



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



9th International Young Scientist Conference on Computational Science (YSC 2020)

# COVID-19 dynamic model: balanced identification of general biological and country specific features

A.V.Sokolov<sup>a\*</sup>, L.A.Sokolova<sup>b</sup>

<sup>a</sup> Institute for information transmission problem (Kharkevitch Institute) RAS, Bolshoy Karetny per. 19, build.1, Moscow 127051 Russia

<sup>b</sup> Federal Research Center "Computer Science and Control" of RAS Institute for Systems Analysis, pr. 60-letiya Oktyabrya 9, Moscow 117312 Russia

---

## Abstract

Typical tasks of scientific research include breaking down a complex phenomenon into its components, considering the processes that determine its dynamics, formalizing the accepted hypotheses in mathematical equations, selecting appropriate experimental and statistical material, and ultimately, constructing a mathematical model. This paper explores a complex bio-social phenomenon (COVID-19 epidemic) using a specific data processing method (balanced identification) as part of data driven modeling approach. Combined with appropriate information technology, the method made it possible to consider a number of models, describe the general biological laws of the virus vs. human interaction (common to all populations), and the country specific social epidemic management in the populations under consideration. As statistical data, only new cases were used. Data from different countries was taken from official sources and processed in a uniform way.

© 2020 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0>)

Peer-review under responsibility of the scientific committee of the 9th International Young Scientist Conference on Computational Science

*Keywords:* mathematical modeling, COVID-19, balanced identification, bio-social system, data-driven modeling

---

## 1. Introduction

The dynamics of epidemics (including COVID-19) is determined both by biological characteristics of the human body vs. virus interaction, and by social aspects of the interaction of man and society.

---

\* Corresponding author. Tel.: +7-903-295-2433;  
e-mail address: [alexander.v.sokolov@gmail.com](mailto:alexander.v.sokolov@gmail.com)

Biological characteristics determine:

- “potential” number of people infected by one person,
- the probability of recovery (without detection and isolation),
- “visibility” (manifestation) of symptoms (for identification and subsequent isolation).

Society can manage an epidemic in three ways:

- limit contacts (self-isolate and quarantine those who had contacts with infected) or make them safer (distancing, masks, etc.),
- identify and isolate those infected,
- vaccination and/or prophylactic of population (out of scope of the paper).

This article forgoes special epidemiological models [1] that imply specific (not always obvious) parameterization of functions. Instead the Data Driven Modeling approach [2] is applied: choosing from a rather wide parametric family a model that describes a (statistical) dataset, satisfies a small set of fairly general assumptions about model (in terms of state variables and functions) and minimizes special error estimate.

As a model family the authors use population-based models of the virus spreading in human population (such as “host-parasite” or “predator-prey”). Particular attention is paid to the separation of the biological processes of interaction between the predator (virus) and the prey (human) and the social mechanisms of society counteracting an epidemic. In the (dynamic) model used, the internal structure corresponds to modern ideas about the biological nature of the object, and external controls reflect the specifics of anti-epidemic measures in different countries.

The epidemic dynamics is determined by *undetected infected* (UDI) people – those identified are isolated (more or less carefully) by the society, and those not identified continue to “multiply”, i.e. infect others. So, only the population of UDI (overlooked by daily new cases statistics) is considered.

Epidemics spreading speed is largely defined by the contagiousness of the infected, which substantially depends on the duration of infection. The typical time from infection to the moment when the host becomes contagious (latent period) is about 4-5 days [3] for COVID-19 (the value not used in the model). Therefore, to describe the dynamics of UDI population it should be divided into groups according to the time elapsed since the infection, which we will call the *duration of infection* (DI). Models of this type have long been known ([4], [5]) and are widely used in demography, ecology and epidemiology.

A single type of statistical information is used to build (identify) the models – the standard official data on the number of new (detected per day) cases of infection. Data is used “as is” without pre-processing. Countries with abnormal data (for example, China due to an abrupt change in new cases counting method in February) are ignored. The number of countries present (7 populations) was determined by visual restrictions of showing the simulation results in one figure. Meanwhile the technology of balanced identification allows to increase the number several times. The constructed model considers the populations of Great Britain (Gbr), Germany (Deu), Italy (Ita), Spain (Esp), France (Fra), Russia excluding the city of Moscow and the Moscow Region (Rus-) and the city of Moscow including the Moscow Region (Mos+). The division of Russia into two parts is caused by the significant difference in the dynamics.

The model under consideration was selected as the one that best fits the volume and quality of the selected statistics. The balanced identification method used [6], [7] allowed to quantify how much the accepted set of hypotheses about the functioning of an object (bio-social system) corresponds to the available factual material (statistics).

We consider the terms “infected” and the “diseased” as synonyms that define contagious or soon to be contagious virus carrier.

