

Model of socio-economic factors in the spread of SARS-CoV-2 across Russian regions

E.V. Sinitsyn^{1,*}, *A.V. Tolmachev*^{1,2}, and *V.M. Laptev*¹

¹Ural Federal University, Mira St., 19, 620002 Ekaterinburg, Russia

²Datatel-Ural LLC, Sibirsky trakt, 12/8, 620100 Ekaterinburg, Russia

Abstract. The worldwide spread of a new infection SARS-CoV-2 makes relevant the analysis of the different factors that lead to the vulnerability of modern civilization to previously unknown diseases. In this regard, the development of mathematical models describing the evolution of epidemics like COVID-19 and the identification of socio-economic factors affecting the epidemiological situation in regions is an important research task. The paper proposes a probabilistic mathematical model for the spread of the COVID-19 epidemic, which allows to analyze the evolution of the main characteristics of the disease and to assess main factors influencing them. The study is based on the official statistical data on the spread of the COVID 19 presented on coronavirus sites in the Russian Federation and other countries, the Yandex Data Lens dataset service, as well as the data from the Federal State Statistics Service. In the research some data mining methods were used for evaluation the model's parameters. The model equations allow to predict the evolution of the disease and estimate the confidence interval of such prognosis. We estimated the ratio of detected and hidden cases of the disease, the distribution of the disease's duration probability and its average value for different regions. It has been mathematically proven that the vaccination is the necessary and sufficient condition of achievement a stationary stable state - the cessation of a pandemic. The regions of Russian Federation were clustered by the course of the disease COVID-19 on the base of k-means method. The analysis of the most important socio-economic factors affecting the epidemiological situation was provided separately for each cluster.

1 Introduction

The worldwide spread of the new SARS-CoV-2 infection raise a question about socio-economic problems that make modern civilization vulnerable to new, previously unknown diseases. To prevent situations like the current in the future, it is worth to model pandemic, analyze it possible development, distinctive features. It is important to assess the influence of various socio-economic factors on the rate of the disease's spread and its course (duration of the disease, mortality). The results of this analysis can be used to make the decisions to constrain epidemic. This study is the second work of the authors devoted to this problem. In a previous work [1] we presented a review of works (see, for example [2-9]) devoted to

* Corresponding author: sinitsyn_ev@mail.ru

modeling of the epidemics. In particular, the last of the mentioned works laid the foundations of the so-called SIR epidemics models (from the English Susceptible - Infected - Recovered). The approach considered in our first work was also close to SIR model. On its basis, we analyzed the initial phase of the COVID-19 disease (until the end of the so-called "first" wave in July 2020), mainly in the regions of the Russian Federation, with regional comparisons. Contemporary development of the epidemic situation in the Russian Federation and in the world, the emergence of new data, the start of the vaccination company, requires adapting both the model and the analysis results to the current data.

In connection with the above, the main goals of our present research are:

— The adaptation of a probabilistic mathematical model [1] describing the course of the epidemic to its current phase.

— Regional comparisons of the current epidemic situation in the regions of the Russian Federation based on the base of adapted model.

— The official statistical data on the spread of the epidemic presented on coronavirus sites in the Russian Federation and other countries, the Yandex Data Lens dataset service, as well as data from the Federal State Statistics Service, were used.

2 Theoretical foundations of the analysis of the COVID-19 spread

As in the previous work of the authors [1] for modeling the COVID-19 epidemic, we divide the population into the following clusters (groups):

I – infected individuals in whom the disease has not been identified: those who are ill without clinical symptoms, who did not apply to medical institutions or were mistakenly diagnosed with the absence of the disease.

H – healthy and susceptible to disease individuals.

A – detected active cases of the disease.

D – deaths.

U – disease-resistant (disease-resistant individuals: recovered, including implicitly, as well as vaccinated).

Hereinafter, the size of each group will be denoted by a letter N with the corresponding index $N_I \dots N_U$. The relative share of each group in the population of the analyzed region N will be denoted as n with an index, denoting the group: $n_I \dots n_U$.

$$N_I + N_H + \dots + N_U = N$$

or

(1)

$$n_I + n_H + \dots + n_U = 1$$

The graph corresponding to the model used herein is shown in Fig. 1.

