

A mathematical model for the coronavirus COVID-19 outbreak

Roman Cherniha ¹ and Vasyl' Davydovych ²

*Institute of Mathematics, National Academy of Sciences of Ukraine,
3, Tereshchenkivs'ka Street, Kyiv 01004, Ukraine*

Abstract

A mathematical model is proposed for quantitative description of the outbreak of novel coronavirus COVID-19 in China. Although the model is relatively simple, the comparison with the public data shows that an exact solution solution of the model (with the correctly-specified parameters) leads to the results, which are in good agreement with the measured data. Prediction of the total number of the COVID-19 cases is discussed and an example is presented using the measured data in Austria.

Keywords: mathematical model; modeling infectious diseases; logistic equation; exact solution.

2010 MSC: 92D30; 34C11.

1 Introduction

The outbreak of novel coronavirus called COVID-19 in China has attracted extensive attention of many scientists, in particular mathematicians working in mathematical modeling. The first papers were already published in February and March 2020 [1–5]. At the present time, there is an oblivious threat that the COVID-19 outbreak will spread over the world as a pandemic. There were almost 1300000 coronavirus cases up to date April 6 [6].

At the present time, there are many mathematical models used to describe epidemic processes and they can be found in any book devoted to mathematical models in biology and medicine (see, e.g., [7–10] and papers cited therein). The paper [11] is one of the first papers in this direction. The authors created a model based on three ODEs, which nowadays is called the SIR model. There are several generalizations of the SIR model and the SEIR model [12, 13], which involves four ODEs, is the most common among them.

Here (Sections 2 and 3) we propose a simple model, which was developed using the data from [6] in the case of the COVID-19 outbreak in China. This case was used because there are obvious indications that this epidemic threat was effectively removed in China. A prediction of the total number of the COVID-19 cases is discussed and an example is presented using the measured data in Austria (Section 4).

¹Corresponding author. E-mail: r.m.cherniha@gmail.com

²E-mail:davydovych@imath.kiev.ua

2 Mathematical model

The first nontrivial biological model used for calculation and the time evolution of the total world population of people was created in 1838 by Verhulst [14]. His model is usually called the logistic model and has the form (in dimensionless variables)

$$\frac{dU}{dt} = U(1 - U), \quad U(0) = N_0 > 0,$$

and is the classical example in any textbook on Mathematical Biology. Its exact solution is well known

$$U(t) = \frac{N_0 e^t}{1 + N_0 (e^t - 1)}$$

and depending on the value N_0 suggests three different scenarios for the population evolution. In particular, the useful curve, the so-called sigmoid, is obtained if $N_0 < 1/2$ (see, e.g., Fig. 1.1 in [15]).

It can be noted that the data for the total COVID-19 cases in China [6] can be approximated by a sigmoid with the correctly-specified parameters. Having this in mind, we introduce a smooth function $u(t)$, which presents the total number of the COVID-19 cases identified up to day t (for any integer number t). We assume that the first case (cases) u_0 was (were) identified at $t = 0$. Obviously, the function $u(t)$ is non-decreasing. So, we obtain

$$\frac{du}{dt} = u(a - bu), \quad u(0) = u_0 \geq 0 \tag{1}$$

where a and b are positive constants. One may define a as $a_0 S$, where $a_0 < 1$ is the infection rate and S is an average number of healthy persons, who was contacted by a fixed infected person. Obviously, each infected person can be in contact only with a limited number of people (usually it is relatives and close friends). The term bu has an opposite meaning to a , because one reflects the efforts B , in order to avoid contacts with infected persons and to make other restrictions defined by the government. The coefficient B should increase with growing $u(t)$. In other words, the government and ordinary people should apply stronger measures in order to stop growing $u(t)$, otherwise the control on the epidemic process will be lost. So, we assume that $B \approx b_* u^{1+\gamma}$ with $\gamma > 0$, therefore the term $b_* u^{1+\gamma}$ (here $b_* > 0$) leading to the equation

$$\frac{du}{dt} = u(a - b_* u^\gamma) \tag{2}$$

is derived. In the case $\gamma = 1$, Eq. (2) coincides with (1). We note that the nonlinearity in (2) was introduced in [16] for describing competition between species, while the logistic equation in epidemiology occurs naturally and it is shown under some general assumptions in [7].

