Reproducing country-wide COVID-19 dynamics can require the usage of a set of SIR systems

Eugene B. Postnikov

Department of Theoretical Physics, Kursk State University, Radishcheva st., 33, 305000 Kursk, Russia

Corresponding author:
E.B. Postnikov

Email address: postnikov@kursksu.ru

ABSTRACT

This work shows that simple compartmental epidemiological models may not reproduce actually reported country-wide statistics since the latter reflects the cumulative amount of infected persons, which in fact is a sum of outbreaks within different patched. It the same time, the multilogistic decomposition of such epidemiological curves reveals components, which are quite close to the solutions of the SIR model in logistic approximations characterised by different sets of parameters including time shifts. This line of reasoning is confirmed by processing data for Spain and Russia in details and, additionally, is illustrated for several other countries.

INTRODUCTION

The COVID-19 pandemic is recently one of the most urgent problems and challenges of public health, which affects almost all countries of the world, see Sohrabi et al. (2020); Wang et al. (2020); He et al. (2020); Gates (2020); Bedford et al. (2020) for a general review. Respectively, this situation states the problem of mathematical modelling of the local and global outbreaks aimed at understanding the epidemic process, revealing factors affecting the course of the outbreak, the effectiveness of applied measures and provided recommendations, which should prevent possible new waves of this disease Cobey (2020).

One of the natural approaches to such modelling is the usage of different variants of compartmental models Brauer et al. (2008), which subdivide the full populations into groups of susceptible S, infected I and recovered/removed R persons (SIR model), which can also be supplied with the group of exposed persons (E), and even with more detailed subdivision. However, it should be pointed out that the expended SEIR model have some practical weak points, when one address the country-wide data. First of all, its solution can be reduced to 2-dimensional system only Weinstein et al. (2020), in contrast to the SIR model, which is has an exact 1-dimensional representation. This results in the necessity of fitting a "hidden" variable E and additional kinetic parameter, which an appropriate change of variables for the SIR model puts its solution in correspondence to the 1-dimensional data reported. At the second, both models are compartmental and, in the strict sense, requires consideration of a well-mixed population. But the compartmental models are still applicable, when an outbreak starts practically synchronous over a wide region. But in this case the summation of inputs from different regions makes the delays realised as the two-step transition $S \to E \to I$ sufficiently less definite than in the case of small well-localised populations, and, in fact, it can be taken into account by using an appropriate effective kinetic coefficient considering the simple $S \to I$ transition.

A number of recent studies Maier and Brockmann (2020); Fanelli and Piazza (2020); Postnikov (2020); Linka et al. (2020); Carcione et al. (2020) demonstrated a reasonable reproduction of the dynamics of the actually reported cases in a variety of countries on the time intervals from the growing stage of the outbreak till a vicinity of its peak as well as the work Dehning et al. (2020) analysed the well-developed outbreak with switching parameters of SIR model. It has been shown even for the case of the simplest SIR model chosen for the reasons mentioned above: at the first, the lack of detailed data on intra-country

46 distributions, movements and contacts in the country-wide reports, which give a strong mean-field picture,
47 makes desirable minimisation of the hidden parameters, and in the second and important, the SIR model
48 provides an opportunity for the direct data-driven approach.

49 It has been shown Postnikov (2020) such model defined by the system of ordinary differential equations

$$\frac{dS}{dt} = -kSI, \tag{1}$$

$$\frac{dI}{dt} = kSI - \tau^{-1}I, \tag{2}$$

$$\frac{dR}{dt} = \tau^{-1}I, \tag{3}$$

where k and τ are the transmission and recovery rates, can be sequentially reduced to one Verhulst's logistic equation with respect to the cumulative number of infected persons

$$\frac{d\Sigma I}{dt} = r\Sigma I \left(1 - \frac{\Sigma I}{K}\right), \tag{4}$$

50 where r is the growth rate, which is directly connected with the basic reproduction number $\mathfrak{R}_0$ via
51 parameters of the differential equations (1)–(3), see the explicit formulae below, in the section "Methods".
52 The parameter K is the saturation value. It depends on both $\mathfrak{R}_0$ and the total number of potentially
53 susceptible people, see "Methods" again; the problem of its proper determination, which can be connected
54 to either spatially or socially separated subpopulations, i.e. there is a difficulty in its estimation *a priori*.
55 Moreover, the cumulative mean-field epidemic curve can be affected by mixing such subpopulations that
56 is the main focus of the present study.

57 At the same time, it should be kept in mind that the logistic function, which is the analytic solution
58 to Eq. (4), is also an approximate solution to the SIR-system of ordinary differential equations (1)–(3)
59 derived on a basis of existing invariants of this ODE system and assumption that $\max I(t) << K$ that fulfils
60 in practice, see Postnikov (2020) for the step-by-step derivation. The analysis carried out for different
61 countries till the first decade of April confirmed quantitative adequacy of this approach for different
62 countries among which were Italy, France, Germany, Spain.

63 However, new data registered to the decaying stage of the outbreak exhibit drastic deviations from
64 such logistic curve. The same behaviour is noted for other compartmental models that induced different
65 attempts to their more complicated modification, e.g. Oliveira (2020); Mamon (2020); Schaback (2020).

66 Note that the preliminary analysis evaluated for the case of Italy Vannucci and Vannucci (2020)
67 demonstrated that reproducibility could be restored if to consider a combination of two logistic curves
68 corresponding to two outbreak waves. The similar combination was applied earlier to the long-time
69 modelling of tuberculosis Lavrova et al. (2017) based on a superposition of SIR models with different
70 (switched with time) control parameters. Note also that a similar behaviour of data curves is already
71 known in the theory of markets' evolution and technological changes known as the logistic substitution
72 model Meyer et al. (1999) and the respective software Loglet Lab 4 is developed to carry out such analysis.
73 It is available online at `https://logletlab.com` and supplied with a variety of examples of such
74 decomposition and details of the numerical algorithms.

