

Mathematical modeling of the development of confirmed daily infection numbers in the COVID-19 pandemic by a special exponential function

Egbert Keller¹, Johannes Keller², and Matthias Selg³

¹ private, non-institutional contribution

Former affiliation: Kristallographie, Albert-Ludwigs-Universität Freiburg i.Br., Germany

E-mail: eukell77@gmail.com

² private, non-institutional contribution

E-mail: Johannes.Keller2@gmx.de

³ private, non-institutional contribution

E-mail: selg.m@gmx.de

Correspondence address: eukell77@gmail.com

Abstract

The developments of confirmed daily Sars-CoV-2 infections can be modeled *a posteriori* in a comparatively simple and satisfactory way by means of a special exponential function: its exponential coefficient varies exponentially with time itself. This property enables the function to directly simulate the „daily case“ curve fragments observed for the first waves of the current epidemics in a number of counties and countries. Linear combinations of two or more of the functions allow for the modeling of the complete curves observed in the vast majority of all regions for which case number developments have been tabulated.

Keywords

COVID-19 pandemic, Sars-CoV-2, confirmed daily infections, mathematical model

Introduction

For the mathematical modeling of the progress of infection numbers in the course of epidemics, e.g., the COVID-19 pandemic, usually theoretical models are employed, which base on systems of differential equations and most often are copies or extensions of the well-known SIR model [1-6] (but see also, e.g. [7,8]). The letter „I“ in SIR (or – expressed as a time-dependent function – „I(t)“) represents the number of infectious persons in a certain pool of people (e.g. the population of a country) at a certain day. Differing from that, the symbol $\hat{I}(t)$ will represent, in this paper, the number of *confirmed* new infections at a certain day t , or the time-dependent development of this number if t is growing continuously.

While there generally are no real $I(t)$ data which is, of course, also valid with respect to COVID-19, $\hat{I}(t)_{\text{real}}$ numbers for the pandemic have been registered and documented for numerous countries and counties [9]. Under the assumption that the relation between the number of *confirmed* cases and the number of the *actually* newly infected individuals doesn't vary too much in time, the development of the former should still allow conclusions about the spreading of the epidemic in a certain country.

As a relation between $I(t_i)$ and $\hat{I}(t_i)$ can – under certain simplifying conditions – be formulated quantitatively (e.g., by assuming a relative dark number), the above mentioned mathematical models can be calibrated by comparison of their predictions with the real case number developments $\hat{I}(t)_{\text{real}}$, which is equivalent to modeling such developments *a posteriori* [see, e.g., 3-6].

A somewhat different strategy is to model the progress of case numbers $\hat{I}(t)_{\text{real}}$ (or the hereof derivable development of the *cumulated* confirmed case numbers $\phi(t)_{\text{real}}$) in a country directly by a single mathematical function with the aim to be able to predict the further development of the epidemic. Modeling attempts by using a modified Gauss error function [10] or a logarithmic normal distribution function [11], for example, led to satisfying matches with the corresponding real case number developments. However, in the first case, the symmetry of the model functions is incompatible with the skewness of almost all documented $\phi(t)_{\text{real}}$ (and $\hat{I}(t)_{\text{real}}$) developments, in the second case the models predicted a soon-to-be end of the epidemics (far in the past, as seen from now) for numerous (e.g. West European) countries, which obviously is inconsistent with reality.

In the following we present the alternative mathematical function (1) to be used for *a posteriori* modeling of $\hat{I}(t)_{\text{real}}$ -developments (COVID-19) with respect to – in principle – arbitrary countries or counties, which doesn't suffer from the two mentioned drawbacks:

$$\hat{I}(t) = \hat{I}(0) e^{(ae^{-bt})t} \quad (1)$$

Results

By means of the Python3 software mentioned below a satisfying fitting of a single function of type (1) to the developments of the numbers of confirmed new infections $\hat{I}(t)_{\text{real}}$ could be obtained for numerous countries and US counties (status: end of June, 2020). The quality of the fit increased (not unexpectedly) when 7-days averages were modeled instead of daily values. Averaging (performed symmetrically in time, see section Methods) smoothes the frequently observable fluctuations of confirmed case numbers in a weekly rhythm and other discontinuities which accompany the execution of tests and the reporting of positively tested cases. A similar (but asymmetric) averaging is, for example, performed for the determination of the so-called reproduction number [12]. For 33 countries und 24 US counties a very good fit ($R^2 > 0,97$) of a function of type (1) to the smoothed case number curve could be obtained (status: end of June, 2020) (Fig. 1 and Supplementary Material: Cov19_1F.pdf). Seen from today (end of October, 2020), the corresponding case number developments represented the „first waves“ of the epidemics in the corresponding countries and counties.