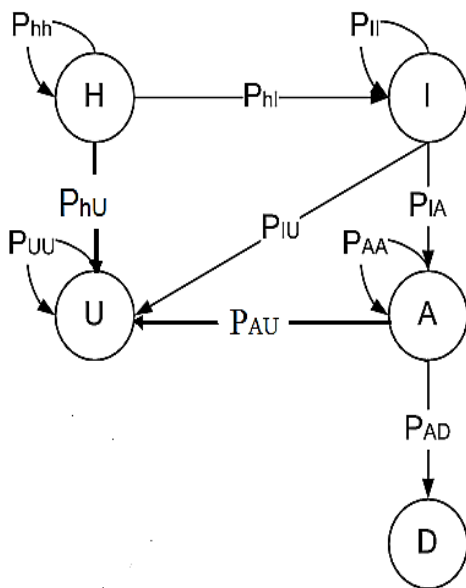


Fig. 1. The graph of possible transitions between groups of participants in the epidemic. The indicators on the arrows are transition probabilities.

New statistical data on the course of the disease caused by the SARS-CoV-2 virus, that appeared since the publication of our previous work, led to some transformations of the graph in Fig. 1, in comparison with that one in [1]:

- The beginning of the vaccination campaign is considered (transition $H \rightarrow U$).
- Statistics, at least for RF, shows that a severe course of the disease in group I, finishing by a death, is usually accompanied by an appeal to medical attention and the isolation of the patient (falling into group A); exceptions are quite rare, in this regard, the transition ($I \rightarrow D$) is not considered because of its negligible effect on the spread of the epidemic.
- Cases of re-infection have been identified so few that they can be neglected in modeling (only about 50 cases of re-infection are officially announced around the world, and most of them are associated with the specifics of the patient's health).

The listed features of the transitions between the vertices of the graph transform it into the tree with one initial (H) and two final states (U, D). The transitions shown in Fig. 1 can be described using a 5×5 matrix \hat{T} which has the form:

$$\hat{T} = \begin{array}{ccccc|c} & \mathbf{H} & \mathbf{I} & \mathbf{A} & \mathbf{D} & \mathbf{U} & \\ \hline & P_{HH} & 0 & 0 & 0 & 0 & \mathbf{H} \\ & P_{HI} & P_{II} & 0 & 0 & 0 & \mathbf{I} \\ & 0 & P_{IA} & P_{AA} & 0 & 0 & \mathbf{A} \\ & 0 & 0 & P_{AD} & P_{DD} & 0 & \mathbf{D} \\ & P_{HU} & P_{IU} & P_{AU} & 0 & P_{UU} & \mathbf{U} \end{array} \quad (2)$$

The columns of the matrix (2) contain the probabilities of transitions from the group indicated in the top row to the group presented in the rightmost column, which implies that the sum of all elements of each column is equal to 1. Such matrices are usually called stochastic [10].

Using the matrix (2), it is easy to formulate the equation for the probability $P(n_I, n_H, \dots, n_U; t) \equiv P(\vec{X}; t) \equiv P$ – the fraction of participants in groups $I - U$ at the time t will be n_I, n_H, \dots, n_U . Here $\vec{X} = \{n_I, \dots, n_U\}$ – is a vector, with the sum of the components equal to 1. After simple transformations performed by the analogy with the approach described in our previous work and, for example, in the work of Feller (1964) one can get:

$$\frac{\partial P}{\partial t} = P \cdot \{T_{kk}N_k - N\} + T_{k \neq j}(N_k + 1)P(\dots, N_k + 1, \dots, N_j - 1, \dots; t) \quad (3)$$

It is implied that repeated indices are summed.

The equation (3) allows to study various aspects of the changes in the probability $P(\vec{X}; t)$, in particular, the scenario analysis of the epidemic evolution. Let's start from the deriving the equations for mathematical expectations n_I, n_H, \dots, n_U on the base of (in the equations (4a, d), for brevity and simplicity of notation, we assumed that a positive effect starts immediately after vaccination and its duration is longer than the duration of analyzing time periods. Now there is not enough statistical data about the time of the immunity formation and its duration for more general consideration) (3):

$$\frac{dn_H}{dt} = -(P_{HI} + P_{HU})n_H(t) \quad (4a)$$

$$\frac{dn_I}{dt} = - \int_0^t f_{IA}(t - \tau) \cdot n_I(\tau) \cdot d\tau - \int_0^t f_{IU}(t - \tau) \cdot n_I(\tau) \cdot d\tau + P_{HI}n_H(t) \quad (4b)$$

$$\begin{aligned} \frac{dn_A}{dt} = & - \int_0^t f_{AD}(t - \tau) \cdot n_A(\tau) \cdot d\tau - \int_0^t f_{AU}(t - \tau) \cdot n_A(\tau) \cdot d\tau + \\ & + \int_0^t f_{IA}(t - \tau) \cdot n_I(\tau) \cdot d\tau \end{aligned} \quad (4c)$$

