medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078154; this version posted May 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Predicting COVID-19 peaks around the world

Constantino Tsallis[1,2,3*] and Ugur Tirnakli[4†]

[1] *Centro Brasileiro de Pesquisas Fisicas and National Institute of Science and Technology for Complex Systems Rua Xavier Sigaud 150, Rio de Janeiro 22290-180, Brazil*

[2] *Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501, USA*

[3] *Complexity Science Hub Vienna, Josefstädter Strasse 39, 1080 Vienna, Austria*

[4] *Department of Physics, Faculty of Science, Ege University, 35100 Izmir, Turkey*

(Dated: April 30, 2020)

The official data for the time evolution of active cases of COVID-19 pandemics around the world are available online. For all countries, a peak has been either observed (China and South Korea) or is expected in near future. The approximate dates and heights of those peaks imply in important epidemiological issues. Inspired by similar complex behaviour of volumes of transactions of stocks at NYSE and NASDAQ, we propose a q-statistical functional form which appears to describe satisfactorily the available data of all countries. Consistently, predictions become possible of the dates and heights of those peaks in severely affected countries unless efficient treatments or vaccines, or sensible modifications of the adopted epidemiological strategies, emerge.

It is possible to predict the thermostatistical properties of uncountable physical systems at thermal equilibrium through the one-body distribution $p(\epsilon) \propto \omega(\epsilon)e^{-\beta\epsilon}$, where $\omega(\epsilon)$ is the density of states as a function of the energy ϵ, multiplied by the celebrated Boltzmann factor $e^{-\beta\epsilon}$, β being the inverse temperature. The function $\omega(\epsilon)$ comes from mechanical considerations (classical, quantum, relativistic) related with the number of degrees of freedom, and does not depend on the temperature; the exponential weight comes instead from standard statistical-mechanical considerations. In many cases it is, either exactly or approximatively, $w(\epsilon) \propto \epsilon^\alpha$ $(\alpha \in R)$. For thermostatistical properties of the stationary- or quasi-stationary-state of wide classes of complex systems, the Boltzmann factor is to be generalised into the q-exponential factor $e_q^{-\beta_q\,\epsilon} \equiv [1-(1-q)\beta_q\,\epsilon]^{1/(1-q)}$ $(q \in R;\, e_1^z = e^z)$ [1–3]. This procedure yielded quite satisfactory results for the high-frequency stock-market in NYSE, NASDAQ and elsewhere [4–6].

Let us focus now on the data available for the COVID-19 pandemics. Briefly after the beginning of the pandemics, several studies analysing the available data, and employing different models and candidate functions, started to appear in the literature [7–12]. Most of them are interested in the behaviour of total cases and fatality curves. We will concentrate here on the analysis of the active cases and deaths *per day*. The inspection of the public data [13] (updated on a daily basis), in particular of the time evolution of the number N of active cases (surely a lower bound of the unknown actual numbers) showed a rather intriguing similarity with the distributions of volumes of stocks. Along this line we adopt the following functional form for each country or region:

$$N = C(t-t_0)^\alpha e_q^{-\beta\,(t-t_0)^\gamma} = \frac{C(t-t_0)^\alpha}{[1+(q-1)\beta\,(t-t_0)^\gamma]^{1/(q-1)}}\,, \tag{1}$$

