Transmission Dynamics of the Global COVID-19 Epidemic: Analytical Modeling and Future Prediction

Keqin Li*
Science Hall 249, State University of New York
1 Hawk Drive, New Paltz, New York 12561, USA
Email: lik@newpaltz.edu

*The author can be reached at phone: (845) 257-3534, fax: (845) 257-3996.

Abstract

Background: The outbreak and epidemic of COVID-19 has created worldwide impact and attracted global attention. Considerable effort has been devoted to the study of the transmission dynamics of COVID-19. However, there are lack of simple and straightforward expressions of the growing curves of important indicators, such as the cumulative number of confirmed cases and the cumulative number of dead cases.

Methods: We adopt two methods. The first method is based on regression analysis. We fit the available data into a curve by the method of least squares. The best curve is obtained by solving a multivariable minimization problem. The second method is based on differential equations. We establish an analytical model of transmission dynamics based on the susceptible-exposed-infectious-recovered-dead (SEIRD) process using a linear system of ordinary differential equations, which characterize the daily change in each compartment. The size of each compartment (i.e., the number of people in each stage of the SEIRD process) is readily available based on the solution to these differential equations.

Results: Both methods are applied to the COVID-19 epidemic data in the world as a case study. Furthermore, predictions of the cumulative number of confirmed cases and the cumulative number of dead cases in April 2020 using our models and methods are also provided. From a global perspective, unless powerful and effective social and medical impacts are made, by the end April of 2020, the cumulative number of confirmed cases is 23.333 and 36.068 millions respectively using regression analysis and analytical model, and the cumulative number of dead cases is 1.148 and 2.528 millions respectively using regression analysis and analytical model, based on the current situation.

Conclusions: In this paper, we make some progress towards analytical expressions of the daily growth of the cumulative number of confirmed cases and the cumulative number of dead cases, two most important and daily reported figures.

Keywords: Analytical model, COVID-19, linear system of ordinary differential equations, prediction, regression analysis, SEIRD process, transmission dynamics.

1 Introduction

1.1 Motivation

The outbreak and epidemic of COVID-19 has created worldwide impact and attracted global attention [6, 13]. The total number of coronavirus cases and the total number of coronavirus deaths in the world are growing exponentially and explosively [2, 4, 5]. Such pandemic has generated significant impact socially, economically, and politically.

Considerable effort has been devoted to the study of the transmission dynamics of COVID-19 [9, 12]. Li *et al.* reported that the mean incubation period was 5.2 days, the mean serial interval was 7.5 days, the basic reproductive number R_0 was 2.2, and the epidemic was doubled in size every 7.4 days [15]. Liu *et al.* reported that the average incubation duration of 2019-nCoV infection was 4.8 days, and estimated the R_0 values of 2019-nCoV as 2.90 and 2.92 using the methods of exponential growth (EG) and maximum likelihood (ML) respectively [16]. Nishiura *et al.* estimated the median serial interval at 4.0 days [19]. Zhang *et al.* mentioned that the ML value of R_0 was 2.28 for COVID-19 outbreak at the early stage on the Diamond Princess cruise ship [21]. Zhao *et al.* pointed out that the early outbreak data largely follow the exponential growth, and estimated that the mean R_0 for the 2019-nCoV ranges from 2.24 to 3.58 [22]. Zhou *et al.* found that the median duration of viral RNA shedding from oropharyngeal specimens was 20 days (range of 8–37 days) [23].

It is clear that all existing studies have only focused on individual quantities such as the basic reproductive number, the mean serial interval, the mean incubation period, and the mean recovery time. However, based on these quantities, it is still not possible to predict how the number of infected people, the number of recovered individuals, and the number of dead cases change daily. There are lack of simple and straightforward expressions of the growing curves of important indicators, such as the cumulative number of confirmed cases and the cumulative number of dead cases.

1.2 Contributions

In this paper, we make some progress towards analytical expressions of the daily growth of the cumulative number of confirmed cases and the cumulative number of dead cases,

two most important and daily reported figures. We adopt two methods.

The first method is based on regression analysis. We fit the available data into a curve by the method of least squares. The best curve is obtained by solving a multivariable minimization problem. The details are presented in Section 2.

The second method is based on differential equations. We establish an analytical model of transmission dynamics based on the susceptible-exposed-infectious-recovered-dead (SEIRD) process using a linear system of ordinary differential equations, which characterize the daily change in each compartment. The size of each compartment (i.e., the number of people in each stage of the SEIRD process) is readily available based on the solution to these differential equations. The details are presented in Section 3.

Both methods are applied to the COVID-19 epidemic data in the world as a case study. Furthermore, predictions of the cumulative number of confirmed cases and the cumulative number of dead cases in April 2020 using our models and methods are also provided. The details are presented in Section 4.

2 Regression Analysis

In this section, we develop a regression analysis method.