## 2. Statistics used

The model is based on statistics on new cases (of infection) as per:

- UK, Germany, Italy, Spain and France <https://ourworldindata.org/coronavirus-source-data>
- Russia and Moscow [https://www.rospotrebnadzor.ru/about/info/news\\_time/](https://www.rospotrebnadzor.ru/about/info/news_time/)

The statistics are normalized by time – for each country, the last date with the number of infected less than 100 was adopted as 0: for Great Britain the analyses starts on 2020.03.05, for Germany – 2020.02.29, for Italy –

2020.02.23, for Spain – 2020.03.02, for France – 2020.02.29, for Russia (without Moscow and Moscow Region) and for Moscow with the region – 2020.03.20.

The entire array of statistics (dataset) is denoted as

$$DS: \{ nC_{t,p}, t \in [0, T_p]; p \in [0, P], P=6 \},$$

where  $nC_{t,p}$  is the (daily) number of new cases of infection at time  $t$  in a population with index  $p$ .

The datasets for Russia (without Moscow and Moscow Region) and for Moscow and the region are illustrated in Fig. 6.

### 3. Dynamics model of undetected infected distributed by duration of infection

This section discusses a (single) population model. First, a verbal description of the hypotheses underlying the model is given. Then, hypotheses formalization is considered in the form of a finite difference (matrix) model and in the form of a continuous model (partial differential equations with an integral boundary condition). The formal model description in a special SvF notation, which is used to translate the problem of identifying unknown model parameters in balanced identification technology, can be found in the file [COVID-19-mng.txt](#). To understand the modeling results given in the following sections, the reader may restrict themselves to a finite-difference model or even just a verbal description of the accepted hypotheses.

#### 3.1. Hypotheses

The hypotheses below are merely stated. For discussion see section 6.

- (1) **The virus does not change.** There are no mutations. The functions that determine the interaction of the virus with the human body do not change over time.
- (2) **Human body does not change over time or from population to population.** Immunity is not taken into account. There are no vaccinations and prophylactic remedies.
- (3) **Identified infected do not infect the healthy.** Quarantine and isolation work: the quantity of infected by identified infected (at the time of infection) is significantly less than that of infected by the unidentified infected.
- (4) **There is no migration.** The populations are considered as isolated. It is assumed that infection is introduced once at the initial moment and migration is minute compared with the number of secondary infected.
- (5) **The time step of the model.** A time step of 1 day corresponds to the data update rate.
- (6) **Maximum DI.**  $\tau_{max} = 16$  days. We assume that any UDI stops being contagious after 16 days. It corresponds to the quarantine duration (2 weeks).
- (7) **Population.** A population of UDI is considered. The individuals can differ from each other (at time  $t$ ) by a single parameter – the time elapsed since the infection (DI),  $\tau$ . Sex and age structure, location, liability to disease, immunity etc. are not considered. The number of UDI at time  $t$  with a DI value of  $\tau$  is denoted by  $n(t, \tau)$  and will be called the *UDI density*.
- (8) **Recovery.** The recovery of UDI (and the cessation of the virus spread) is one of the causes for the UDI population decline. The number of recoveries equals  $\mu(\tau) * n_p(t, \tau)$ , where  $\mu(\tau) \leq 1$  is the probability of recovery for (DI)  $\tau$ . In the initial period of the disease, the probability of recovery is 0.
- (9) **Case detection.** Another reason for the UDI population decrease is case detection and isolation. Case detection depends on the virus visibility with DI equal to  $\tau$  and on the society's efforts to identify the infected people at the moment  $t$ . It is assumed that the dependence can be described by the multiplicative function  $F(t) * f(\tau) * n(t, \tau)$ , where  $F(t) * f(\tau) \leq 1$  is the relative detectability (manifestation) of the virus (a common biological factor, depending only on the DI), and  $F(t)$  is the detection index – the social management of detection (a factor that differs from country to country and varies over time). In the initial period of the disease, the probability of detection is 0. The efforts of society to combat the epidemic are increasing every day – the detection index  $F(t)$  is assumed to be a nondecreasing function.
- (10) **Infectiousness** (addition to the UDI population). The infectiousness of UDI is determined by the duration of infection  $\tau$  and the society's efforts to reduce the possibility of infection transmission (self-isolation, distance, masks, sanitary measures, etc.) at time  $t$ . It is assumed that the dependence is described by a multiplicative function, so one UDI generates (infects)  $I(t) * b(\tau)$  of new UDI, where  $b(\tau)$  is the relative infectivity (contagiousness,

a common biological factor that depends only on the DI), and  $I(t)$ , the infectivity index, is the social management of infection (a factor that differs from country to country and varies over time). At the beginning and at the end of the disease, the relative UDI infectiousness is 0.

