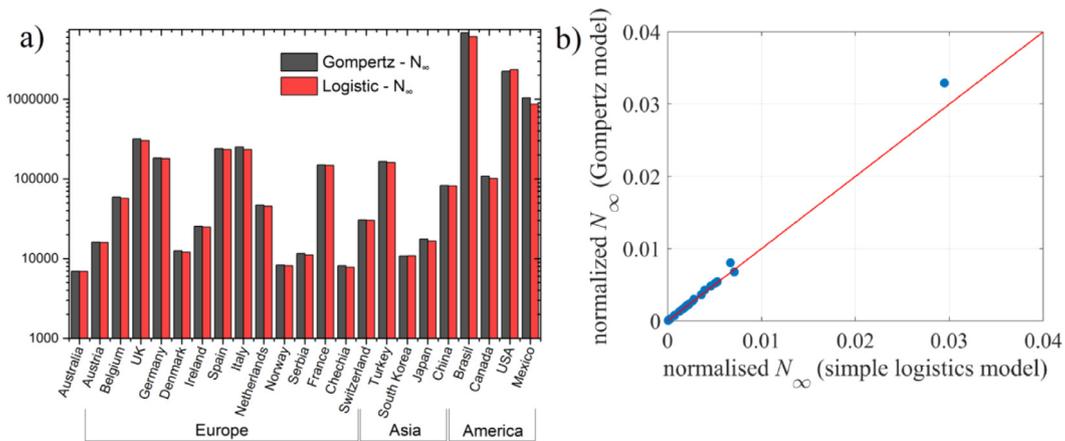


Fig. 3. Maximal number of infected persons N_∞ in simple logistic model and Gompertz equation for all considered countries (a) and correlation of the N_∞ normalized for total population N_p for the two considered models (b).

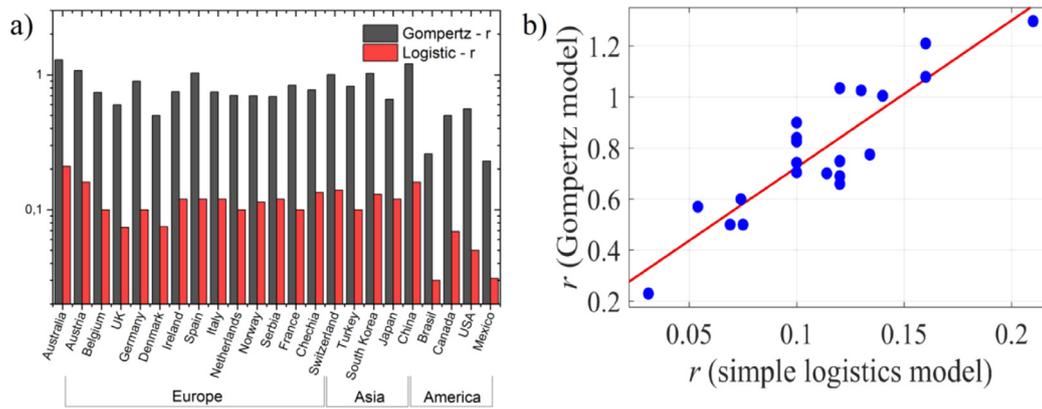


Fig. 4. Infection rate r in simple logistic model and Gompertz equation (a) and r values acquired with Gompertz models versus logistic equation (b) for all considered countries. The red line shows regression line with coefficient of 5.9 (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

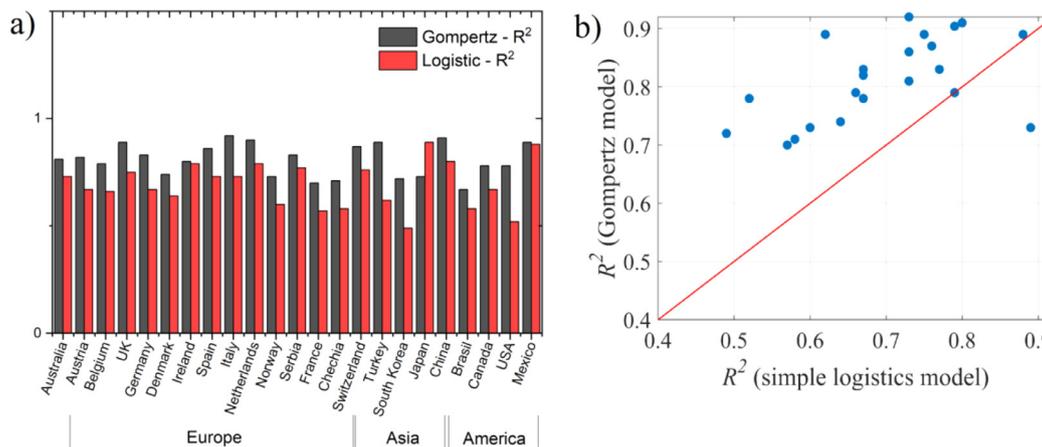


Fig. 5. Comparison of the determination coefficients R^2 for the logistic equation and the Gompertz model.

all, it is worth noting that the estimations of the total number of infected persons for the two models are in good agreement, which is quite expected [see Fig. 3(a,b)].

