Calibrating the parameters of the cholera epidemic spread model

Natalia Baturina[1*]

[1]Don State Technical University, 1, Gagarin Square, 344003 Rostov-on-Don, Russia

Abstract. A system-dynamic model of cholera spread, which takes into account different ways of infection and restrictive measures, is developed. The model allows calibration of a large number of system parameters based on data on the number of infected real observations. Three scenarios of epidemic development (blurred peak, pronounced peak, plateau and two peaks) are investigated. The developed model can be modified for other scenarios and epidemics. Its implementation does not require large computing and human resources. It can be used to pre-predict the dynamics of the epidemic, as well as to organize effective measures to prevent threats associated with a pandemic.

1 Introduction

The past pandemic COVID-19 has caused an increase in research related to both pandemics in general and specific aspects of specific infections. A global trend is the improvement of data access technologies around worldwide and the application of artificial intelligence methods. Currently, it is possible to analyze and process large amounts of data, which is necessary for researchers and allows them to give more accurate pandemic forecasts, to apply effective strategies to prevent pandemic threats.

Along with filtering methods, regression and network models used in the study of the spread of infections earlier, machine learning methods, convolutional deep learning neural networks are used. They allow analyzing large epidemiological data, carrying out short-term and long-term forecasting [1-5]. The paper [6] reviewed agent-based models of epidemics, analyzed their architecture, identified the main design concepts and key components for modeling epidemic processes. Many studies are characterized by a combination of models of different types. The papers [7-11] present the results obtained using a software package created at the Siberian Branch of the Russian Academy of Sciences. Both forward and backward epidymology problems are considered. Models based on systems of differential equations, agent-oriented and models of middle field games are used. Such an integral approach makes it possible to very accurately predict the dynamics of the development of a pandemic, to take into account heterogeneity in the behavior and strategies of people in the population. In addition, sensitivity analysis of model parameters to data variation is carried out, as well as sensitivity analysis of model states in relation to parameter variation [12]. In research of this scale, a huge amount of human and material resources are involved.

* Corresponding author: baturinata@mail.ru

In addition to new types of models, the use of classic SEIR approach and its modifications. In the system of differential equations, time delays are introduced to accurately reproduce the duration of infectious processes with several waves of virus spread, or a piecewise constant function for controlling restrictive measures is added [13-14]. The paper [15] proposes modifications of the SEIR model (model with a long plateau and several waves). In the study [16], a modified SIRS model of the spread of epidemics in the form of a lattice of stochastic cellular automata is constructed. The model uses dynamic population regulation. Depending on the control parameters, steady-state modes are defined. It has been shown that the behavior of the model of cellular automata generally corresponds to the model of the middle field.

Cholera is one of the most dangerous infections, therefore research institutes dealing with the problems of especially dangerous infections annually analyze the epidemiological situation in cholera in Russia and the world, make a forecast for the future [17-19]. In the work [17], an assessment of the situation with cholera in Russia and the world in 2013-2022 was carried out. In particular, it is noted that the risks of importing infection into the territory of the Russian Federation associated with the intensification of the epidemic process in Asia, Africa and the Caribbean region remain. The long border with Ukraine, where cholera is likely to be imported from endemic countries, also contributes to the increase in the degree of risks, and bioterrorism is not excluded. The emergence of new Vibrio cholerae mutations is also noted, which gives a potential threat to the spread of epidemic cholera in Russia and other territories. Despite the absence of real epidemic threats [17], studies of cholera as a particularly dangerous infection are relevant for the development of strategies and measures to control and prevent infection. Along with epidemiological, bacteriological, molecular genetic, clinical, immunological studies of cholera, model studies are important.

Simulation modeling is one of the methods of research on models. For these purposes, AnyLogic can be used, which is a multifunctional modeling tool [20]. It combines flexibility, usability, including without the use of large computing power. It enables the integration of various types of models: agent, macroscopic and discrete-event. AnyLogic can be used to represent the epidemiological situation, pre-predict the spread of the disease, assess the effectiveness of restrictive measures, calibrate the model using real data.

In [21], a combined model of the spatial spread of epidemic diseases (using the example of cholera) based on a probabilistic cellular automaton and a compartment model is considered. The use of cellular automata in the study of cholera makes it possible to create a discrete model that can take into account various spatial and temporal factors.

Despite the significant advantages that modeling provides, the variety of approaches, there are also disadvantages in existing methods for researching pandemics, including cholera. One of the main drawbacks of research is the availability and quality of data. Along with complex modeling systems, the development of models that do not require large computing power that can take into account changing epidemiological data remains relevant.