During the epidemic process there are two possibilities for the infected persons. A majority, say w , among them will recover, while some people, v , will die. Obviously, the equality

$$u = v + w$$

takes place at any time t . A typical equation for the time evolution of v (see the last equation in the SIR model) is

$$\frac{dv}{dt} = k(t)u, \quad v(0) = v_0 \geq 0 \quad (3)$$

(a similar equation can be written for w but there is no need to use more equations), where v_0 is the number of deaths at $t = 0$. Here the coefficient $k(t) > 0$ reflects the effectiveness of the health system of the country (or a region) in question. From mathematical point of view, this coefficient should have the asymptotic behavior $k(t) \rightarrow 0$, if $t \rightarrow \infty$, otherwise all infected people will die. In particular, the useful form is $k(t) = k_0 \exp(-\alpha t)$, $\alpha > 0$.

3 Application for the COVID-19 outbreak in China

The general solution of Eq. (1) is well-known, so that Eq. (3) with the given function $k(t)$ can be easily integrated. So, setting $k(t) = k_0 \exp(-\alpha t)$, $\alpha > 0$, we arrive at the exact solution of the model (1) and (3)

$$\begin{aligned} u(t) &= \frac{au_0 e^{at}}{a + bu_0(e^{at} - 1)}, \\ v(t) &= ak_0 u_0 \int_0^t \frac{e^{(a-\alpha)\tau}}{a + bu_0(e^{a\tau} - 1)} d\tau + v_0. \end{aligned} \quad (4)$$

Remark 1 *The integral in (4) leads to an expression involving the special function LerchPhi(y, c, ν) $\equiv \sum_{n=0}^{\infty} \frac{y^n}{(\nu+n)^c}$, which cannot be expressed in terms of elementary functions for arbitrary parameters α and a . However, it can be done in some specific cases. For example, one obtains*

$$v(t) = \frac{2k_0 \sqrt{u_0}}{\sqrt{b(a - bu_0)}} \left(\arctan \left(\frac{\sqrt{bu_0}}{\sqrt{a - bu_0}} e^{\frac{at}{2}} \right) - \arctan \left(\frac{\sqrt{bu_0}}{\sqrt{a - bu_0}} \right) \right) + v_0$$

in the case $2\alpha = a$.

Now we need to specify all the parameters in (4) using the data for the COVID-19 outbreak in China. It follows from [6] that the earliest well-founded data were fixed on Jan.22, hence we fix this date as $t = 0$ and immediately obtain $u_0 = 571$ and $v_0 = 17$. The parameter b can be found from the known asymptotic behavior of the function $u(t)$ in (4) and information from [6], therefore $b \approx \frac{a}{80000}$. The plausible interval for parameter a can be estimated by using

