Predicting the number of total COVID-19 cases and deaths in Brazil by the Gompertz model

Jemy A. Mandujano Valle

Abstract In this work, we estimate the total number of infected and deaths by COVID-19 in Brazil and two Brazilian States (Rio de Janeiro and Sao Paulo). To obtain the unknown data, we use an iterative method in the Gompertz model, whose formulation is well known in the field of biology. Based on data collected from the Ministry of Health from February 26, 2020, to July 2, 2020, we predict, from July 3 to 9 and at the end of the epidemic, the number of infected and killed for the whole country and for the Brazilian states of Sao Paulo and Rio de Janeiro. We estimate, until July 9, 2020, a total of 1,709,755 cases and 65,384 deaths in Brazil, 331,718 cases and 15,621 deaths in Sao Paulo, 134,454 cases and 11,574 deaths in Rio de Janeiro. We also estimate the basic reproduction number $R_0$ for Brazil and its two states. The estimated values ($R_0$) were 1.3, 1.3, and 1.4 for Brazil, Sao Paulo, and Rio de Janeiro, respectively. The results show a good fit between the observed data and those obtained by the Gompertz. The proposed methodology can also be applied to other countries and Brazilian states, and we provide an executable as well as the source code for a straightforward application of the method on such data.

Keywords Model Gompertz · Minimal error method · Inverse problem · Covid-19

1 Introduction

The World Health Organization declared the coronavirus disease 2019 a pandemic on March 11, pointing to the over 118,000 cases in over 110 countries and territories around the world at that time [18]. The COVID-19 pandemic in Brazil began on February 26, 2020, when a man from Sao Paulo who returned from Italy tested positive for the virus. The first case of COVID-19 in Rio de Janeiro was confirmed on March 5, 2020. Nowadays, Brazil is considered the epidemic center of Latin America, occupying the second place in the total number of cases and, more recently, in the total number of deaths. Currently (July 9, 2020), Sao Paulo and Rio de Janeiro are the states with the highest number of deaths by the new coronavirus in Brazil, according to data from the Brazilian Health Ministry.

In the past few months, a considerable number of studies related to the evolution of COVID-19 in the world have been submitted and published. In the following, we describe some of these works. Ahmadi et al. [1] developed mathematical models to predict the number of COVID-19 cases in Iran from April 3, 2020, to May 13, 2020. The unknown parameters in these models were estimated by running the fminsearch, a MATLAB function, which is a least-squares algorithm. Torrealba et al. [22] analyzed the modeling and prediction of COVID-19 in Mexico, from an initial approximation, and using the Gauss–Newton algorithm, the authors estimated parameters in...
the Gompertz and Logistic models. Articles [12, 15, 16] estimated the number of total COVID-19 cases and deaths in the world using the Gompertz model. In [11], they analyzed the number of deaths by COVID-19 with social distancing in Brazil, Sao Paulo. Using the SEIR model, the authors recommend temporary lockdowns, however, great the economic and social costs. In [2], the authors studied the impact on the evolution of cases in Rio de Janeiro, Brazil, considering a Susceptible-Infectious-Quarantined-Recovered (SIQR) model with containment.

In this work, we model the near-future trajectory of the cumulative number of infections and deaths by COVID-19 for Brazil and two Brazilian states, given the cumulative number of infected cases and deaths from February 26, 2020 to July 2, 2020. Data were obtained from the Brazilian Ministry of Health until July 2 (https://covid.saude.gov.br), and we considered the cumulative reports from the date on which the first case was notified in Brazil and in each analyzed state. We also estimate the basic reproduction number ($R_0$), which represents the average number of secondary infections generated by each infected person. To estimate the unknown parameters, we use the Gompertz model and propose a gradient type iterative method (Minimal error method).

The Gompertz model is one of the particular cases of the Richards model, as well as Brody, negative exponential, logistic and von Bertalanffy models, see [20, 21].

