

Two-scale Phenomenological and Statistical Analysis of the COVID-19 Pandemic

Vassilis Pontikis^{1,*}, Theodoros Karakostas², and Philomela Komninou²

¹Commissariat à l'Energie Atomique et aux Energies Alternatives, Université Paris-Saclay, Institut Rayonnement-Matière de Saclay (IRAMIS) & Département des Matériaux Nucléaires, 91191 Gif-sur-Yvette, Cedex, France

²Aristotle University of Thessaloniki, Department of Physics, Thessaloniki, GR-54124, Greece

*vassilis.pontikis@cea.fr

+these authors contributed equally to this work

ABSTRACT

In this work, the daily reported new infection cases and deaths globally due to COVID-19 are approximated by Gaussian functions of the time elapsed since the beginning of the pandemic. This approximation is shown to consistently designate the time evolution of daily rates of new cases and the total duration of the epidemic in any given region or country. At the pandemic scale, this phenomenological analysis is applied for condensing statistical data of the epidemic evolution in twenty-four countries into two “universal master plots” involving in total five adjustable parameters per country. The analysis reveals intrinsic features of the pandemic and local specificities of the epidemic as well, helps in anticipating the epidemic peak date in each country and yields confidence intervals for the average numbers of incidences. Finally, the phenomenology facilitates comparisons of the epidemic evolution in different countries and evaluations of the effectiveness of related health policies.

Introduction

The mathematical and statistical analyses of epidemic data are considered as substantial tools for national health organizations to develop prevention policies limiting the propagation of an epidemic while sustaining health systems and lowering human and economic losses. Known models, statistical, mathematical, mechanistic space-state and empirical (machine learning) have been thoroughly reviewed recently¹. Since the emergence of COVID-19, several newly published research articles rely on mathematical models to predict the spread of the virus, to estimate the basic reproduction factor, R_0 and, to highlight the influence of public policies on the outcome of the epidemic²⁻¹¹. Statistical analyses of the epidemic data from different countries, by means of distribution and error functions, have also been performed with the aim of predicting the epidemic peak^{12,13} some of which make use of machine learning techniques^{14,15}. Although statistical information from official health organizations and various other information channels¹⁶⁻¹⁸ becomes increasingly available and trusty, modeling seems difficult because collected data are affected by large fluctuations, especially in the neighborhood of the epidemic peak. In addition, comparing the respective efficiency of the public health policies adopted in different countries/regions is delicate since the cardinal numbers of affected population sets and the distributions of primary seeds are very different. Nevertheless, examining the statistical databases shows qualitatively that the total number of infections/deaths as a function of the elapsed time since the first case occurrence shapes sigmoidal whereas its derivative, i.e. the number of new cases per day as a function of the elapsed time, τ , looks reasonably Gaussian. It should be thereby assumed that the virus evolves similarly in different countries. Common to the regions experiencing the virus spread, are also the adopted countermeasures, lockdown policies, travel restrictions and quarantine, that justify viewing countries as closed, non-interacting systems.

Alike others, the present work relies on statistical datasets from twenty four countries¹⁶⁻¹⁸, and postulates that in all of them the number of daily new cases is a Gaussian function of the elapsed time. The time evolution of the corresponding total number of cases, $n(\tau)$, is then obtained by integration yielding an error function of, τ , which sigmoidal shape consistently describes the data. It is shown in addition that $n(\tau)$ and its time derivatives, easily transform in to non-dimensional quantities in terms of which time series of statistical data from any country superimpose with very little dispersion on master plots. The latter represent the virus evolution in any country. Comparisons of the epidemic evolution in different countries/regions is thereby simplified, consisting therefrom in comparing the respective values of the parameters spanning this homothetic transformation. It is found that by reaching the epidemic peak the parameters yielding the master plots have converged toward stationary values and estimations of the epidemic duration become faithful.

In the following, the model and thereby obtained results are first presented. Therefrom, conclusions can be drawn about the

epidemic evolution and the limits of the adopted analysis are revealed. The methodology and procedures used in this work, are detailed in the methods and computations section aiming at adequately address issues important for others in the field to reproduce the results¹⁹. The discussion section, lists the main conclusions of the present work, comments on limitations and envisages directions for future work for further improving the methodology and agreement with the statistical data.

Model

As it has been already proposed in the literature^{13,20}, the number of daily new cases, infections or deaths, are supposed evolving as Gaussian functions of the time elapsed since the occurrence of the first case, τ , expressed in number of days:

$$\frac{dn(\tau)}{d\tau} = A_0 \exp\left(-\frac{(\tau - \tau_m)^2}{2\sigma^2}\right) \text{ with } A_0 = \frac{N_0}{\sqrt{2\pi}\sigma} \quad (1)$$

where, τ_m , locates in time the epidemic peak, σ , is the standard deviation and A_0 , is a normalization constant proportional to the total number of cases, N_0 that should be reached upon vanishing the epidemic at elapsed time, τ_{end} . Integration of eq.1 over the time interval $[0, \tau]$ yields the number of predicted cases as a function of time, τ :

$$n(\tau) = \frac{N_0}{2} \left(1 + \operatorname{erf}\left(\frac{\tau - \tau_m}{\sqrt{2}\sigma}\right)\right) \quad (2)$$