### 3.2. Finite-difference (matrix) model

Let’s formalize the above hypotheses. We use the Leslie matrix model [4], [5], that describes the UDI population dynamics (with 1 day step), divided into groups by DI (with 1 day step).

$$n(t + 1, \tau + 1) = n(t, \tau) - \mu(\tau) n(t, \tau) - F(t) f(\tau) n(t, \tau); \tag{1}$$

$$\mu(0)=0; \tag{2}$$

$$f(0) = 0; F(t + 1) \geq F(t); \sum_{\tau=0}^{16} f(\tau) = 1; \tag{3}$$

$$n(t + 1, 0) = I(t) \sum_{\tau'=0}^{16} b(\tau')n(t, \tau'); \quad b(0) = 0; b(1) = 0; b(16) = 0; \sum_{\tau'=0}^{16} b(\tau') = 1; \tag{4}$$

$$n(0, \tau)=n_0(\tau); \tag{5}$$

where  $n(t, \tau)$  is the number of UDI at time  $t$  with DI equal  $\tau$  (more accurately from  $\tau$  to  $\tau+1$ ),  $\mu(\tau)$  is the probability of recovery for the UDI with DI  $\tau$ ,  $F(t)$  is the detection index,  $f(\tau)$  is the relative detectability,  $I(t)$  is the infectivity index,  $b(\tau)$  is the relative infectivity,  $n_0(\tau)$  is the initial condition that defines the number of UDI at time 0.

Equation (1) describes the dynamics of UDI by “DI parameter shift”: the number of UDI  $n(t, \tau)$  at the next moment  $t+1$  is shifted by 1 (to  $\tau+1$ ) and decreases due to recovery and detection. For example, the number of people with a DI equal to 7 days on May 24 equals the number with a DI of 6 days on the previous day, May 23, minus recovered and identified from May 23 to May 24.

For the unambiguity of the multiplicative presentation of detectability and contagiousness, one of the factors is normalized: the sums at the end of the sets of formulas 3 and 4 determine the normalization of the functions  $f(\tau)$  and  $b(\tau)$ . Equation (4) determines the “birth” of new ( $\tau=0$ ) UDI as a result of weighted (by infectivity  $I(t)*b(\tau)$ ) summation of UDI at the previous time. Finally, equation (5) defines the initial condition – the distribution of UDI over DI at time 0.

### 3.3. Continuous model

A verbal description of the hypotheses can also be formalized in the form of a mathematical model of the dynamics of the UDI with continuous time and a continuous distribution over the DI. For this, the McKendrick Von Foerster model [8] is used.

$$\frac{\partial n}{\partial t} + \frac{\partial n}{\partial \tau} = -\mu(\tau)n(t, \tau) - F(t)f(\tau)n(t, \tau); \tag{6}$$

$$\mu(0) = 0; \tag{7}$$

$$f(0) = 0; \frac{dF}{dt} \geq 0; \int_0^{\tau_{max}} f(\tau') d\tau' = 1; \tag{8}$$

$$n(t, 0) = I(t) \int_0^{\tau_{max}} b(\tau')n(t, \tau') d\tau'; \quad b(\tau) = 0, \tau \in [0,2]; \quad b(16) = 0; \int_0^{\tau_{max}} b(\tau') d\tau' = 1; \tag{9}$$

$$n_p(0, \tau) = n_{0p}(\tau); \tag{10}$$

In the case of matching boundary (9) and initial (10) conditions and the continuity of the functions included in the statement, the solution  $n(t, \tau)$  exists and is unique [9].

### 3.4. Summing up

The accepted hypotheses and their mathematical formalization determine the dynamic model, which for given general biological functions  $b(\tau)$ ,  $\mu(\tau)$ ,  $f(\tau)$ , (country-specific) social managements  $I(t)$ ,  $F(t)$  and initial distribution of the UDI  $n_p(0, \tau)$  allows to calculate  $n(t, \tau)$ , the temporal dynamics of the UDI distributed over the DI.

To use the model formally designated as  $n(t, \tau) = M(b(\tau), \mu(\tau), f(\tau), I(t), F(t), n_0(\tau))$  we need to identify  $b(\tau)$ ,  $\mu(\tau)$ ,  $f(\tau)$ ,  $I(t)$ ,  $F(t)$ ,  $n_0(\tau)$  on the base of statistic data.

### 4. Balanced identification

The identification problem consists in finding unknown common biological functions  $b(\tau)$ ,  $\mu(\tau)$ ,  $f(\tau)$  and the population specific functions  $F_p(t)$ ,  $I_p(t)$ ,  $p \in [0, P]$ ,  $P=6$ , so that the model new cases satisfactorily describe the new cases from the DI.

