

Monte Carlo approach to model COVID-19 deaths and infections using Gompertz functions

Tulio Rodrigues * and Otaviano Helene 

Experimental Physics Department, Physics Institute, University of São Paulo, P. O. Box 66318, CEP 05315-970, São Paulo, Brazil



(Received 13 August 2020; accepted 25 November 2020; published 16 December 2020)

This study provides a phenomenological method to describe the exponential growth, saturation, and decay of coronavirus disease 2019 (COVID-19) deaths and infections via a Monte Carlo approach. The calculations connect Gompertz-type trial distributions of infected people per day with the distribution of deaths adopting two gamma distributions to account for the elapsed time that encompass the incubation and symptom onset to death periods. The analyses include death data from the USA, Brazil, Mexico, the United Kingdom (UK), India, and Russia, which comprise the four countries with the highest number of deaths and the four countries with the highest number of confirmed cases, as of August 07, 2020, according to the World Health Organization webpage. The Gompertz functions were fitted to the data of weekly averaged confirmed deaths per day by mapping the χ^2 values. The uncertainties, variances, and covariances of the model parameters were calculated by propagation, taking into account the standard errors of the data for each epidemiological week. The fitted functions for the average deaths per day for the USA and India have an upward trend, with the former having a higher growth rate and quite huge uncertainties. For Mexico, the UK, and Russia, the fits are consistent with a downward-sloping pattern. For Brazil we found a subtle trend down but with significant uncertainties. The USA, UK, and India data showed first peaks with higher growth rates compared with the second ones (4.2, 2.2, and 3.5 times higher, respectively), demonstrating the benefits of nonpharmacological interventions of sanitary measures and social distance flattening the secondary peaks of the pandemic. For the case of the USA, however, a third peak seems quite plausible, most likely related with the recent relaxation policies. Brazil's data are satisfactorily described by two highly overlapped Gompertz functions with similar growth rates, suggesting a two-step process for the pandemic spreading. For the cases of Mexico and Russia single peaks with smoother slopes fitted the data satisfactorily. The 95% confidence intervals for the total number of deaths ($\times 10^3$) predicted by the model for August 31, 2020, are 160 to 220, 110 to 130, 59 to 62, 41.3 to 41.4, 54 to 63, and 16.0 to 16.7 for the USA, Brazil, Mexico, the UK, India, and Russia, respectively. Our estimates for the point prevalences of infections are compared with some preliminary data from serological studies and/or model calculations focused on the USA, Brazil, and UK scenarios. The point prevalences and 95% confidence intervals for August 1, 2020, were estimated to be 5.7 (3.9–7.5)%, 8.9 (7.4–10.3)%, 9.3 (8.3–10.3)%, 5.7 (4.5–6.9)%, 0.9 (0.8–1.0%), and 1.2 (1.0–1.3)% for the USA, Brazil, Mexico, the UK, India, and Russia, respectively. The method represents an effective few-parameter framework to estimate the line shape of the infection curves and the uncertainties of the relevant parameters based on the actual death data of a pandemic.

DOI: [10.1103/PhysRevResearch.2.043381](https://doi.org/10.1103/PhysRevResearch.2.043381)

I. INTRODUCTION

The outbreak of the new coronavirus disease 2019 (COVID-19) brought a challenging scenario worldwide [1–3], urging timely and effective responses from the authorities regarding the availability of intensive care units [4,5], as well as the implementation of nonpharmacological interventions of social distance and protective sanitary measures [6–8]. Epidemiological models [9–18] and other statistical approaches [19–21] have been very useful to guide actions to manage

this crisis and to shed light on how to safely and gradually resume economics and social activities [22]. On the other hand, quantitative analyses are strongly susceptible to several uncertainties, such as underreporting of confirmed cases and deaths [23–25], lack of massive tests in some countries [26], changes in policies and methods for reporting confirmed cases and deaths as time evolves during the pandemic growth, and very distinct socioeconomic patterns and health facility capabilities among different countries and also among different focuses of the disease within the same country. Some analyses of the overall mortality during the COVID-19 pandemic have also been quite useful to shed light on the outbreak of the disease. Weinberger *et al.* [27] have found that the number of excess all-cause deaths in the USA from March 01 through May 30, 2020, was 28% higher than the number of deaths officially attributed to COVID-19. Another study based on a time-series analysis [28] found an excess mortality in Italy correlated in time with the official COVID-19 deaths but a

*tulio@if.usp.br

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

factor 1.5 higher. Such a complex and puzzling scenario directly influences the forecast capabilities of the calculations, supporting the need for a multidisciplinary cooperation of the scientific community and the development of mathematical models to provide plausible estimates for the uncertainties of the relevant parameters [29–32]. Therefore, the present analysis provides an effective phenomenological method to estimate the magnitude and the relevant uncertainties of some important quantities as the pandemic evolves, such as (1) the peak time(s), growth rate(s), and total number of infected people for the reconstructed infection curves; (2) forecast distribution of deaths per day; and (3) forecast total number of deaths until August 31, 2020. It is worth mentioning, however, that the predictions of the model strongly depend on the prevailing conditions already in place for the specific country which data were analyzed, as any substantial change in governmental policies, either in the direction of loosening or tightening social distance, will generate different dynamics for the spread of the virus.

The best-fit parameters are generally strongly correlated, but its uncertainties reflect the dispersion of the data at each epidemiological week, making the present analysis a suitable quantitative method to describe the pandemic line-shape behavior with few parameters. Moreover, such approach could also be useful for the identification of upward trends and its correlations with relaxation policies and procedures as global social and economics activities are gradually resumed [33].