2.1 The Method

Assume that a group of n available data points $(x_1,y_1),(x_2,y_2),...,(x_n,y_n)$ are to be fit into a function $y = f(a_1,a_2,...,a_k,x)$, where $a_1,a_2,...,a_k$ are parameters of f to be decided. The method of least squares [1] is used to find $a_1,a_2,...,a_k$. The sum of squared residuals is

$$E(a_1,a_2,...,a_k) = \sum_{i=1}^{n}(y_i - f(a_1,a_2,...,a_k,x_i))^2,$$

where E is viewed as a function of $a_1,a_2,...,a_k$. To minimize $E(a_1,a_2,...,a_k)$, we only need to find $a_1,a_2,...,a_k$, such that $\nabla E(a_1,a_2,...,a_k) = 0$, i.e., $\partial E/\partial a_j = 0$, for all $1 \leq j \leq k$. This is a multivariable minimization problem.

For transmission dynamics of coronavirus, we consider an exponential function in the form of $f(x) = ab^x + c$, where a and b are parameters and c is a given constant, such that $y_i \approx ab^{x_i} + c$, for all $1 \leq i \leq n$. To use the method of least squares to find a and b, the

sum of squared residuals is

$$E(a,b) = \sum_{i=1}^{n} (y_i - (ab^{x_i} + c))^2,$$

where E is viewed as a function of a and b. To minimize $E(a,b)$, we need

$$\frac{\partial E}{\partial a} = \sum_{i=1}^{n} 2(y_i - (ab^{x_i} + c))(-b^{x_i}) = 0,$$

and

$$\frac{\partial E}{\partial b} = \sum_{i=1}^{n} 2(y_i - (ab^{x_i} + c))(-ax_i b^{x_i - 1}) = 0.$$

To solve the above equations, we notice that the first equation implies that

$$a = \frac{s_1}{s_2} = \frac{\sum_{i=1}^{n} (y_i - c)b^{x_i}}{\sum_{i=1}^{n} b^{2x_i}},$$

and the second equation implies that

$$a = \frac{s_3}{s_4} = \frac{\sum_{i=1}^{n} (y_i - c)x_i b^{x_i - 1}}{\sum_{i=1}^{n} x_i b^{2x_i - 1}}.$$

To find b, we need $s_1/s_2 = s_3/s_4$. Notice that $F(b) = s_1 s_4 - s_2 s_3$ is an increasing function of b. Hence, b can be found by using the standard bisection method ([8], pp. 22), such that $F(b) = 0$.

The quality of the above regression analysis can be evaluated by the *adjusted relative error* (ARE), defined for y_i as

$$\text{ARE}_i = \left(\frac{y_i}{\max_{1 \le i \le n}(y_i)}\right) \times \left(\frac{ab^{x_i} + c - y_i}{y_i}\right) \times 100\%.$$

Notice that

$$\left(\frac{ab^{x_i} + c - y_i}{y_i}\right) \times 100\%$$

is the relative error for y_i. Since the magnitudes of $y_1, y_2, \ldots, y_n$ can differ dramatically, simply taking the maximum relative error does not seem appropriate, since a large relative error for a very small y_i does not seem significant. Therefore, the relative error for y_i is adjusted by a factor of

$$\frac{y_i}{\max_{1 \le i \le n}(y_i)}.$$

Table 1: Quality of Regression Analysis

	$C_{\text{World}}(t)$			$D_{\text{World}}(t)$		
t	Analytical	Actual	ARE	Analytical	Actual	ARE
21	304874	304979	-0.012%	12931	13011	-0.189%
22	336362	337459	-0.128%	14465	14640	-0.414%
23	372070	378830	-0.787%	16225	16513	-0.682%
24	412565	422574	-1.166%	18244	18894	-1.541%
25	458490	471035	-1.461%	20560	21282	-1.712%
26	510571	531865	-2.481%	23217	24073	-2.031%
27	569634	596366	-3.114%	26264	27343	-2.559%
28	636614	663127	-3.089%	29759	30861	-2.612%
29	712575	723390	-1.260%	33769	34065	-0.701%
30	798718	784738	1.629%	38369	37773	1.415%
31	896410	858377	4.431%	43645	42146	3.557%

2.2 A Case Study

In this section, we apply our method to a case study.

Let $C_{\text{World}}(t)$ be the cumulative number of confirmed cases in the world by the tth day (i.e., $x_t = t$ and $y_t = C_{\text{World}}(t)$). Based on the data published by Worldometer [5] during March 1–31, 2020 (with March 1, 2020 as the first day), we get

$$C_{\text{World}}(t) = 16729.4550655 \times 1.1340597^t + 70000.$$

Let $D_{\text{World}}(t)$ be the cumulative number of dead cases in the world by the tth day (i.e., $x_t = t$ and $y_t = D_{\text{World}}(t)$). Based on the data published by Worldometer [5] during March 1–31, 2020 (with March 1, 2020 as the first day), we get

$$D_{\text{World}}(t) = 584.4834711 \times 1.1470922^t + 2500.$$

In Table 1, we display the analytical data, the actual data, and the ARE for $C_{\text{World}}(t)$ and $D_{\text{World}}(t)$ during March 21–31, 2020, with ARE_i in the range $[-3.114\%, 4.431\%]$ for $C_{\text{World}}(t)$, and ARE_i in the range $[-2.612\%, 3.557\%]$ for $D_{\text{World}}(t)$.