75 Thus, the main goal of this work is to explore if it is possible to decompose the data reported for
76 COVID-19 outbreaks at the country-wide scale into a set of curves, each of which satisfies the SIR-
77 Verhulst model. In other words, if it is possible to consider the dynamics of such data as a superposition
78 of outbreaks satisfying the simple compartmental model but operating with different subpopulations and
79 parameters and, respectively, to get a tool for revealing a number of such individual characteristic modes
80 from unique general data.

MATERIALS AND METHODS

82 The project "Our world in Data" aggregating the data related to COVID-19 pandemics from different
83 sources Roser et al. (2020) was used as the data source. These data were downloaded in an automatic
84 regime with a home-written function, which belongs to the developed tool for MATLAB/GNU Octave
85 made available in the GitHub repository
86 `https://github.com/postnicov/owid_loglet_interface` and their subset containing

the cumulative number of infected persons for a chosen country (total cases) was processed to form a text file suitable for uploading to the online software Loglet Lab 4 and carrying out the multilogistic decomposition. Comments on the particular issues of the program realization, processed data structure and the workflow can be found also in the mentioned repository, the full set of the data and the processing code for countries considered in this work is placed there in the "examples" directory under the name "10countries_data_August".

The results of this decomposition report the components of decomposition. Each such component is the logistic curve satisfying the SIR-Verhulst equation (4) and has the form

$$\Sigma I(t) = \frac{K}{1 + e^{-r(t - t_m)}}, \tag{5}$$

where t_m defines the time moment when the cumulative number of infected persons reaches half of the saturated values. It should be pointed out that the growth and saturation parameters r and K are directly connected with the parameters of the SIR model (1)–(3) as $r = (\Re_0 - 1)\tau^{-1}$ and $K = \Re_0^2 (2(\Re_0 - 1)N)^{-1}$, where $\Re_0 = k\tau$ is the basic reproductive number and $N = S + I + R = \text{const}$, see Postnikov (2020) for details.

An additional output is the so-called Fisher-Pry plot Fisher and Pry (1971) for each component. Its utility if based on its construction as the plot of the logarithm of the cumulative registered cases divided by the expected maximum as a function of time. In the case of the solution (5) this representation gives

$$\ln \frac{\Sigma I / K}{1 - \Sigma I / K} = -r(t - t_m), \tag{6}$$

i.e. one can easily check if the data already satisfy the logistic curve looking do they follow a straight line or not. If linearity of the Fisher-Pry plot fulfils, this means that the SIR model provides a sufficient accuracy to reproduce the basic features of the analysed data, if not – one need to change the model.

In addition, the average relative absolute deviation (*AAD*),

$$AAD = 100\,\% \left\langle \frac{\Sigma I_{rec} - \Sigma I_{calc}}{\Sigma I_{rec}} \right\rangle$$

is used in this work as a quantitative criterion. Here ΣI_{rec} and ΣI_{calc} are the actually registered and the model-based calculated values of the cumulative number of infected persons, respectively. The average value is taken over the whole time interval covering values $\Sigma I_{rec} > 0.05 \cdot \max(\Sigma I_{rec})$.

The particular considerations of the cases of Spain and Russia use also more geographically detailed data, which were obtained from from the web-site of the National Epidemiological Center of Spain CNE (2020) and accumulated by the Federal Service for Surveillance on Consumer Rights Protection and Human Wellbeing accessed via Yandex's web-service DataLens (2020), respectively. The data and the program code, which carried out their processing, can be be found in the GitHub repository mentioned above, too. In particular, the illustrated worklow's description for the detailed analysis of data for two cumulative data from two regions of Spain is placed there in the zip-archive named "regionsSpain" as well as the data and the source code used for their processing.

STATEMENT OF THE PARTICULAR PROBLEM

As two principal illustrative examples, which illustrate the failure of reproducing the long-term country-wide dynamics of COVID-19 outbreaks, the cases of two countries, Spain and Russia, can be considered. The work Postnikov (2020) showed that it satisfied the Verhulst-SIR model (4) with high accuracy for the state-of-the- art up to the first decade of April. However, the actual longevity and cumulative effects of outbreaks exceed the predictions made using the data available for the early stages of the process.

Fig. 1 demonstrates the data downloaded from "Our World in Data", uploaded for processing into Loglet Lab and post-processed again with the Octave script (see Materials & Methods). Here the unique logistic curve was applied for the cumulative number of infected persons (total cases), i.e. the "composite plot" is automatically composed of one component.