II. METHODS

A. Modelling the infections and deaths of a pandemic with a Monte Carlo algorithm

The calculations assume that the total number of infected people increases according to a Gompertz-type function, which is a sigmoid curve with a lower growth rate at the beginning and at the end, such that:

$$I(t) = Ne^{-e^{-\lambda(t-t_0)}}, \quad (1)$$

where N represents the asymptotic number of infected people for $t \rightarrow \infty$, λ is the growth rate, and t_0 the peak time of the derivative of $I(t)$. In such model, the number of infected people per time period can be written as:

$$G(t) = \frac{d}{dt}I(t) = N\lambda e^{-e^{-\lambda(t-t_0)}} e^{-\lambda(t-t_0)}. \quad (2)$$

For the specific case of COVID-19, the data of confirmed cases depend very strongly on testing and reporting policies for the related country. These policies may also vary along time, distorting the shape of the distributions. This complicated scenario disfavors the use of confirmed cases as a reliable source of information to describe the pandemic dynamics, which is crucial to guide government actions and decisions. In order to overcome these difficulties, we have adopted the data of deaths per day, as they should be more consistent with the actual spreading mechanism of the virus. The connection between the trial function of the number of infected people per day and the number of deaths per day takes into account the probability distribution function for the elapsed time between the infection and the death, which can be satisfactorily described by the sum of two time

periods, namely the incubation period, t_{inc} , and the symptom onset to death period, t_{s-d} . Both periods are generated in the Monte Carlo algorithm, assuming that they are independent and gamma distributed with averages of 5.1 and 17.8 days and coefficients of variation of 0.86 and 0.45, respectively, as proposed elsewhere [34] based on previous studies of Wuhan data [6,35]. In fact, the average elapsed time from symptom onset to death was corrected to 17.8 days in Ref. [35], superseding the previous parameter of 18.8 days used in Ref. [34]. So the time of the death t_d can be written as $t_d = t_{\text{inf}} + t_{\text{inc}} + t_{s-d}$, with t_{inf} representing the time of infection.

B. Data collection and fitting procedure

The analyses of the death data¹ were done considering the weekly averaged deaths per day and its corresponding standard error for the respective epidemiological week, counted retrospectively from the end date of August 7, 2020. The time counting considered the day of the first confirmed case for each country and the first epidemiological weeks were chosen in such a way that all days of the week had at least one death, meaning that the present calculations do consider the early stages of the pandemic, including from 19 to 22 weeks depending on the country (see Table I). The mean day of each epidemiological week was chosen with a time bin of ± 3.5 days, and the mean of each day was taken at the half of the day. The trial line shapes of the number of infected people per day were generated considering a single Gompertz function for the case of Mexico and Russia, a sum of two Gompertz functions for Brazil, the United Kingdom, and India, and three Gompertz functions for the USA. Such an approach allows the inclusion of two or more superimposed dynamics of the disease as one would expect considering the changes in policies as time evolves (which might remarkably reduce the growth rates) and also the cluster structure expected for large countries with multiple focuses of the disease (not observed for Mexico and Russia so far). Such a method would also allow the inclusion of other peaks as the countries start their processes of reopening social and economic activities, which seems to be the case for the USA.

The fitting procedure was performed by mapping the χ^2 , defined as:

$$\chi^2 = \sum_{iw=1}^n \frac{(\tilde{F}_{iw} - y_{iw})^2}{\sigma_{y_{iw}}^2}, \quad (3)$$

where \tilde{F}_{iw} is the trial function for deaths calculated at time t_{iw} (the mean day of the corresponding week), y_{iw} the weekly averaged deaths per day, and $\sigma_{y_{iw}}$ its standard error. In that sense, the average values with higher standard errors had lower weights in the fitting procedure. Our particularly choices for the most suitable number of Gompertz functions adopted for each country were based on the statistical relevance of adding one more Gompertz function to the previous

¹The deaths data included in the analysis were taken from the European Centre for Disease Prevention and Control webpage: <https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide>

TABLE I. Best-fit parameters of the Gompertz functions of the reconstructed infection curves and fitting results of the death data obtained for all six countries. Also shown the 95% CI for the total number of deaths for August 31, 2020, the adopted *ifr* values for each country and the estimated point prevalences at the first day of each month.