The equations for the derivatives $\frac{dn_D}{dt}, \frac{dn_U}{dt}$ are:

$$\frac{dn_D}{dt} = \int_0^t f_{AD}(t - \tau) \cdot n_A(\tau) \cdot d\tau \quad (4d)$$

$$\frac{dn_U}{dt} = \int_0^t f_{AU}(t - \tau) \cdot n_A(\tau) \cdot d\tau + \int_0^t f_{IU}(t - \tau) \cdot n_I(\tau) \cdot d\tau + P_{HU}n_H(t). \quad (4e)$$

Here time $t=0$ corresponds to the beginning of the pandemic, and t to the considered current moment. The main feature of integra-differential equations (4) is their temporal nonlocality (delay) associated with the following factors:

— $f_{IA}(t - \tau)$ – is connected with incubation (latent) period, due to which the symptoms of the disease and its detection occurs with a delay on some time distributed near it most probable value.

— $f_{AD}(t - \tau), f_{AU}(t - \tau), f_{IU}(t - \tau)$ – is caused by the finite time of the disease's course (in general, different for its various outcomes). These times are also distributed in the neighborhoods of the most probable values.

The features of the functions $f(t - \tau)$, characterizing the time delay, will be described in more details in the next section. Here we only note that the distribution of the delay times mentioned above is rather "narrow", so the integrands in (4) are different from zero in narrow intervals near current time t . These intervals are small compared to the characteristic scales of significant changes in the unknowns n_I, n_H, n_U . So, let's expand the unknowns $n_A(\tau), n_I(\tau)$ in integrands (4) into the Taylor series in powers $(t-\tau)$ and left only the linear terms. This gives:

$$\frac{dn_H}{dt} = -(P_{HI} + P_{HU})n_H(t) \quad (5a)$$

$$\frac{dn_I}{dt} (1 + P_{IA} \overline{T_A} + P_{IU} \overline{T_C}) = -(P_{IA} + P_{IU})n_I(t) + P_{HI}n_H(t) \tag{5b}$$

$$\frac{dn_A}{dt} (1 + P_{AD} \overline{T_C} + P_{AU} \overline{T_C} - P_{IA} \overline{T_A}) = -(P_{AD} + P_{AU})n_A(t) + P_{IA}n_I(t) \tag{5c}$$

$$\frac{dn_D}{dt} (1 + P_{AD} \overline{T_C}) = P_{AD}n_A(t) \tag{5i}$$

$$\frac{dn_U}{dt} (1 + P_{AU} \overline{T_C} + P_{IU} \overline{T_C}) = (P_{AU} + P_{IU})n_A(t) + P_{HU}n_H(t) \tag{5f}$$

Note that from equations (4) and (5), it follows:

$$\frac{dn_H}{dt} + \frac{dn_I}{dt} + \frac{dn_A}{dt} + \frac{dn_D}{dt} + \frac{dn_U}{dt} = 0 \tag{6}$$

In equation (the extension of the limits of integration in (7) and (8) to ∞ is possible due to the aforementioned "narrowness" of the distributions of the corresponding times) (5)

$$P_{IA} = \int_0^\infty f_{IA}(\tau) \cdot d\tau \tag{7a}$$

$$P_{IU} = \int_0^\infty f_{IU}(\tau) \cdot d\tau \tag{7b}$$

$$P_{AU,D} = \int_0^\infty f_{AU,D}(\tau) \cdot d\tau \tag{7c}$$

As will be shown in the next section, the functions $f_{AU}(\tau)/P_{AU}$ and $f_{AD}(\tau)/P_{AD}$ are practically identical. It is reasonable to assume that the function $f_{IU}(\tau)$, which characterizes the distribution of recovery times for group I, also slightly differs from $f_{IA}(\tau)$.

The values $\langle T \rangle$ in (5) are the corresponding average times:

$$\overline{T_A} = \frac{1}{P_{IA}} \int_0^\infty f_{IA}(\tau) \cdot \tau \cdot d\tau \tag{8a}$$

$$\overline{T_C} = \overline{T_U} \approx \overline{T_D} = \frac{1}{P_{AU,D}} \int_0^\infty f_{AU,D}(\tau) \cdot \tau \cdot d\tau \tag{8b}$$

As it was mentioned in the paper [11], the probability P_{IA} can be represented as:

$$P_{IA} = P_\alpha \left(d + \frac{T}{N} \right) \tag{9}$$

Here d is the likelihood of an infected person self-referring to a doctor per time unit, T/N is the number of tests per capita carried out per time unit to detect COVID-19, and finally, P_α is the reliability of detecting the disease – the probability of a type I error in diagnosis (the disease is present, but not detected).