with $C>0; \alpha>0;\, \beta>0; \gamma>0,\, q>1$ and $t_0 \geq 0$. The constant t_0 indicates the first day of appearance of the epidemics in that particular region; it is conventionally chosen to be zero for China; for the other countries it is the number of days elapsed between the appearance of the first case in China and the first case in that country. The normalising constant C reflects the total population of that particular country. For $\alpha = 0$, if $\gamma = 1$ we recover the standard q-exponential expression; if $\gamma = 2$, it is currently referred in the literature as q-Gaussian; for other values of γ, it is referred to as stretched q-exponential. Through the inspection of the roles played by the four nontrivial parameters, namely $(\alpha, \beta, \gamma, q)$, it became rather transparent that (α, β) depend strongly on the epidemiological strategy implemented in that region, in addition to the biological behaviour of the coronavirus in that geographical climate. In contrast, the parameters (γ, q) appear to be sort of universal, mainly depending on the coronavirus. Therefore, we investigate several countries, which do not reach their peaks yet, with the basic assumption that these two parameter would not change much from one country to another and we fixed these values at the values that we determine for China since this country has already had nearly the full evolution of the pandemics. This assumption seems to be working. For other contries whose peaks are already reached, we use the same functional form (1) but adjusting all parameters for a better fitting. The results of China and South Korea are given in Fig. 1. It is evident that, although the functional form (1) does yield satisfactory results for both China and South Korea, the (γ, q)

*Electronic address: tsallis@cbpf.br

†Electronic address: ugur.tirnakli@ege.edu.tr

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078154; this version posted May 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

parameter values differ somewhat for each of these countries. On the other hand, as can be easily seen from Fig. 2, our assumption is corroborated by several countries that we have numerically analysed. In Table 1 we present the forecasted dates and heights of the peaks, as well as the values of the fitting parameters using the data accumulated until 28 of April, 2020.