Let us use the method of balanced identification [6], [7]. We define the procedure for selecting the entire set of unknown functions by minimizing the functional depending on the vector of regularization parameters  $\alpha$

$$K(\bar{\alpha}) = \min_{b, f, \mu, I, F, n_0} \left( \sum_{p=0}^P \sum_{t \in T_p} \left( nC_{t,p} - F_p(t) \sum_{\tau'=0}^{16} f(\tau') n_p(t, \tau') \right)^2 \right. \\ \left. + \alpha_1 \cdot Curv(b) + \alpha_2 \cdot Curv(f) + \alpha_3 \cdot Curv(\mu) + \sum_{p=0}^P \alpha_4 \cdot Curv(I_p) + \sum_{p=0}^P \alpha_5 \cdot Curv(F_p) + \alpha_6 \cdot \sum_{p=0}^P \sum_{t \in T_p} N_p(t) \right),$$

where  $n_p(t, \tau) = M(b(\tau), \mu(\tau), f(\tau), I_p(t), F_p(t), n_{0p}(\tau))$  are model solutions,  $nC_p(t) = F_p(t) \sum_{\tau'=0}^{16} f(\tau') n_p(t, \tau') -$  model new cases,  $Curv(y(x)) = \int_a^b (y''(x))^2 dx -$  curvature of the  $y(x)$  function,  $N_p(t) = \sum_{\tau'=0}^{16} n_p(t, \tau') -$  total number of UDI.

The first term of the functional is a measure of the model's fit to measurements (standard deviation), addends 2-6 are the regularization additives, the measure of complexity of the model, which is the curvature of the functions  $b(\tau)$ ,  $\mu(\tau)$ ,  $f(\tau)$ ,  $F_p(t)$  и  $I_p(t)$ . Finally, the last addend minimizes the total number of UDI, so that the solution found is as small as possible (lower bound). This is explained by the fact that the measured variables of the dynamic system  $nC_p(t)$  allow us to determine the solutions  $n_p(t, \tau)$  up to a factor.

The regularization parameters  $\alpha$  are selected by minimizing the mean square error of cross-validation [6], [7].

### 5. Modeling results

As a result of model identification based on new cases data, the general biological functions ( $b(\tau)$ ,  $\mu(\tau)$ ,  $f(\tau)$ ) and social controls ( $F_p(t)$ ,  $I_p(t)$ ,  $p \in [0, P]$ ,  $P=6$ ) for each of 7 populations were determined. In addition, solutions ( $n_p(t, \tau)$ ) and a number of additional indicators convenient for analysis of the results were calculated (specifically for Moscow and Russia).

Common biological functions of the model are shown in Fig. 1

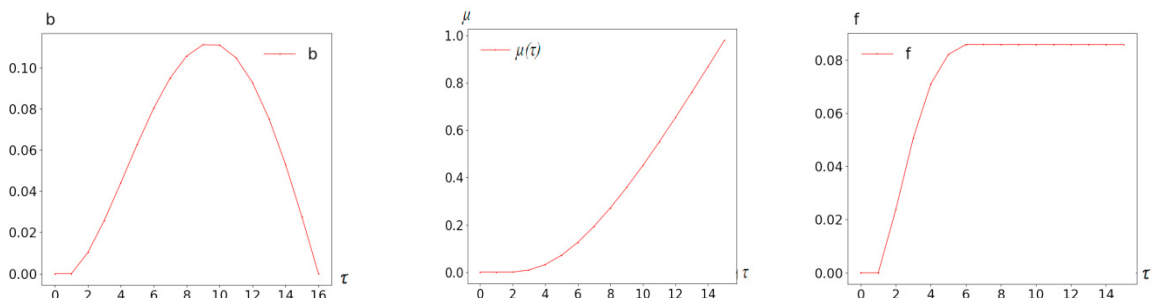


Fig. 1. Biological functions of the virus vs. human interaction:  $b(\tau)$  is the relative infectivity (contagiousness),  $\mu(\tau)$  is the possibility of recovery and  $f(\tau)$  is the relative detection,  $\tau$  is the duration of infection (DI)

Generally accepted and more obvious modifications of biological functions of infectivity and detectability (factoring in the reduction in the number of UDI cohort due to recovery) are shown in Fig. 2:

$$b^*(\tau) = b(\tau)\exp\left(-\int_0^\tau \mu(\tau') d\tau'\right)$$

$$f^*(\tau) = f(\tau)\exp\left(-\int_0^\tau \mu(\tau') d\tau'\right)$$

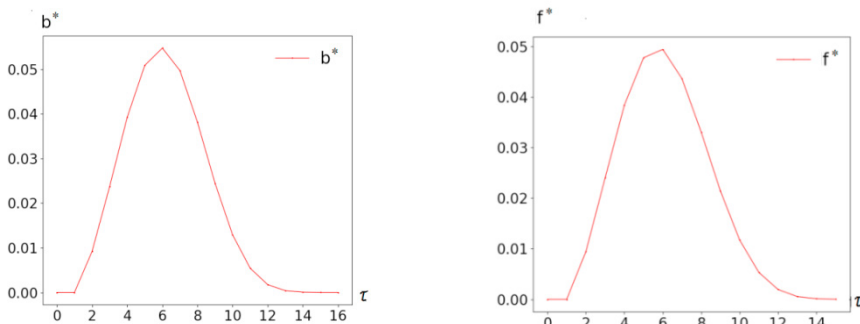


Figure 2. Modified biological functions of the virus vs. human interaction:  $b^*(\tau)$  is the modified relative infectivity (contagiousness),  $f^*(\tau)$  is the modified relative detection,  $\tau$  is the infection duration (DI)

Social management functions (infectivity and detection indices) for the countries under study are shown in Fig. 3.