	USA	Brazil	Mexico	UK	India ^a	Russia ^a
$N_1(10^6)$	7.9 ± 0.8	13 ± 17	16.3 ± 1.0	2.8 ± 0.9	0.5 ± 0.6	2.01 ± 0.10
$\lambda_1(d^{-1})$	0.109 ± 0.012	0.030 ± 0.010	0.0220 ± 0.0010	0.110 ± 0.023	0.06 ± 0.03	0.0260 ± 0.0011
$t_{01}(d)$	65.6 ± 1.0	76 ± 26	103 ± 3	50.5 ± 2.8	68 ± 13	112.8 ± 2.2
$N_2(10^6)$	8.8 ± 0.8	12 ± 26	—	1.0 ± 0.9	30 ± 12	—
$\lambda_2(d^{-1})$	0.026 ± 0.004	0.023 ± 0.023	—	0.051 ± 0.006	0.017 ± 0.005	—
$t_{02}(d)$	113 ± 11	135 ± 29	—	80 ± 22	180 ± 21	—
$N_3(10^6)$	4 ± 6	—	—	—	—	—
$\lambda_3(d^{-1})$	0.07 ± 0.07	—	—	—	—	—
$t_{03}(d)$	179 ± 18	—	—	—	—	—
N_d (95% CI) (10^3) ^b	180 (160–220)	120 (110–130)	60 (59–62)	41.4 (41.3–41.4)	59 (54–63)	16.3 (16.0–16.7)
<i>ifr</i> (%) ^c	0.96	0.59	0.48	1.09	0.41	0.92
Point prevalences(%)						
March 1, 2020	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1	< 0.1
April 1, 2020	1.6 (1.3–1.8)	0.2 (0.2–0.3)	0.1 (0.09–0.13)	3.1 (2.4–3.8)	< 0.1	< 0.1
May 1, 2020	3.0 (2.6–3.5)	1.6 (1.4–1.8)	1.1 (1.0–1.2)	5.0 (3.9–6.1)	< 0.1	0.2 (0.21–0.26)
June 1, 2020	3.9 (3.3–4.5)	4.1 (3.7–4.6)	3.7 (3.3–4.0)	5.5 (4.3–6.7)	0.2 (0.18–0.24)	0.6 (0.57–0.68)
July 1, 2020	4.5 (3.7–5.3)	6.7 (5.9–7.5)	6.7 (6.1–7.3)	5.7 (4.5–6.9)	0.5 (0.45–0.57)	1.0 (0.9–1.0)
August 1, 2020	5.7 (3.9–7.5)	8.9 (7.4–10.3)	9.3 (8.3–10.3)	5.7 (4.5–6.9)	0.9 (0.8–1.0)	1.2 (1.0–1.3)
<i>n/n.d.f.</i> ^d	22/13	20/14	19/16	21/15	19/13	19/16
χ^2/p ^e	10.3/0.670	11.8/0.625	12.3/0.722	9.6/0.846	17.7/0.171	23.5/0.101
χ^2/p ^f	$42.3/3.5 \times 10^{-4}$	29.3/0.0325	—	$75.3/4.3 \times 10^{-9}$	$43.2/2.6 \times 10^{-4}$	—
Begin date ^g	March 07, 2020	March 21, 2020	March 28, 2020	March 14, 2020	March 28, 2020	March 28, 2020

^aFive percent error was added to the death data to achieve a successful fitting.

^bModel estimates for August 31, 2020.

^cThe infection-fatality-ratios were estimated assuming an uniform susceptibility across all age groups and weighting the age-dependent *ifr* values reported by Verity *et al.* [35] with the age composition of the respective country taken from the United Nations World Population Prospects 2019.

^d*n.d.f.*: number of degrees of freedom.

^e*p*: probability of exceeding the χ^2 (*p* value)

^fThese results refer to the fittings performed with just one Gompertz function for Brazil, UK, and India and two Gompertz functions for the USA.

^gThese dates correspond to the first day of the first epidemiological week included in the fitting.

fitting until reaching a probability of exceeding χ^2 (*p* value) greater than 5%.

III. RESULTS

A. Best-fit parameters of the Gompertz functions

The best-fit parameters of the Gompertz functions (N , λ , and t_0) were obtained by sorting 5000 random sets of parameters around plausible guessing values and calculating the respective trial function and χ^2 for each candidate, assuming a total of 10 million infections using a time bin of one day. The trial functions \tilde{F}_{iw} were calculated for each set of parameters by the connection between the 10 million infection events with the respective death events using the probability distribution function of $t_{inc} + t_{s-d}$, herein denoted as $Prob(\Delta t) = Prob(t_d - t_{inf})$. This distribution is shown by the blue histogram of Fig. 1, together with the gamma distributions for the incubation (black histogram) and symptom onset to death (red histogram) periods.

The sorting procedure was done several times with progressively narrow bins for each parameter’s increments until the respective χ^2 converged to the minimal value (the conver-

gence criteria required that the χ^2 obtained for 5000 random sets of parameters is lower than the one achieved for 4000 sets and their difference is lower than 0.05 units).

The normalization of the number of infections given the number of deaths was done assuming specific infection-fatality-ratios (*ifr*) for each country, as shown in Table I. These parameters took into account the age-dependent *ifr* values found by Verity *et al.* [35] and the specific demographic age distributions for each country taken from the United Nations World Population Prospects 2019. We assumed a uniform susceptibility across all age groups in our calculations (see Discussion). For the fittings, we have included three Gompertz functions for the USA; two for Brazil, UK, and India; and one for Mexico and Russia. These choices correspond to the minimal number of parameters that suitably fit the data with $p > 0.05$. Any increase in the number of Gompertz functions of the analyses does not change the statistical relevance of the χ^2 and *p* value of the fittings.

B. Propagation of uncertainties

A least-squares method [36] was applied for the calculation of the uncertainties of the best-fit parameters, which covari-

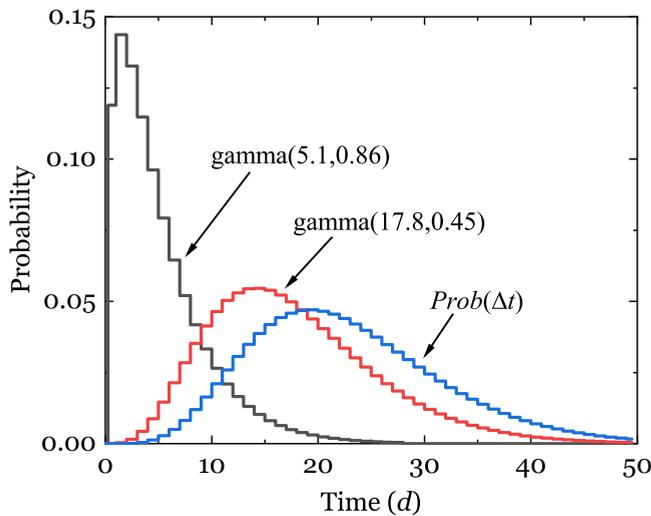


FIG. 1. The Monte Carlo generated $Prob(\Delta t)$ distribution (blue histogram) and the gamma distributions for the incubation (black histogram) and symptom onset to death (red histogram) periods.