In more than 180 additional cases (56 countries, 125 US counties), for which a model consisting of a single function of type (1) had led to unsatisfactory results, modeling of the smoothed case number curves succeeded in a simple and optically satisfactory way ($R^2 > 0,97$) for the sum of two or three of such functions, in principle corresponding to the occurrence of a second and third infection wave (status: end of June, 2020; see also Supplementary Material: Cov19_2F.pdf and Cov19_3F.pdf).

The number of successful fittings can be expanded to far beyond 238, when sums of more than three functions of type (1) are allowed for the models. By end of October, 2020, this measure is now actually required for almost all countries (including those represented in Fig. 1) if modeling of the complete case number developments is desired (Fig. 2). An answer to the question, how far the single curves in the different cases actually correspond to real processes, still requires additional investigations.

It should be noted here, finally, that (1) is suited to model, e.g., „daily deaths“-developments $D(t)$ in a similarly satisfying way as compared to $\hat{I}(t)$, an aspect which however will not further be discussed in this paper.

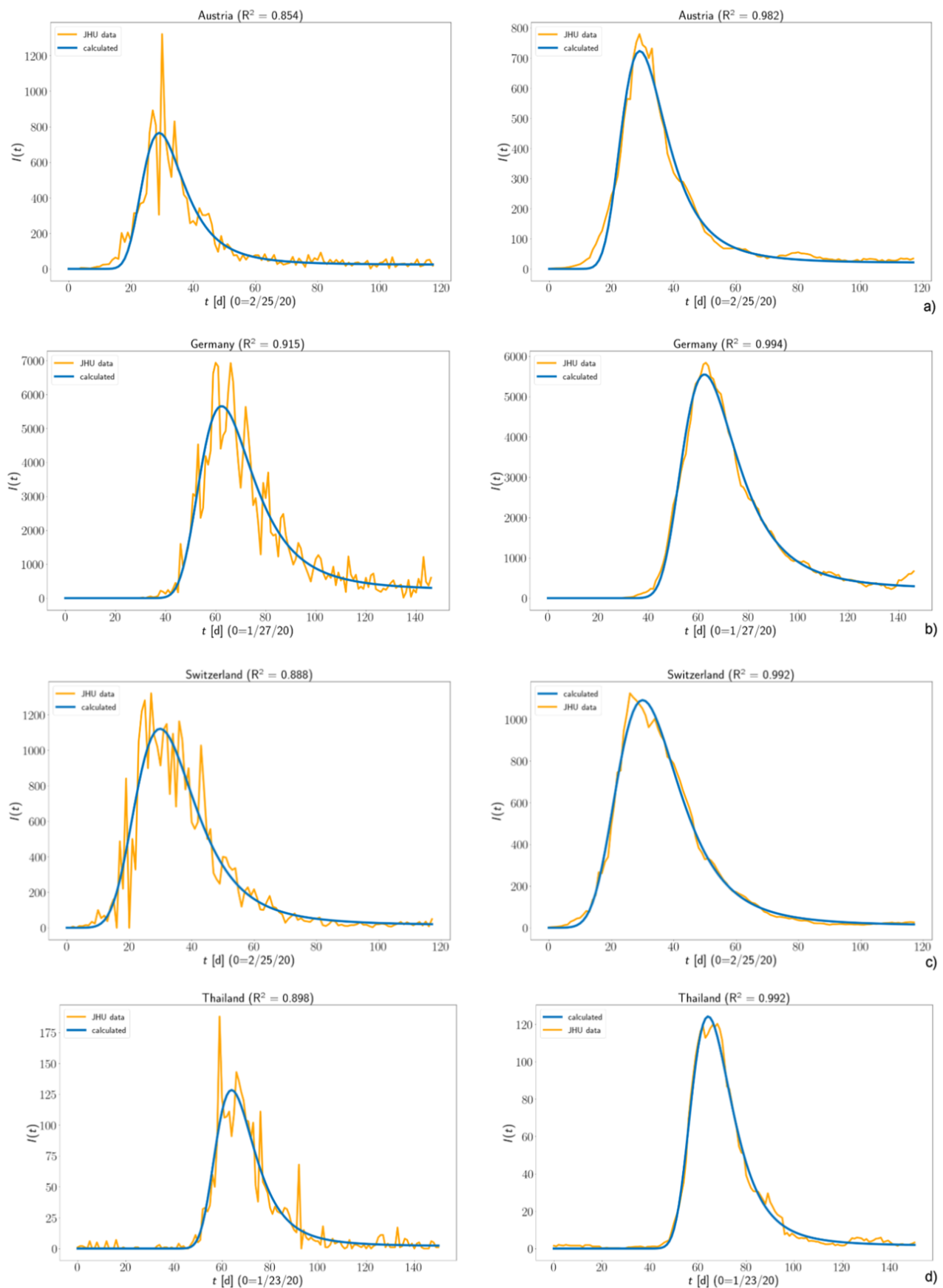


Fig. 1. Examples of „first wave“ case number developments (orange), and fitted functions of type (1) (blue). Left-hand side: daily JHU data [9], right-hand side: data after time-symmetric 7-days averaging. Status: end of June, 2020. a) Austria, b) Germany, c) Switzerland, d) Thailand.