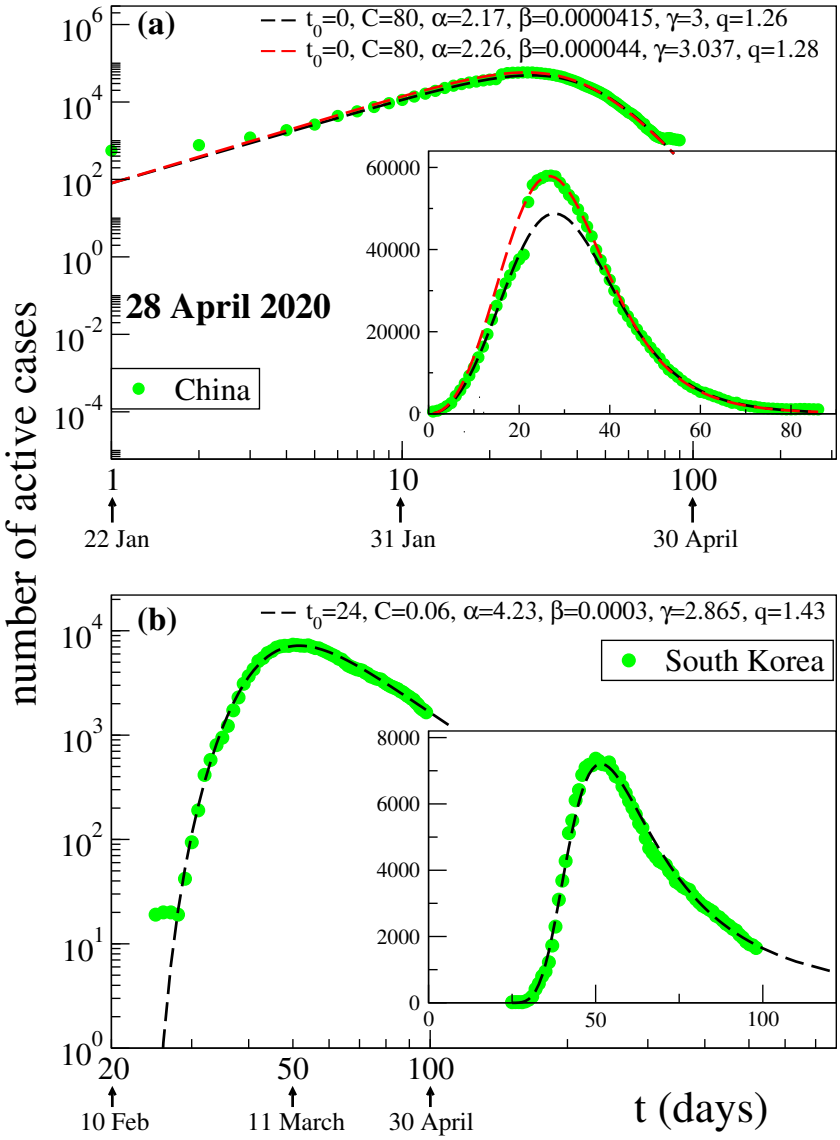

FIG. 1: Two possible fits for the COVID-19 active cases evolution in (a) China and (b) South Korea: the dots are the available data at 28 April 2020, and the lines are fittings using Eq. (1). For China a strange kink is present in the ascending part of the data curves, which makes that it is not possible to make a single fitting which would satisfactorily account for both increasing and decreasing parts. We present here two log-log representation of different curves which describe either the increasing part or the decreasing part but not both. *Inset:* linear-linear representation of the same data.

.

We also test our formula (1) and assumption for the evolution of deaths per day. Again we fixed the (γ, q) parameter values at China's results, which can be seen in Fig. 3, and tried to fit the data of the same 6 countries. It is quite surprising to see that the results given in Fig. 4 seem to suggest that the evolution of death per day is also possible to be fitted without changing these two parameters.

We may summarise as follows. The death curve of South Korea is atypical, in the sense that it sensibly differs from all those that we analysed in the present work. Because of that, we have not included it here. We remind that the values for (γ, q) used for the South Korea evolution curve of active cases also differ from those used for all the other countries, which reinforces the fact that some sort of exceptionality exists there, for reasons that are unknown to us. For all countries that do not reach their peak values yet, we have adopted the values of (γ, q) obtained from

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078154; this version posted May 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