ance matrix can be written as:

$$Vb = (\tilde{F}'^T V^{-1} \tilde{F}')^{-1}, \quad (4)$$

where $\tilde{F}' = \tilde{F}'_{iw,j}$ stands for the partial derivative of \tilde{F}_{iw} at any t_{iw} in respect to the P_j parameter of each Gompertz function (N, λ , and t_0). The variance matrix of the death data is diagonal such that $V_{iw,iw} = \sigma y_{iw}^2$. Given the lower number of deaths per day for India and Russia and the huge variation of the data in each week, we have included an additional uncertainty of 5% ($\sigma y_{iw} \rightarrow \sigma y_{iw} + 0.05 y_{iw}$) in order to achieve a successful fitting. For the calculation of $\tilde{F}'_{iw,j}$ we have used the resulting convolution between the reconstructed infection curve $G(t)$ and the probability density function $Prob(\Delta t)$, such that:

$$\tilde{F}'_{iw,j} = \frac{d}{dP_j} [F_C(t_{iw}, N_k, \lambda_k, t_{0k})], \quad \text{with} \quad (5)$$

$$F_C(t) = ifr_s \int_0^t \sum_{k=1}^{k_{\max}} [G(\tau, N_k, \lambda_k, t_{0k}) Prob(t - \tau) d\tau], \quad (6)$$

where $k_{\max} = 1$ for Mexico and Russia; 2 for Brazil, the UK, and India; and 3 for the USA with ifr_s representing the ifr for the specific country. The propagation of the uncertainties of the best-fit parameters to the reconstructed infection curve took into account the full covariance matrix Vb , as similarly described in Ref. [37], with the vector G'_m being defined as:

$$G'_m = \begin{pmatrix} G'_{m,1} \\ \vdots \\ G'_{m,j_{\max}} \end{pmatrix}, \quad (7)$$

where $G'_{m,1}, \dots, G'_{m,j_{\max}}$ are the partial derivatives of the infection curve $G(t_m, N_k, \lambda_k, t_{0k})$ in respect to the parameter $P_1, \dots, P_{j_{\max}}$ calculated at each day t_m ($j_{\max} = 3$ for Mexico and Russia; 6 for Brazil, the UK, and India; and 9 for the USA). Consequently, the uncertainty of the infection curve at each point can be written as:

$$\sigma_m = \sqrt{G_m'^T Vb G'_m}. \quad (8)$$

Assuming a 95% confidence interval (CI) as $\sim 2\sigma_m$, one can write the upper and lower limits of the convoluted functions as:

$$F_C^\pm(t) = F_C(t) \pm ifr_s \int_0^t [2\sigma(\tau) Prob(t - \tau) d\tau], \quad (9)$$

where $\sigma(\tau)$ is obtained by the interpolation of σ_m . Figure 2 shows the weekly averaged deaths per day distributions for all six countries (data points) and its respective convoluted functions $F_C(t)$ (dashed-dotted gray lines). The upper and lower estimates $F_C^\pm(t)$ are presented by the red and blue dashed-dotted lines, respectively. The total number of deaths and its 95% CI at any given time t_f can be written as:

$$N_d(t_f) = N_d(t_i) + \int_{t_i+1}^{t_f} F_C(t') dt', \quad \text{and} \quad (10)$$

$$N_d^\pm(t_f) = N_d(t_i) + \int_{t_i+1}^{t_f} F_C^\pm(t') dt', \quad (11)$$

where $N_d(t_i)$ corresponds to the actual data of accumulated deaths until the day (t_i) for each country.

Table I summarizes all the results, including the model predictions for the total number of deaths for August 31, 2020, as well as our estimates for the point prevalences (%) at the beginning of each month. All the intervals in parentheses consider a 95% CI.

The upper panel of Fig. 3 shows the model predictions for the accumulated number of infections (solid lines) and the lower panel shows the corresponding model estimates for the accumulated deaths, in comparison with the available data (data points) at each 5 day time interval. Once again a nice agreement between the data and the model is verified, with some discrepancies found in the early stages of the pandemic (accumulated number of deaths $\lesssim 100$).

IV. DISCUSSIONS

The countries with well-defined first peaks (USA and UK) present the highest initial growth rates ($0.109 \pm 0.012 \text{ d}^{-1}$ and $0.110 \pm 0.023 \text{ d}^{-1}$) and first peak times with uncertainties of 1.0 and 2.8 days, respectively. The second peaks in both cases have much lower growth rates, demonstrating the flattening of the secondary infection curves probably related with nonpharmacological interventions of social distance and sanitary measures. Besides the UK, which clearly shows a well-controlled scenario so far, Mexico and Russia also have a trend down with modest uncertainties, which is a consequence of the fitting being successfully performed with a single Gompertz function. For the case of India, the first Gompertz function has a quite small contribution ($\sim 2\%$) in the total number of infections and the second curve dominates the distribution of deaths per day, resulting in modest uncertainties. For the case of Brazil, there are few data points to properly constraint the second Gompertz function, which has a similar growth rate compared with the first one. The large overlap between the two functions leads to high correlation coefficients between N_1 and λ_1 (-0.989), N_2 and λ_2 (-0.973), λ_1 and λ_2 (-0.847), and N_1 and N_2 (-0.974), influencing for the large uncertainties. A similar situation also plays a role for the huge uncertainties found in the parameters of the third