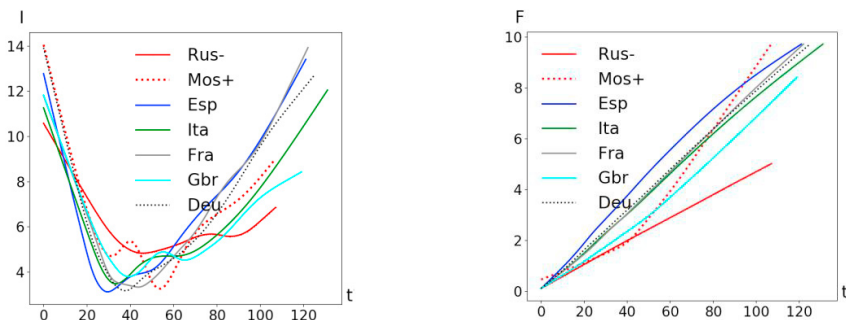


Fig. 3. Social management that determines the UDI dynamics as of July 5:  $I(t)$  is the infectivity index (restriction of contacts, self-isolation, masks, etc.);  $F(t)$  is the detection index (temperature measurement, tests, etc. and subsequent isolation)

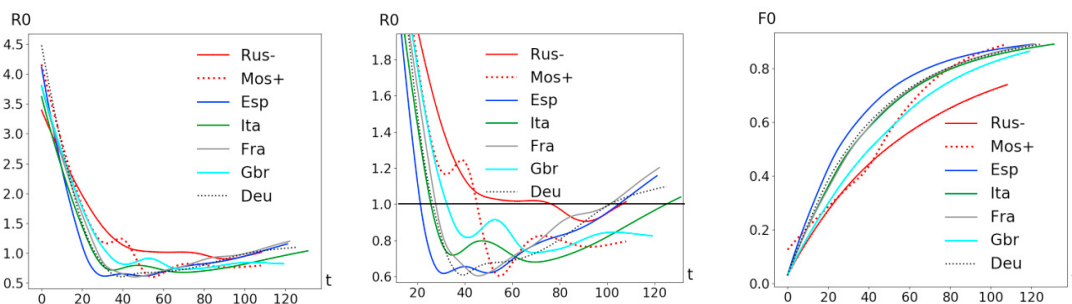


Fig. 4. Reproduction index  $R_0$  (basic reproductive number) and Detection index  $F_0$ . As of July 5

Complex indicators are shown in Fig. 4: the frequently used indicator  $R_0$  (the reproduction index or basic reproductive number) – the average number of people infected by one infected over the entire duration of the disease, and the isolation index  $F_0$  – the probability of an infected to be detected during the entire duration of the disease:

$$R0_p(t) = I_p(t) \int_0^{\tau_{max}} b(\tau) \exp\left(-\int_0^{\tau} (\mu(\tau') + F_p(t)f(\tau')) d\tau'\right) d\tau$$

$$F0_p(t) = F_p(t) \int_0^{\tau_{max}} f(\tau) \exp\left(-\int_0^{\tau} (\mu(\tau') + F_p(t)f(\tau')) d\tau'\right) d\tau$$

These indices give an integral idea of the effectiveness of anti-epidemic measures.  $R0 = 1$  (see middle Fig. 4) is a threshold value: values less than 1 indicate attenuation of the epidemic, values greater than 1 indicate its growth.

Total number of UDI and new cases are shown in Fig. 5.  $nC(t)$  is the parameter that connects the model with the corresponding statistical indicator. In Figure 6  $nC(t)$  are shown with statistics used for Rus- and Mos+.

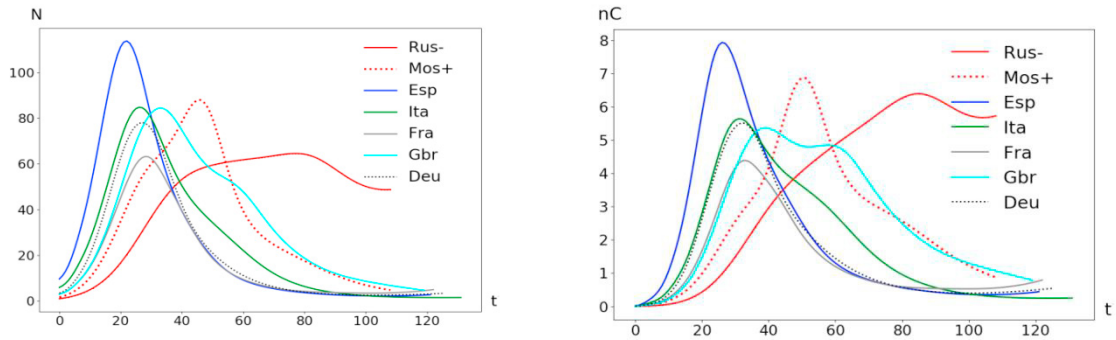


Fig. 5. Dynamics of the total number of UDI –  $N(t)$  (thousand people), lower estimate. The dynamics of new cases  $nC(t)$  (thousand people) As of July 5.