At the same time, the infection rate is about 6-fold higher for the Gompertz model as compared to logistic equation (Fig. 4), which originates from a more accurate description on the left part of the official statistics data (initial development of epidemic) in Fig. 1, which is faster than the relaxation period (right part in Fig. 1). Note, that for this case the regional effect is mostly pronounced for American countries demonstrating lower infection rate as compared to other continents.

The coefficient of determination R^2 appears to be higher for Gompertz model as compared to the logistic equation for about 13% (Fig. 5) indicating better approximation of the official statistics curve. The results of the comparison allow concluding that the Gompertz model is more suitable for describing the COVID-19 pandemic spread owing to asymmetric shape of the provided solutions allowing to more accurately reproduce the real dependencies. From all the considered countries only Japan shows better agreement with the logistic model as compared to the Gompertz model (Table 1), since its official statistic demonstrates a symmetric picture (see Fig. 1, Chart 21).

4. Temporal dynamics of the total infected cases in the Gompertz model

The Gompertz model also allows to reproduce the temporal dynamics of the total new cases. Corresponding dynamics for all the considered countries were calculated using formula (3) employing the coefficient values shown in Table 1. Fig. 6 confirms good agreement of the Gompertz model with official statistics. However, in several countries (Netherlands, Serbia, Turkey, and Czech) a discrepancy is observed in the end of the considered period. In these countries, the first wave has evolved into the second one without an evident break. In this connection, the saturation of the total number of new cases predicted by the model is not observed in official statistics. Nevertheless, a good agreement is observed for these countries in the active development phase of the first wave. Moreover, the point where the official statistics deviates from the model can be treated as the start of the second wave, and the parameters of the second wave could be derived starting from this point.

Analysis of the Gompertz model with the delay term [Eqs. (11) and (12)] demonstrated (we omit details of calculations), that the best fit solution for all the considered countries is provided in the