These values were defined in the paper [1]: $P_\alpha d = 0,12$; $P_\alpha = 65.3\%$

To evaluate the probability P_{HI} in (4, 5) we also use the model of the work [11], in which:

$$P_{HI} = \frac{f \cdot n_I(t) \cdot n_H(t)}{n_H(t) + n_I(t)}, \tag{10}$$

where f is the average frequency of contacts in the population multiplied by the probability of infection by contact. Thus, the presented formula assumes that a person can be infected with a probability proportional to the fraction of infected persons among the all individuals with whom she or he has contacted. The identified (and therefore isolated) infected persons, as well as persons from group U, do not influence on the probability of infection. In the work [1], the parameter $f = 0.14$ was determined on the base of statistical data on the COVID-19 for the Russian Federation.

Let us provide a general analysis of equations (5). Firstly, we'll define rigorously the popular term "plateau" (the reports that it is either already achieved, or not yet in the mass media are very diverse and contradictory). We will assume that the "plateau" corresponds to the situation:

$$\frac{dn_A}{dt} = 0, \tag{11a}$$

this, in accordance with (5c), gives:

$$\frac{n_A(t)}{n_I(t)} = \frac{P_{IA}}{P_{AD}+P_{AU}} \tag{11b}$$

Note that evaluation the ratio of the detected cases of the disease to their total number (including latent cases o) is an urgent and important problem in the planning of anti-epidemic measures. The equation (11b) can be used to assess latent morbidity, since all parameters in it can be determined based on data, presented in the official medical statistics. The illustrations will be given in the next section.

It would be tempting to expand definition (11a) by extending the concept of "plateau" to stationary points of the equations (4, 5) defined by the relations:

$$\frac{dn_A}{dt} = \frac{dn_I}{dt} = \frac{dn_H}{dt} = \frac{dn_D}{dt} = \frac{dn_U}{dt} = 0. \tag{12}$$

It is easy to verify that the stationary point defined by (12) is stable [12]. However, from (5d, e) it follows that in this case: $n_A(t) = 0$, and hence: $n_H(t) = 0$. Then from (5b) one can get $n_I(t) = 0$. Thus, (12) describes not a plateau, but the end of the epidemic. The analysis of (5) "from top to bottom", starting with (5a), shows that the achievement of the condition $n_H(t) = 0$ is necessary and sufficient to reach the stationary point (12) – the end of the epidemic. Moreover, state $n_H = n_A = n_I = 0, n_D + n_U = 1$ is the only stationary point of equations (4,5). This means, that the pandemic can be finished only if all the population become either dead (group D) or immune (group U)

We would like to emphasize That this conclusion from the equation (4,5) is especially important in connection with the vaccination company. The stable state (12) (the transition $H \rightarrow U, D$ for every person), is achievable by either the vaccination or the COVID-19 disease. The choice of the way to achieve it – at the discretion of citizens.

3 Methods of analysis

To analyze the entire set of collected epidemic and the socio-economic data we used:

1. methods of statistical analysis [13,14]:
 - correlation analysis of parameters characterizing morbidity and factors describing the socio-economic state of the regions.
 - multivariate regression.
2. methods of data mining - [10, 15], for example, clustering by k-means method.
3. methods of statistical modeling.

For the analysis, we used data from the website on coronavirus in the Russian Federation¹and in the World², the service with data sets Yandex DataLens³, as well as data from Rosstat for the constituent entities of the Russian Federation [16].

In accordance with the Data mining standard [17], the initial data underwent preprocessing. We have excluded from the analysis those countries where recoveries were not registered, or the reported values differed significantly from their typical values for most countries. Subjects with a small number (less than 2000) of closed disease cases (recoveries and deaths)- were also excluded. Finally, the errors in data were corrected, (for example:

negative numbers of new or closed cases). Data analysis was performed by means of scripts, developed by the authors in the Python in Anaconda data analysis package.

4 The analysis of factors affecting the epidemic of COVID-19

In this section, the mathematical model, presented above (section 2), will be used for a comparative analysis of the regions of the Russian Federation by characteristics of the COVID-19 disease.

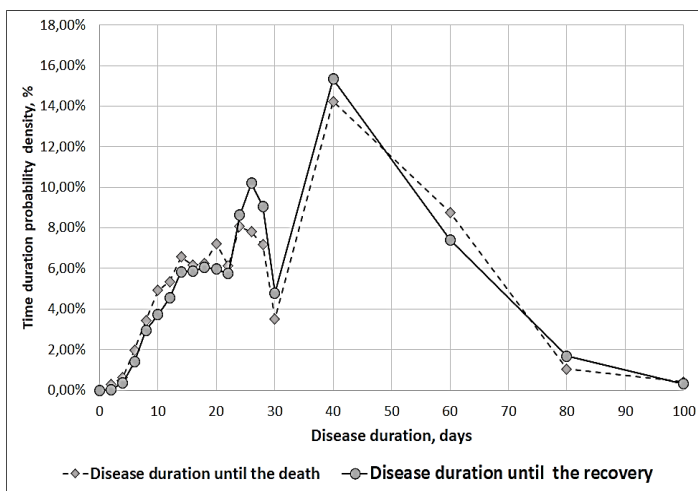
4.1 The course duration of the disease COVID-19 in the Russian Federation

The initial data for analyzing the probability distribution of recovery (death) times from COVID-19 disease were obtained using the Yandex DataLens service. We analyzed the period from the start of the pandemic to the current moment 13.05.2021.