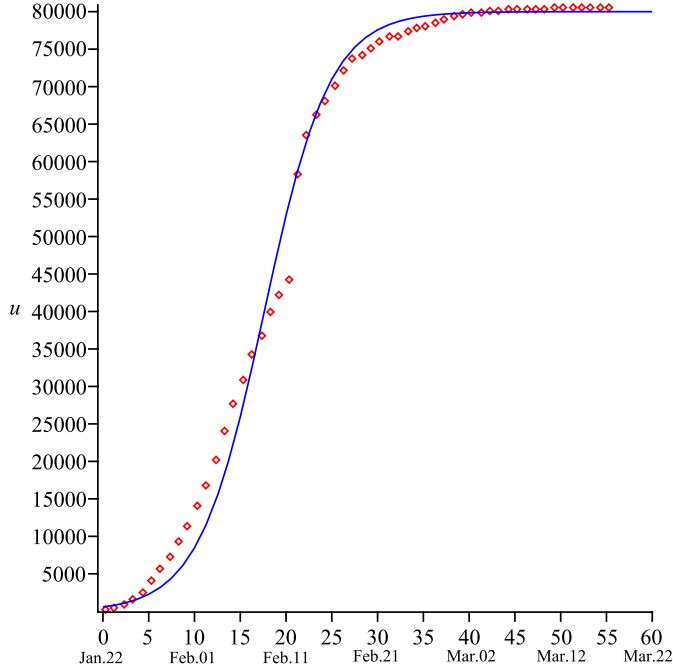


Figure 1: The comparison of the exact solution $u(t)$ (4) with $a = 0.28$, $b = \frac{7}{2000000}$, $u_0 = 571$ (blue curve) and the measured data of the COVID-19 cases (red dots).

option ‘animation’ in MAPLE in order to fit plot of the function $u(t)$ to the measured data after Jan.22. So, we have numerically proved that $a \in [0.25, 0.30]$.

Because the function $v(t)$ should be monotonic non-decreasing function (we remind that it is the number of total deaths), we conclude that $a > \alpha$. It was identified that a good choice is $4\alpha = a$. Finally, the coefficient k_0 was found from the formula

$$v|_{t=T} \equiv 571k_0 \int_0^T \frac{e^{0.21\tau}}{1 + \frac{571}{80000}(e^{0.28\tau} - 1)} d\tau + 17 = V$$

(here V is the number of total deaths in the time $t = T$ presented in [6]) for the fixed value of $a = 0.28$. The value of the coefficient k_0 is slowly varied from 0.0092385 to 0.0096878 if T is changed from 65 to 45, respectively. So, the value $k_0 = 0.0094$ was chosen.

Fig. 1 and Fig. 2 present the comparison of the results obtained from the model (1) and (3) (with the parameters specified above) and the measured data for the COVID-19 outbreak in China [6]. One may note that there is a good agreement between the total number of the COVID-19 cases and that predicted by our model. Of course, one may claim that exactness is not sufficiently good in the interval $t \in [10, 25]$ in Fig. 1. However, we assume that either the method of measurement of the COVID-19 cases was corrected, or an unpredictable spike of

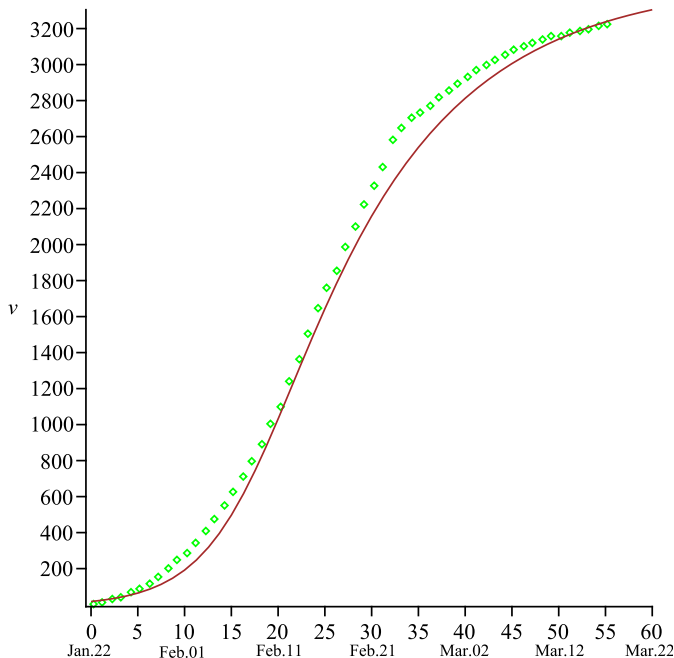


Figure 2: The comparison of the exact solution $v(t)$ (4) with $a = 0.28$, $b = \frac{7}{2000000}$, $u_0 = 571$, $k_0 = 0.0094$, $\alpha = 0.07$, $v_0 = 17$ (brown curve) and the measured data the total number of deaths (green dots).

such cases occurred around date $t = 25$ (there is a jump from 44653 cases on Feb.11 to 58761 cases on Feb.12 [6]).