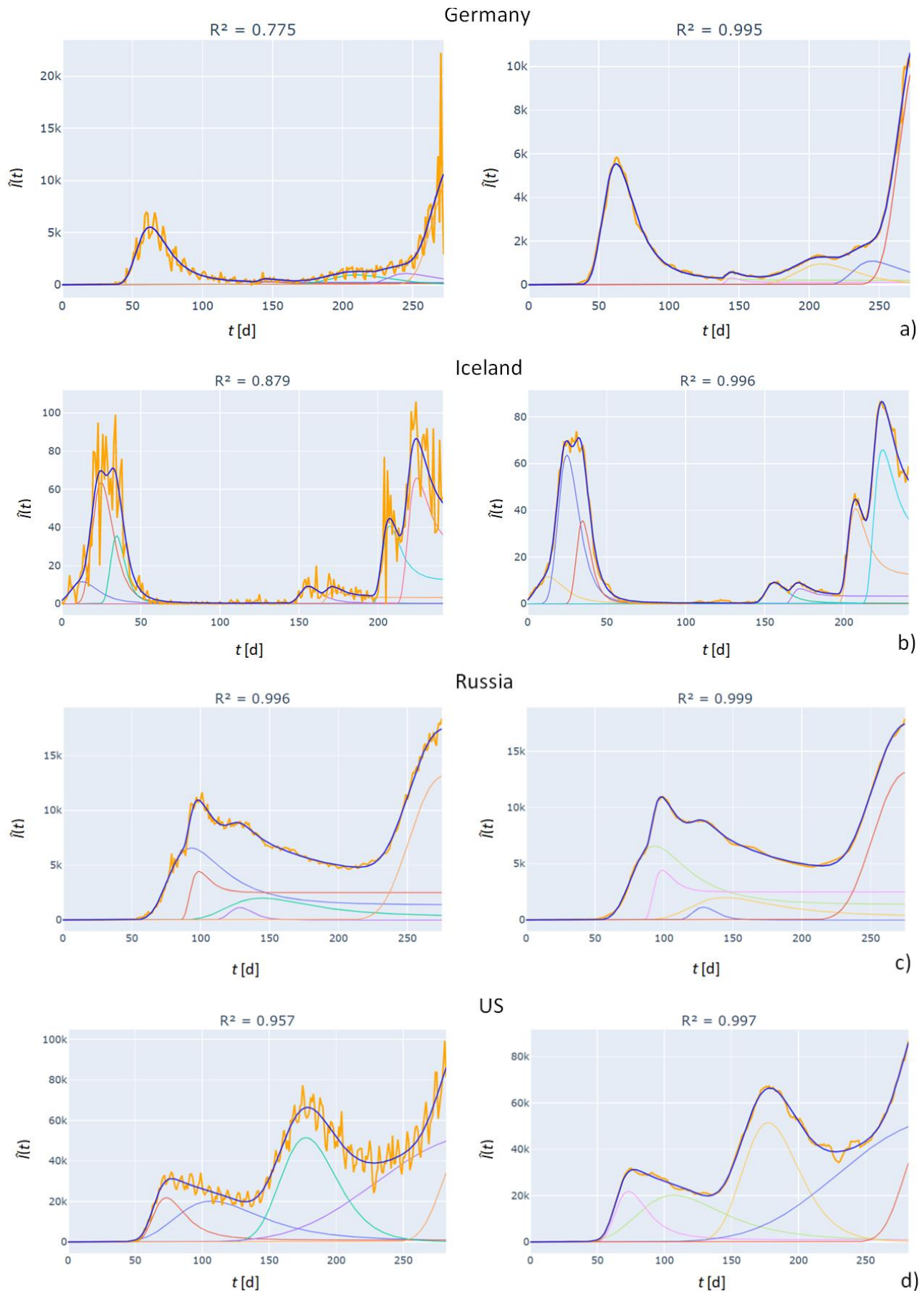


Fig. 2. Some examples for modeling of case number developments with sums of n functions of type (1): a) Germany ($n=5$), b) Iceland (7), c) Russia (5), d) US (5). (status: end of October, 2020). Curves representing the n functions are also shown (in different colours) in the diagrams.

Discussion

One advantage of the modeling method presented here is that it doesn't base upon differential equations, which means that it is comparatively easy to understand and comparatively simple to apply. Our model function (1) is an exponential function with another exponential function in the exponent. It shows the asymmetry (relative to the curve maximum) which could be observed for almost all real pure „first wave“ case number developments. When the foothills of the waves had been reached (as to be seen in Fig. 1), the function did not predict a soon-to-come end of the epidemics. As (1) doesn't fall below $\hat{I}(0)$ at its right-hand side, it predicted instead – hitherto undisproved for the majority of all countries – a long persistence of the epidemics.

$\hat{I}(0)$ represents the function value at $t = 0$ and plays the role of a proportionality factor. An $\hat{I}(0)$ value of 1 could formally be assigned to a single infected person who starts the infection wave, $\hat{I}(0) > 1$ then would represent the more or less „synchronous“ start of more than one „individual“ waves superimposing each other within the population of question. Of the two additional parameters in (1) a seems to quantify the initial unrestricted exponential growth of the number of daily confirmed infection cases due to (2):

$$\hat{I}(t) = \hat{I}(0) e^{at} \quad (2)$$

In (1), the constant a from (2) is replaced by the time-dependent ae^{-bt} , which means that the initial growth of the function is increasingly weakened and finally – after having passed the curve maximum – replaced by a decrease. The amount of the weakening of (2) (and thus – at first sight – the success of the national efforts to stop the epidemic) is quantified by b . An irritating fact in this context is, however, that b is a constant which doesn't change in time. Measures like „social distancing“ or „lockdown“ obviously don't lead to an increase of b in the course of the development (but see below).