The growth functions can be grouped in three main categories: those without inflection point (Brody and negative models), those with sigmoidal shape and a fixed inflection point (Gompertz, logistic, and von Bertalanffy models), and those with a flexible inflection point (Richards model). The logistic, Gompertz, and von Bertalanffy models exhibit inflection points at about 50, 37, and 30% of the upper asymptote, respectively, which means that the Gompertz and von Bertalanffy processes are asymmetric, whereas the logistic is a symmetric process, see [19].

The Gompertz and the logistic models are the most frequently used sigmoid functions, and the literature on these models is extensive [3, 21]. In general, the cumulative number of deaths and cases by COVID-19 presents an asymmetrical sigmoidal growth curve. Therefore, using an inappropriate growth curve can have a substantial impact on forecasting [3].

The Gompertz model [5] was proposed by Benjamin Gompertz in 1825. Since then, this exponential model has been used to describe growth in plants, animals, bacteria, and cancer cells [22]. The Gompertz differential equation has the following form,

$$N_t = r N \ln \left( \frac{K}{N} \right), \quad \text{with} \quad N(t_0) = N^1. \quad (1)$$

where $t$ represents time; $N(t)$ is the cumulative population size at time $t$; $r$ the intrinsic growth rate of model; $K$ is the maximum value of model ($N$) when $t$ goes to infinity; $t_0$ represents the initial time; and $N^1$ is the initial population or condition.

The analytic solution from (1) is

$$N(t) = Ke^{\ln((N^1/K) \times e^{-r(t-t_0)})}, \quad (2)$$

where $\lim_{t \to \infty} N(t) = K$.

The turning point is the time at which the rate of accumulation changes from increasing to decreasing or vice versa and can be easily located by finding the inflection point of the epidemic curve, that is, the moment at which the trajectory begins to decline. Clearly, this quantity is of epidemiological importance, indicating either the beginning (i.e., the moment of acceleration after deceleration) or end (i.e., the moment of deceleration after acceleration) of a phase [6]. From Eq. (2), it is trivial to show that the inflection point is given by

$$t_{i,p} = t_0 + \frac{\ln((K/N^1))}{r}. \quad (3)$$

At the inflection point, the number of infected or killed is given by $N(t_{i,p}) = K/e$.

In this paper, we consider the discrete function from Eq. (2) and for $t_0 = 1$,

$$N_i = Ke^{\ln((N^1/K) \times e^{-r(i-1)})}, \quad i = 1, 2, \ldots, m; \quad (4)$$

where $i$ represents time in days and $N_i$ is the cumulative number of infected cases (or deaths) at day $i$.

In this paper, we consider that equality (3) is the inflection point of Eq. (4).

We denote $M = (N^1, N^2, \ldots, N^m) \in \mathbb{R}^m$. The inverse problem is to estimate $x = (K, r)$, from (4), given $M$.

2 Method: inverse problem to the Gompertz model

Here, we consider an application of the minimal error method (MEM) to the inverse problem at the Gompertz
model. The MEM is a variant of the conjugate gradient method [4,10,17], and this approach was used to estimate unknown parameters in a computational neuroscience model [24,25]. Also, [14] solves the Cauchy problem in linear elasticity with MEM.

Knowing the cumulative number of infections or deaths from COVID-19 (M), we want to determine x assuming that Eq. (4) holds.

The transpose of a vector y is denoted by yT. Let the nonlinear operator \( F : \mathbb{R}^2 \rightarrow \mathbb{R}^m \), defined by

\[
F(x) = \begin{bmatrix} f^1(x) \\ f^2(x) \\ \vdots \\ f^m(x) \end{bmatrix}^T = \begin{bmatrix} N^1 \\ N^2 \\ \vdots \\ N^m \end{bmatrix} = M,
\]

where \( f^1, f^2, \ldots, f^m : \mathbb{R}^2 \rightarrow \mathbb{R} \) are multivariable functions with real value and M solves (4).