3 An Analytical Model

In this section, we establish an analytical model of transmission dynamics.

3.1 The SEIRD Process

In this section, we describe the SEIRD process (see Figure 1).

The susceptible-exposed-infectious-recovered-dead (SEIRD) model belongs to the class of compartmental disease models [7]. “Susceptible", “Exposed", “Infectious", “Recovered", and “Dead" are compartments, and each individual in the population is assigned to one of these compartments. The model is extended from the SIR model [10] and the SEIR model [11, 14, 20] (other variations include SZR, SIZR, and SIZRQ models [18]).

- Susceptible individuals have no immunity to the disease, who may be exposed to the disease and move into the “Exposed" compartment through contact with an exposed or infectious person.
- Exposed individuals have been exposed to the disease, but have not shown sign or illness and thus behave like normal persons; however, they can transmit disease to others, and will move into the “Infectious" compartment.
- Infectious individuals show clear symptom of sickness, and have been tested and confirmed to be patients, who are typically quarantined or hospitalized, can still transmit disease to family members, medical staff, and other people, and eventually move into the “Recovered" or the “Dead" compartments.
- Recovered individuals can no longer become infected, typically because they have immunity from a prior exposure, which is often appropriate if immunity is long-lasting or the disease is being modeled over a relatively short time period.
- Dead individuals remain in the “Dead" compartment forever.

3.2 A System of Differential Equations

In this section, we establish an analytical model of transmission dynamics using a system of ordinary differential equations.

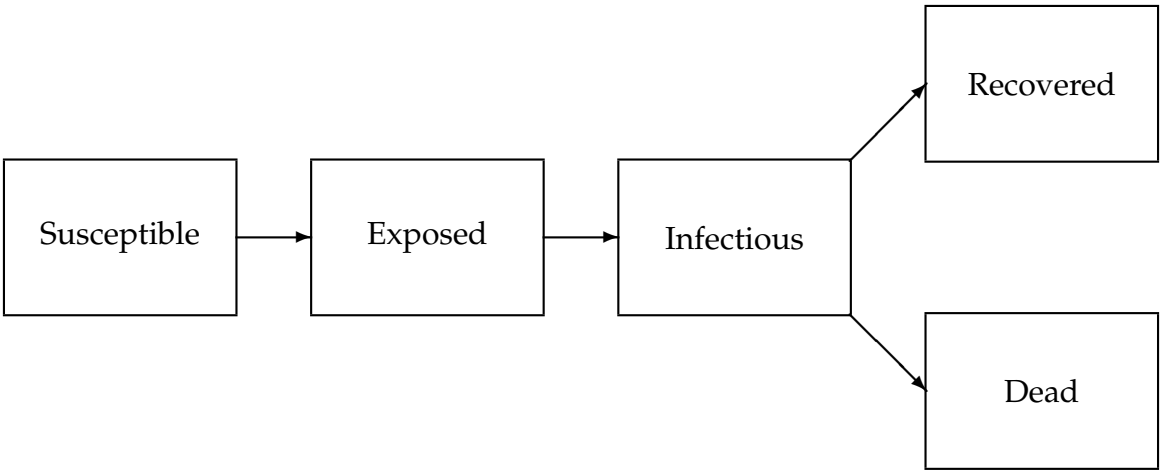

Figure 1: The SEIRD process

Let $S(t), E(t), I(t), V(t), D(t)$ be respectively the numbers of individuals in the susceptible, exposed, infectious, recovered, and dead compartments at time t (measured in days). $S(t), E(t), I(t), V(t), D(t)$ are nonnegative functions of $t \geq 1$, which satisfy

$$S(t) + E(t) + I(t) + V(t) + D(t) = N,$$

and

$$dS(t)/dt + dE(t)/dt + dI(t)/dt + dV(t)/dt + dD(t)/dt = 0.$$

The daily reported cumulative number of confirmed cases is $C(t) = I(t) + V(t) + D(t)$. The daily reported cumulative number of dead cases is $D(t)$.

Let N, R_E, R_I, T_E, T_I, α be positive constants defined below.

- N is the size of (i.e., the number of people in) the population.
- R_E is the reproductive number of an exposed individual.
- R_I is the reproductive number of an infectious individual.
- T_E is the average number of days in the “Exposed" compartment.
- T_I is the average number of days in the “Infectious" compartment.
- α is the percentage of infectious individuals who eventually die.

Notice that the basic reproductive number is actually $R_0 = R_E + R_I$.