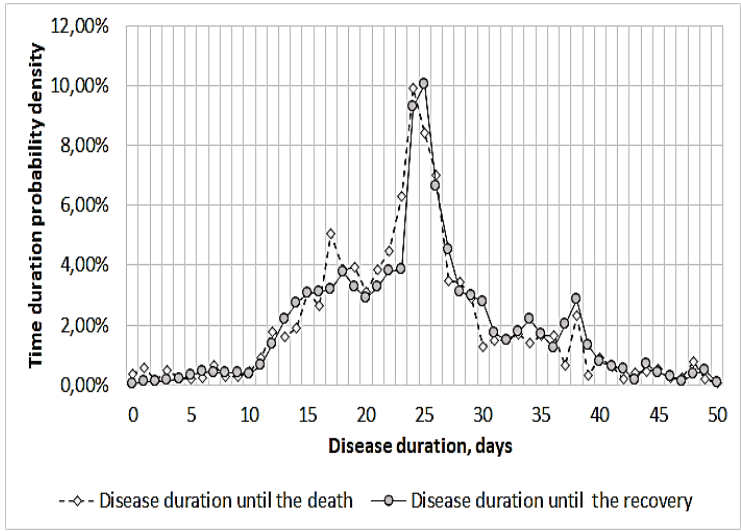
To determine the time of the disease's, duration we used the following supposition - the person, who entered the group of active cases first was the first who left it (after recovery or death). This is analog of well-known accounting FIFO method (First In - First Out). This method is practically the only possible way who estimate the duration of disease on the base of purely official statistics without detailed analysis of individual disease histories.

To share closed disease cases between groups "Recovered" and "Dead" we applied Monte Carlo procedure: closed cases were distributed between mentioned groups with a probability of recovery or death at the moment when the case of disease was closed.

The obtained probability distributions of the COVID-19 duration up to recovery or death for the Russian Federation are presented in Fig. 2. For comparison, we presented the same probability distribution evaluated near one year ago in our work [1] (Fig. 2b)



a)



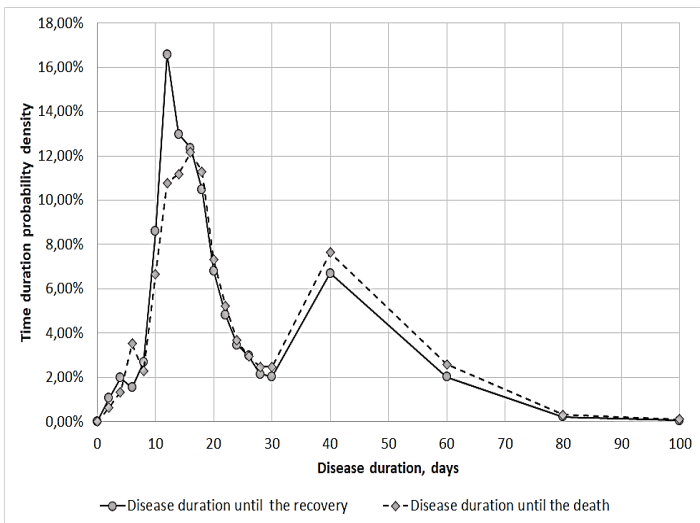
b)

Source: the authors' calculations based on the dataset from the Yandex DataLens service.

Fig. 2. Time duration (recovered and dead cases of COVID-19) probability density, Russian Federation: a) – on the date 13.05.2021, b) – 18.07.2020.

Note that Fig. 2a shows an additional peak of the probability density at the point near 40 days. Highly likely this reflects the experience gained during the year passed in the treatment of severe cases of the disease, which have a relatively long duration.

We calculated also shows the same probability distributions as in Fig. 2, but for the whole World (Fig. 3).



Source: the authors' calculations based on the dataset from the Yandex DataLens service.

Fig. 3. Time duration (recovered and dead cases of COVID-19) probability density, the whole World.

We can also see (Fig. 3) the presence of additional peak in the region of 40 days (like that in Fig. 2), that was absent one year ago [1].