One can see that the single logistic curve fits the total cases only qualitatively on average, see Fig. 1(A) for Spain, *AAD* = 6.66 % and Fig. 1(B) for Russia, *AAD* = 8.19 %. The quantitative details of approximation do not accurately reproduce the dynamics, black circles go under and over the red
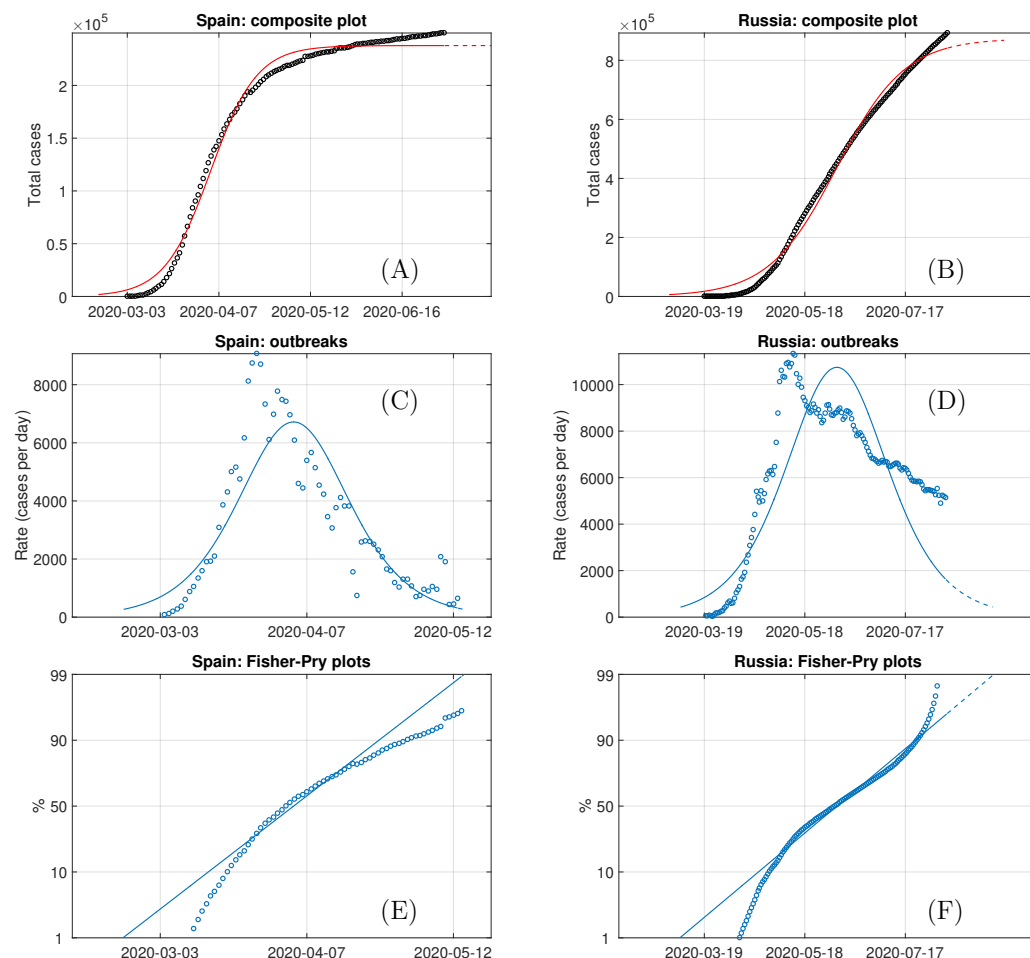
Figure 1. The total cases of persons infected by COVID-19 according to "Our World in Data" for Spain (from March, 2 to July, 1) and Russia on August, 11 (circles) and their logistic fit (solid curves) as well as growth rates and Fisher-Pry plots for these data and their fitting.

curve within relatively long intervals as well as there are significant displacements of recorded and model curves on the bell plots, see Figs. Fig1log(C) and (D), respectively. In addition, this inconsistency is most obvious looking at the Fisher-Pry plots (Figs. Fig1log(E) and (F)): the markers do not follow straight lines as it should be from (6) if the recorded data with a single defined K satisfy the logistic growth law. An attempt to improve the correspondence superposing maxima manually do not really improve the situation since there will emerge "long tails" of recorded data drastically overcoming fast-decaying model rate curve and underestimating the curves of total cases.

These facts argue in favour of the idea that the data accumulated over the whole country may in fact consist of different maxima and different growth rates for different regions, which start to be significant after the middle of April (before this date there was a quite accurate correspondence between he data and the SIR-Verhulst model, see Postnikov (2020)). This hypothesis will be tested considering both multilogistic decomposition and the outbreak's dynamics in spatially separated regions within each country.

RESULTS

As the first example, let us consider the case of Spain since the registered and the model curves in Fig. 1(C) visually differ mainly by the location of their maxima. Fig. 2(A) and (C) represents the examples of regional statistics for two separated regions of Spain, Madrid and Catalonia. One can see that before the
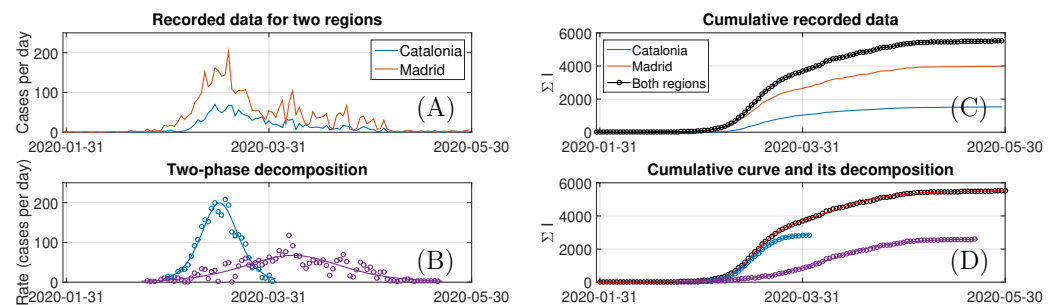
Figure 2. The plots demonstrating daily registered incidences of COVID-19 (A) and their cumulative sums (C) in two valuable spatially separated regions of Spain and the decomposition of the registered data curve as well as the data decomposition into two logistic curves (D) with the growth rates shown in the subpanel (B).

first decade of April both curves had the similar shape, similar location of their maxima and differ in the magnitude only. Therefore, such a synchronization leads to a possibility to represent the summed data as a single curve with some unique set of effective parameters of the SIR-Verhust model. The same is true for the rest of the regions too and, as a results, one system (1)–(3) reduced to one equation (4) with one logistic functions as the solution is completely enough as it has been already demonstrated earlier.

On the contrary, the data for Madrid demonstrates the secondary outbreak between the first decades of April and May, which can not be predicted directly from the data recorded earlier. Moreover, since the SIR system (1)–(3) has an unimodal solution for $I(t)$, it can be hypothesized that the disease spread over a group of susceptible persons, which differ from the initial one.