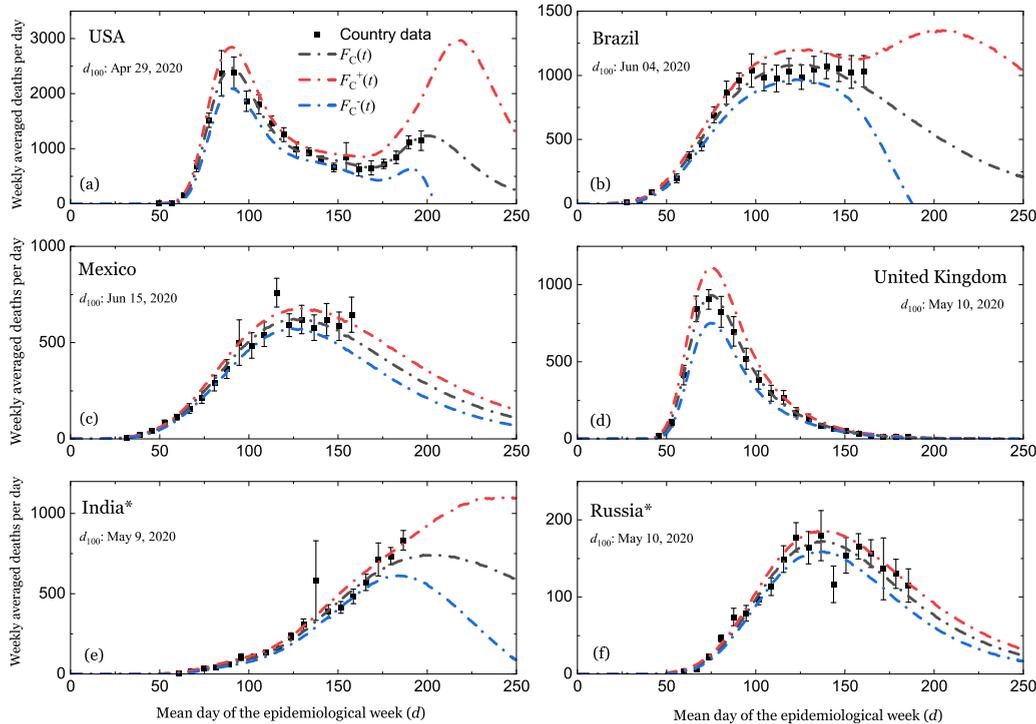


FIG. 2. Weekly averaged deaths per day for the USA (a), Brazil (b), Mexico (c), the United Kingdom (d), India (e), and Russia (f) (data points) and the respective model estimates given by $F_C(t)$ (gray dotted-dashed lines) and its 95% CI upper and lower limits (red and blue dashed dotted lines, respectively). Also shown for clarity the hundredths calendar day (d_{100}) since the first confirmed case for each country. (*) For the case of India and Russia a 5% error was added to the death data to achieve a successful fitting.

Gompertz function for the USA, which is weakly constrained with few data points. The death peak times have an average shift of 22.9 days from the corresponding infection’s peaks, which is the average of $Prob(\Delta t)$ (the sum of the averages of the two gamma functions). In order to evaluate the statis-

tical relevance of the adopted number of Gompertz functions for each country, we performed other fittings decreasing one Gompertz function from the optimal results for the USA, Brazil, the UK, and India. The resulting χ^2 and p values for these fittings are presented in Table I, showing the remarkable statistical relevance that is achieved adding one more Gompertz function. For the case of the USA, the optimal fitting with three Gompertz function has a χ^2 of 10.3 with a p value of 0.670. Decreasing one Gompertz function increases the χ^2 to 42.3 and decreases the probability of exceeding the χ^2 to less than 0.1%. Similar situations occur for Brazil, the UK, and India, showing that the structures in the death data for these countries do reflect the dynamics of the virus spreading in terms of superimposed focuses of the disease.

Our estimates for the point prevalences of infections at the beginning of each month are also shown in Table I for all six countries. For the case of the USA, our estimated prevalence for April 3 and 4, 2020 [1.7% (1.4–2.0%)], is in reasonable agreement with a preliminary serological study carried out in Santa Clara County [38] 2.8% (1.3–4.7%). For the case of Brazil, we found point prevalences of 2.9% (2.5–3.3%) within May 14–21, 2020, and 4.6% (4.0–5.1%) within June 4–7, 2020, which are higher than the overall prevalences found in a survey that included 133 Brazilian cities [1.9% (1.7–2.1%) within May 14–21 and 3.1% (2.8–3.4%) within June 4–7, 2020] [39]. In contrast, our prediction for May 06, 2020 [2.0% (1.7–2.2%)], is significantly lower than previous model estimates [40], where they found infections rates of 3.30% (2.83–3.68%) and 3.35% (2.83–3.78%) for the states of São Paulo and Rio de Janeiro, respectively. According

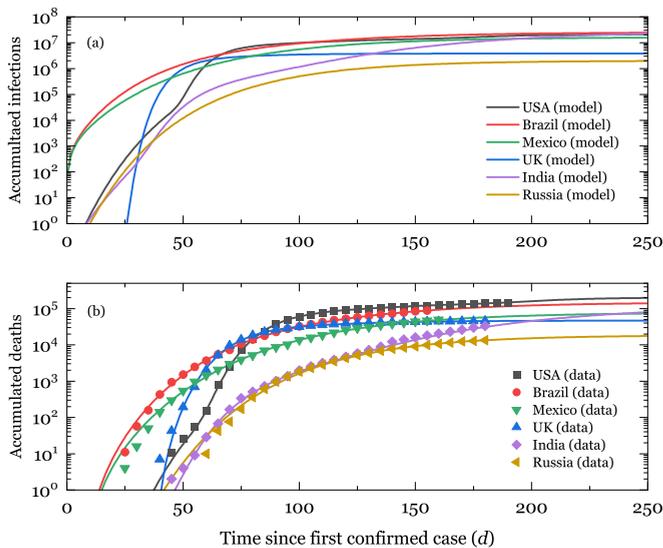


FIG. 3. Upper panel: Model predictions for the accumulated number of infections for all six countries (solid lines). Lower panel: Model predictions for the accumulated deaths (solid lines) versus the available data (data points), which are presented within 5 days’ time intervals for clarity.