To obtain an approximation for x, given M and x1, we used the minimal error iteration

\[
x_{k+1} = x_k + w_k F'(x_k)^*(M - F(x_k)),
\]

where \( F'(x_k) \) is the directional derivative of \( F \) computed at \( x_k \), and \( F'(x_k)^* \) is its adjoint. We also define

\[
w_k = \frac{\|M - F(x_k)\|_{\mathbb{R}^m}^2}{\|F'(x_k)^*(M - F(x_k))\|_{\mathbb{R}^2}^2},
\]

where \( F(x_k) = M_k \) solves (4), replacing \( x = (K, r) \) by \( x_k = (K_k, r_k) \). Note that \( M_k = (N_k^1, N_k^2, \ldots, N_k^m) \).

It is possible to show that, under certain conditions (we assume that is the case), \( x_k \) converges to a solution of \( F(x) = M \); see [9, Theorem 3.21].

From Eq. (5), we obtain an approximation \( x_k \) for \( x \), but the adjoint \( F'(x_k) \) is not known. In this paper, we obtain the unknown operator (see Appendix) together with the following iteration

\[
K_{k+1} = K_k + w_k \sum_{i=1}^m (N^i - N^i_k) \frac{\partial f^i(x_k)}{\partial K},
\]

\[
r_{k+1} = r_k + w_k \sum_{i=1}^m (N^i - N^i_k) \frac{\partial f^i(x_k)}{\partial r},
\]

where

\[
\frac{\partial f^i(x_k)}{\partial K} = \frac{f^i(x_k)}{K_k} \left(1 - e^{-r_k(i-1)}\right),
\]

\[
\frac{\partial f^i(x_k)}{\partial r} = -f^i(x_k) \left(N^1 - \ln(K_k)\right) (i - 1) e^{-r_k(i-1)},
\]

\[
w_k = \frac{\|M - M_k\|_{\mathbb{R}^m}^2}{\sum_{i=1}^m \left((N^i - N^i_k)^2 + (N^i - \bar{N})^2\right)^2}.
\]

The regression coefficient \( R^2 \) is used to evaluate the fitting ability of various methods and can be obtained by the following equation,

\[
R^2_k = 1 - \frac{\sum_{i=1}^m (N^i - N^i_k)^2}{\sum_{i=1}^m (N^i - \bar{N})^2},
\]

where

\[
\bar{N} = \frac{\sum_{i=1}^m N^i}{m}
\]

is the average of the cumulative confirmed COVID-19 cases (or deaths).

| Table 1 | Estimates of parameters unknown |
|---------|-------------------------------|
|         | K                | r               | R^2              | Peak day (t_i,p) | R_0   |
| Brazil  | 4,618,893  | 0.0204         | 0.999           | July 9          | 1.3   |
| Cases   | 91,440    | 0.0309         | 0.999           | June 4          | –     |
| Deaths  | 820,202   | 0.0202         | 0.997           | July 4          | 1.3   |
| Sao Paulo | 20,802   | 0.0311         | 0.994           | May 30          | –     |
| Cases   | 255,854   | 0.0235         | 0.998           | June 20         | 1.4   |
| Deaths  | 17,585    | 0.0275         | 0.998           | June 7          | –     |

Second and third columns are estimated values for the Gompertz model (4). Note that K is the number of infected (or killed) at the end of the pandemic. The fourth column describes the regression coefficient according to Eq. (8). The fifth column presents the date when the number of cases or deaths reaches its peak (i.e., the moment of deceleration after acceleration), see Eq. (3). Finally, the sixth column is the basic reproduction number, see Remark (2).
Table 2  Estimated number of cases and deaths by days in Brazil and its two states

| Date    | Total cases | Total deaths |
|---------|-------------|--------------|
|         | Prediction  | Real         | Prediction | Real |
| Brazil  |             |              |            |      |
| July 3  | 1,501,970   | 1,539,081    | 61,059     | 63,174 |
| July 4  | 1,536,474   | 1,577,004    | 61,815     | 64,265 |
| July 5  | 1,571,050   | 1,603,055    | 62,557     | 64,867 |
| July 6  | 1,605,681   | 1,623,284    | 63,285     | 65,487 |
| July 7  | 1,640,353   | 1,668,589    | 63,998     | 66,741 |
| July 8  | 1,675,049   | 1,713,160    | 64,698     | 67,964 |
| July 9  | 1,709,755   | 1,755,779    | 65,384     | 69,184 |