Therefore, it is natural idea to carry out more detailed analysis, which includes consideration of an additional independent SIR system. As follows from the reducibility of the SIR ODE system to one Verhulst's equation, this means that it is required to consider two logistic curves whose sum should fit the data not fitted by one curve only. This decomposition in practice can be easily carried out by means of Loglet Lab software switching in the mode of several logistic curves (the detailed workflow of this procedure is described within the supplementary GitHub-deposited data, see Materials & Methods section). Applying this procedure to the sum of regional curves (it is shown as black circles in Fig. 2(B)), it revealed that enlarging number of possible curves to two instead of one results in a sufficiently accurate reproducing the analysed data: compare the upper curves in Fig. 2(B) and Fig. 2(D), quantitatively $AAD = 0.86$ % for the two-component model. These logistic components are shown in Fig. 2(D) too, and their derivatives, which have a meaning of the daily registered cases for each of these components are depicted in Fig. 2(C). Considering the latter, it should be pointed out that the multicomponent decomposition algorithm does not distinguish the synchronized curves but is reveals outbreaks, which are are separated in time. But in this case the method reveals the principal features of temporal dynamics: one can see comparing maxima in Fig. 2(C) and curves in Fig. 2(A) that the first revealed component detected the common maximum of outbreaks in both Madrid and Catalonia, and the second – the summer maximum seen in the epidemic curve for Madrid.

Therefore, it is possible to apply this method to data for the whole Spain. The respective result is shown in Figs. 3. Fig. 3 indicates that, in the case of Spain, the assumption of the existence of two logistic components gives sufficiently better reproducing the observed dynamics, $AAD = 1.73$ %. One component corresponds to the more fast and strong outbreak with the peak on March, 27 and the second is slower with the peak on April, 18 that is agreed with the discussion of Fig. 2 above. Both components follow the solution of the Verhulst equation as seen from the respective plot in Fig. 3 and reasonable linearity of partial Fisher-Pry plots there. The last fact confirms the validity of logistic approximation for the both components considered as independent. The full set of parameters for both logistic functions are listed in Table. 1.

The next example, the case of Russia, which is sufficiently more spatially extended country than Spain,
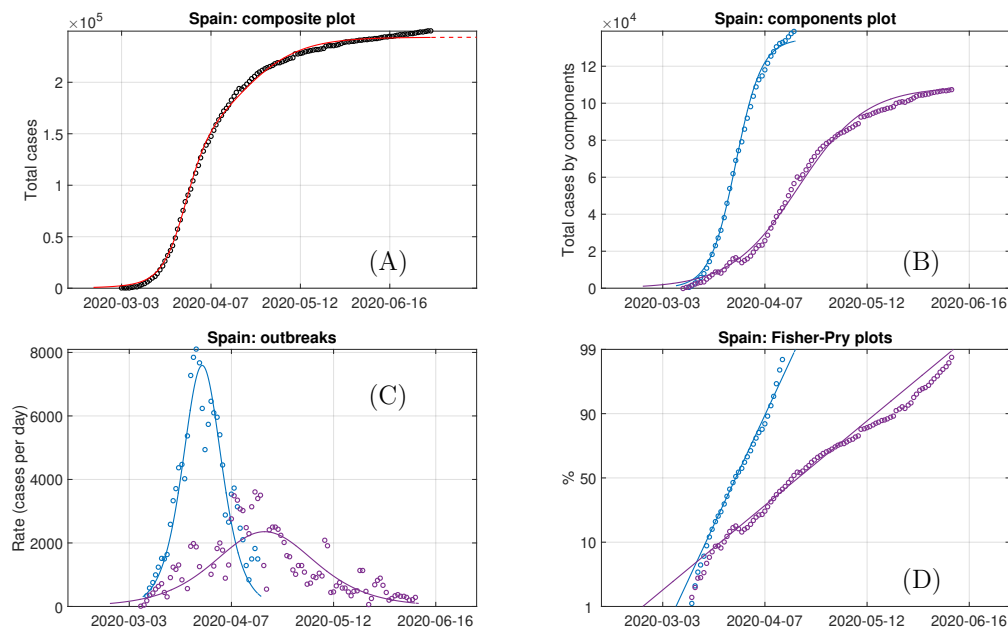
Figure 3. The plots generated after the downloading data from "Our World in Data" for Spain on July, 1, their processing with Loglet Lab and postprocessing. Different components revealed as comprising the total epidemic curve (cumulative number of revealed infected persons) are marked with different colours. The circles are recorded data and the lines are their reconstruction by the solutions of the SIR-based logistic equation (4).

Table 1. The parameters of the separated logistic components (5) for the case of Spain, where r (days^{-1}) is the components growth rate, t_m (days) is the time location of the inflection point for the logistic function, and K (persons) is the logistic curve's maximum. The first day of the time series is March, 2. R^2 denotes the coefficient of determination.