Additional results for Russia and Moscow are shown in Fig. 6 and Fig.7. Fig. 6 shows the initial statistics for new cases of infection and the corresponding model curves, Fig. 7 – UDI dynamic ( $N(t)$ ) and the relative increase ( $B(t)$ ) and the relative decrease ( $D(t)$ ) in the population of UDI.

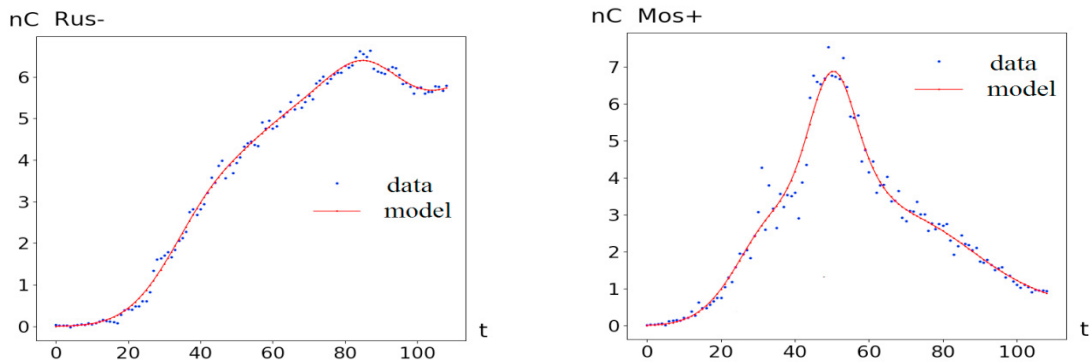


Fig. 6. New cases of infection for Russia excluding Moscow and the region (nC Rus-) and the city of Moscow with the region (nC Mos+). Source data  $nC$ , and model curves  $nC(t)$  (thousand people).

The epidemic can be analyzed in more detail by looking at the reproductive index  $R0$  (see Fig. 6) or at Fig. 7, that shows the graphs of the relative (as a percentage of the total number of UDI) increase and the relative decrease in the population of UDI. The relative increase (due to new cases of infection) is calculated by the formula:

$$B_p(t) = n_p(t,0)/N(t),$$

and the relative decrease in the population of UDI occurs due to recovery and detection and is calculated as:

$$D_p(t) = \left( \sum_{\tau'=0}^{16} \mu(\tau')n_p(t, \tau') + F_p(t) \sum_{\tau'=0}^{16} f(\tau')n_p(t, \tau') \right) / N(t)$$



So, for example, the intersection points of curves  $B$  and  $D$  (zero growth point), for Moscow and the region happened on May 3, for the rest of Russia on June 2 and July 3. A more complete list of social events (with a time reference) for Moscow is shown in Table 1.

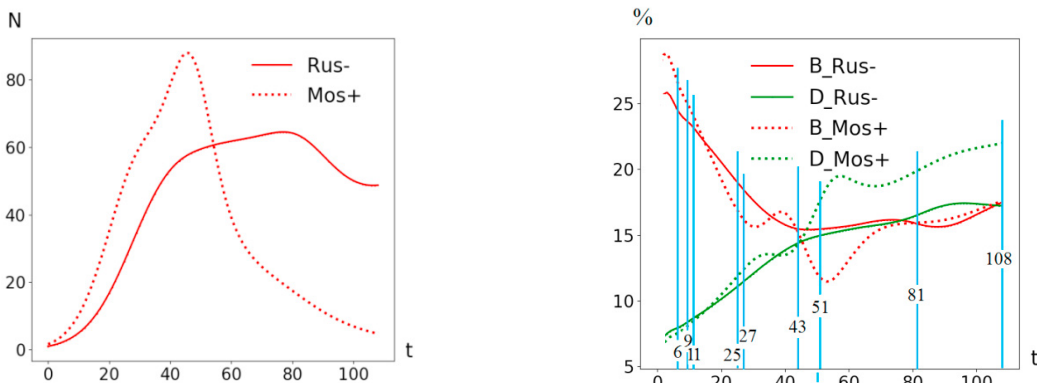


Fig. 7.  $N(t)$  is the dynamics of the total number of UDI for Russia without Moscow and the region and for Moscow and the region (thousand people, lower estimate). The increase ( $B$ ) – the total number of new UDI and the decrease ( $D$ ) – the total amount of UDI decline, as a percentage of the total number of the UDI. As of July 5. The difference gives an increase in percent. The intersection points are zero growth. Specific Moscow events corresponding to the points on Fig. 7 (right) – see Table 1