Sao Paulo
July 3  | 295,141     | 310,702      | 14,729     | 15,694 |
July 4  | 301,245     | 312,530      | 14,886     | 15,996 |
July 5  | 307,349     | 320,179      | 15,039     | 16,078 |
July 6  | 313,451     | 323,070      | 15,190     | 16,134 |
July 7  | 319,548     | 332,708      | 15,337     | 16,475 |
July 8  | 325,637     | 341,365      | 15,480     | 16,788 |
July 9  | 331,718     | 349,715      | 15,621     | 17,118 |

Rio de Janeiro
July 3  | 121,965     | 118,956      | 10,738     | 10,500 |
July 4  | 124,083     | 120,440      | 10,882     | 10,624 |
July 5  | 126,188     | 121,292      | 11,025     | 10,667 |
July 6  | 128,278     | 121,879      | 11,165     | 10,698 |
July 7  | 130,353     | 124,086      | 11,304     | 10,881 |
July 8  | 132,412     | 126,329      | 11,440     | 10,970 |
July 9  | 134,454     | 128,324      | 11,574     | 11,115 |

The cases and deaths are cumulative. The mean absolute percentage error (MAPE) values for the cumulative number of cases and deaths in Brazil are 2.0% and 3.9%, respectively. Considering the data from Sao Paulo, the MAPEs for the cumulative number of cases and deaths are 4.3% and 6.9%, respectively. Finally, the MAPEs for the cumulative number of cases and deaths in Rio de Janeiro are 4.4% and 3.4%, respectively.

Remark 1  From differential Eq. (1) and since $M$ is known, we consider the following initial approximation for $x$

$$x_1 = \left(1.2N^m, \frac{N^m - N^{m-1}}{N^m \ln(1.2N^m/N^{m-1})}\right).$$

(9)

The numerical scheme for our method poposed is in Algorithm 1.

Remark 2  The basic reproduction number $R_0$ is the most important parameter to analyze any epidemic model for any disease. In a practical sense, $R_0$ is the average number of secondary infections caused by one infected individual during his/her entire infectious period. An epidemic occurs only if a single individual can spread its illness to more than one individual ($R_0 > 1$). If $R_0 < 1$, on average an infectious individual infects less than one person and the contagion is expected to stop spreading.

Our result can be used to compute the basic reproduction number

$$R_0 = \exp(rt),$$

see [6–8,13], where $r$ denotes the intrinsic growth rate in the Gompertz model and $T$ is the generation time of
disease transmission. In this paper, we consider $T = 14$.

3 Results

In this section, we will provide the results obtained for the whole country (Brazil) and for the Brazilian states of Sao Paulo and Rio de Janeiro. Table 1 describes the estimated values for the Gompertz model and also presents the regression coefficient, the inflection point, and the basic reproduction number.

The estimated daily values for the cumulative number of confirmed cases and deaths, from July 3 to 9, are illustrated in Table 2. Figures 1, 2, 3 illustrate the relationship between observed values and those predicted by the Gompertz model. The tables and figures show the efficiency of the proposed iterative method.

4 Conclusion

In this article, we propose an iterative method to estimate the unknown parameters in the Gompertz model. The results of the modeling show a good fit between the estimated and the observed data. We obtained an estimate of 99.9% ($R^2$) for the cumulative number of infected and killed in Brazil. For Sao Paulo, we obtained an approximation of 99.7% and 99.4% for the cumulative number of infected and deaths, respec-
maximum number of infected and deaths. We obtained the cumulative number of infected and deaths.

Regarding the methodology used, there are several methods for estimating or determining parameters in a mathematical model. We can divide these methods into two groups, iterative and non-iterative methods. In this paper, we used a gradient type iterative method. This method can recover unknown parameters with a non-uniform distribution (non-constant functions) given noisy data. On the other hand, like all iterative methods, it depends on an initial estimate. If this initial approximation is far from the solution, the iteration will diverge. Although it is difficult to determine an initial estimate for some problems, in this work we found an appropriate initial guess, see Remark 1.