Introducing the reduced variable, $\tau^* = \frac{\tau - \tau_m}{\sqrt{2}\sigma}$, eqs.(1–2) read as follows:

$$\frac{dn^*(\tau)}{d\tau^*} = \frac{\sqrt{2\pi}\sigma}{N_0} \frac{dn(\tau)}{d\tau} = \exp(-\tau^{*2}) \quad (3)$$

$$n^*(\tau^*) = \frac{2n(\tau)}{N_0} - 1 = \operatorname{erf}(\tau^*) \quad (4)$$

Parameters, $(N_0^k, \tau_m^k, \sigma^k)_i$, with indices, k and i (i =inf/dec) labelling respectively the k^{th} ($k=1-24$) country-dataset and its kind, infections or deaths, can be determined via least-squares fits of eq.(2) to the data (see Methods and Computations below). The relationship between populations of infected and deceased individuals is accounted for in the present analysis by choosing identical standard deviations for both of them and reducing thereby the number of adjustable parameters to five. Data from different origins (countries or regions) transformed via eqs.(3,4) and plotted all together, yield non-dimensional master plots condensing all the available information on the graphs of the righthand-side functions in these equations. Unlike processing of individual datasets at the country epidemic scale, their combination yielding the master plots targets the description of the virus propagation at the upper scale of the pandemic.

Remarkable and common to all the datasets is the presence of large fluctuations that seem amplifying upon approaching the epidemic peak^{16,17}. These do certainly reflect counting errors affecting the data but may also originate from causes intrinsic to the virus propagation. However, there is little hope to extract useful information from the data scatter present in country-specific datasets given the low values of the corresponding cardinal numbers that equals the number of days of the lasting epidemic. For the master plots however, the situation is different despite the moderate, yet persisting, data dispersion. Indeed, the related dataset size is now proportional to the number of the contributing countries/regions. Considering the master plots as representative of the pandemic behavior on average, confidence intervals can be determined at any desired value of τ^* . This is equivalent to considering the daily new cases observed in each country as resulting from the realization of the same stochastic process, with, $\tilde{n} = \frac{dn^*(\tau)}{d\tau^*}$, the associated random variable. This view of the pandemic emerging from the lower scale as the superimposition of the parallel and shifted in time epidemic evolutions in different countries is reasonable, provided active regions in each country have not reached the 2-dimensional percolation threshold and that countries correspond to the image invoked in the introduction of closed and non-interacting systems. Noteworthy, country-specific confidence intervals are easy to obtain via eqs. (3,4) and the appropriate to each country/region parameter set, $(N_0^k, \tau_m^k, \sigma^k)_i$. Of great practical importance, the main duration, $\delta\tau_{duration}^* = 2\tau_{end}^*$, of the pandemic can be defined as the time delay separating time instances at $\alpha\%$ of the maximum of the master plot gaussian (eq.4), with the α value conveniently defined. In the following, the value $\alpha=1$ has been arbitrarily chosen. Thereby, an effective elapsed time is obtained marking the end of the epidemic in any given country. In real time units this reads for both, infections and deaths, $\delta\tau_{duration} = \tau_{end} - \tau_m = 2\sqrt{2}\sigma\tau_{end}^*$ ($\tau_{end}^* \approx 2.628$).