Table 1. The list of epidemic related events in Moscow.

| Day # | Date  | Event                               |
|-------|-------|-------------------------------------|
| 0     | 19.03 | starting point                      |
| 6     | 25.03 | self-isolation for 65+ citizen      |
| 9     | 28.03 | parks, restaurants, etc. closed     |
| 11    | 30.03 | city-wide self-isolation            |
| 25    | 13.04 | passes introduced                   |
| 27    | 15.04 | crowds in the metro                 |
| 43    | 01.05 | May 1 festivities                   |
| 51    | 09.05 | May 9 festivities                   |
| 81    | 08.06 | self-isolation and passes cancelled |
| 108   | 05.07 | end of modeling period              |

### 6. Discussion

*Discussing the results.* The perception of the COVID-19 epidemic is constantly changing, as are the ways in which various statistics are calculated [11]. Consequently, we can only talk about the qualitative correspondence of the results obtained to "modern scientific concepts as of July 2020".

Fig. 2: The calculated function of modified relative infectivity (contagiousness) with a maximum on the 6th day corresponds to the epidemiological ideas on COVID-19 contagiousness [3]. It should be noted that the functions depicted in Fig. 2 are similar. Presumably, the greater the relative infectivity, the greater the relative detectability (symptom noticeability) of UDI. Apparently, with proliferation of tests that allow asymptomatic (at the time of testing) patients' detection, the shape of relative detectability is expected to change significantly.

Fig. 3 and Fig. 4: To interpret the graphs of social management functions (complex indicators), it is essential to have an understanding of the situation in a given country. For example, the discrepancy in the detection index for Rus- and Mos + can be explained by the beginning of mass testing in Moscow (about 40th point). Similarly, you can link other features (extremes) of charts for Moscow to specific events. Fig. 3 shows the infectivity and detection

indices for the countries under the study. Analysis of the graphs suggests that countries that have overcome the epidemic (or its first wave) returned to the usual (pre-epidemic) values of the infectivity index, and their relative stability can be explained by intensive cases detection measures and isolation.

Fig. 5: The indicators presented here for different countries allow for a qualitative comparison of the dynamics of the epidemic in different countries. It appears that their analysis together with other indicators by experts familiar with the situation in countries can lead to an assessment of the effectiveness of various strategies for managing the epidemic. It should be borne in mind that different countries use different (case) registration methods

Fig. 6. It can be seen that the model curves approximate datasets for Rus- and Mos+ quite well. The graphs show distinctly different dynamics of the epidemic in different regions of Russia: in Moscow and the region, the epidemic is on the wane, and the rest of Russia is about to reach the plateau.

Fig. 6. One can see that the model curves approximate datasets for Rus- and Mos + fairly well. The graphs show distinctly different dynamics of the epidemic in different regions of Russia: in Moscow and the region, the epidemic is on the wane, and the rest of Russia is about to reach the plateau.

Fig. 7. Similar conclusions can be drawn by analyzing the total numbers of UDI (see Fig. 7). The maximum of UDI for Moscow and the region corresponds to May 4, for the rest of Russia – to June 3, with a month between them. That is why Moscow and the region were separated from Russia. The figure on the right is an example of a detailed analysis of the dynamics of the epidemic in Mos +. So, for example, the intersection points of curves *B* and *D* (zero growth point), for Moscow and the region happened on May 3, for the rest of Russia on June 2 and July 3. A more complete list of social events (with a time reference) for Moscow is shown in table 1.

*Discussing the hypotheses.* Simplification of a complex real phenomenon to a mathematical model inevitably leads to errors. In our case, such a simplification is based on a set of hypotheses that only partially (imprecisely, not always or everywhere) reflects reality. Therefore, it is pointless to say that the hypothesis is incorrect – a hypothesis is always incorrect. It only makes sense to discuss how acceptable is the hypothesis and under what conditions (and time periods) its postulation leads to acceptable errors. We will discuss the acceptability of hypotheses as of July 2, 2020

Hypothesis 2 (Human body does not change) seems to us acceptable, since there are no confirmed means of prevention and the percentage of population with immunity is small (in Moscow, the estimate for May was 14%)

Hypothesis 3 (on the isolation of the identified infected) unfortunately cannot be assessed due to the lack of systematic data on the violation of quarantine. Admittedly among the medical staff there are a lot of cases, especially in the beginning of the epidemic. However, the role of hospital transmissions in our study can be neglected. If hospital transmissions were to a large extent determined the dynamics, then in order to provide hundreds and thousands of diseases a day, almost all personnel had to get sick, and this is not the case.

Hypothesis 4 (no migration), apparently, is questionable for the beginning of the modeling interval. However, this inaccuracy is imperceptible against data errors in the initial period of modeling.