Acknowledgements The author would like to thank the National Council for Scientific and Technological Development of Brazil - CNPq (Grant No. 301330/2020-4) and PCI/LNCC (Programa de Capacitação Institucional do Laboratório Nacional de Computação Científica) for the financial support of this work.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Appendix

Theorem 1 Consider the iteration in Eq. (5). Then, Eqs. (6) and (7) holds.

Proof Given \( x_k = (K_r, r_k) \in \mathbb{R}^2 \) and unit vector \( \theta \in \mathbb{R}^2 \), the directional derivative of \( F \) at \( x_k \) in the direction \( \theta \) is given by

\[
F'(x_k)(\theta) = \lim_{\lambda \to 0} \left( \frac{F(x_k + \lambda \theta) - F(x_k)}{\lambda} \right) = J_F(x_k) \cdot \theta, \tag{11}
\]

where \( J_F(x_k) = (\nabla f_1(x_k), \nabla f_2(x_k), \ldots, \nabla f_m(x_k))^T \in \mathbb{R}^{m \times 2} \) is the Jacobian matrix of \( F \) at \( x_k \), \( \nabla f^i(x_k) \) represents the gradient of function \( f^i \) at \( x_k \), and \( J_F(x_k) \cdot \theta \) denotes the multiplication between matrix \( J_F(x_k) \) and unit vector \( \theta \).

From the minimal error iteration in Eq. (5), we gather that

\[
\langle x_{k+1} - x_k, \theta \rangle_{\mathbb{R}^2} = w_k \langle F'(x_k)^*(M - F(x_k)), \theta \rangle_{\mathbb{R}^2},
\]

By the definition of adjoint operator,

\[
\langle x_{k+1} - x_k, \theta \rangle_{\mathbb{R}^2} = w_k \langle M - M_k, F'(x_k)(\theta) \rangle_{\mathbb{R}^m},
\]

From previous equation and by the definition of Euclidean inner product on \( \mathbb{R}^m \), we obtain

\[
\langle x_{k+1} - x_k, \theta \rangle_{\mathbb{R}^2} = \sum_{i=1}^{m} (N_i - N_k) \langle \nabla f^i, \theta \rangle,
\]
by definition of Euclidean inner product on $\mathbb{R}^2$, we have
text{where $w \in \mathbb{R}^2$ is arbitrary, we gather that the following iteration holds:}

$x_{k+1} = x_k + w_k \sum_{i=1}^{m} (N^i - N_k^i) \nabla f_i(x_k),$

where

$w_k = \frac{\| M - M_k \|^2_{\mathbb{R}^m}}{\sum_{i=1}^{m} (N^i - N_k^i) \nabla f_i(x_k)}.$