The comparison between the total number of deaths and that predicted by our model shows that exactness is sufficiently good for any time (see Fig. 2). One may also note that the function $v(t)$ is still increasing beyond the time $t = 60$. Such behavior reflects the real situation in the epidemic process, namely: some people will die even in absence of new COVID-19 cases because they were infected earlier. So, the final number of total deaths will be fixed later than that of the COVID-19 cases.

4 Discussion

In this work, a mathematical model is proposed for quantitative description of the outbreak of novel coronavirus COVID-19 in China. Although the model is relatively simple, the comparison with the data listed in [6] shows that the analytical solution of the model (with the correctly-specified parameters) leads to the results, which are in good agreement with the measured data.

Some well-known recommendations naturally follow from the model. It follows from the exact solution (4) that one needs to reduce the coefficient $a = a_0 S$ as much as possibly. It means that the number of contacts S should be minimized. On the other hand, the government should make more efforts (to close shops, restaurants, to restrict transport traffic etc.) in order to increase the function $B(u)$. These efforts should increase with growing of the total number of the COVID-19 cases. The government restrictions can be stopped only under condition that that the number of new COVID-19 cases per day already began to decrease from day to day. It means mathematically that the second order derivative of the function $u(t)$ takes negative values. In order to find the so-called critical number u^* , we analyze the function $u(t)$ from (4). Calculating the second order derivative, one obtains

$$u'' = \frac{u_0 a^3 (a - bu_0) e^{at} (a - bu_0 - bu_0 e^{at})}{(a - bu_0 + bu_0 e^{at})^3}.$$

Solving the algebraic equation $u'' = 0$ with respect to the time, we arrive at

$$t^* = \frac{1}{a} \ln \left(\frac{a}{bu_0} - 1 \right),$$

hence $u^* = u(t^*)$. On the other hand, formula $u'' = 0$ allows to find the parameter b provided the time t^* is known from the measured data. Assuming that a is known one calculates

$$b = \frac{a}{u_0 (e^{at^*} + 1)}. \quad (5)$$

Taking into account interpretation of the parameters, we believe that the parameter a varies not so much as b and can be specified (at least estimated with a sufficient exactness) as follows. Obviously, the total number of the COVID-19 cases in the initial period of epidemic process can be approximated as $u(t) \approx u_0 e^{at}$ (see $u(t)$ in (4) for small time). So, having the measured data in the initial period, we may specify the parameter a . It means that our model can allow to predict the total numbers of the COVID-19 cases if the data for t^* and u^* are known.

Let us consider, an example. It can be noted from the public data [6] that the COVID-19 outbreak in Austria had the maximum number of new daily cases on March 26. So, $u^* = 6909$. If we fix March 8 as the initial point $t = 0$, then $t^* = 18$ and $u_0 = 104$ (we think that there are essential errors in measuring at the very beginning of the epidemic process, so that it is unreasonable to start from very small numbers of u_0). Now we make approximation of the measured COVID-19 cases using the formula $u(t) \approx 104e^{at}$ during the first 15 days. It turns out that the parameter $a = 0.27$ provides very good approximation during the first 12 days (see Fig. 3). So, using formula (5) we define $b = 0.000020$. Now we may predict that the total number the COVID-19 cases in Austria should be $u_{max} = a/b \approx 13500$. Taking into account that this number was calculated under some assumptions, the real number can be larger. For example, if one takes March 10 as the initial point $t = 0$ then $u_{max} \approx 13800$. We estimate the maximum error in 10 percent.