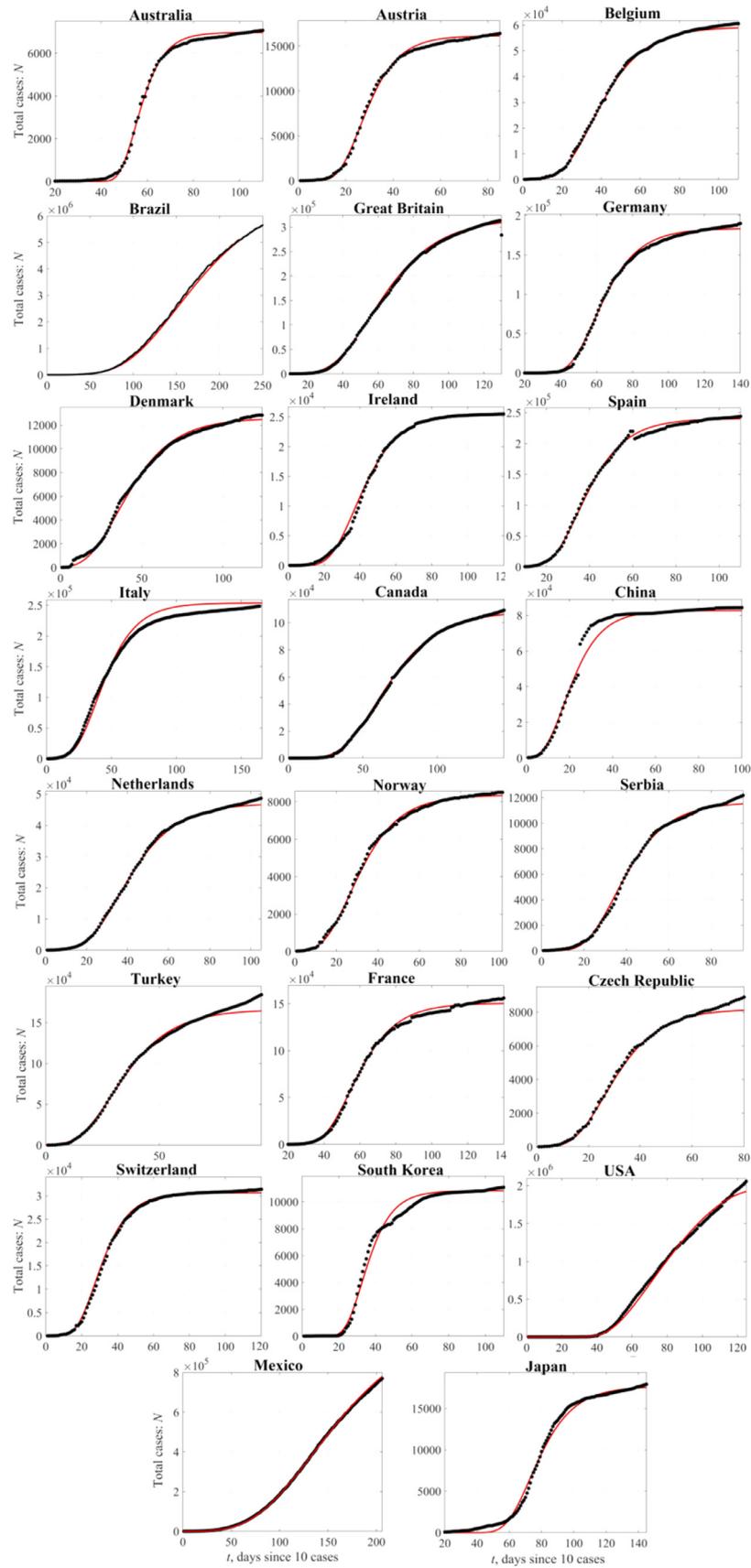


Fig. 6. Total number of new cases versus time: official statistics (black dots) and solution of the Gompertz equation (red line) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

case without delay, moreover, the presence of T_2 delay results in higher deviation from the solution without delay as compared to T_1 .

5. Conclusion

The paper reported the analysis of the temporal dynamics of the number of infected persons during the first wave of COVID-19 pandemic in 23 countries: Australia, Austria, Belgium, Brazil, Great Britain, Germany, Denmark, Ireland, Spain, Italy, Canada, China, Netherlands, Norway, Serbia, Turkey, France, Czech Republic, Switzerland, South Korea, USA, Mexico, Japan. Observed data were analyzed in the frames of classical logistic model and Gompertz model, both models have minimum number of the parameters (two) as compared to SIR-like models. The Gompertz model was demonstrated to provide better fit of official statistics data compared to the solutions provided by the classical logistic equation in all considered countries except Japan, which is presumably determined by local features. The total new cases dynamics derived from the Gompertz equation using the parameters extracted from fitting dependencies of daily new cases versus total case number provided good approximation of the official statistics, except countries where the first and the second waves are not strictly separated. For these countries a discrepancy was observed in the end of the considered period, which can be employed as an approach for revealing the start of the second wave. Employment of the Gompertz model with delay for fitting official statistics data did not reveal any effect of the delay time in providing better fit of the official statistics.

In conclusion we would like to mention again, the discrete Gompertz model for small discretization time gives the same solutions as the continuous model. In this connection, the calculations in the frames of the discrete model were not performed. The same can be pointed for actively discussed effects associated with the delay time. Our computations with time delays in few days lead to worse agreement with official data as compared to the model without delay.

Declaration of Competing Interest

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

Acknowledgment

The study is supported by joint RFBR (project no. 20–51–80004), CNPq (project no. 441016/2020–0), and NSFC (project no. 82161148012).

References

- [1] Carletti T, Fanelli D, Piazza F. COVID-19: the unreasonable effectiveness of simple models. *Chaos Solitons Fractals X* 2020;5:100034. doi:10.1016/j.csf.2020.100034.