Results

Model validation

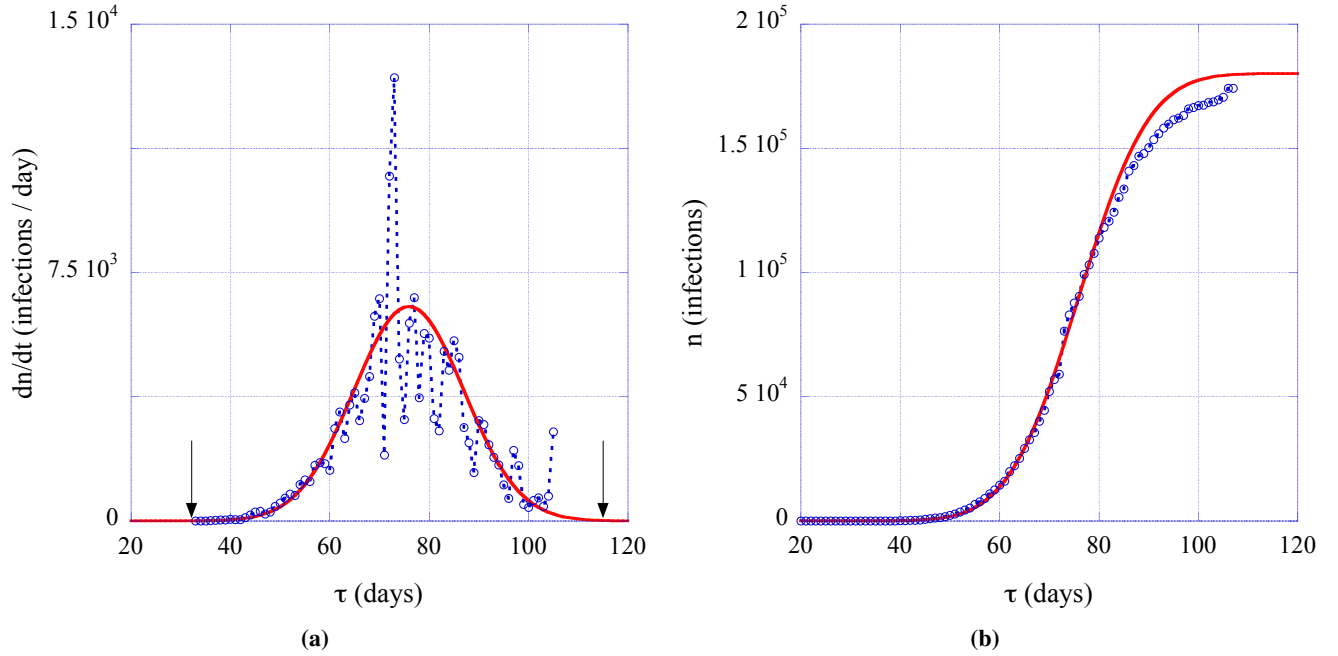


Figure 1. Daily new infections (a) and cumulated cases (b) in France as functions of the time elapsed since January 22, 2020 expressed in days. Full lines are fits of Eqs. (1, 2) to the data with parameters listed in Table 1. Dotted lines are guides for the eye aiming at highlighting the data scatter. The arrows mark the instances of time corresponding to the 1‰ levels of the predicted maximum values of daily infections/deaths (full lines, see section Model).

This phenomenological analysis has been applied here to reported datasets from France^{16–18}, taken as a working example of a country that has passed the epidemic peak at present time (Fig.1, a and b). In Fig. 1a, arrows mark time instances with daily rates at 1‰ of the maxima of the theoretical graphs, thus defining an effective relaxation time, characteristic of the epidemic duration in France, $\delta\tau_{duration}^{inf} = \delta\tau_{duration}^{dec} = 83$ days (Table 1). In this table, columns $Shift^{inf}$ and $Shift^{dec}$ represent the time delay since January 2020 till the occurrence of the first case (infection/death). Accordingly, the position of the epidemic peak in this figure (infections) is obtained by adding to, $\tau_m^{inf} = 73.9$, the corresponding, $Shift^{inf} = 3$ days. The relaxation time is a kind of "mean field" parameter, emerging from the present analysis, which may help refining comparisons of the epidemic evolution in various countries. With no exceptions, all the country datasets used in the present work are fitted with equal quality by the Gaussian and error function forms, demonstrating as a first result of the present work, that these consistently designate the evolution of the epidemic.

Fig.1a shows that with increasing the elapsed time the scatter of the data around the theoretical graph increases. As it has already been mentioned above, this is a feature common to all the datasets forming the statistical database used in the present work. No indications are available about the causes of these fluctuations though it is reasonable to admit that counting errors are superimposed to other causes, possibly intrinsic to the epidemic. Interestingly, the graph representing the time evolution of cumulated numbers of infections is much less affected by fluctuations, which should be considered as the usual smoothing of integrating over time the daily rates of new infections and a strong indication that counting errors should predominantly cause the observed data dispersion (Fig. 1b). In the following, the usual prescription of statistical data analyses is used, considering data dispersion as the consequence of a Gaussian noise, which yields confidence intervals around the mean values of the numbers of infections or deaths²¹.

Master plots

Having determined $(N_0^k, \tau_m^k, \sigma^k)_i$ separately for each country, $k=1-24$, whereas parameter values are given in Table 1, raw datasets have been converted in reduced coordinates via eqs. (3,4) and are displayed all together in Figs.2 (a-d). It can be seen that cumulated numbers of events condense with little dispersion on the master graphs excepted for two of them departing significantly from the average behavior (Fig. 2b). More specifically, the infection/death datasets of China do not

relate consistently as, unlike the first, the last departs markedly from the master plot prescription. Though less marked, a similar trend is observed for the South Korean datasets. Obviously the expected correlation that should exist between infections and deaths does not hold in these cases. Reasons behind such discrepancies are not clear and deserve further investigation, well beyond the scope of the present work, in which pragmatism has led eliminating these two singular datasets from the subsequent statistical analysis. This exemplifies the discriminating power of the master plots analysis and constitutes a main, significant result of the present work: master plots are essential in checking consistency of the statistical datasets of the virus spreading.

Error bars shown in Figs. (3) represent confidence intervals at a 99% confidence level, computed under the usual assumption that data dispersion stems from a random process with Gaussian distribution of probability. It can be seen that error bars represent faithfully the scatter of data displayed therein. More precisely, the reduced time interval $[-4,4]$ has been divided in contiguous elements of equal length with values, ≈ 0.18 and ≈ 0.14 , for infections and deaths respectively. To the data contained in these intervals is then assigned a time instance corresponding to the interval center whereas, $\langle n^* \rangle$ and $\langle \frac{dn^*}{d\tau^*} \rangle$, are set equal to the mean values of the data contained in the given time interval (full circles in these figures). Because the time intervals do contain a modest number of elements (≈ 36), error bars have been determined by resorting to Student's t-distribution²², replacing appropriately the Gaussian distribution of probability when dealing with finite datasets. It can be also seen from these figures that the theoretical graphs of cumulated numbers and daily rates of events (full lines) fit satisfactorily the data points, although the former systematically underestimate daily rates above, $\tau^* > 0.8$ (Figs. (3a–3b)). This trend highlights the fact that the Gaussian is a useful yet non perfect approximation of daily rates and points to the need of further investigation targeting its improvement. Finally, it is worth noting that error bars for cumulated events are quite small, as is the related noise and, would not be visible without the amplification factor (x5) being used in Figs. (3c–3d).