Of the almost 60 countries and counties, of which the smoothed case number developments could be described in a satisfactory manner ($R^2 > 0,97$) by a single function (1), t_{\max} was clearly exceeded in 27 cases (status: end of June, 2020). Of these, the nine West European countries are presented in Tab. 1 together with the corresponding parameter values and additional specific data. As can be seen, the highest determined values for a as well as for b differ from the lowest ones by about a factor of 3. The quotients a/b , however, scatter only slightly, i.e., with a standard deviation of 1 about their common mean value 10. If the additional 18 countries/counties (not in Tab. 1) are also accounted for, the scattering of the quotient values grows significantly, thus, a value of about 10 cannot be expected in general for a/b .

If (1) is fitted to the first parts of the curves of the nine countries only (the fragments ending at about half of the maximum value of the smoothed curves) this generally results in a larger a and a smaller b value as compared to the values in Tab. 1. Thus, national antiviral measures become visible at least in this comparison. Furthermore, this findings insinuate that, in contrast to the initial approach given above, a and b can not be discussed separately with respect to their epidemiological meanings.

The quotient a/b determines via $\hat{I}(0) e^{a/be}$ the maximum value \hat{I}_{\max} of the function (see Appendix: Function maximum). Because of the relatively small variation width of the a/b values the differences between the \hat{I}_{\max} values in Tab. 1 are mainly caused by the differences in the $\hat{I}(0)$ values. $\hat{I}(0)$ itself is positively correlated with the population number Z . Via $\hat{I}(0) / Z$ as a new proportionality factor normalized graphs of the modeled case number developments can be obtained (Fig. 3), which allow a direct comparison of the progress of the epidemics in the different countries.