TABLE II. COVID-19 *ifr* from Verity *et al.* [35] and the equivalent gamma distribution coefficient of variation (CV) for each age group.

Age groups (years)	<i>ifr</i> (%)	CV (%)
0–9	0.00161 (0.000185–0.0249)	82.0
10–19	0.00695 (0.00149–0.0502)	65.1
20–29	0.0309 (0.0138–0.0923)	39.7
30–39	0.0844 (0.0408–0.185)	35.8
40–49	0.161 (0.0764–0.323)	35.4
50–59	0.595 (0.344–1.28)	28.6
60–69	1.93 (1.11–3.89)	28.7
70–79	4.28 (2.45–8.44)	28.7
≥ 80	7.80 (3.80–13.3)	31.2

to their figures, these states encompass 50% of the deaths and 51.6% of the infections in all the 16 states included in the analysis. Another recent preliminary report based on a serological survey [41] found much higher prevalences in São Paulo [22.4% (19.9–27.6%) within August 10–21, 2020] and Manaus [66.1% (60.8–79.9%), within August 8–19, 2020], in comparison with our estimates for Brazil of 9.6% (7.4–11.8%) within August 8–21, 2020. On the other hand, our result within May 13–15, 2020 [(2.6% (2.3–2.9%)] is in reasonable agreement with the figures reported in a preliminary research performed in the Brazilian state of Espírito Santo [42] 2.1% (1.67–2.52%). For UK, we found a point prevalence of 2.5% (2.0–3.1%) for March 28, 2020, which is in close agreement with some previous model estimates [34] [2.7% (1.2–5.4%)].

It is worth mentioning, however, that our point prevalences are strictly connected with the estimated *ifr* values for each country, which, in turn, carry significant uncertainties. In fact, the age-dependent 95% confidence intervals for the *ifr* values reported by Verity *et al.* [35] can be satisfactorily described in terms of gamma distributions with specific coefficients of variation for each age group, as depicted in Table II. Under this approximation, the corresponding coefficients of variation are typically within 30 to 40% for ages above 20 years, demonstrating the huge uncertainties of this parameter. For the specific case of Brazil, a global *ifr* average of 0.85% (weighted by the number of infections) can be derived from the results of 16 Brazilian states reported elsewhere [40]. Additional *ifr* estimates of 0.46% and 0.72% (0.17% and 0.28%) were recently reported [41] for São Paulo (Manaus) city considering confirmed COVID-19 deaths and probable COVID-19 deaths based on syndromic identification, respectively.

The sensitivity of our *ifr* parameters with respect to the assumption of equal prevalences across all age groups was investigated considering the age-dependent effects in transmission of COVID-19 reported by Davies *et al.* [43]. According with this study, people under 20 years old have a susceptibility approximately half that of adults aged over 20 years. So, taking this susceptibility parameter, we found *ifr* values 0.001% to 0.004% higher than our average values obtained under the assumption of equal susceptibility across all age groups, showing that this assumption does not intro-

duce further uncertainties given the current understanding of the COVID-19 dynamics.

Considering the significant uncertainties of the *ifr* values for COVID-19 reported in the literature so far, and the major differences between countries related with healthcare facilities, socioeconomic profiles, and governmental policies to face the pandemic, the achievement of accurate *ifr* values for each country is beyond the scope of this work. Despite of this limitation, strictly related with our estimates for the absolute number of infected people, our model is indeed a powerful tool for the identification of trends and structures in the death data, as they do not depend on the global normalization given by *ifr* but are quite sensitive to changes in the prevailing circumstances, either in the direction of strengthening or relaxing nonpharmacological interventions of social distance, school closure, and sanitary measures. In this regard, the method presented here can be satisfactorily applied for other epidemiological scenarios, and the accuracy of the results will be closely related with the knowledge of the infection-fatality-ratio parameter.

V. CONCLUSIONS

We have developed a Monte Carlo algorithm to model pandemic evolutions in terms of superimposed Gompertz-type functions. The method was successfully applied for the outbreak of COVID-19, providing a consistent interpretation of the time evolution of deaths in the USA, Brazil, Mexico, the UK, India, and Russia. The uncertainties of the best-fit parameters were obtained by propagation, using a least-squares method weighted by the dispersion of the death data in each epidemiological week. The structures of the death data were statistically consistent with more than one single focus of the disease for the case of the USA, Brazil, UK, and India, evidencing the cluster structure of the virus spreading and the powerfulness of the model in predicting trends and secondary peaks, as the world economics and social activities are being resumed. The USA, UK, and India data have a first peak with much higher growth rates, when compared with the second one, showing the positive effects of nonpharmacological interventions of social distance and sanitary measures flattening the second propagation of the virus. Our results for the point prevalences strictly depend on the *ifr* parameters for each country, which have been calculated assuming an equal prevalence across all age groups. Such limitation directly influences the accuracy of the absolute number of infected people, which can be significantly improved with more accurate inputs from further serological studies worldwide.

ACKNOWLEDGMENT

The authors appreciate the fruitful comments of the anonymous referees, which helped with the improvement of our work. T.R. contributed to the conceptualization of the work, data survey, development of the Monte Carlo algorithm, calculations, data analyses, and interpretation of the results. O.H. contributed to the conceptualization of the work, data survey, calculations, advices on the statistical methods, data analyses, and interpretation of the results.