We have the following linear system of ordinary differential equations:

$$\begin{aligned}
\frac{dS(t)}{dt} &= -\left(\frac{R_E}{T_E}\cdot E(t) + \frac{R_I}{T_I}\cdot I(t)\right),\\
\frac{dE(t)}{dt} &= \frac{R_E - 1}{T_E}\cdot E(t) + \frac{R_I}{T_I}\cdot I(t),\\
\frac{dI(t)}{dt} &= \frac{E(t)}{T_E} - \frac{I(t)}{T_I},\\
\frac{dV(t)}{dt} &= (1-\alpha)\frac{I(t)}{T_I},\\
\frac{dD(t)}{dt} &= \alpha\frac{I(t)}{T_I},
\end{aligned}$$

with given initial values $S(1), E(1), I(1), V(1), D(1)$. These equations are elaborated as follows.

- $dS(t)/dt$: R_E/T_E is the number of newly exposed people contacted by an exposed individual every day, and $(R_E/T_E)E(t)$ is the total number of such people on the tth day. Similarly, R_I/T_I is the number of newly exposed people contacted by an infectious individual every day, and $(R_I/T_I)I(t)$ is the total number of such people on the tth day. Therefore, the total number of newly exposed people (who leave the "Susceptible" compartment) is $(R_E/T_E)E(t) + (R_I/T_I)I(t)$ on the tth day.

- $dE(t)/dt$: The number of newly exposed people (who move into the "Exposed" compartment) is $(R_E/T_E)E(t) + (R_I/T_I)I(t)$ on the tth day. On the other hand, the number of exposed people who move into the "Infectious" compartment is $E(t)/T_E$ on the tth day. Therefore, the change in the "Exposed" compartment is $((R_E - 1)/T_E)E(t) + (R_I/T_I)I(t)$ on the tth day.

- $dI(t)/dt$: The number of exposed people who move into the "Infectious" compartment is $E(t)/T_E$ on the tth day. On the other hand, the number of infectious people who leave the "Infectious" compartment is $I(t)/T_I$ on the tth day. Therefore, the change in the "Infectious" compartment is $E(t)/T_E - I(t)/T_I$ on the tth day.

- $dV(t)/dt$: The number of infectious people who move into the "Recovered" compartment is $(1-\alpha)(I(t)/T_I)$ on the tth day.

- $dD(t)/dt$: The number of infectious people who move into the "Dead" compartment is $\alpha(I(t)/T_I)$ on the tth day.

3.3 An Analytical Solution

In this section, we provide an analytical solution to the linear system of ordinary differential equations.

Notice that $dE(t)/dt$ and $dI(t)/dt$ form an autonomous linear system of differential equations:

$$\frac{dE(t)}{dt} = \left(\frac{R_E - 1}{T_E}\right)E(t) + \left(\frac{R_I}{T_I}\right)I(t),$$
$$\frac{dI(t)}{dt} = \left(\frac{1}{T_E}\right)E(t) - \left(\frac{1}{T_I}\right)I(t),$$

It is well known that such differential equations accommodate analytical solutions [3].

Let

$$A = \begin{bmatrix} \frac{R_E-1}{T_E} & \frac{R_I}{T_I} \\ \\ \frac{1}{T_E} & -\frac{1}{T_I}. \end{bmatrix}$$

If matrix A has eigenvalues λ_1 and λ_2 with corresponding eigenvectors $v_1 = (x_1, y_1)$ and $v_2 = (x_2, y_2)$, then we have

$$E(t) = c_1 x_1 e^{\lambda_1 t} + c_2 x_2 e^{\lambda_2 t},$$
$$I(t) = c_1 y_1 e^{\lambda_1 t} + c_2 y_2 e^{\lambda_2 t},$$

for some constants c_1 and c_2, which can be decided based on the initial conditions $E(1) = b_1$, and $I(1) = b_2$, by solving the following equations:

$$c_1 x_1 e^{\lambda_1} + c_2 x_2 e^{\lambda_2} = b_1,$$
$$c_1 y_1 e^{\lambda_1} + c_2 y_2 e^{\lambda_2} = b_2,$$

whose solution is

$$c_1 = \begin{vmatrix} x_2 e^{\lambda_2} & b_1 \\ y_2 e^{\lambda_2} & b_2 \end{vmatrix} \Big/ \begin{vmatrix} x_2 e^{\lambda_2} & x_1 e^{\lambda_1} \\ y_2 e^{\lambda_2} & y_1 e^{\lambda_1} \end{vmatrix},$$
$$c_2 = \begin{vmatrix} x_1 e^{\lambda_1} & b_1 \\ y_1 e^{\lambda_1} & b_2 \end{vmatrix} \Big/ \begin{vmatrix} x_1 e^{\lambda_1} & x_2 e^{\lambda_2} \\ y_1 e^{\lambda_1} & y_2 e^{\lambda_2} \end{vmatrix}.$$

The eigenvalues of the matrix A are values of λ that satisfy the equation

$$\begin{vmatrix} \frac{R_E-1}{T_E}-\lambda & \frac{R_I}{T_I} \\ \frac{1}{T_E} & -\frac{1}{T_I}-\lambda \end{vmatrix} = 0,$$

or,

$$\begin{vmatrix} -(\lambda-\frac{R_E-1}{T_E}) & \frac{R_I}{T_I} \\ \frac{1}{T_E} & -(\lambda+\frac{1}{T_I}) \end{vmatrix} = 0,$$

that is,

$$\lambda^2-\left(\frac{R_E-1}{T_E}-\frac{1}{T_I}\right)\lambda-\frac{R_E+R_I-1}{T_ET_I}=0.$$