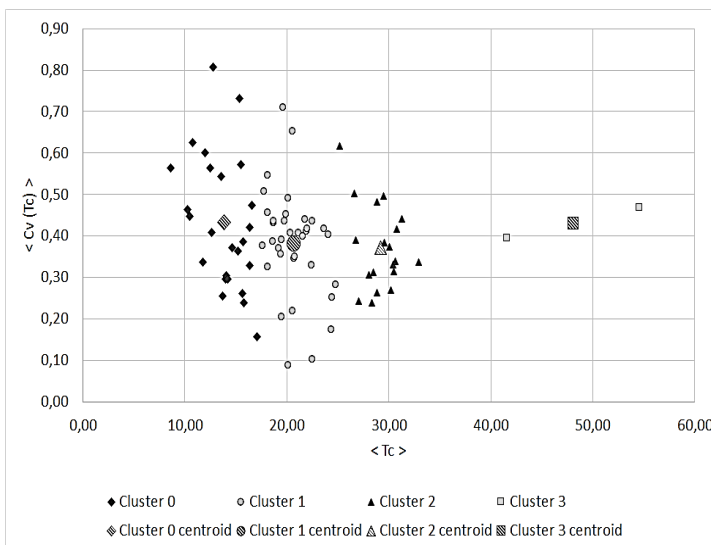
It can be seen from the presented figures (Fig. 2, Fig. 3) that the probability distributions of the disease durations up to recovery or death do not differ statistically significantly. So, one can use only one time $-T_C$ instead of two times of recovery and death. The data on the time of illness for different countries and different regions of Russia are significantly different. Note that, as was mentioned in Section 3, the probability distribution of these times really can be treated as "narrow" in comparison with characteristic times of pandemic course.

The duration of the COVID-19 disease may depend on a number of factors, in particular: the virus strain, that is predominantly common in the region, the level of medical care, the organization of disease detection, measures to limit it spread, etc.

It seems useful to cluster the regions of the Russian Federation and countries of the world (Fig. 4, Fig. 5), according to such parameters as the average time of disease $-T_C$ and the coefficient of its variation:

$$C_v(T_C) = \frac{\sigma(T_C)}{\langle T_C \rangle}, \tag{13}$$

here $\sigma(T_C)$ is the standard deviation, $\langle T_C \rangle$ the mathematical expectation (average) of the time of disease course.



Source: the authors' calculations based on the dataset from the Yandex DataLens service.

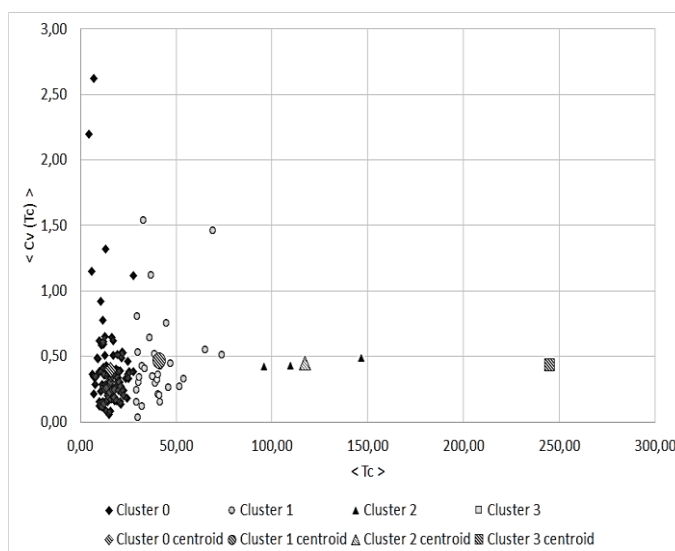
Fig. 4. Clustering the regions of the Russian Federation by the parameters $\langle T_C \rangle$ and $C_v(T_C)$ (13).

There are 4 clusters in the RF. Typical representatives of cluster 0 are: the Komi Republic, Volgograd region, Tula region, Altai Territory, Belgorod region. Cluster 1: Kursk Region, the Republic of Chuvashia, Karelia, Oryol, Kursk and Sverdlovsk Regions, Perm Territory. Cluster 2 is represented by such subjects as: Primorsky Territory, the Republic of Tatarstan, Karachay-Cherkessia, the Chelyabinsk region, the cities of St. Petersburg and Moscow. Finally, cluster three is represented by only two subjects: the Moscow and Pskov regions. Some indicators describing the socio-economic state in the above-mentioned clusters 1-4 are presented below (Table 1). These indicators have statistically significant correlation coefficients with the average duration of COVID-19 disease in one or more clusters.