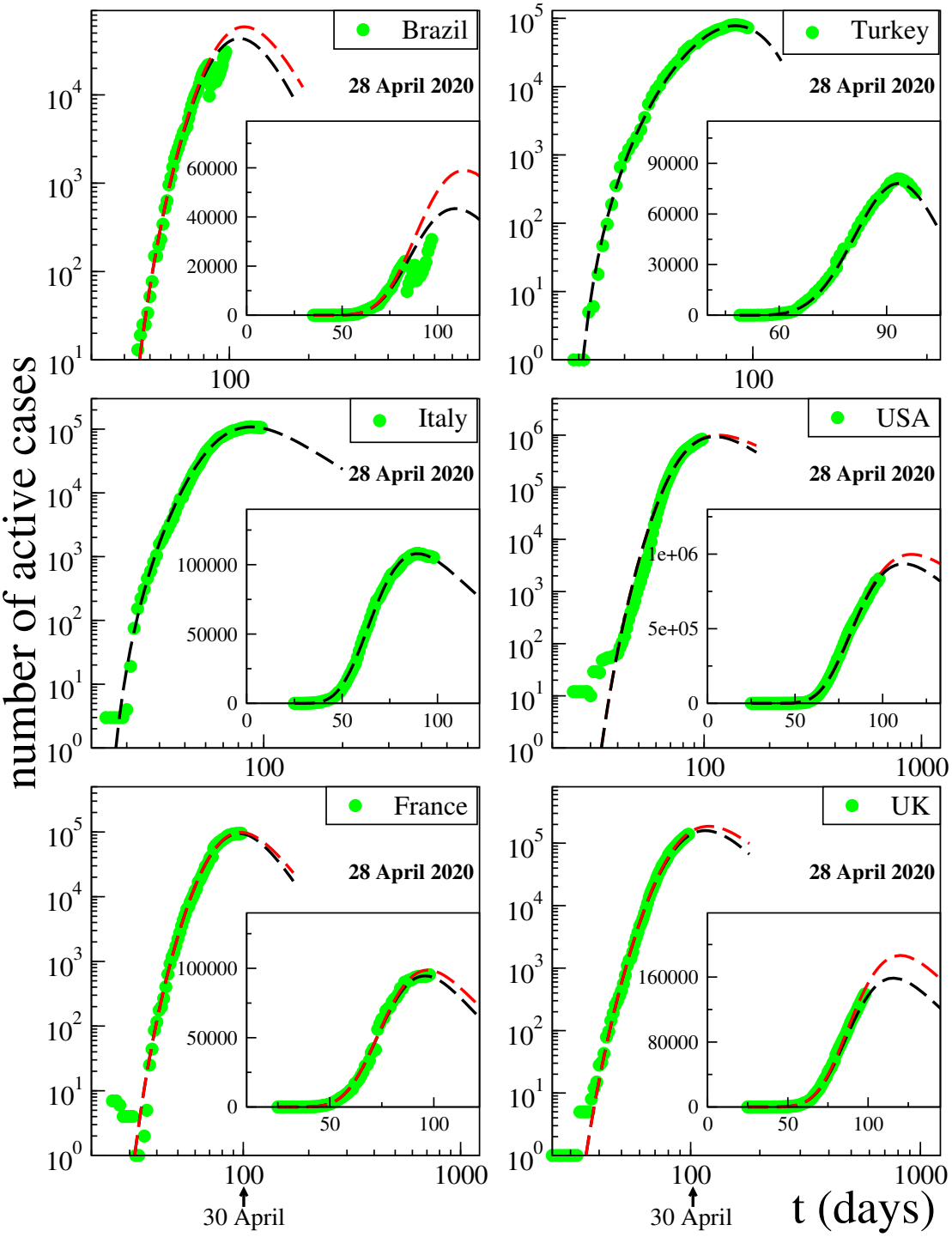

FIG. 2: Fittings of the active cases data available in 28 April 2020 for various severely affected countries around the world with Eq. (1). The fitting parameters γ and q are fixed at the values of China for all countries which do not reach their peak values yet. Notice that, in the case of Brazil, a disruption occurred in the publicly available data in middle April. We do not know the cause of this. Coincidentally, however, precisely in those days the President of Brazil decided to change his Minister for Health.

.

the inspection of the entire curve of China.

We verify that the present work appears to belong to the realm of complex systems which includes not only, as mentioned above, high-frequency financial transactions (with $\alpha > 0$ [4, 5], but also anthropological issues such as

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078154; this version posted May 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

γ and q parameters are fixed at China's values if peak is not achieved yet								
Country	t_0	C	α	β	γ	q	Prediction interval of peak height	Prediction interval of peak day
Brazil	34	1.55×10^{-5}	5.65	8.00×10^{-6}	3	1.26	$[43400, 58965]$	$[109, 114]$
				8.00×10^{-6}	3.037	1.28	[0.00021,0.00028]	[09 May, 14 May]
Turkey	48	8.0×10^{-3}	4.625	1.52×10^{-5}	3.025	1.02	achieved at 80808	achieved on day 93 23 April
Italy	24	4.0×10^{-3}	4.68	1.10×10^{-5}	3.083	1.43	achieved at 106100	achieved on day 96 26 April
USA	24	3.4×10^{-9}	8.60	1.65×10^{-5}	3	1.26	$[934680, 997830]$	$[112, 117]$
				1.46×10^{-5}	3.037	1.28	[0.0028,0.0030]	[12 May, 17 May]
France	24	1.6×10^{-6}	6.56	9.20×10^{-6}	3	1.26	$[94551, 98726]$	$[96, 97]$
				1.19×10^{-5}	3.037	1.28	[0.0014,0.0015]	[26 April, 27 April]
UK	24	1.2×10^{-8}	7.60	9.60×10^{-6}	3	1.26	$[158700, 186610]$	$[116, 120]$
				8.00×10^{-6}	3.037	1.28	[0.0024,0.0028]	[16 May, 20 May]