Phase	K	t_m	r	R^2
1	134912	26.0	0.225	0.973
2	108805	47.4	0.087	0.973

requires a larger number of logistic components. Fig. 4 demonstrates this five-modes decomposition valid to the current state of still progressing epidemics in Russia, which reproduces the actual data with a pretty good accuracy, $AAD = 0.334$ %. Each component curve corresponds to the logistic growth process with high accuracy too that is seen from all plot representations, for the components themselves, their growth rates, and the Fisher-Pry decompositions as well as from the high value of the coefficient of determination. Notably, that the growth coefficient values k given in Table 2 can be subdivided into two groups. At the first stage (from the middle of March till the middle of May), there are two subsequent component with large k, and then (since the middle of May up to date) three components with practically equal k again time-shifted. This is also demonstrably visible in the plot representing the Fisher-Pry transform. This division can be put in line with the governmental measures: the period March, 30 – May, 11 was non-working nationwide, the public places lockdown and strict travel restrictions were applied. Notably, this period parenthetically coincides with the period of existence of the major part of the first two components in Fig. 4. Therefore, it can be concluded that the measures applied localised the number of potentially susceptible persons by several spheres of close contacts of infected persons and led to the fast reaching the saturated values for these two first logistic components. Certainly, the contacts were not broken completely and the outbreak continued to spread countrywide but with sufficiently slower growth rate even over the background of more mild restrictions. The three subsequent logistic components can be
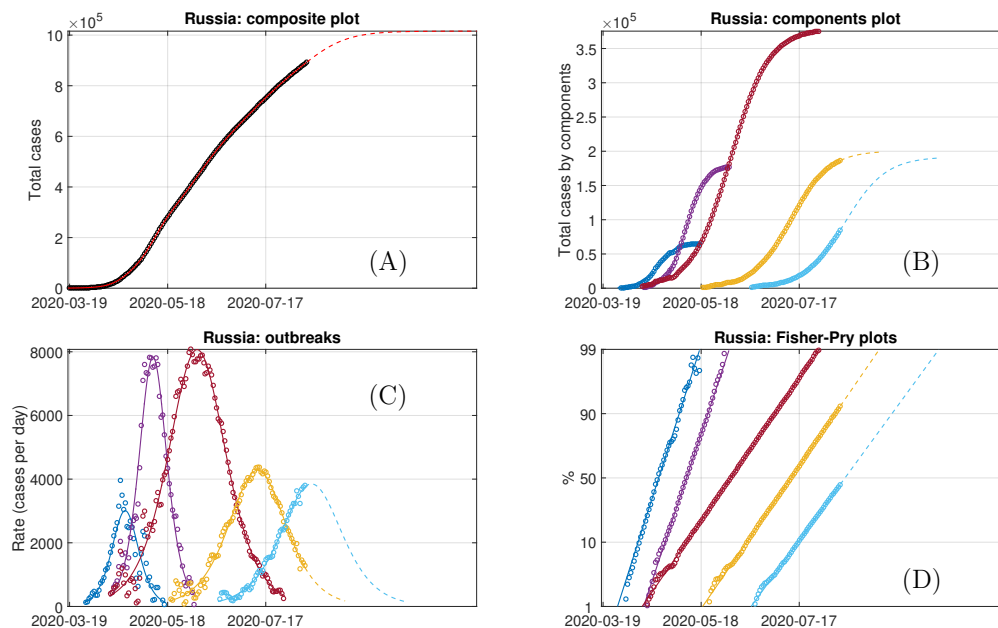
Figure 4. The plots generated after the downloading data from "Our World in Data" for Russian on August, 11, their processing with Loglet Lab and postprocessing. Different components revealed as comprising the total epidemic curve (cumulative number of revealed infected persons) are marked with different colours. The circles are recorded data and the lines are their reconstruction by the solutions of the SIR-based logistic equation (4). Dashed lines correspond to the model-based projections.

associated with such a spatial spread.

Table 2. The parameters of the separated logistic components (5) for the case of Russia. The units of r, t_m, K are days, days^{-1}, persons, respectively. The first day of the time series is March, 18. R^2 denotes the coefficient of determination.

Phase	K	t_m	r	R^2
1	65744	35.0	0.183	0.979
2	178111	52.0	0.176	0.996
3	379133	79.0	0.085	0.999
4	200709	116	0.085	0.997
5	192037	149	0.080	0.997

As above for Spain, the origin of such a behaviour for Russia can be explained exploring the geographically subdivision as it is shown in Fig. 5 using the regional data for Russia accumulated by the Federal Service for Surveillance on Consumer Rights Protection and Human Wellbeing accessed via Yandex's web-service DataLens (2020). Moscow as the capital, the most populated city, and the largest transport hub is considered separately along with the European (without Moscow) and Asian (including the borderline Ural region) parts of country. From epidemiological point of view, one can see that Fig. 5 demonstrates that such a subdivision provides comparable amount of infected persons. Thus, it is possible to put in correspondence features in Figs. 4 and 5. First of all, it is visible that the outbreak in Moscow is the leading one during the first stage of the time course. In addition, the respective curve is bimodal. Therefore, the April's peak can be associated with the primary introduced virus spreaders (Moscow is a huge transport hub); this peak's behaviour also corresponds to the model results provided in Postnikov (2020), where only the data up to the first decade of April were used. Further, the spreaders simultaneously distributed the virus over Moscow (secondary peak of the red curve in Fig. 5) and over the country with a prevalence to regions closer to Moscow (the European part, blue curve in in Fig. 5). Note, however, that
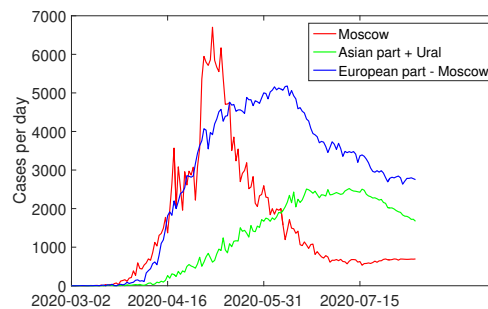
Figure 5. The plots demonstrating daily registered incidences of COVID-19 in Russia subdivided accordingly to principal socio-geographical division.

this region consists of multiple administrative units and, whence, the blue blue curve has not so accurate bell-like shape as one city, Moscow. The third (green) curve corresponds to more distant from Moscow and more weakly connected Ural, Siberian, and Far East part of Russia, and it demonstrates the larger time delay in the outbreaks' development. At the same time, more-or-less uniform restrictive measures "synchronize" the parameters of SIR model even for distinct weakly or non-connected locations whereby a sum of almost independent data provides an opportunity to use the mean-field description resulting in the accurate sequence of logistic representations for the countrywide data shown in Fig. 4. At the same time, it should be kept in mind that the parameters K from Table 2 are not directly associated with particular geographic regions within such a multiple-SIR but mean-field approach, they represent a characteristic mixture of potentially susceptible groups distributed between different regions but simultaneously actual for the particular stage of the epidemic spread.