- [2] Consolini G, Materassi MA. Stretched logistic equation for pandemic spreading. *Chaos Solitons Fractals* 2020;140:110113. doi:10.1016/j.chaos.2020.110113.
- [3] Pelinovsky E, Kurkin A, Kurkina O, Kokoulina M, Epifanova A. Logistic equation and COVID-19. *Chaos Solitons Fractals* 2020;140:110241. doi:10.1016/j.chaos.2020.110241.
- [4] Kokoulina MV, Epifanova AS, Pelinovsky EN, Kurkina OE, Kurkin AA. Analysis of coronavirus dynamics using the generalized logistic model. *Trans NNTU n.a. R E Alekseev* 2020;3:28–41. doi:10.46960/1816-210X_2020_3_28.
- [5] Duhon J, Bragazzi N, Kong JD. The impact of non-pharmaceutical interventions, demographic, social, and climatic factors on the initial growth rate of COVID-19: a cross-country study. *Sci Total Environ* 2021;760:144325. doi:10.1016/j.scitotenv.2020.144325.
- [6] Romanovski MA. Locally sequential refinement of the growth dynamics identification. *Inverse Probl Sci Eng* 2021;1–38. doi:10.1080/17415977.2021.1948025.
- [7] Verhulst PF. Notice sur la loi que la population suit dans son accroissement. *Corresp Math Phys* 1838;10:113–26.
- [8] Lee SY, Lei B, Mallick B. Estimation of COVID-19 spread curves integrating global data and borrowing information. *PLoS One* 2020;15:e0236860. doi:10.1371/journal.pone.0236860.
- [9] Yang Z, Zeng Z, Wang K, Wong SS, Liang W, Zanin M, He J. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *J Thorac Dis* 2020;12:165 3. doi:10.21037/jtd.2020.02.64.
- [10] Liang K. Mathematical model of infection kinetics and its analysis for COVID-19, SARS and MERS. *Infect Genet Evol* 2020;82:104306. doi:10.1016/j.meegid.2020.104306.
- [11] Long YS, Zhai ZM, Han LL, Kang J, Li YL, Lin ZH, Lai YC. Quantitative assessment of the role of undocumented infection in the 2019 novel coronavirus (COVID-19) pandemic. *arXiv preprint*; 2020. arXiv:2003.12028 .
- [12] Zhai ZM, Long YS, Tang M, Liu Z, Lai YC. Optimal inference of the start of COVID-19. *Phys Rev Res* 2021;3:013155 1. doi:10.1103/PhysRevResearch.3.013155.
- [13] Pacheco CC, de Lacerda CR. Function estimation and regularization in the SIRD model applied to the COVID-19 pandemics. *Inverse Probl Sci Eng* 2021;1–16. doi:10.1080/17415977.2021.1872563.
- [14] Memarabashi R, Mahmoudi SMA. Dynamic model for the COVID-19 with direct and indirect transmission pathways. *Math Methods Appl Sci* 2021;44:5873–87. doi:10.1002/mma.7154.
- [15] Turkyilmazoglu M. Explicit formulae for the peak time of an epidemic from the SIR model. *Phys D Nonlinear Phenom* 2021;422:132902. doi:10.1016/j.physd.2021.132902.
- [16] Hethcote HW. The mathematics of infectious diseases. *SIAM Rev* 2000;42:599–653 4. doi:10.1137/S0036144500371907.
- [17] Pastor J. *Mathematical ecology of populations and ecosystems*. Wiley-Blackwell; 2008. p. 344.
- [18] Rubin A, Riznichenko GY. *Mathematical biophysics* 2014;15:273. doi:10.1007/978-1-4614-8702-9.
- [19] Chakraborty B, Bhattacharya S, Basu A, Bandyopadhyay S, Bhattacharjee A. Goodness-of-fit testing for the Gompertz growth curve model. *Metron* 2014;72:45–64 1. doi:10.1007/s40300-013-0030-z.
- [20] Tjørve KMC, Tjørve E. The use of Gompertz models in growth analyses, and new Gompertz-model approach: an addition to the Unified-Richards family. *PLoS One* 2017;12:e0178691 6. doi:10.1371/journal.pone.0178691.
- [21] Paul A, Reja S, Kundu S, Bhattacharya S. COVID-19 pandemic models revisited with a new proposal: plenty of epidemiological models outcast the simple population dynamics solution. *Chaos Solitons Fractals* 2021;144:110697. doi:10.1016/j.chaos.2021.110697.
- [22] Strogatz SH. *Nonlinear dynamics and chaos: with applications to physics, biology and chemistry*. Perseus; 2001. doi:101063/PT32751.
- [23] Jackson T, Radunskaya A. *Applications of dynamical systems in biology and medicine*, 158. ed. Springer; 2015. doi:10.1007/978-1-4939-2782-1.
- [24] Bodnar M, Piotrowska MJ, Forys U. Gompertz model with delays and treatment: mathematical analysis. *Math Biosci Eng* 2013;10:551 3. doi:10.3934/mbe.2013.10.551.
- [25] Qian Y, Qi W. Numerical oscillation analysis for Gompertz equation with one delay. *Fundam J Math Appl* 2020;3:1–7 1. doi:10.33401/fujma.623500.