The real time delay corresponding to the time intervals in the above described 'coarse-graining' procedure yielding the error bars, amounts $\approx 2–3$ days. Close examination of Figs. (3) shows also that the coarse-graining smoothens the data sufficiently to eliminate the fluctuations otherwise present in all the datasets of daily rates (e.g. Fig.1). Since this delay is much shorter than any characteristic time of the epidemic, incubation period or total duration, the hypothesis made above, that victims counting modes are principally responsible of the fluctuations present in the daily rates, is favored.

The capability of estimating error bars from the evolution of the pandemic is a direct consequence of the present two-scale phenomenological analysis that yields the master plots and should be considered as the third substantial result of this work.

It is generally accepted that phenomenology serves in classifying data and testing them for internal consistency rather than constituting a predictive tool. This statement fully applies to the present work, which however departs slightly from the general case when considering two particular aspects, the first is related to the symmetry properties of Gaussian functions and the second, empirical, emerges from the present study: beyond the epidemic peak the country parameters converge toward stationary values reflecting the existence of a robust minimum of the least-squares functional (see below, section Methods and Computations).

Stationary parameters, σ^i and τ_m^i ($i=\text{inf/dec}$), imply that widths and positions of the Gaussian rates distributions are also stationary, which yields their widths at any height values below their maxima. In particular, by arbitrarily defining the dates of beginning and end of the epidemic at the time instances such as the events (infections or deaths) amount, 1%, of the respective values at the corresponding maxima, one obtains the total duration values of the epidemic in terms of infection or death events. Duration values of the epidemic, defined in terms of infection daily rates, are collected in Table 1. By averaging the values for countries having passed the epidemic peak one obtains, $\langle \tau_{duration}^{inf/dec} \rangle = 94$ days, which represents the characteristic, estimated relaxation time of the pandemic. Increasing confidence in this value will be gained by averaging over more and more countries having crossed the epidemic peak, with main interest to provide a sound basis for defining how long should last confinement restrictions.

Methods and Computations

Data in the form of time series of cumulated numbers of infections/deaths, updated on a daily basis, have been collected from various official sources for 24 selected countries experiencing the endemic since January 22 till May 8, 2020^{16–18}. Therefrom, first derivatives are numerically obtained via the Richardson finite differences scheme working at order $O(h^6)$ ^{23,24}. At the date above, no epidemic events were announced in the countries entering the present study but China, for which the starting date of the epidemic is still badly known. The parameters, $(N_0^k, \tau_m^k, \sigma^k)_i$, have then been numerically estimated for each country by using a least-squares minimization scheme applied to the populations of infected and deceased individuals while adopting a common value for the corresponding standard deviations. The home-made computer program driving the minimization¹⁹, interfaces with the multidimensional minimization package MERLIN²⁵. In principle, the parameters can be adjusted via one among eqs.(1,3) or (2,4) indifferently. In practice however, due to the large fluctuations affecting the time derivatives (Fig.1a), the smoother integral forms reveal preferable (Fig. 1b). In this work, eq.(4) has been used together with the objective function

given by:

$$f_{lsq} = \sum_{j=1}^m w_j \left(1 - \frac{n^*(\tau_j)}{n_p^*(\tau_j)} \right)^2 \quad (5)$$

where, the index, j , runs over the events in any given dataset, w_j , are weights ($w_j=10$ if $|\tau_j^*|>1$ and $w_j=1$ otherwise) and, $n^*(\tau_j^*)$, $n_p^*(\tau_j^*)$, are respectively the observed and predicted numbers of events expressed in reduced units (eq. 3).

The statistical analysis of the reduced data fluctuations around the mean values, $\langle n^* \rangle$ and $\langle \frac{dn^*}{dt} \rangle$ at any τ^* (master curves), has been made under the assumption that these follow normal distributions of known mean but of unknown variance. A reasonable compromise between the number of master plot data carrying error bars (confidence intervals of the mean) and the need for the highest possible statistical significance of the estimated errors, leads to defining classes with about 50 contiguous time intervals, $\delta\tau^* \in [0.14, 0.18]$. Otherwise, the computations proceeded along the standard guidelines of statistical data analyses²¹.

The minimization procedure yields stationary values of the parameters for all the countries that are close or have crossed the pandemic peak (Table 1). If this condition is not fulfilled, the quality of the fit improves at the expense of increasingly large, yet unacceptable, N_0 and τ_m values. In these cases, τ_m and σ 's were determined by limiting, N_0 , to the values of the actual numbers of infections/deaths within a 5% tolerance interval, by using a native tool contained in the package MERLIN that allows setting lower and upper limits to guesses by the minimization package of the adjustable parameters values. Figure 1 illustrates the successful application of this procedure to the data from France.