Finally, Fig. 6 provides demonstrations additional to the considered above and covers the most affected countries during the spring–summer outbreak. One can see (not only visually but also basing on the values of the average relative absolute deviations) that the proposed decomposition reproduces the specific features of the cumulative number of cases with high accuracy and the multilogistic representation (i.e. superposition of several SIR systems associated with different sub-communities) is a widespread phenomenon. A single curve corresponds in the cases of Columbia (Fig. 6(C)) and India (Fig. 6(D)) to the situation, when the outbreak has not reach the peak yet (or, equivalently, the middle level with respect to the saturation estimated from the exiting data fitting). However, it should be pointed out that the forecast of the maximum total number of infected in such a case has a high level of uncertainty, see numerical tests in Alberti and Faranda (2020); Postnikov (2020) and an additional analysis of sensibility of sigmoid, including logistic, curves to perturbations in Faranda et al. (2020). This caveat is possibly true also for the second stage of the two-component representation of data for Peru (Fig. 6(F)).

Among the rest countries considered, where the epidemic curve looks overcoming the middle point, three of six demonstrate the multilogistic character: two components for Brazil (Fig. 6(A)) and Mexico (Fig. 6(E)), and three for the USA (Fig. 6(H)). The latter is more spatially spread and, additionally, social turbulence of last months may influence the parameters of the third component there.

DISCUSSION

The examples given above reveals that the simple logistic regression with a unique fitting function corresponding to the solution of the SIR model can not give a fully adequate picture of the outbreak in its full time course even if it was accurately applicable during almost complete initial growth phase, see Postnikov (2020) for the same countries. The main reasons for this conclusion are the following: i) the epidemics spreads over spatially distributed countries while the reported data curves gives summed values, i.e. and the validity level of a mean-field model depends on a possible contrast in the course of developed epidemics in different part of the country (more "compact" countries are more uniform in the response that more "elongated" as seem from the comparison of Spain and Russia above and other examples discussed below); ii) effects of counter-epidemic measures can change the outbreak size and prolong its continuity that is reflected in a less value of the parameter k in Eqs. (1)–(2), or, equivalently
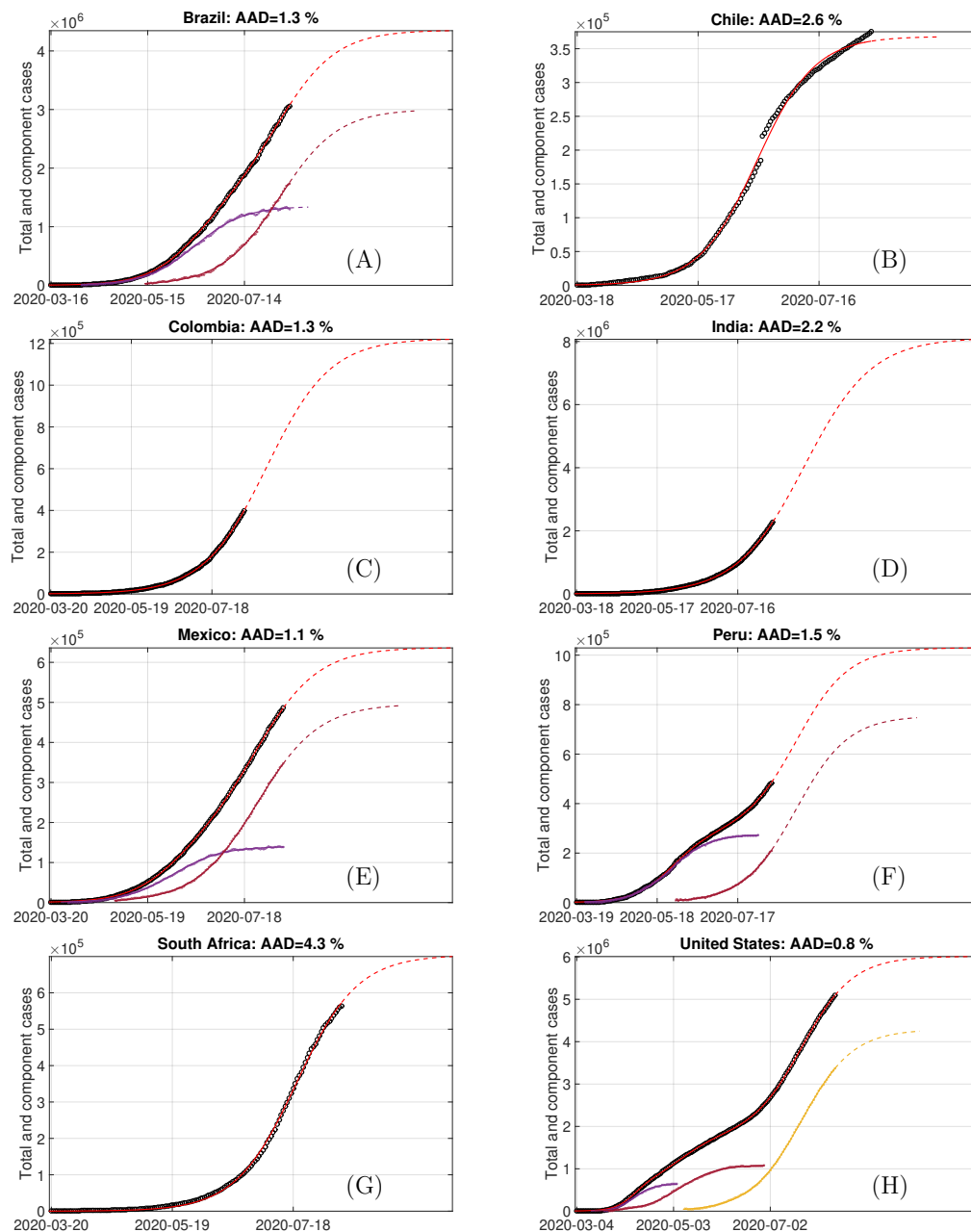
Figure 6. The plots demonstrating the cumulative numbers of cases (circles), the respective solutions of logistic and multilogistic models (red solid line, the dashed part corresponds to the projection based on the simple continuation with the present state parameters) and their components (if any; dots denotes the sorted data and curves are the simulated continuous solutions) for eight countries among top-10 most affected ones. The parameters for all models/components are given in supplementary material on GitHub project (see "Methods") as well as the data and code for their processing.