Hypothesis 6 (Maximum DI) proved to be nonsignificant. Changing  $\tau_{max} = 21$  made no difference to the results (These calculations are omitted). This result was expected: an analysis of the possibility recovery function (Fig. 1) and modified biological functions (Fig. 2) allows us to conclude that by the 13th day of illness, most of the UDI either recovers or goes into quarantine. Thus, the limitation  $\tau_{max} = 16$  is given with a margin.

Hypothesis 9 (detection index - nondecreasing function). During the timeframe of the study (up until July, 5) the efforts of society to detect infected (test, temperature control etc.) increased.

*What to do next.* The considered model satisfactorily describes the epidemic up until July, 5. To increase the timeframe, it is necessary to modify the model and its underlying hypotheses. First of all, to take into account the immunization of the population, especially since in many countries screening studies are carried out in order to assess the percentage of immunized.

For the practical use of the constructed model, it is essential that abstract indices of infectivity and detection are associated with specific anti-epidemic measures of society, such as the number of tests, the number of isolated ones, indices of population mobility, etc. It is not until such connections are identified that it will be possible to set and solve epidemic optimization management problems and forecasting problems.

Prediction epidemic dynamics requires a separate study, with a different procedure for estimating modeling errors. However, within the framework of the model used, a forecast of (external) social administrations – infectiousness ( $I(t)$ ) and detection ( $F(t)$ ) indices – is needed to forecast the epidemics progression. If the forecast

(through society's specific anti-epidemic measures) with known accuracy exists, then the UDI prediction becomes straightforward. Moreover, the error of such forecast can also be estimated.

## 7. Conclusion

The combination of knowledge about the interaction of the virus vs. man and man vs society with statistical data based on the balanced identification method, made it possible to choose a model with complexity corresponding to the volume and quality of the experimental material. The resulting functions are consistent with current ideas on the processes that determine the epidemics dynamics.

The technology of balanced identification supports the (evolutionary) modification of the model – a new (and usually more complex) model is built on the basis of the previous one. The previously found solution is used as an initial approximation in finding a solution for a new model. As new data or knowledge becomes available, the model can be improved and developed. For example, new information on a considerable percent of Moscow citizens with immunity leads to a change in initial hypotheses and possibly model modification (В наших планах).

Comparison of possible models (and the choice of a model that is in good agreement with available observations) may be carried out in a very short time (in fact, primary models were built in a few days, and then modified as new data arrived). This was made possible through the use of balanced identification (SvF) technology [7], and the SSOP service designed to solve sets of independent optimization problems on computing resources connected to the optimization modeling subsystem <https://optmod.distcomp.org>, deployed on the platform Everest [10]

## Acknowledgements

This research:

- has been carried out using computing resources of the federal collective usage center Complex for Simulation and Data Processing for Mega-science Facilities at NRC “Kurchatov Institute”, <http://ckp.nrcki.ru/>.
- was supported in part through computational resources of HPC facilities at NRU HSE ”.
- was financially supported by the Russian Science Foundation (grant No. 16-11-10352).

## References

- [1] Wynants, Laure et al. (2020) “Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal”, *BMJ* **2020**;369: m1328.
- [2] Solomatine D.P., Ostfeld A. (2008) “Data-driven modelling: some past experiences and new approaches”, *J of Hydroinformatics*, **10(1)**: 3-22.
- [3] McAloon, Conor G et al. (2020) “The incubation period of COVID-19: A rapid systematic review and meta-analysis of observational research”, *medRxiv* 2020.04.24.20073957.
- [4] Svirezhev Yu.M., Logofet D.O. (1978) “Sustainability of biological communities” *Moscow: Nauka*: 352.
- [5] Hansen P.E. (1986) “Leslie matrix models”, in: Csetenyi, A. (ed) *Papers on Mathematical Ecology*, **1**, Budapest: Department of Mathematics, Kari Marx University of Economics: 1-139.
- [6] Sokolov A.V., Voloshinov V.V. (2018) “Choice of mathematical model: balance between complexity and proximity to measurements.” *International Journal of Open Information Technologies* **6(9)**: 33–41.
- [7] Sokolov A.V., Voloshinov V.V. (2019) “Balanced Identification as an Intersection of Optimization and Distributed Computing”, URL: <https://arxiv.org/pdf/1907.13444.pdf>
- [8] Nakhshuev A.M. (1995) “The equations of mathematical biology. Textbook manual for universities.” *M.: Higher School*: 301.
- [9] Berezgova R.Z. (2017) “On a nonlocal boundary-value problem for the McKendrick Von Foerster loaded equation with Caputo operator”, *Vestnik KRAUNC. Fiz.-mat. nauki*, **19(3)**: 5-9.
- [10] Sukhoroslov O., Volkov S., Afanasiev A. (2015) “A Web-Based Platform for Publication and Distributed Execution of Computing Applications” *14th International Symposium on Parallel and Distributed Computing (ISPDC)*. IEEE: 175-184.
- [11] Rahul Shetty (2020) “Rapid Response: Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal”, <https://www.bmj.com/content/369/bmj.m1328/rr-4>