Discussion

Unlike other works, the present study takes advantage of a phenomenological approach to evidence characters common to the epidemic in different countries. The approach relies on the assumption that time rates of events are Gaussian functions of the elapsed time, which is proven to satisfactorily designate the shapes of the data from different origins. However, the developed methodology is not unconditionally bound to this hypothesis since any other distribution, better approximating the observations or stemming from justified modeling principles, will perfectly adapt to the phenomenological framework of the present work. Defining this framework has been motivated by the observation that cumulated numbers and daily rates of events of different origins condense in to uniquely defined master graphs, via a linear transformation of the coordinates. This is the major result of the present work, susceptible of general applicability to cases other than the COVID-19 pandemic.

The early applied travel and mobility restrictions, aiming at hindering the virus propagation, have practically transformed the countries experiencing the epidemic in to closed and non-interacting population ensembles. It follows that the global virus evolution at the pandemic scale may be viewed as resulting from the linear superimposition of, $k=1-24$, shifted in time replicas of the same stochastic process evolving in, k , different countries. Thereby, data dispersion overwhelmingly present in the master plots of the datasets of daily rates (Figs. 2c, 2d), is used for evaluating error bars, which provide expectations for the upper limits of daily rates as functions of the elapsed time at the desired confidence level. Although the mathematical projection of the duration of the epidemic between countries may differ substantially, as is seen in Table 1, the corresponding reduced data are within the estimated error bars (Fig. 3), thus leading to the conclusion that the intrinsic characteristics of the pandemic are well obeyed. This result is of interest for national health authorities. In this context, the coarse-graining procedure yielding the estimation of errors has revealed useful in eliminating data fluctuations (Fig. 3). However, clarifying their origin is matter for further investigations.

Finally, the present phenomenological analysis can be used as a tool in checking for consistency the datasets referring to the time series of cumulated numbers, the reduced data of infections and deaths from different countries. Without addressing the details of the correlations between them, the phenomenology discriminates datasets that do importantly depart from the master plots. Thereby, singularities are shed in light that justify the elimination of the incriminated datasets from the error estimation in the present analysis, while encouraging further investigations for better understanding their origins. Work in progress focuses on improving the data description, by using better justified time distributions of events.

References

1. Siettos, C.I and Russo, L. Mathematical modelling of infectious disease dynamics. *Virulence* **4:4**, 295–306, DOI: <https://doi.org/10.4161/viru.24041> (2013).
2. Adam, D. Modelling the pandemic. *Nature* **580**, 316–318, DOI: <https://doi.org/10.1038/d41586-020-01003-6> (2020).

3. Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., Di Filippo, A., Di Matteo, A. and Colaneri, M. Modelling the covid-19 epidemic and implementation of population-wide interventions in italy. *Nat. medicine* **580**, 317–318, DOI: <https://doi.org/10.1038/s41591-020-0883-7> (2020).
4. Heesterbeek, H. et al. Modeling infectious disease dynamics in the complex landscape of global health. *Science* **347**, 6227, DOI: <https://doi.org/10.1126/science.aaa4339> (2015).
5. Russo, L. Tracing day-zero and forecasting the fade out of the covid-19 outbreak in lombardy, italy: A compartmental modelling and numerical optimization approach. *medRxiv preprint* DOI: <https://doi.org/10.1101/2020.03.17.20037689> (2020).
6. Anastassopoulou, C., Russo L., Tsakris A., Siettos C. Data-based analysis, modelling and forecasting of the covid-19 outbreak. *PLoS ONE* **15**, 3): e0230405, DOI: <https://doi.org/10.1371/journal.pone.0230405> (2020).
7. Cobey, S. Modeling infectious disease dynamics. *Science* **eabb5659**, DOI: <https://doi.org/10.1126/science.abb5659> (2020).
8. Zhu, Y. and Chen, Y. Q. On a statistical transmission model in analysis of the early phase of covid-19 outbreak. *Stat Biosci* DOI: <https://doi.org/10.1007/s12561-020-09277-0> (2020).
9. Shaikh, A. S., Shaikh, I. N. and Nisar, K. S. A mathematical model of covid-19 using fractional derivative: Outbreak in india with dynamics of transmission and control. *Preprints* DOI: <https://doi.org/10.20944/preprints202004.0140.v1> (2020).
10. Lee, M. J. and Lee, D-S. Understanding the temporal pattern of spreading in heterogeneous networks: Theory of the mean infection time. *Phys. Rev. E* **99**, 032309–1 –9, DOI: <https://doi.org/10.1103/PhysRevE.99.032309> (2019).
11. Fang, Y. et al. Transmission dynamics of the covid-19 outbreak and effectiveness of government interventions: A data-driven analysis. *J. Med. Virol.* **92**, 645–659, DOI: <https://doi.org/10.1002/jmv.25750> (2019).
12. Khrapov, P. V. and Loginova, A. A. Comparative analysis of the mathematical models of the dynamics of the coronavirus covid-19 epidemic development in the different countries. *Int. j. open inf. technol.* **8**, **5**, 17–22 (2020). [ISSN: 2307-8162].
13. Ciufolini, I. and Paolozzi, A. Mathematical prediction of the time evolution of the covid-19 pandemic in italy by a gauss error function and monte carlo simulations. *Eur. Phys. J. Plus* **135:355**, 1–8, DOI: <https://doi.org/10.1140/epjp/s13360-020-00383-y> (2020).
14. Ardabili, S. F. et al. Outbreak prediction with machine learning. *Preprint* DOI: <https://doi.org/10.2139/ssrn.3580188> (2020).
15. Nemati, M., Ansary, J. and Nemati, N. Covid-19 machine learning based survival analysis and discharge time likelihood prediction using clinical data. *Preprint* DOI: <https://doi.org/10.2139/ssrn.3584518> (2020).
16. Global-projections. <https://covid19-projections.com/#global-projections> (2020). [Online; last accessed May 8, 2020].
17. Todays data. geographic-distribution-covid-19-cases-worldwide. <https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide> (2020). [Online; last accessed May 8, 2020].
18. Covid deaths. <https://ourworldindata.org/covid-cases>, and <https://ourworldindata.org/covid-deaths> (2020). [Online; last accessed May 8, 2020].
19. Data and computer programs used in the present work are available upon demand sent to: vassilis.pontikis@cea.fr.
20. Barmparis, G. D. and Tsironis, G. P. Estimating the infection of covid-19 in eight countries with a data-driven approach. *Chaos, Solitons Fractals* **138**, 109842, DOI: <https://doi.org/10.1016/j.chaos.2020.109842> (2020).
21. Kreyszig, E. *Statistische Methoden und ihre Anwendungen* (Vandenhoeck & Ruprecht, Göttingen, 1973), 4 edn. ISBN: 3-525-40713-0.
22. Bonamente, M. *Statistics and Analysis of Scientific Data* (Springer, Graduate Texts in Physics, 2013), 2 edn. ISBN: 978-1-4939-6570-0.
23. Richardson, L. F. The approximate arithmetical solution by finite differences of physical problems including differential equations, with an application to the stresses in a masonry dam. *Philos. Transactions Royal Soc. A.* **210** (459–470), 307–357, DOI: <https://doi.org/10.1098/rsta.1911.0009> (1911).
24. Richardson, L. F.; Gaunt, J. A. The deferred approach to the limit. *Philos. Transactions Royal Soc. A.* **226** (636–646), 299–349, DOI: <https://doi.org/10.1098/rsta.1927.0008> (1927).
25. Papageorgiou D. G., Demetropoulos I. N. and Lagaris I. E. Merlin-3.1.1. a new version of the merlin optimization environment. *Comput. Phys. Commun.* **159**, 70–71, DOI: <https://doi.org/10.1016/j.cpc.2003.12.005> (2004).