Tab.1. Parameters of (1)¹⁾ and other data for some West European countries with $R^2 > 0.97$

Country	\hat{I}_0 ²⁾	a [d ⁻¹]	b [d ⁻¹]	a/b	$e^{a/be}$	\hat{I}_{\max}	Z ³⁾ *10 ⁻⁶	(\hat{I}_0/Z) *10 ⁶	(\hat{I}_{\max}/Z) *10 ⁶	t_{\max} [d]	t_1 [d]	t_2 [d]	Δt_h [d]
Austria	22.3	0.706	0.075	9.41	31.91	712	8.86	2.52	80	13	7	25	18
Belgium	23.7	0.380	0.034	11.18	61.04	1447	11.43	2.07	127	29	15	52	37
Germany	261.0	0.440	0.053	8.30	21.20	5534	83.17	3.14	67	19	9	36	27
Italy	159.0	0.333	0.035	9.51	33.12	5266	60.26	2.64	87	29	14	52	38
Luxembg	5.6	0.641	0.069	9.29	30.49	171	0.63	8.89	271	14	7	27	20
Netherlds	30.5	0.345	0.035	9.86	37.57	1146	17.29	1.76	66	29	14	52	38
Spain	224.0	0.484	0.051	9.49	32.83	7353	47.10	4.76	156	20	10	36	26
Switzerld	12.2	0.583	0.048	12.15	87.20	1064	8.60	1.42	124	21	11	36	25
UK	293.0	0.226	0.028	8.07	19.48	5707	66.44	4.41	86	36	16	69	53

¹⁾Status: end of June, 2020; ²⁾ non-integer values result from the parameter optimization process;

³⁾population number [13] (shown are values of Z after multiplication by 10^{-6})

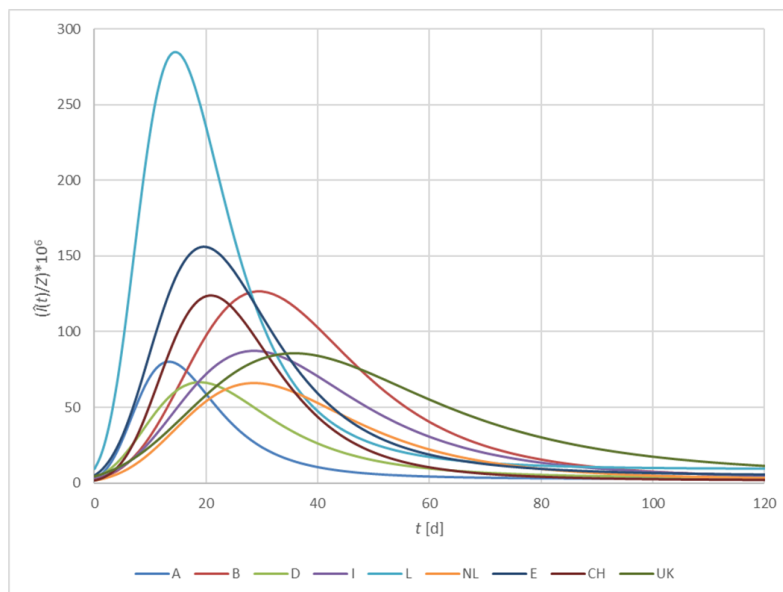


Fig. 3. Synchronized (i.e., with all $t(0)$ values set to 0) and normalized (to a population number of 10^6) functions (1) for the nine countries from Tab. 1 (A = Austria, B = Belgium, D = Germany, I = Italy, L = Luxembourg, NL = Netherlands, E = Spain, CH = Switzerland, UK = United Kingdom).

Another characteristic quantity, which is determined by *both* parameters a and b , is the full width at half maximum Δt_h of (1). It can be obtained as the difference of the two solutions of equation (3) (see Appendix: Full width at half maximum):

$$\ln t - bt - \ln(a/b e - \ln 2) + \ln a = 0 \quad (3)$$

Δt_h can be interpreted as a (reciprocal) measure of the velocity at which – in retrospect – the fight against the virus succeeded. At a constant quotient a/b (i.e., at a constant \hat{I}_{\max}), Δt_h grows with decreasing values of a and b . From Tab. 1 one can see, that, e.g., for Austria and Italy with similar \hat{I}_{\max}/Z values the Δt_h values differ by a factor of 2. The same is valid for the comparison of the data for Germany and the United Kingdom.