TABLE I: The values of the parameters and predictions for the maximum number of active cases and for the day at which the maximum will be achieved. The second line in the column of prediction interval of active cases peak height shows the same data divided by the total population of the country.

medieval trading networks and biotech intercorporate networks (with $\alpha < 0$) [14], relaxation in spin-glasses [15], as well as q-Weibull distribution-like systems [16–19], which correspond to the particular case $\alpha = \gamma - 1$. It remains open as a highly desirable goal to formulate a model which, along lines somewhat similar to epidemiological models such as the SIR one, would predict an evolution curve such as the present Eq. (1). This would clarify important issues such as the conditions under which the values of (γ, q) could indeed be universal and essentially determined by the biology of the infecting agent, such as the present COVID-19.

Acknowledgments

We have benefitted from useful remarks by E.P. Borges and partial support by CNPq and Faperj (Brazilian Agencies).

[1] C. Tsallis, *Possible generalization of Boltzmann-Gibbs statistics*, J. Stat. Phys. **52**, 479 (1988).
[2] M. Gell-Mann and C. Tsallis, eds., *Nonextensive Entropy - Interdisciplinary Applications*, (Oxford University Press, New York, 2004).
[3] C. Tsallis, *Introduction to Nonextensive Statistical Mechanics - Approaching a Complex World* (Springer, New York, 2009).
[4] C. Tsallis, C. Anteneodo, L. Borland and R. Osorio, *Nonextensive statistical mechanics and economics*, Physica A **324**, 89 (2003).
[5] R. Osorio, L. Borland and C. Tsallis, *Distributions of high-frequency stock-market observables*, in [2].
[6] G. Ruiz and A.F. de Marcos, *Evidence for criticality in financial data*, Eur. Phys. J. B **91**, 1 (2018).
[7] V. Zlatic *et al.*, *Bi-stability of SUDR+K model of epidemics and test kits applied to COVID-19*, arXiv:2003.08479, (2020). year = "2020",
[8] G. L. Vasconcelos *et al.*, *Modelling fatality curves of COVID-19 and the effectiveness of intervention strategies*, doi = https://doi.org/10.1101/2020.04.02.20051557, (2020), preprint.
[9] E. L. Piccolomini and F. Zama, *Preliminary analysis of COVID-19 spread in Italy with an adaptive SEIRD model*, arXiv:2003.09909, (2020).
[10] M. Perc *et al.*, *Forecasting COVID-19*, Front. Phys. **8**, 127 (2020).
[11] A. Pluchino *et al.*, *A Novel Methodology for Epidemic Risk Assessment: the case of COVID-19 outbreak in Italy*, arXiv:2004.02739, (2020).
[12] S. B. Bastos and D. O. Cajueiro, *Modeling and forecasting the early evolution of the Covid-19 pandemic in Brazil*, arXiv:2003.14288v2, (2020).
[13] `https://www.worldometers.info/coronavirus/#countries`

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078154; this version posted May 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