The solutions are

$$\lambda_1=\frac{1}{2}\left(\left(\frac{R_E-1}{T_E}-\frac{1}{T_I}\right)+\sqrt{\left(\frac{R_E-1}{T_E}-\frac{1}{T_I}\right)^2+4\left(\frac{R_E+R_I-1}{T_ET_I}\right)}\right),$$

and

$$\lambda_2=\frac{1}{2}\left(\left(\frac{R_E-1}{T_E}-\frac{1}{T_I}\right)-\sqrt{\left(\frac{R_E-1}{T_E}-\frac{1}{T_I}\right)^2+4\left(\frac{R_E+R_I-1}{T_ET_I}\right)}\right).$$

The eigenvector $v_i=(x_i,y_i)$ satisfies

$$\begin{bmatrix} -(\lambda_i-\frac{R_E-1}{T_E}) & \frac{R_I}{T_I} \\ \frac{1}{T_E} & -(\lambda_i+\frac{1}{T_I}) \end{bmatrix}\begin{bmatrix} x_i \\ y_i \end{bmatrix}=0.$$

We can set $x_i=1$, and $y_i=(\lambda_i-\frac{R_E-1}{T_E})/(\frac{R_I}{T_I})$, for $i=1,2$.

Based on $E(t)$ and $I(t)$, we can get

$$V(t)=\left(\frac{1-\alpha}{\lambda_1T_I}\right)c_1y_1(e^{\lambda_1t}-e^{\lambda_1})+\left(\frac{1-\alpha}{\lambda_2T_I}\right)c_2y_2(e^{\lambda_2t}-e^{\lambda_2})+V(1),$$

and

$$D(t)=\left(\frac{\alpha}{\lambda_1T_I}\right)c_1y_1(e^{\lambda_1t}-e^{\lambda_1})+\left(\frac{\alpha}{\lambda_2T_I}\right)c_2y_2(e^{\lambda_2t}-e^{\lambda_2})+D(1),$$

and

$$\begin{aligned} C(t)&=I(t)+V(t)+D(t)\\ &=\left(1+\frac{1}{\lambda_1T_I}\right)c_1y_1e^{\lambda_1t}+\left(1+\frac{1}{\lambda_2T_I}\right)c_2y_2e^{\lambda_2t}-\frac{c_1y_1e^{\lambda_1}}{\lambda_1T_I}-\frac{c_2y_2e^{\lambda_2}}{\lambda_2T_I}+V(1)+D(1), \end{aligned}$$

and

$$\begin{aligned}S(t) &= N - E(t) - I(t) - V(t) - D(t)\\ &= N - c_1x_1e^{\lambda_1 t} - c_2x_2e^{\lambda_2 t}\\ &\quad -\left(1+\frac{1}{\lambda_1 T_I}\right)c_1y_1e^{\lambda_1 t} - \left(1+\frac{1}{\lambda_2 T_I}\right)c_2y_2e^{\lambda_2 t} + \frac{c_1y_1e^{\lambda_1}}{\lambda_1 T_I} + \frac{c_2y_2e^{\lambda_2}}{\lambda_2 T_I} - V(1) - D(1).\end{aligned}$$

Notice that since $\lambda_2 < 0$, we have $e^{\lambda_2} < 1$, and $e^{\lambda_2 t} \to 0$ as $t \to \infty$. Hence, we have the following approximations:

$$C^*(t) = \left(1+\frac{1}{\lambda_1 T_I}\right)c_1y_1e^{\lambda_1 t},$$

and

$$D^*(t) = \left(\frac{\alpha}{\lambda_1 T_I}\right)c_1y_1e^{\lambda_1 t},$$

which are accurate for large t.

Now, we can explain why R_0 (or any other quantity) alone cannot predict how $C(t)$ and $D(t)$ grow. It is clear that $C(t)$ is determined by R_E, R_I, T_E, T_I, $E(1)$, $I(1)$, $V(1)$, $D(1)$, collectively, and $D(t)$ is determined by R_E, R_I, T_E, T_I, α, $E(1)$, $I(1)$, $D(1)$, collectively.

3.4 A Case Study

In this section, we apply our method to a case study.

The parameters in our model are set as follows. The incubation period means the time between catching the virus and beginning to have symptoms of the disease. Most estimates of the incubation period for COVID-19 range from 1–14 days, most commonly around five days. We set $D_E = 5$ [15, 16]. Using available preliminary data, the median time from onset to clinical recovery is approximately 2 weeks for mild cases and 3–6 weeks for severe or critical cases. We set $D_I = 20$ [17, 23].