By the possibility to expand the model to a sum of n functions of type (1) the case number developments of most of the countries and counties for which corresponding data are documented become possible objects for modeling. In the first instance, mainly those cases are of interest where a satisfactory modeling can be achieved for a comparatively small number n , as is the case for the examples shown in Figs. 2 and 3, with $n = 3$ to 7. In such cases, the spread of the virus which is mirrored in the development of the case numbers can be thought of as being composed of n spatially and/or temporally separated infection waves. We cannot say, however, how far this attribution is generally justified.

Methods

a) Development of the function equation

Summing up the daily published $\hat{I}(t)_{\text{real}}$ values in the past leads to the number of *cumulated* confirmed new infections $\phi(t)_{\text{real}}$. For an *unrestricted* spread of an epidemic one would expect for $\phi(t)$ (and – with a smaller growth – also for $\hat{I}(t)$) an exponential initial increase as given in (4):

$$\phi(t) = \phi(0) e^{\alpha t} \quad (4)$$

For any two function values $\phi(t)$ and $\phi(t+1)$ equation (2) implies equation (5) (see Appendix: Parameter α):

$$\alpha = \ln[\phi(t+1) / \phi(t)] \quad (5)$$

From the numbers published by the Robert-Koch-Institut (RKI) for Germany [14] one could realize as early as March 2020, that the progress of $\phi(t)_{\text{real}}$ stayed behind the one to be expected according to (4).

Therefore, one of us (MS) investigated, how α changes in time, when it is calculated (via (5)) by employing the real cumulated case numbers $\phi(t+1)_{\text{real}}$ und $\phi(t)_{\text{real}}$. The result was an exponential decrease (per average) of α due to $\alpha = ae^{-bt}$ (Fig. 4).

Following this result, the „behaviour“ of a simple exponential function $C = C(0) e^{ct}$ was investigated, when the constant c in the exponent is replaced by the variable, time-dependent expression ae^{-bt} .

The resulting function graphs strongly resembled the development of daily cases $\hat{I}(t)_{\text{real}}$ in Germany, which immediately led to the tentative formulation of (1).

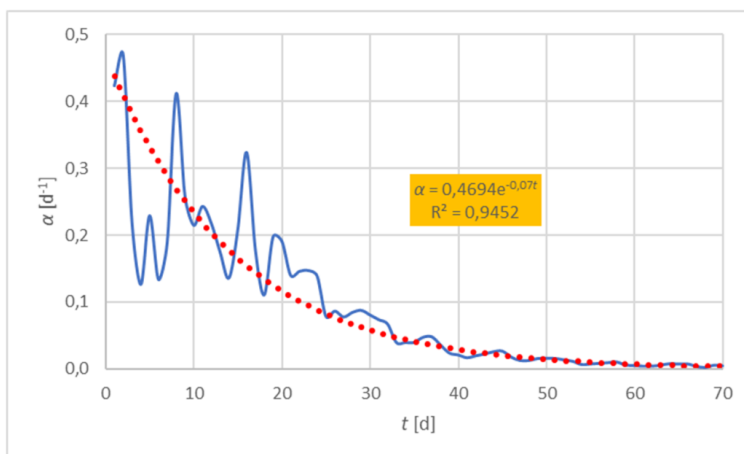


Fig. 4. $\alpha = \ln[\phi(t+1)_{\text{real}} / \phi(t)_{\text{real}}]$ in dependance of t (see text)