Table 1. Some socio-economic indicators for clusters 0-3.

Indicators for clusters	Average values				Variation coefficients			
	0	1	2	3	0	1	2	3
Population density, people / sq. Km	40,1	34,6	26,4	11,5	2,4	0,8	1,0	7,9
The fraction of the employed people (average annual number) in the population	0,5	0,5	0,5	0,4	0,1	0,1	0,2	0,01
Gross regional product, mln. RUB per capita	0,4	0,4	0,7	0,4	0,9	0,7	1,0	0,5
Population per hospital bed, people	118,1	125,0	123,8	131,6	0,1	0,2	0,2	278,0
Population per doctor, people	219,9	216,8	209,7	0,1	0,2	0,2	0,2	0,1
Number of diseases PER 1000 people	830,6	771,9	772,9	713,1	0,2	0,2	0,2	0,02

For the whole World "k-means" method gives four clusters also (Fig. 5). The most numerous cluster is 0: representatives – Kazakhstan, Slovakia, Armenia, Russia. Next cluster 1, represented by Bulgaria, Ukraine, Egypt, Hungary, Spain. The remaining clusters have few members: cluster 2 (Cyprus, Honduras, France) and 3 (Netherlands).



Source: the authors' calculations based on the dataset from the Yandex DataLens service.

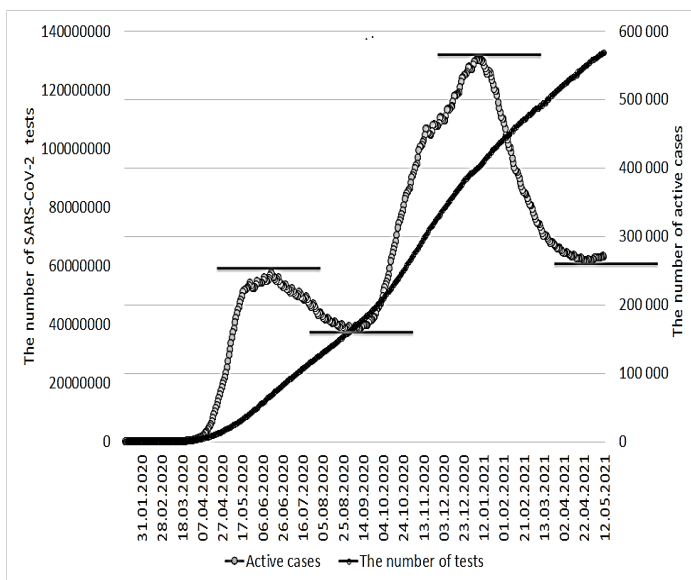
Fig. 5. Clustering the countries of the World by the parameters $\langle T_c \rangle$ и $\langle C_v(T_c) \rangle$ (13).

The performed clustering was used for the correlation analysis of the influence of socio-economic factors on the course of the disease in the Russian regions. Such analysis was carried separately for each cluster, over the regions included into it. The subjects of Cluster 3 were excluded from the correlation analysis.

4.2 Assessment the ratio of detected and latent cases of COVID-19 disease in the Russian Federation

Let's use the theoretical model of Section 3, namely, formula (11b) to assess the ratio between detected (active) and latent cases of COVID-19 that did not registered by the official statistics, for the Russian Federation. For this purpose, we'll use the data on active cases and tests on SARS-CoV-2 performed (Fig. 6).

The horizontal lines (Fig. 6) mark the dates when $\frac{dn_A}{dt} = 0$. With the help of (9) and the data [1], on the parameters included in it, the value of P_{IA} . Probabilities P_{AU}, P_{AD} for the disease COVID-19 in the Russian Federation, were calculated on the base of data presented in the Yandex DataLens service³. The results for the dates marked in Fig. 5 are presented in Table 2.



Source: the authors' calculations based on the dataset from the Yandex DataLens service.

Fig. 6. The Data for calculating the ratio between detected and latent cases of the disease COVID-19 according to the formula (11b).

Table 1. The ratio between detected and latent cases of COVID-19, calculated according to (11b).

Dates	22.04.2021	01.01.2021	03.09.2020	12.06.2020
Value	19,7	15,1	8,3	5,3
Confidence interval (Significance 10%)	±9,5	±7,3	±4,0	±2,6

Unfortunately, the data on the number of tests in the Russian Federation and other countries contain many omissions and inaccuracies, which reduces the reliability of the estimations. Nevertheless, the estimations (Table 2), even for the lower limit of the

confidence interval, are impressive. In RF we have near five million detected cases of the COVID-19 disease, most of them has finished by recovery. In addition, according to table 2, there are about 50 million of latent cases of the disease, and about 21 million citizens have get the first vaccination. Thus, the total number of people immune to SARS-CoV-2 infection is approaching to 75 million. This is already close to the notorious 70% of the population.