- [1] F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.-W. Tao, J.-H. Tian, Y.-Y. Pei *et al.*, A new coronavirus associated with human respiratory disease in China, *Nature* **579**, 265 (2020).
- [2] J. Bedford, J. Farrar, C. Ihekweazu, G. Kang, M. Koopmans, and J. Nkengasong, A new twenty-first century science for effective epidemic response, *Nature* **575**, 130 (2019).
- [3] T. Lancet, Emerging understandings of 2019-ncov, *Lancet (Lond.)* **395**, 311 (2020).
- [4] J. Phua, L. Weng, L. Ling, M. Egi, C.-M. Lim, J. V. Divatia, B. R. Shrestha, Y. M. Arabi, J. Ng, C. D. Gomersall *et al.*, Intensive care management of coronavirus disease 2019 (COVID-19): Challenges and recommendations, *Lancet Respiratory Medicine* **8**, 506 (2020).
- [5] P. Christen, J. D'Aeth, A. Lochen, R. McCabe, D. Rizmie, N. Schmit, A. Nayagam, M. Miraldo, P. White, P. Aylin *et al.*, Report 15: Strengthening hospital capacity for the COVID-19 pandemic, Technical Report (2020).
- [6] N. M. Ferguson, D. Laydon, G. Nedjati-Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunubá, G. Cuomo-Dannenburg *et al.*, Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand (unpublished).
- [7] M. U. Kraemer, C.-H. Yang, B. Gutierrez, C.-H. Wu, B. Klein, D. M. Pigott, L. Du Plessis, N. R. Faria, R. Li, W. P. Hanage *et al.*, The effect of human mobility and control measures on the COVID-19 epidemic in China, *Science* **368**, 493 (2020).
- [8] C. J. Wang, C. Y. Ng, and R. H. Brook, Response to COVID-19 in Taiwan: Big data analytics, new technology, and proactive testing, *J. Am. Med. Assoc.* **323**, 1341 (2020).
- [9] A. J. Kucharski, T. W. Russell, C. Diamond, Y. Liu, J. Edmunds, S. Funk, R. M. Eggo, F. Sun, M. Jit, J. D. Munday *et al.*, Early dynamics of transmission and control of COVID-19: A mathematical modelling study, *Lancet Infectious Disease* **20**, 553 (2020).
- [10] C. E. Overton, H. B. Stage, S. Ahmad, J. Curran-Sebastian, P. Dark, R. Das, E. Fearon, T. Felton, M. Fyles, N. Gent *et al.*, Using statistics and mathematical modeling to understand infectious disease outbreaks: COVID-19 as an example, *Infectious Disease Modelling* **5**, 409 (2020).
- [11] G. Giordano, F. Blanchini, R. Bruno, P. Colaneri, A. Di Filippo, A. Di Matteo, and M. Colaneri, Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy, *Nat. Med.* **26**, 855 (2020).
- [12] K. Prem, Y. Liu, T. W. Russell, A. J. Kucharski, R. M. Eggo, N. Davies, S. Flasche, S. Clifford, C. A. Pearson, J. D. Munday *et al.*, The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: A modelling study, *Lancet Public Health* **5**, e261 (2020).
- [13] G. Gaeta, A simple SIR model with a large set of asymptomatic infectives, *Mathematics in Engineering* **3**(2), 1 (2021).
- [14] L. F. Scabini, L. C. Ribas, M. B. Neiva, A. G. Junior, A. J. Farfán, and O. M. Bruno, Social interaction layers in complex networks for the dynamical epidemic modeling of COVID-19 in Brazil, *arXiv:2005.08125* (2020).
- [15] K. Leung, J. T. Wu, D. Liu, and G. M. Leung, First-wave COVID-19 transmissibility and severity in China outside Hubei after control measures, and second-wave scenario planning: a modelling impact assessment, *Lancet* **395**, 1382 (2020).
- [16] D. S. Candido, I. M. Claro, J. G. de Jesus, W. M. Souza, F. R. Moreira, S. Dellicour, T. A. Mellan, L. du Plessis, R. H. Pereira, F. C. Sales *et al.*, Evolution and epidemic spread of sars-cov-2 in Brazil, *Science* **369**, 1255 (2020).
- [17] Y. Fang, Y. Nie, and M. Penny, Transmission dynamics of the COVID-19 outbreak and effectiveness of government interventions: A data-driven analysis, *J. Med. Virol.* **92**, 645 (2020).
- [18] E. Massad, M. Amaku, A. Wilder-Smith, P. C. C. dos Santos, C. J. Struchiner, and F. A. B. Coutinho, Two complementary model-based methods for calculating the risk of international spreading of a novel virus from the outbreak epicentre: The case of COVID-19, *Epidemiol. Infect.* **148**, e109 (2020).
- [19] J. Schüttler, R. Schlickeiser, F. Schlickeiser, and M. Kröger, Covid-19 predictions using a Gauss model, based on data from april 2, *Physics* **2**(2), 197 (2020).
- [20] R. H. Pedrosa, The dynamics of covid-19: Weather, demographics and infection timeline, *medRxiv* 2020.04.21.20074450 (2020).
- [21] A. I. ul Hassan, Understanding the COVID-19 pandemic curve through statistical approach [<https://www.medrxiv.org/content/medrxiv/early/2020/04/08/2020.04.06.20055426>].
- [22] X. Xu, S. Wang, J. Dong, Z. Shen, and S. Xu, An analysis of the domestic resumption of social production and life under the COVID-19 epidemic, *PLoS ONE* **15**, e0236387 (2020).
- [23] B. Paixão, L. Baroni, R. Salles, L. Escobar, C. de Sousa, M. Pedroso, R. Saldanha, R. Coutinho, F. Porto, and E. Ogasawara, Estimation of COVID-19 under-reporting in Brazilian States through sari, *arXiv:2006.12759* (2020).
- [24] S. G. Krantz and A. S. S. Rao, Level of underreporting including underdiagnosis before the first peak of COVID-19 in various countries: Preliminary retrospective results based on wavelets and deterministic modeling, *Infect. Contr. Hosp. Epidemiol.* **41**, 857 (2020).
- [25] K. Jagodnik, F. Ray, F. M. Giorgi, and A. Lachmann, Correcting under-reported COVID-19 case numbers: Estimating the true scale of the pandemic, *medRxiv* 14 (2020).
- [26] J. Cohen and K. Kupferschmidt, Countries test tactics in 'war' against COVID-19, *Science* **367**, 1287 (2020).
- [27] D. M. Weinberger, J. Chen, T. Cohen, F. W. Crawford, F. Mostashari, D. Olson, V. E. Pitzer, N. G. Reich, M. Russi, L. Simonsen, A. Watkins, and C. Viboud, Estimation of Excess Deaths Associated With the COVID-19 Pandemic in the United States, March to May 2020, *JAMA Internal Medicine* **180**, 1336 (2020).
- [28] C. Modi, V. Boehm, S. Ferraro, G. Stein, and U. Seljak, How deadly is covid-19? A rigorous analysis of excess mortality and age-dependent fatality rates in italy, *medRxiv*, 10.1101/2020.04.15.20067074 (2020).
- [29] A. Capaldi, S. Behrend, B. Berman, J. Smith, J. Wright, and A. L. Lloyd, Parameter estimation and uncertainty quantification for an epidemic model, *Math. Biosci. Eng.* **9**, 553 (2012).
- [30] R. H. Gaffey and C. Viboud, Application of the CDC Ebola Response modeling tool to disease predictions, *Epidemics* **22**, 22 (2018).
- [31] T. Kuniya, Prediction of the epidemic peak of coronavirus disease in japan, 2020, *J. Clin. Med.* **9**, 789 (2020).
- [32] A. Smirnova, B. Sirb, and G. Chowell, On stable parameter estimation and forecasting in epidemiology by the Levenberg-Marquardt Algorithm with Broyden's Rank-one updates for the Jacobian Operator, *Bull. Math. Biol.* **81**, 4210 (2019).