248 Eq. (4) that is also visible in Tables 1 and 2: the values of k for the second Spanish component and the
249 third-fifth Russian components are practically the same (the governmental restrictive measures were more
250 or less similar in the both countries).

251   Note also that the proposed method indicates time-shifted maxima, while synchronous outbreaks are
252 identified as a single one. This situation can emerge when initial conditions for the outbreak's start fulfil
253 in geographically different sub-communities (a widespread initial introduction of a new type of infections
254 agent is a plausible reason for such a behaviour). In addition, this observation explains why compartment
255 models developed for well-mixed populations may reproduce epidemics in spatially-extended regions.

256   Thus, these complications do not mean that the considered disease does not always follow the classic
257 well-established SIR model of mathematical epidemiology (and its logistic-form) and the mean-field
258 consideration, which addresses the country-wide data. The analysis carried out in the present work
259 revealed that a more realistic data regression requires the decomposition over multiple logistic curves.
260 Each of such curve corresponds to its own SIR system with its own set of parameters. Among these
261 parameters, the possible maximal number of susceptible person within weakly connected communities
262 (on even approximated as uncoupled ones for simplicity of fitting) is one of the most important. Such
263 situation can have an interpretation in terms of metapopulation approach Colizza and Vespignani (2008);
264 Bichara et al. (2015); Ball et al. (2015); Chowell et al. (2016). In fact, the country-wide statistics unifies
265 data from different regions and even from different social strata within each region. Due to differences
266 and different connectivity between such sub-populations, the local outbreaks can be shifted in time and
267 intensity. This problem especially arise under conditions of different measures of travel restrictions and
268 social distancing applied practically in all countries during COVID-19 pandemics.

CONCLUSION

270 The main message of this work is aimed at attracting attention to the importance of multimodal considera-
271 tion of epidemic curves given at the macroscopic (country-wide) level and developing a method, which
272 allows revealing time intervals of qualitative changes in dynamics as well as their quantitative estimations.

273   The complicated non-standard shape of epidemic curves not necessarily originates from some special
274 features of the current pandemics providing a challenge in developing new very complicated multicom-
275 ponent models, which contain a large number of parameters inaccessible directly. In fact, when one
276 addresses the global mean-field data on the "macroscopic" scale of a country, a set of simple SIR model
277 is plausible. In its turn, the number and temporal localisation of the revealed principal modes provide
278 perspectives for building more detailed "microscopic" models, which should take into account spatial,
279 social, etc. features, i.e. to reflect phenomena originated from the metapopulational dynamics. Note that
280 the summation of individual growth curves attracted attention as an empiric approach from the classical
281 problems of pure population dynamics of heterogeneous populations Reed and Pearl (1927) up to the
282 mathematical problems of epidemic spread in metapopulations Chowell et al. (2016). In this work, it is
283 shown that the effective logistic curve-based approach could play a role of an accurate tool for studying
284 the recent COVID-19 pandemics in a tight connection with the classic SIR model approach. In contrast to
285 the direct simulation of ODE's including the case of parameter switching in *a priory* known time points
286 (e.g. due to lockdown measures, see Dehning et al. (2020)), the proposed method solves the inverse
287 problem revealing actual parameters and inflection points originated from the actual data.

288   Finally, this opens certain perspectives for future investigations in modelling by analogy to other
289 field of science, which use dynamical systems as a principal tool. As some examples, the dimension
290 reduction to several principal modes is used for simplified modelling of turbulence Sapsis and Majda
291 (2013), distributed elastic mechanical systems Kerschen and Golinval (2002), chemical processes Qing
292 et al. (2020), etc. Thus, the revealed possibility of describing the cumulative dynamics of the epidemics
293 with a finite number of components calls for development of spatio-temporal models whose full-scale
294 dynamics is reducible to the dynamics of several principal modes only reflected in the generalized data.

ACKNOWLEDGEMENTS

296 I thank Dr. Irina L. Zhirnova for the motivational discussions of the epidemic growth curves dynamics
297 and composition.

REFERENCES

Alberti, T. and Faranda, D. (2020). On the uncertainty of real-time predictions of epidemic growths: a COVID-19 case study for China and Italy. *Communications in Nonlinear Science and Numerical Simulation*, page 105372.

Ball, F., Britton, T., House, T., Isham, V., Mollison, D., Pellis, L., and Tomba, G. S. (2015). Seven challenges for metapopulation models of epidemics, including households models. *Epidemics*, 10:63–67.

Bedford, J., Enria, D., Giesecke, J., Heymann, D. L., Ihekweazu, C., Kobinger, G., Lane, H. C., Memish, Z., Oh, M.-d., Schuchat, A., Ungchusak, K., and Wieler, L. H. (2020). COVID-19: towards controlling of a pandemic. *Lancet*, 395(10229):1015–1018.