For the same data in Section 2.2 for the world, we get

$$C^*_{\text{World}}(t) = 29539.4430255 \times 1.1155783^t,$$

and

$$D^*_{\text{World}}(t) = 1454.9810657 \times 1.1155783^t,$$

where we set $R_E = 1.39$, $R_I = 0.50$, and $\alpha = 0.157$. The initial values are $E(1) = 20,000$, $I(1) = 10,000$, $V(1) = 20,000$, $D(1) = 2,000$.

Table 2: Quality of Analytical Model

	$C_{\text{World}}(t)$			$D_{\text{World}}(t)$		
t	Analytical	Actual	ARE	Analytical	Actual	ARE
21	296654	304979	-0.970%	13865	13011	2.026%
22	330674	337459	-0.790%	15518	14640	2.083%
23	368612	378830	-1.190%	17366	16513	2.024%
24	410921	422574	-1.358%	19431	18894	1.274%
25	458107	471035	-1.506%	21738	21282	1.082%
26	510736	531865	-2.462%	24314	24073	0.572%
27	569438	596366	-3.137%	27190	27343	-0.363%
28	634914	663127	-3.287%	30402	30861	-1.089%
29	707949	723390	-1.799%	33986	34065	-0.187%
30	789417	784738	0.545%	37988	37773	0.510%
31	880294	858377	2.553%	42453	42146	0.728%

In Table 2, we display the analytical data, the actual data, and the ARE for $C_{\text{World}}(t)$ and $D_{\text{World}}(t)$ during March 21–31, 2020, with ARE_i in the range $[-3.287\%, 2.553\%]$ for $C_{\text{World}}(t)$, and ARE_i in the range $[-1.089\%, 2.083\%]$ for $D_{\text{World}}(t)$.

4 Future Prediction

Using our analytical results in Sections 2 and 3, we display predictions of $C_{\text{World}}(t)$ and $D_{\text{World}}(t)$ on April t, 2020 in Figures 2 and 3 respectively. The regression analysis method gives higher predictions than the analytical model method, since the analytical model results in a smaller base of exponentiation.

From a global perspective, unless powerful and effective social and medical impacts are made, by the end April of 2020, the cumulative number of confirmed cases is 23.333 and 36.068 millions respectively using regression analysis and analytical model, and the cumulative number of dead cases is 1.148 and 2.528 millions respectively using regression analysis and analytical model, based on the current situation.