SN	Country	N_0^{inf}	τ_m^{inf}	$Shift^{inf}$	N_0^{dec}	τ_m^{dec}	$Shift^{dec}$	σ	$\delta\tau_{duration}^{inf/dec}$
1	AR	5888.8	48.0	42	322.6	47.3	46	17.6	132
2	AU	6833.1	63.5	4	92.1	38.3	40	8.4	62
3	BE	49966.4	66.0	14	8314.0	35.0	50	11.3	84
4	BR	155684.0	64.9	35	10011.7	43.0	56	10.4	77
5	CA	73658.5	87.0	4	7971.3	57.7	48	17.5	130
6	CN	82439.2	19.3	-	4426.0	29.9	1	9.6	71
7	DE	165393.1	47.7	25	7418.7	37.7	48	11.7	87
8	ES	246592.4	49.7	25	25857.0	34.5	42	13.5	100
9	FR	180077.8	73.9	3	25462.4	55.6	25	11.1	83
10	GB	206180.9	62.0	25	31884.2	43.0	44	13.8	102
11	GR	2609.7	34.9	36	36143.8	24.8	51	11.9	88
12	HU	3303.9	44.0	42	446.8	37.9	54	14.1	105
13	IT	213013.0	64.1	7	29910.1	43.5	31	16.4	122
14	IN	86944.9	78.0	25	2915.9	52.0	51	12.3	92
15	IR	103135.0	44.5	29	6486.0	44.6	29	16.9	126
16	JP	15289.5	83.0	1	656.1	71.8	23	11.4	85
17	NL	42510.2	42.5	37	5268.0	36.9	45	14.4	107
18	PK	33633.6	65.0	36	914.0	60.0	44	14.2	106
19	PL	15200.0	43.0	43	840.1	41.7	51	14.2	106
20	RU	311074.8	96.0	10	2145.7	38.0	63	10.6	79
21	SE	23300.3	59.3	25	2971.9	37.8	50	12.4	92
22	KR	10749.2	18.6	25	246.8	34.7	30	9.9	74
23	TR	135000.5	35.6	49	3776.9	32.0	55	12.1	90
24	US	1292623.0	61.0	25	90332.7	53.0	39	15.0	112

Table 1. Per country alphabetic list of the phenomenological analysis parameters (country abbreviation code: ISO 3166-2 alpha-2). Codes in boldface mark countries having passed the epidemic peak. Columns labeled $Shift^{inf/dec}$ list the number of days separating the common starting date of country datasets (January 22, 2020) and the occurrence of the first infection/death case. Standard deviation values, σ , are the same for infections and deaths (superscripts 'inf' and 'dec'). The delays, $\delta\tau_{duration}^{inf/dec}$, are the widths of the theoretical Gaussians representing rates of infections/deaths at 1% of the maximum and represent the estimated effective duration of the epidemic in the given country. For countries that have not yet reached the epidemic peak values displayed here are projections likely to change in the near future. Time-delay values are rounded to the nearest integer number of days.