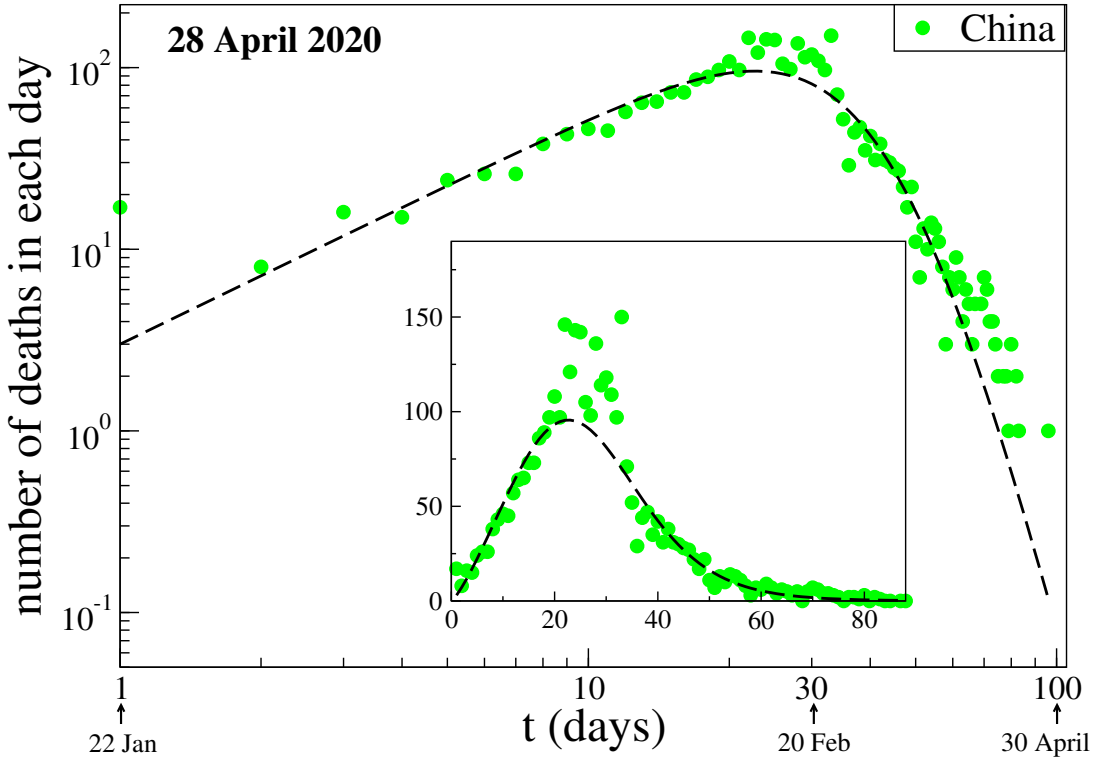

FIG. 3: Evolution of deaths per day due to COVID-19 in China. The data are those available at 28 April 2020. The fitting parameters γ and q are fixed at one of the two choices of the values for China, namely $(\gamma, q) = (3, 1.26)$. The other fitting parameters, namely $(C_{death}, \alpha_{death}, \beta_{death})$, have been chosen to better fit the available data for China.

[14] D.R. White, N. Kejzar, C. Tsallis, D. Farmer and S. White, *A generative model for feedback networks*, Phys. Rev. E **73**, 016119 (2006).

[15] R.M. Pickup, R. Cywinski, C. Pappas, B. Farago and P. Fouquet, *Generalized spin glass relaxation*, Phys. Rev. Lett. **102**, 097202 (2009).

[16] S. Picoli Jr., R.S. Mendes and L.C. Malacarne, *q-exponential, Weibull, and q-Weibull distributions: An empirical analysis*, Physica A **324**, 678 (2003).

[17] E.M. Assis, E.P. Borges and S.A.B. Vieira de Melo, *Generalized q-Weibull model and the bathtub curve*, International Journal of Quality & Reliability Management **30**, 720 2013.

[18] T.S.V. Podesta, T.V. Rosembach, A.A. Santos and M.L. Martins, *Anomalous diffusion and q-Weibull velocity distributions in epithelial cell migration*, Plos One (2017).

[19] M. Xu, *Using the q-Weibull distribution for reliability engineering modeling and applications*, Doctor Thesis (University of Maryland, College Park, 2019).

medRxiv preprint doi: https://doi.org/10.1101/2020.04.24.20078154; this version posted May 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