□

References

1. Ahmadi, A., Fadai, Y., Shirani, M., Rahmani, F.: Modeling and forecasting trend of covid-19 epidemic in Iran until May 13, 2020. Med. J. Islam. Repub. Iran (MJIRI) 34(1), 183–195 (2020)
2. Crokidakis, N.: Covid-19 spreading in Rio de Janeiro, Brazil: do the policies of social isolation really work? Chaos Solitons Fract. 136, 109930 (2020)
3. Franses, P.H.: A method to select between Gompertz and Logistic trend curves. Technol. Forecast. Soc. Change 46(1), 45–49 (1994)
4. George, S., Sabari, M.: Convergence rate results for steepest descent type method for nonlinear ill-posed equations. Appl. Math. Comput. 294, 169–179 (2017)
5. Gompertz, B.: On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. In a letter to Francis Baily, Esq F.R.S. & co. Philos. Trans. R. Soc. Lond. 115, 513–583 (1825)
6. Hsieh, Y.H.: Richards model: a simple procedure for real-time prediction of outbreak severity. In: Ma, Z., Zhou, Y., Wu, J. (eds.) Modeling and Dynamics of Infectious Diseases, pp. 216–236. World Scientific, Singapore (2009)
7. Hsieh, Y.H., Ma, S.: Intervention measures, turning point, and reproduction number for dengue, Singapore, 2005. Am. J. Trop. Med. Hyg. 80(1), 66–71 (2009)
8. Hsieh, Y.H., Ma, S., Hernandez, J.X.V., Lee, V.J., Lim, W.Y.: Early outbreak of 2009 influenza a (H1N1) in Mexico prior to identification of ph1n1 virus. PLoS ONE 6(8), e23853 (2011)
9. Kaltenbacher, B., Neubauer, A., Scherzer, O.: Iterative Regularization Methods for Nonlinear Ill-posed Problems, vol. 6. Walter de Gruyter, Berlin (2008)
10. King, J.: A minimal error conjugate gradient method for ill-posed problems. J. Optim. Theory Appl. 60(2), 297–304 (1989)
11. Komatsu, B.K., Menezes-Filho, N.: Simulando os efeitos de políticas de distanciamento social em são paulo usando um modelo seir. Insper, Centro de Políticas Públicas, Policy Paper no 44, 2020. Available at https://www.insper.edu.br/wp-content/uploads/2020/06/Covid-1.pdf
12. Lee, S.Y., Lei, B., Mallick, B.K.: Estimation of covid-19 spread curves integrating global data and borrowing information. arXiv:2005.00662 (2020)
13. Liu, W., Tang, S., Xiao, Y.: Model selection and evaluation based on emerging infectious disease data sets including A/H1N1 and ebola. Comput. Math. Methods Med 2015, 207105 (2015). https://doi.org/10.1155/2015/207105
14. Marin, L.: The minimal error method for the Cauchy problem in linear elasticity. Numerical implementation for two-dimensional homogeneous isotropic linear elasticity. Int. J. Solids Struct. 46(5), 957–974 (2009)
15. Mazurek, J., Nenickova, Z.: Predicting the number of total covid-19 cases and deaths in the USA by the gompertz curve. (2020). https://doi.org/10.13140/RG.2.2.19841.81761
16. Mazurek, J., Rico, C.P., Garcia, C.F.: Forecasting the number of covid-19 cases and deaths in the world, uk, russia and turkey by the gompertz curve. (2020). https://doi.org/10.13140/RG.2.2.32701.61925
17. Neubauer, A., Scherzer, O.: A convergence rate result for the steepest descent method and a minimal error method for the solution of nonlinear ill-posed problems. Zeitschrift für Analysis und ihre Anwendungen 14(2), 369–377 (1995)
18. Perc, M., Gorišek Miksič, N., Slavínek, M., Stožer, A.: Forecasting covid-19. Front. Phys. 8, 127 (2020)
19. Teleken, J.T., Galvão, A.C., Robazza, W.D.S.: Comparing non-linear mathematical models to describe growth of different animals. Acta Sci. Animal Sci. 39(1), 73–81 (2017)
20. Tjørve, E., Tjørve, K.M.: A unified approach to the Richards-model family for use in growth analyses: why we need only two model forms. J. Theor. Biol. 267(3), 417–425 (2010)
21. Tjørve, K.M., Tjørve, E.: The use of Gompertz models in growth analyses, and new Gompertz-model approach: an addition to the unified-Richards family. PloS ONE 12(6), e0178691 (2017)
22. Torrealba-Rodriguez, O., Conde-Gutiérrez, R., Hernández-Javier, A.: Modeling and prediction of covid-19 in Mexico applying mathematical and computational models. Chaos Solitons Fract. 138, 109946 (2020)
23. Valle, J.A.M.: Code to estimate the number of infected and killed by covid-19: Program. https://github.com/MandujanoValle/COVID-19 (2020)
24. Valle, J.A.M., Madureira, A.L.: Parameter identification problem in the Hodgkin and Huxley model. arXiv:1903.10286 (2019)
25. Valle, J.A.M., Madureira, A.L., Leitão, A.: A computational approach for the inverse problem of neuronal conductances determination. J. Comput. Neurosci. 48, 1–19 (2020)