- [33] J. Peto, Covid-19 mass testing facilities could end the epidemic rapidly, *Bmj* **368**, m1163 (2020).
- [34] S. Flaxman, S. Mishra, A. Gandy, H. Unwin, H. Coupland, T. Mellan, H. Zhu, T. Berah, J. Eaton, P. Perez Guzman *et al.*, Report 13: Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries, Technical Report (2020).
- [35] R. Verity, L. C. Okell, I. Dorigatti, P. Winskill, C. Whittaker, N. Imai, G. Cuomo-Dannenburg, H. Thompson, P. G. Walker, H. Fu *et al.*, Estimates of the severity of coronavirus disease 2019: A model-based analysis (unpublished).
- [36] V. Vanin and O. Helene, Covariance analysis by means of the least squares method, in *Update of X Ray and Gamma Ray Decay Data Standards for Detector Calibration and Other Applications. V. 2: Data Selection, Assessment and Evaluation Procedures* (IAEA, Vienna, 2007).
- [37] O. Helene, L. Mariano, and Z. Guimaraes-Filho, Useful and little-known applications of the least square method and some consequences of covariances, *Nucl. Instrum. Methods A* **833**, 82 (2016).
- [38] E. Bendavid, B. Mulaney, N. Sood, S. Shah, E. Ling, R. Bromley-Dulfano, C. Lai, Z. Weissberg, R. Saavedra, J. Tedrow *et al.*, COVID-19 antibody seroprevalence in Santa Clara County, California, *MedRxiv* (2020).
- [39] P. C. Hallal, F. P. Hartwig, B. L. Horta, M. F. Silveira, C. J. Struchiner, L. P. Vidaletti, N. A. Neumann, L. C. Pellanda, O. A. Dellagostin, M. N. Burattini *et al.*, Sars-cov-2 antibody serological prevalence in brazil: Results from two successive nationwide serological household surveys, *Lancet Global Health* **8**, e1390 (2020).
- [40] T. A. Mellan, H. H. Hoeltgebaum, S. Mishra, C. Whittaker, R. P. Schnekenberg, A. Gandy, H. J. T. Unwin, M. A. Vollmer, H. Coupland, I. Hawryluk *et al.*, Report 21: Estimating covid-19 cases and reproduction number in Brazil, *medRxiv* (2020).
- [41] L. F. Buss, C. A. Prete, C. M. Abraham, A. Mendrone, T. Salomon, C. de Almeida-Neto, R. F. França, M. C. Belotti, M. P. Carvalho, A. G. Costa *et al.*, Covid-19 herd immunity in the brazilian amazon, *medRxiv* (2020).
- [42] C. C. Gomes, C. Cerutti, E. Zandonade, E. L. N. Maciel, F. E. C. de Alencar, G. L. Almada, O. A. Cardoso, P. M. Jabor, R. L. Zanotti, T. Q. Reuter *et al.*, A population-based study of the prevalence of COVID-19 infection in Espirito Santo, Brazil: Methodology and results of the first stage, *medRxiv* (2020).
- [43] N. Davies, P. Klepac, Y. Liu, K. Prem, M. Jit, and R. Eggo, Age-dependent effects in the transmission and control of COVID-19 epidemics, *Nat. Med.* **26**, 1205 (2020).