Bichara, D., Kang, Y., Castillo-Chavez, C., Horan, R., and Perrings, C. (2015). SIS and SIR epidemic models under virtual dispersal. *Bulletin of Mathematical Biology*, 77:2004–2034.

Brauer, F., van den Driesche, P. V., and Wu, J., editors (2008). *Mathematical Epidemiology*. Springer, Berlin.

Carcione, J. M., Santos, J. E., Bagaini, C., and Ba, J. (2020). A simulation of a covid-19 epidemic based on a deterministic seir model. *arXiv:2004.03575*.

Chowell, G., Sattenspiel, L., Bansal, S., and Viboud, C. (2016). Mathematical models to characterize early epidemic growth: A review. *Physics of Life Reviews*, 18:66–97.

CNE (2020). https://cnecovid.isciii.es/covid19/.

Cobey, S. (2020). Modeling infectious disease dynamics. *Science*, 368(6492):713–714.

Colizza, V. and Vespignani, A. (2008). Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations. *Journal of Theoretical Biology*, 251:450–467.

DataLens, Y. (2020). https://datalens.yandex/7o7is1q6ikh23.

Dehning, J., Zierenberg, J., Spitzner, F. P., Wibral, M., Neto, J. P., Wilczek, M., and Priesemann, V. (2020). Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science*, 369:eabb9789.

Fanelli, D. and Piazza, F. (2020). Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos, Solitons & Fractals*, 134:109761.

Faranda, D., Castillo, I. P., Hulme, O., Jezequel, A., Lamb, J. S., Sato, Y., and Thompson, E. L. (2020). Asymptotic estimates of SARS-CoV-2 infection counts and their sensitivity to stochastic perturbation. *Chaos*, 30:051107.

Fisher, J. C. and Pry, R. H. (1971). A simple substitution model of technological change. *Technological forecasting and social change*, 3:75–88.

Gates, B. (2020). Responding to Covid-19 – a once-in-a-century pandemic? *New England Journal of Medicine*, 382(18):1677–1679.

He, F., Deng, Y., and Li, W. (2020). Coronavirus Disease 2019 (COVID-19): What we know? *Journal of Medical Virology*, 76(71-76).

Kerschen, G. and Golinval, J.-C. (2002). Non-linear generalization of principal component analysis: from a global to a local approach. *Journal of Sound and Vibration*, 254:867–876.

Lavrova, A. I., Postnikov, E. B., Manicheva, O. A., and Vishnevsky, B. I. (2017). Bi-logistic model for disease dynamics caused by Mycobacterium tuberculosis in Russia. *Royal Society open science*, 4:171033.

Linka, K., Peirlinck, M., Sahli Costabal, F., and Kuhl, E. (2020). Outbreak dynamics of covid-19 in europe and the effect of travel restrictions. *Computer Methods in Biomechanics and Biomedical Engineering*, pages 1–8.

Maier, B. F. and Brockmann, D. (2020). Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. *Science*, 368:742–746.

Mamon, G. A. (2020). Fit of french covid-19 hospital data with different evolutionary models: regional measures of r0 before and during lockdown. *arXiv:2005.06552*.

Meyer, P. S., Yung, J. W., and Ausubel, J. H. (1999). A primer on logistic growth and substitution: the mathematics of the Loglet Lab software. *Technological forecasting and social change*, 61:247–271.

Oliveira, G. (2020). Refined compartmental models, asymptomatic carriers and covid-19. *arXiv:2004.14780*.

Postnikov, E. B. (2020). Estimation of COVID-19 dynamics "on a back-of-envelope": Does the simplest SIR model provide quantitative parameters and predictions? *Chaos, Solitons & Fractals*, 135:109841.

353 Qing, X., Jin, J., Niu, Y., and Zhao, S. (2020). Time–space coupled learning method for model reduction
354   of distributed parameter systems with encoder-decoder and RNN. *AIChE Journal*, page e16251.
355 Reed, L. J. and Pearl, R. (1927). On the summation of logistic curves. *Journal of the Royal Statistical
356   Society*, 90:729–746.
357 Roser, M., Ritchie, H., Ortiz-Ospina, E., and Hasell, J. (2020). Coronavirus Pandemic (COVID-19). *Our
358   World in Data*. https://ourworldindata.org/coronavirus.
359 Sapsis, T. P. and Majda, A. J. (2013). Statistically accurate low-order models for uncertainty quantification
360   in turbulent dynamical systems. *Proceedings of the National Academy of Sciences USA*, 110:13705–
361   13710.
362 Schaback, R. (2020). On covid-19 modelling. *arXiv:2005.07004*.
363 Sohrabi, C., Alsafi, Z., O'Neill, N., Khan, M., Kerwan, A., Al-Jabir, A., Iosifidis, C., and Agha, R.
364   (2020). World Health Organization declares global emergency: A review of the 2019 novel coronavirus
365   (COVID-19). *International Journal of Surgery*, 76:71–76.
366 Vannucci, V. and Vannucci, L. (2020). Estimation of the ending time of Covid-19 emergence in Italy: a
367   2-waves model as sum of logistics. (doi:10.13140/RG.2.2.26675.58403).
368 Wang, L.-s., Wang, Y.-r., Ye, D.-w., and Liu, Q.-q. (2020). A review of the 2019 Novel Coronavirus
369   (COVID-19) based on current evidence. *International Journal of Antimicrobial Agents*, page 105948.
370 Weinstein, S. J., Holland, M. S., Rogers, K. E., and Barlow, N. S. (2020). Analytic solution of the SEIR
371   epidemic model via asymptotic approximant. *Physica D.*, 411:132633.