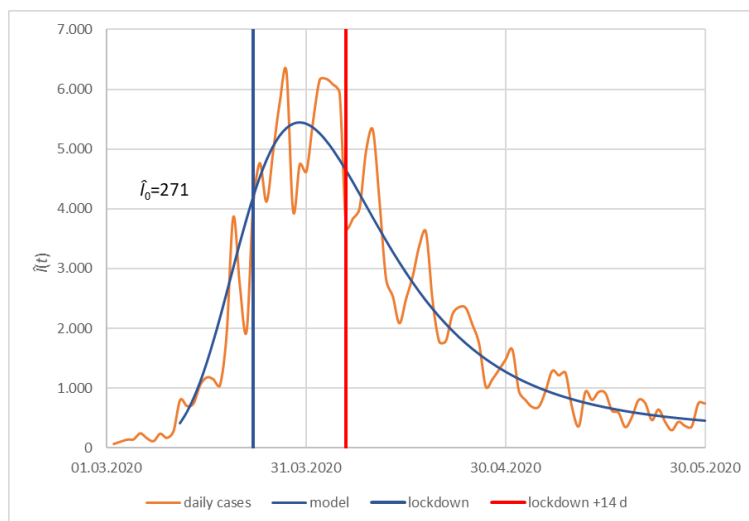


Fig. 5. Confirmed daily cases in Germany up to end of May 2020 (orange, data source: RKI [14]), and calculated curve (blue).

In (1) a and b determine the shape of the function curve (blue in Fig. 5), while $\hat{I}(0)$ acts as a proportionality factor. The function exhibits a single maximum at $t=1/b$ (see Appendix: Function maximum). With an estimated value of t_{max} (according to

the real data) b could be determined. b and an estimated value for \hat{I}_{\max} then allowed the calculation of a . By an adjustment of the parameters $\hat{I}(0)$, a and b with the coefficient of determination, R^2 , as the only criterion, (1) could be made to model the real case number development up to end of May 2020 in a satisfactory way (Fig. 5).

b) Conception and development of used software

Starting from the above results, two of us (EK and JK) developed concepts for the automatic modelling of COVID-19 case number developments by means of (1) and implemented them (partially) in the form of FORTRAN- [15] and Python3 software [16]. Especially the latter allows modeling of the case number developments for most countries and counties of earth, for which daily case number data are available. This software has the following properties:

- Facility to import *ad hoc* up-to-date case number data for all countries from a .csv file which is supplied in the internet by Johns-Hopkins-University, Baltimore, USA [9].
- Optional smoothing of the data curve by calculation of 7-days average values; averaging for day t_j is performed symmetric-in-time across the days t_j-3 , t_j-2 , ... t_j+3 .
- Optimization of coarsely estimated starting values for the parameters $\hat{I}(0)$, $t(0)$, a , b by means of a least-squares method [17]. The additional parameter $t(0)$ allows positioning of the model on the t axis and is introduced by replacing all instances of t in the right side of (1) by the expression $(t - t(0))$.
- Optional extension of the model towards a sum of different functions of type (1).
- Optionally adjustable temporal limitation of the case number developments at one or both sides.
- Output of diagrams and refined (= optimized) parameter values

Conclusion

For quite a number of countries and counties distributed across the globe a satisfactory *a posteriori* modeling of the development of the „first wave“ fragments of the confirmed daily new infections by Sars-CoV-2 is possible by means of a single function of type (1). The number of these countries and counties can be increased substantially, and the restriction to the first wave can be given up, when the models are extended to a sum of few functions of type (1) (which formally corresponds to a superposition of few different infection waves). Thus, our modeling approach seems to be generally applicable as far as the COVID-19 pandemic is concerned. Insofar, the mathematical structure of (1) which is characterized by an exponentially changing factor in the exponent seems to express an up to now unnamed feature of the pandemic.