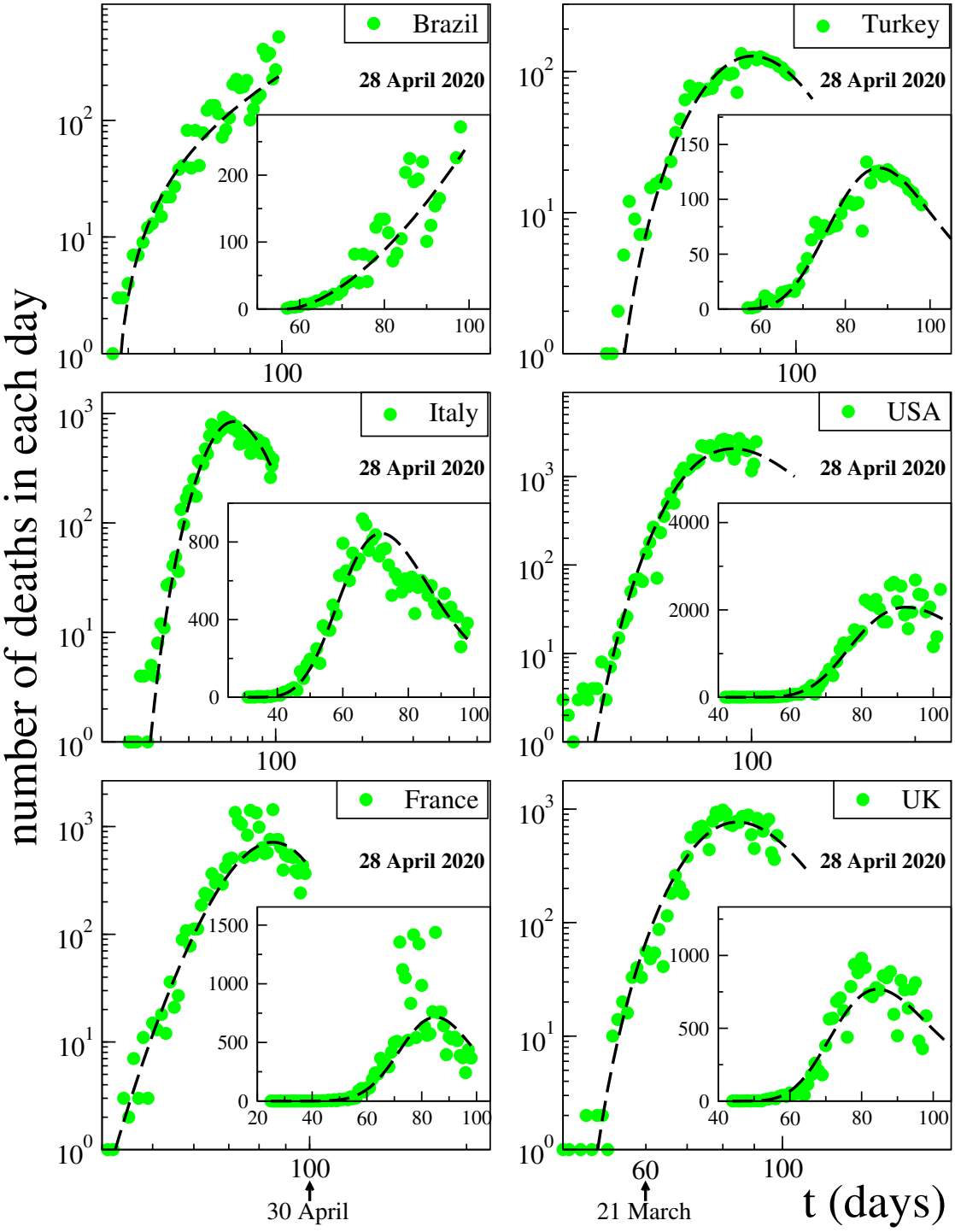

FIG. 4: Fittings of deaths per day data for various strongly affected countries around the world with Eq. (1). The data are those available at 28 April 2020. The fitting parameters γ and q are fixed at one of the two choices of the values for China, namely $(\gamma, q) = (3, 1.26)$. The other fitting parameters, namely $(C_{death}, \alpha_{death}, \beta_{death})$, have been chosen to better fit the available data for each country.

.