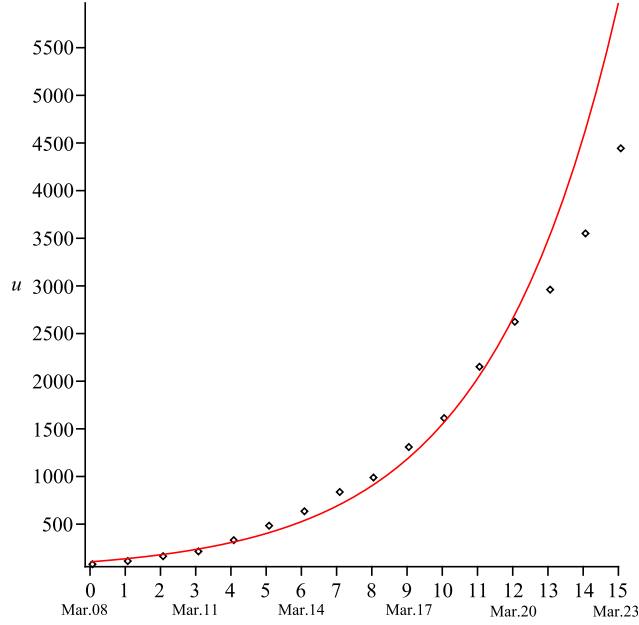


Figure 3: An exponential growing of the total number of the COVID-19 cases in Austria at the initial stage of the coronavirus spread (red curve) and the measured data (black dots)

It should be noted that the parameter γ plays essential role if one uses Eq. (2) instead of Eq. (1). In order to highlight this, we present exact solutions of Eq. (2) with different values of γ in Fig. 4 (all other parameters are the same as in Fig. 1). One may see that $\gamma = 1$ is a good choice in the case of China. On the other hand, taking into account the known data [6], we conclude that $\gamma < 1$ in the case of S. Korea.

Obviously, the model cannot be thought as such that is applicable for the COVID-19 outbreak in each country. For example, the outbreak in China was mostly localized in the province Hubei. The size and population of this province are very small comparing with total those of China. A similar situation is in USA, where two states, New-York and New-Jersey are affected by the coronavirus much more than other states (up to date April 5, 2020).

On the other hand, if we take the epidemic process in Italy then one notes that the size and population of Northern Italy (8 provinces, Lombardia is the largest among them) are comparable with those all of Italy. So, we propose that the space distribution of the infected population should be taken into account in such cases. The simplest generalization of the basic equations of our model are

$$\begin{aligned} \frac{\partial u}{\partial t} &= d_1 \Delta u + u(a - bu^\gamma), \\ \frac{\partial v}{\partial t} &= d_2 \Delta u + k(t)u, \end{aligned} \tag{6}$$

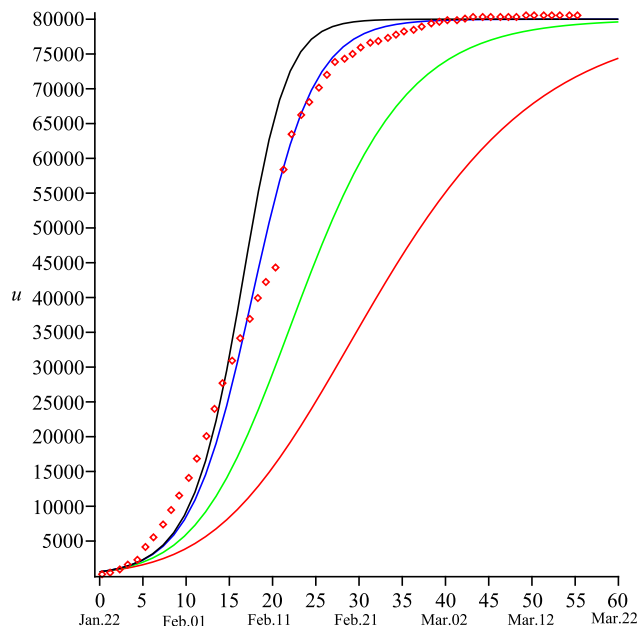


Figure 4: The solution $u(t)$ of Eq. (2) for $\gamma = 0.3$ (red curve), $\gamma = 0.5$ (green curve), $\gamma = 1$ (blue curve), $\gamma = 1.5$ (black curve) and the measured data of the COVID-19 cases in China (red dots).