References

1. Wang N, Fu Y, Zhang H, Shi H. An evaluation of mathematical models for the outbreak of COVID-19. *Prec Clin Med*. 2020;3(2):85–93. Available from: <https://doi.org/10.1093/pcmedi/pbaa016>
2. Mohamadou Y, Halidou A, Kapen PT. A review of mathematical modeling, artificial intelligence and datasets used in the study, prediction and management of COVID-19. *Appl Intell*. 2020. Available from: <https://doi.org/10.1007/s10489-020-01770-9>
3. Ivorra B, Ferrández MR, Vela-Pérez M, Ramos M. Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the undetected infections. The case of China. *Commun Nonlinear Sc Numer Simul*. 2020 Sep; 88;105303. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7190554/>
4. Giordano G, Blanchini F, Bruno R, Colaneri P, Di Filippo A, Colaneri M. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nature Med*. 2020;26:855–860. DOI:10.1038/s41591-020-0883-7
5. Ndaïrou F, Area I, Nieto JJ, Torres DFM. Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan. *Chaos Solit Fract*. 2020;135:109846. Available from: <https://doi.org/10.1016/j.chaos.2020.109846>
6. Wang H, Xu K, Li Z, Pang K, He H, Improved Epidemic Dynamics Model and Its Prediction for COVID-19 in Italy. *Appl Sci*. 2020;10(14):4930. Available from: <https://doi.org/10.3390/app10144930>
7. Cherniha R, Davydovych V, A Mathematical Model for the COVID-19 Outbreak and Its Applications. *Symm*. 2020;12(6):990. Available from: <https://doi.org/10.3390/sym12060990>
8. Postnikov EB. Estimation of COVID-19 dynamics “on a back-of-envelope”: Does the simplest SIR model provide quantitative parameters and predictions? *Chaos Solit Fract*. 2020;135:10984. Available from: <https://doi.org/10.1016/j.chaos.2020.109841>
9. https://raw.githubusercontent.com/CSSEGISandData/COVID-19/master/csse_covid_19_data/csse_covid_19_time_series/time_series_covid19_confirmed_global.csv ; see also <https://github.com/CSSEGISandData/COVID-19>
10. Ciufolini I, 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*. 2020;135:355. Available from: <https://doi.org/10.1140/epjp/s13360-020-00383-y>
11. Olsson S, Zhang, J. The ongoing COVID-19 epidemic curves indicate initial point spread in China with log-normal distribution of new cases per day with a predictable last date of the outbreak version 4: Predictions for selected European countries, USA and the World as a whole and try to predict the end of the outbreak including a discussion of a possible “new normal”. *Preprints* 2020;2020030077. doi:10.20944/preprints202003.0077.v4

12. an der Heiden M, Hamouda O. Schätzung der aktuellen Entwicklung der SARS-CoV-2-Epidemie in Deutschland – Nowcasting. *Epid Bull.* 2020;17:10-16. Available from: https://www.rki.de/DE/Content/Infekt/EpidBull/Archiv/2020/Ausgaben/17_20.pdf?__blob=publicationFile
13. <https://de.wikipedia.org/wiki/>
14. Robert-Koch-Institut, Berlin. Täglicher Lagebericht des RKI zur Coronavirus-Krankheit-2019 (COVID-19), 27.05.2020. Available from: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/2020-05-27-de.pdf?__blob=publicationFile
15. <http://www.silverfrost.com> (homepage Silverfrost FTN95 system)
16. <https://www.python.org/> (homepage programming language Python3)
17. https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.least_squares.html (description of a modul of the „SciPy library“, a Python-based collection of software for scientific calculations)

Appendix: Mathematical Calculations

Function maximum

The first derivative of equation (1) is:

$$\hat{l}'(t) = \hat{l}(0) (ae^{-bt} - abte^{-bt}) e^{(ae^{-bt})t}$$

From $\hat{l}'(t) = 0$ we obtain for the position $T = t_{\max}$ of the singular curve maximum:

$$ae^{-bT} = abTe^{-bT}$$

and thus: $T = t_{\max} = 1/b$ and $\hat{l}_{\max} = \hat{l}(0) e^{a/be}$

Full width at half maximum

The determination of the position of the half maximum function values with respect to the t axis is possible via (see section Function maximum):

$$0.5 \hat{l}(0) e^{a/be} = \hat{l}(0) e^{(ae^{-bt})t}$$

$$0.5 e^{a/be} = e^{(ae^{-bt})t}$$

$$-\ln 2 + a/be = ate^{-bt}$$

$$\ln(a/be - \ln 2) = \ln a + \ln t - bt$$

$$\ln t - bt - \ln(a/be - \ln 2) + \ln a = 0 \quad (3)$$

Equation (3) normally (i.e., if $\hat{l}(0) < \hat{l}_{\max} / 2$) has two solutions (t_1 and t_2) which can be determined numerically.

Parameter α

$$\phi(t) = \phi(0) e^{\alpha t} \quad (4)$$

For $\phi(t+1)$ we get

$$\phi(t+1) = \phi(0) e^{\alpha t + \alpha} = \phi(0) e^{\alpha t} e^{\alpha}$$

$$\phi(t+1) = \phi(t) e^{\alpha} \quad \text{or} \quad \alpha = \ln[\phi(t+1) / \phi(t)] \quad (5)$$