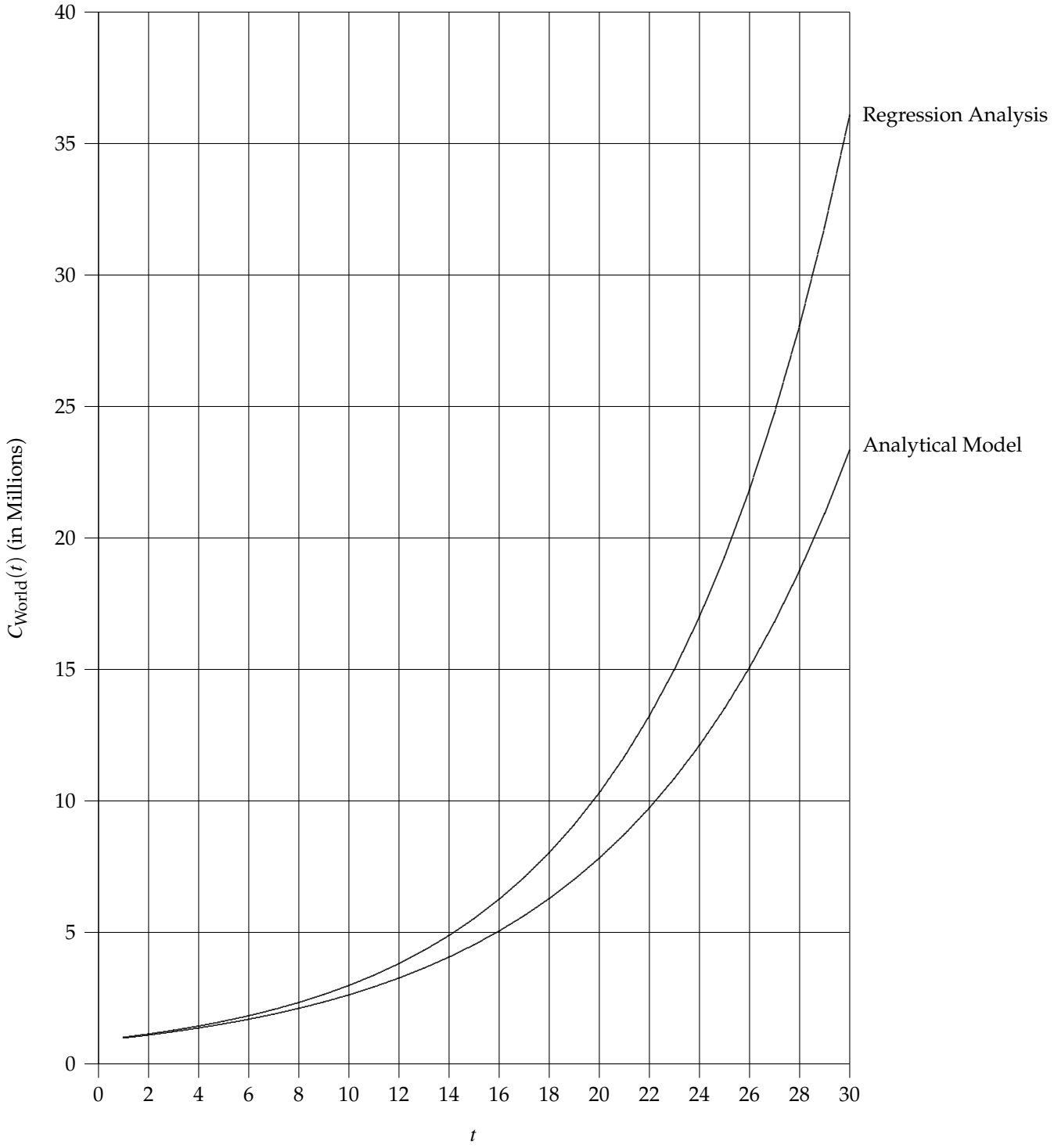

Figure 2: Prediction of $C_{\text{World}}(t)$ on April t, 2020.

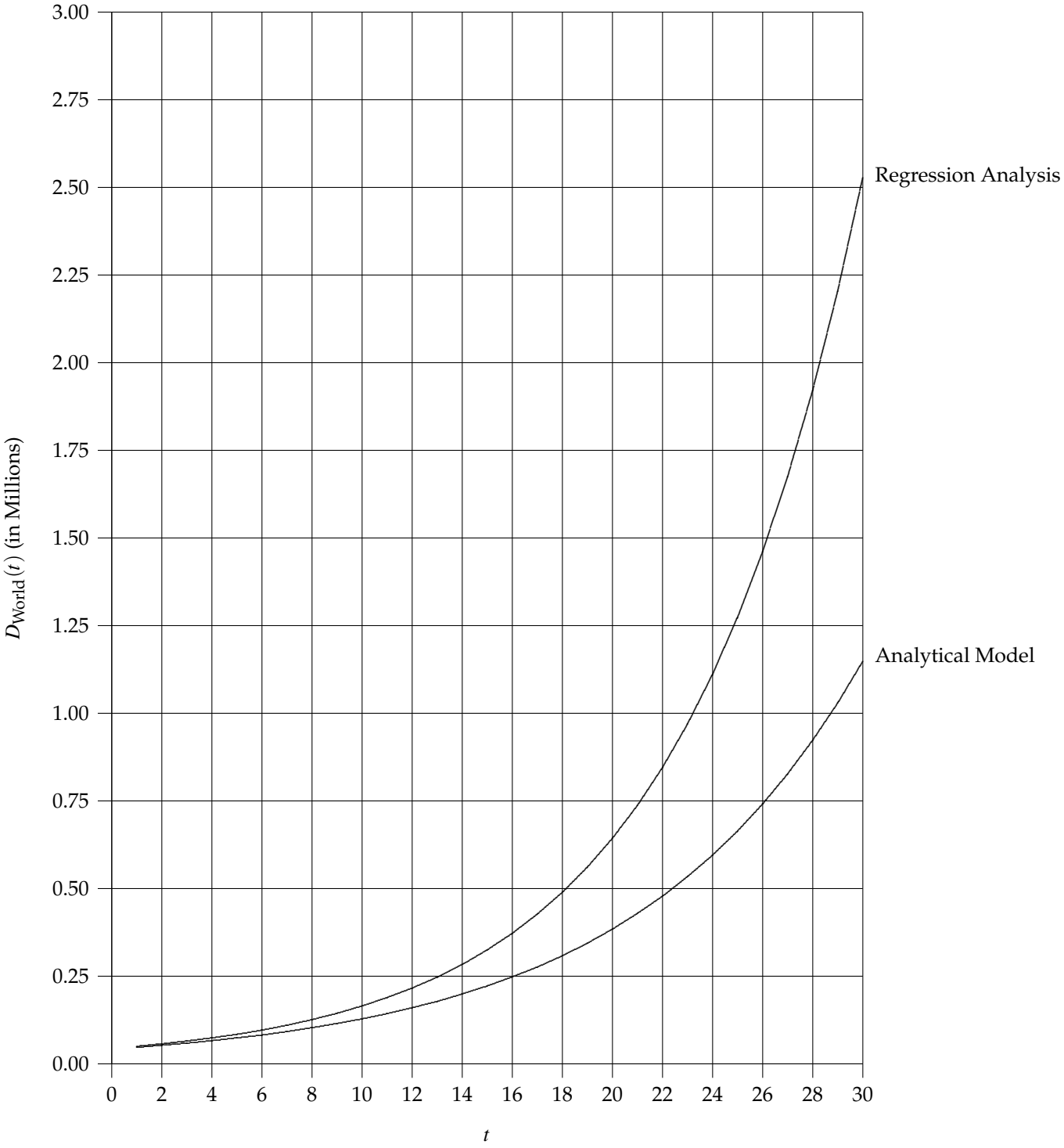

Figure 3: Prediction of $D_{\text{World}}(t)$ on April t, 2020.