where Δ is the Laplace operator, d_1 and d_2 are diffusivities, the functions $u(t, x, y)$ and $v(t, x, y)$ are analogs of $u(t)$ and $v(t)$. Of course, the generalized model based on the system of (6) and relevant boundary conditions (for example, zero flux conditions at the boundary) is much more complicated problem and cannot be solved analytically in contrast to the model (1) and (3). Here we only note that the first equation in (6) with $\gamma = 1$ is the classical Fisher equation [17], which was extensively studied in many works (see, e.g., the monographs [9, 18] and papers cited therein).

References

- [1] X. Luo, Sh. Feng, J. Yang et al., Analysis of potential risk of COVID-19 infections in China based on a pairwise epidemic model, *doi:10.20944/preprints202002.0398.v1* (2020).
- [2] L. Peng, W. Yang, D. Zhang et al., Epidemic analysis of COVID-19 in China by dynamical modeling, *arXiv:2002.06563* (2020).

- [3] N. Shao, M. Zhong, Y. Yan et al., Dynamic models for Coronavirus Disease 2019 and data analysis, *Math. Meth. Appl. Sci.* (2020) 1–7.
- [4] J. Tian, J. Wu, Y. Bao et al., Modeling analysis of COVID-19 based on morbidity data in Anhui, China. *MBE* 17 (2020) 2842–2852.
- [5] C. Weston, W. Roda, M. Varugheseb et al., Why is it difficult to accurately predict the COVID-19 epidemic? *Infectious Disease Modelling* 5 (2020) 271–281.
- [6] <https://www.worldometers.info/coronavirus>
- [7] F. Brauer, C. Castillo-Chavez, *Mathematical models in population biology and epidemiology*, Springer, New York, 2012.
- [8] M. J. Keeling, P. Rohani, *Modeling infectious diseases in humans and animals*, Princeton University Press, Princeton, 2008.
- [9] J. D. Murray, *Mathematical biology*, Springer, Berlin, 1989.
- [10] J. D. Murray, *Mathematical biology II: spatial models and biomedical applications*, Springer, Berlin, 2003.
- [11] W. O. Kermack, A. G. McKendrick, A contribution to the mathematical theory of epidemics, *Proc. Roy. Soc. A* 115 (1927) 700–721.
- [12] R. M. Anderson, R. M. May, Directly transmitted infectious diseases: Control by vaccination, *Science* 215 (1982) 1053–1060.
- [13] K. Dietz, The incidence of infectious diseases under the influence of seasonal fluctuations, *Lecture Notes in Biomathematics* 11, Springer, Berlin, 1976, pp. 1–15.
- [14] P. F. Verhulst, Notice sur la loi que la population suit dans son accroissement, *Corr. Math. Physics.* 10 (1838) 113.
- [15] R. Cherniha, V. Davydovych, *Nonlinear reaction-diffusion systems — conditional symmetry, exact solutions and their applications in biology*, *Lecture Notes in Math.* 2196, Springer, Cham, 2017.
- [16] F. J. Ayala, M. E. Gilpin, J. G. Ehrenfeld, Competition between species: Theoretical models and experimental tests, *Theoretical Pop. Biol.* 4 (1973) 331–356.
- [17] R. A. Fisher, The wave of advance of advantageous genes, *Ann. Eugenics.* 7 (1937) 353–369.
- [18] R. Cherniha, M. Serov, O. Pliukhin, *Nonlinear reaction-diffusion-convection equations: Lie and conditional symmetry, exact solutions and their applications*, Chapman and Hall/CRC, New York, 2018.