Acknowledgements

Lina Prakoura, and Renée Quillivic are gracefully acknowledged for stimulating discussions, constant support and encouragements.

Author contributions statement

V.P. conceived the theoretical model after enriching discussions with Ph.K and T.K, V.P. has developed the computer programs and has performed the calculations, T.K. and Ph.K. have organized and cross-checked the collection of the statistical information that has been used in this work. All authors analyzed the results and reviewed the manuscript.

Additional information

Accession codes (non applicable); **Competing interests** The authors have no competing interests as defined by Nature Research, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

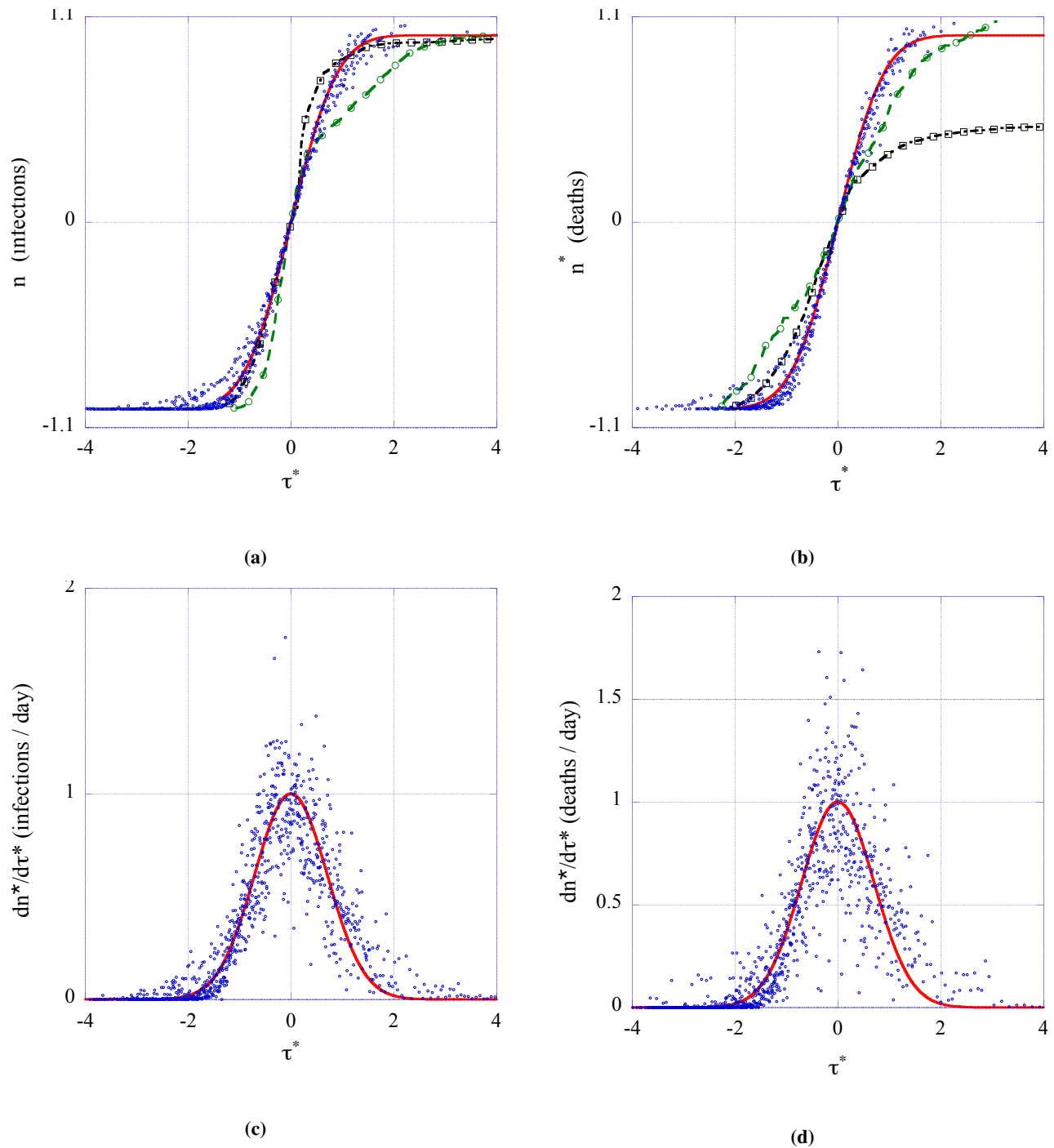
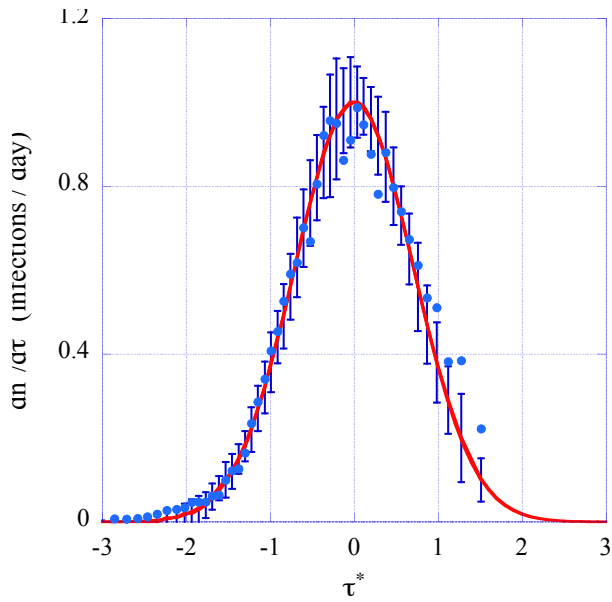
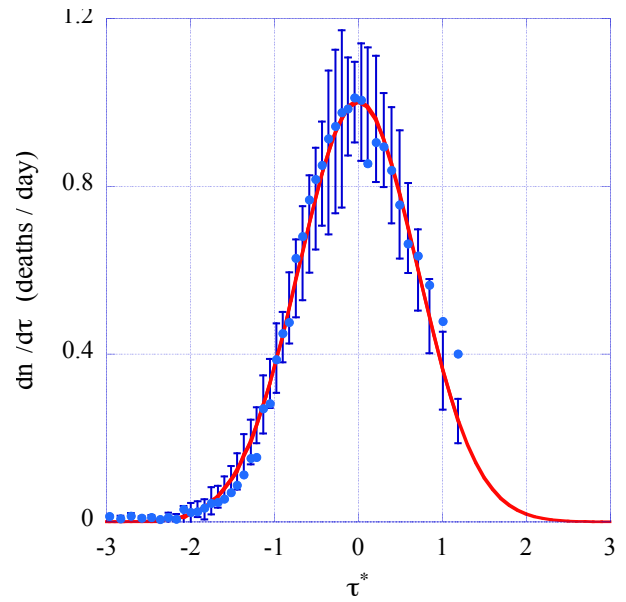