5 Conclusions

We have made some progress towards analytical expressions of the daily growth of the cumulative number of confirmed cases and the cumulative number of dead cases, two most important and daily reported figures. Our analytical methods and results have been tested using the COVID-19 epidemic data in the world, and proven to be effective and accurate. We have also predicted the cumulative number of confirmed cases and the cumulative number of dead cases in April 2020 using our models and methods.

Declarations

Ethics approval and consent to participate: Not applicable.
Consent for publication: Not applicable.
Availability of data and material: Yes.
Competing interests: None.
Funding: None.
Authors' contributions: Single author.
Acknowledgements: None.

References

[1] https://en.wikipedia.org/wiki/Least_squares

[2] https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html

[3] https://www.jirka.org/diffyqs/html/eigenmethod_section.html

[4] https://www.who.int/emergencies/diseases/novel-coronavirus-2019

[5] https://www.worldometers.info/coronavirus/country/us/

[6] F. A. Binti Hamzah, *et al.*, "CoronaTracker: world-wide COVID-19 outbreak data analysis and prediction," *Bulletin of the World Health Organization*, 19 March 2020. https://www.who.int/bulletin/online_first/20-255695.pdf

[7] J. C. Blackwood and L. M. Childs, "An introduction to compartmental modeling for the budding infectious disease modeler," *Letters in Biomathematics*, vol. 5, no. 1, pp. 195-221, 2018.

[8] R. L. Burden, J. D. Faires, and A. C. Reynolds, *Numerical Analysis*, 2nd edition, Prindle, Weber & Schmidt, Boston, MA., 1981.

[9] X. Chen and B. Yu, "First two months of the 2019 coronavirus disease (COVID-19) epidemic in China: realtime surveillance and evaluation with a second derivative model," *Global Health Research and Policy*, 5:7, 9 pp., 2 March 2020.

[10] J. Ciarochi, "How COVID-19 and other infectious diseases spread: mathematical modeling," March 12, 2020. https://triplebyte.com/blog/modeling-infectious-diseases.

[11] Y. Fang, Y. Nie, and M. Penny, "Transmission dynamics of the COVID-19 outbreak and effectiveness of government interventions: A data-driven analysis," *Journal of Medical Virology*, published on 6 March 2020.

[12] Z. Hu, Q. Ge, S. Li, L. Jin, and M. Xiong, "Artificial intelligence forecasting of Covid-19 in China," https://arxiv.org/ftp/arxiv/papers/2002/2002.07112.pdf.

[13] D. S. Hui, "The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health – The latest 2019 novel coronavirus outbreak in Wuhan, China," *International Journal of Infectious Diseases*, vol. 91, pp. 264-266, 2020.

[14] A. J. Kucharski, T. W. Russell, C. Diamond, Y. Liu, J. Edmunds, S. Funk, and R. M. Eggo, "Early dynamics of transmission and control of COVID-19: a mathematical modelling study," *The Lancet*, published online March 11, 2020.

[15] Q. Li, *et al.*, "Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia," *The New England Journal of Medicine*, vol. 382, no. 13, pp. 1199-1207, March 26, 2020.

[16] T. Liu, *et al.*, "Transmission dynamics of 2019 novel coronavirus (2019-nCoV)," bioRxiv, https://www.biorxiv.org/content/10.1101/2020.01.25.919787v1.

[17] K. McIntosh, “Coronavirus disease 2019 (COVID-19)," https://www.uptodate.com/contents/coronavirus-disease-2019-covid-19?source=history_widget.

[18] P. Munz, I. Hudea, J. Imad, and R. J. Smith, “When zombies attack!: mathematical modelling of an outbreak of zombie infection," *Infectious Disease Modelling Research Progress*, J. M. Tchuenche and C. Chiyaka, eds., pp. 133-150, Nova Science Publishers, Inc., 2009.

[19] H. Nishiura, N. M. Linton, and A. R. Akhmetzhanov, “Serial interval of novel coronavirus (COVID-19) infections," *International Journal of Infectious Diseases*, vol. 93, pp. 284-286, 2020.

[20] J. T. Wu, K. Leung, and G. M. Leung, “Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study," *The Lancet*, vol. 395, pp. 689-697, February 29, 2020.

[21] S. Zhang, M. Diao, W. Yu, L. Pei, Z. Lin, and D. Chen, “Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: A data-driven analysis," *International Journal of Infectious Diseases*, vol. 93, pp. 201-204, 2020.

[22] S. Zhao, *et al.*, “Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak," *International Journal of Infectious Diseases*, vol. 92, pp. 214-217, 2020.

[23] F. Zhou, *et al.*, “Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study," *The Lancet*, vol. 395, pp. 1054-1062, March 28, 2020.