4.3 Discussion. Correlation and regression analysis of socio-economic factors affecting the characteristics of the epidemic

The modeling of the spread of COVID-19 disease proposed in this work shows that there is the only one way to finish this pandemic. It is the complete immunization of the population, either through vaccination or through disease. Significant differences in socio-economic conditions, quality of medical care, cultural traditions in different regions, as well as mutations of the virus and the emergence of new strains, will lead to the fact that the end of the pandemic will not be instantaneous. The most likely scenario is the migration of disease flashes from region to region, and a gradual decrease in the number of cases with the increasing of the immunized population.

In this regard, the analysis of socio-economic factors affecting the spread of the epidemic in its initial phase provided in [1], remains relevant. We have updated, its results by treating the new data on the Covid-19 pandemic and the socio-economic state of the regions of the Russian Federation. This allows to confirm the conclusions made in [1]. The main of them are listed below.

The most significant factors affecting the spread of the COVID-19 epidemic (these indicators have the greatest positive correlation coefficients with the number of cases, the rate of increase in incidence, and the duration of the course of the disease) are:

- Average annual number of employed, thousand people.
- Gross regional product.
- Retail turnover.
- Number of unemployed.
- The number of foreign citizens who have a patent for employment.
- The number of enterprises operating in the field of culture, sports, leisure and entertainment, organizations.
- The number of students in educational programs of primary, basic, and secondary general education.
- Passenger transportation by public buses.
- Passenger traffic by public buses.

These factors indicate the adequacy of the measures taken to prevent the explosive spread of the pandemic, although they may not have been consistent and timely everywhere. On the other hand, the listed risk factors should be used as the basis for providing the standard anti-epidemic measures in the cases of the flashes of SARS-CoV-2 - type diseases.

5 Conclusions

The mathematical model proposed in this work gives logically consistent description of the COVID-19 pandemic spread. It should be emphasized that the model relies on data that can be obtained by means of national system of macro-medical statistics and does not require the analysis of individual case histories. The last data is much more difficult to obtain.

The analysis of the various factors influencing on the spread of the disease indicates that it is worth to develop the methods of mathematical probability modeling of epidemics, especially since there are good traditions of such researches in the Russian Federation. This

will make possible to develop a reliable scientific base for various anti-epidemic measures and decisions. It could be recommended to analyze from this point of view the strategies for the socio-economic development of Russian regions for the period up to 2035 and make the necessary adjustments. Such activities and adjustments should be based on the models similar to the model outlined in this paper.

References

1. E.V. Sinitsyn, A.V. Tolmachev, A. S. Ovchinnikov, R-Economy, **6(3)**, 129 (2020)
2. L. A. Rvachev, Experiment on machine forecasting of influenza epidemic, **198(1)**, 68 (1971)
3. P. Whittle, Biometrika, **42(1-2)**, 116 (1955)
4. N. T. J. Bailey, The Mathematical Theory of Infectious Diseases and Its Application (1975)
5. N. S. Barlow, S. J. Weinstein, Physica D: Nonlinear Phenomena, **408**, 132540 (2020)
6. F. Ndairou, I. Area, J. J. Nieto, D. F. M. Torres, Chaos, Solitons & Fractals, **135**, 109846 (2020)
7. M. S. Abdo, K. Shah, H. A. Wahash, S. K. Panchal, Chaos, Solitons & Fractals, **135**, 109867 (2020)
8. X. Zhang, R. Ma, L. Wang, Chaos, Solitons & Fractals, **135**, 109829 (2020)
9. T. Chakraborty, I. Ghosh, Chaos, Solitons & Fractals, **135**, 109850 (2020)
10. J. Leskovec, A. Rajaraman, J. D. Ullman, Mining of Massive Datasets (2020)
11. Yu. M. Svirezhev, Nonlinear waves, dissipative structures and catastrophes in ecology (1987)
12. L. E. Elsgolz, Differential Equations (2020)
13. G. I. Ivchenko, Yu. I. Medvedev, Mathematical Statistics (2014)
14. N. N. Amosova, B. A. Kuklin, S. B. Makarova, Yu. D. Maksimov, Probabilistic sections of mathematics (2001)
15. A. A. Barsegyan, M. S. Kupriyanov, V. V. Stepanenko, I. I. Kholod, Data analysis technologies, Data mining, Visual mining, Text mining, OLAP (2007)
16. *Regions of Russia, Socio-economic indicators*, 32 (2018)
17. P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, R. Wirth, *CRISP-DM 1.0 Step-by-step data mining guide* (2000)