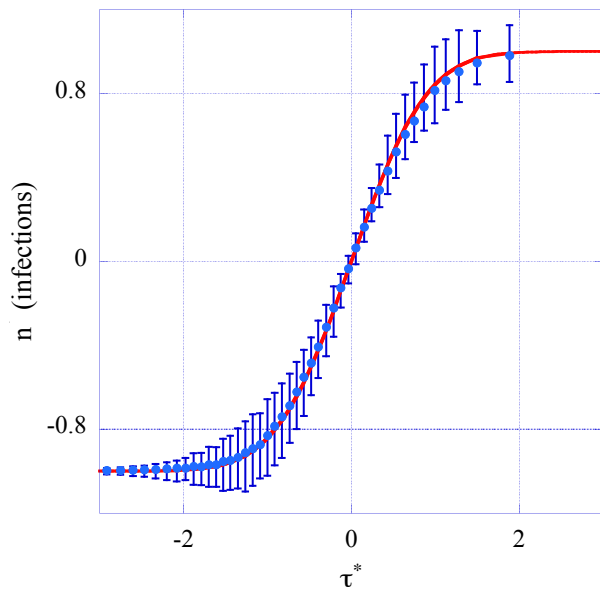
Figure 2. Master plots of reduced cumulated numbers of infections (a) and deaths (b) and of the corresponding reduced daily rates (c, d) as functions of the reduced elapsed time. Full lines represent the theoretical plots (eqs.(3,4)). Dashed (open circles (South Korea)) and dashed-dotted (open squares (China)) lines in (a,b) represent datasets departing significantly from the master plots.



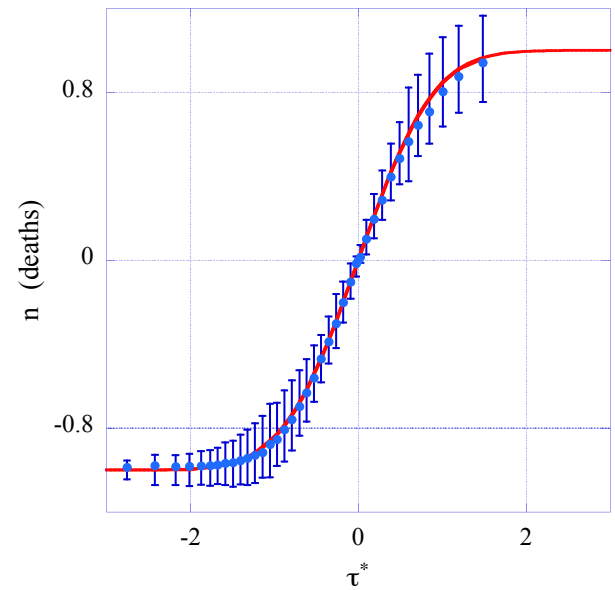
(a)



(b)



(c)



(d)

Figure 3. Master plots with error bars: reduced daily rates of infections (a) and deaths (b) and reduced cumulated numbers (c, d) as functions of the reduced elapsed time. For the sake of readability, the length of error bars in the two graphs at the bottom is magnified by a factor (x5). Full lines represent the theoretical plots (eqs.(3,4)), whereas full dots correspond to the average values of the data in the given time interval (see text). No data are available ahead the present time situation of the pandemic ($\tau^* > 2$).