In works [22-23], the authors considered simulation models (agent and system-dynamic) of cholera distribution created in AnyLogic. The models took into account various routes of cholera infection, sampling from infected water or food sources, vaccination processes and external importation of infection. This work is a development of previous studies, based on studies by other authors.

2 Materials and methods

Let us consider a model including two parallel flow scheemes. The first scheme models the dynamics of the epidemic based on real statistical data. It allows you to incrementally refine the model parameters, which are then used in the second streaming scheme. The second

scheme is used to long-term predict the development of the epidemic, based on current observations.

To develop scheme 1, a system of differential equations of the SEIR system dynamics model was used:

$$\frac{dS}{dt} = -r_E, \quad \frac{dE}{dt} = r_E - r_I, \quad \frac{dI}{dt} = r_I - r_R, \quad \frac{dR}{dt} = r_R.$$

Let us add the second and third equations and enter the notation

$$EI = E + I, \quad r_{EI} = r_E - r_R. \tag{1}$$

As a result, we get a system

$$\frac{dS}{dt} = -r_{EI} - r_R, \quad \frac{d(EI)}{dt} = r_{EI}, \quad \frac{dR}{dt} = r_R, \tag{2}$$

where $r_{EI}-$ is the infection flow rate.

Let us denote $r_{EI}(t_j)$ – rate of infection flow at time t_j , $j = \overline{1,m}$, where $m -$ number of measurements. At the initial stage of infection spread, this rate is approximately by real statistical data. This data (number of infected $I(t_j)$) is preliminarily recorded in the db_table_inf_real table of the AnyLogic database. Further, $r_{EI}(t_j)$ are determined by the formula

$$r_{EI}(t_j) = \frac{I(t_{j+1}) - I(t_j)}{t_{j+1} - t_j}, j = \overline{1, m-1}.$$

The results of the calculations are recorded in the database db_table_speed_inf_real table, which creates the tableSpeedInfRealFunction as a result of interpolation.

The database tables and functions are described in Table 1.

The flow diagram 1 corresponding to system (2) is shown in Figure 1.

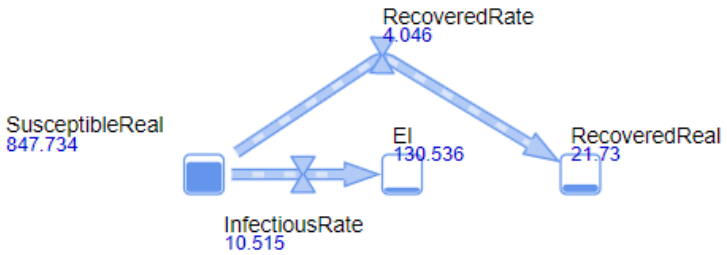

Fig. 1. Scheme 1.

Flow values are determined by formulas (description of variables is given in Table 1):

InfectiousRate=tableSpeedInfRealFunction(time());

RecoveredRate=EI/IverageIllnessDuration;

Flow diagram 2, shown in Figure 2 is a modified version given in [23]. Here, restrictive measures are additionally taken into account, flow formulas for water and food methods of infection are clarified. Scheme 2 is based on the following system of differential equations:

$$\frac{dS}{dt} = r_{SI} - r_{EC} - r_{EW} - r_{EF} - r_{SO} - r_L,$$

$$\frac{dE_C}{dt} = r_{EC} - r_{IC}, \quad \frac{dE_W}{dt} = r_{EW} - r_{IW}, \quad \frac{dE_F}{dt} = r_{EF} - r_{IF}, \quad \frac{dE_D}{dt} = r_{ED} - r_{ID},$$

$$\frac{dI_C}{dt} = r_{IC} - r_{RC}, \quad \frac{dI_W}{dt} = r_{IW} - r_{RW}, \quad \frac{dI_F}{dt} = r_{IF} - r_{RF}, \quad \frac{dI_C}{dt} = r_{ID} - r_{RD},$$

$$\frac{dR}{dt} = r_{RC} + r_{RW} + r_{RF} + r_{RD}, \qquad \frac{dL}{dt} = r_L,$$

$$S(0) = N - I_C(0) - I_W(0) - I_F(0) - I_D(0), \; R(0) = 0,$$

$$E_C(0) = E_F(0) = R(0) = 0, E_W(0) = 1, \; I_C(0) = I_F(0) = I_D(0) = I_W(0) = 0.$$

Here

$$r_{SI} = i_I(1 - p_D), \; r_{SO} = i_O, \; r_L = \frac{S}{N} \cdot i_L,$$

$$r_{EC} = \frac{S^{kC}}{N} \cdot I \cdot n_C, \quad I = I_C + I_W + I_F + I_D,$$

$$r_{EW} = \frac{S^{kw}}{N} \cdot n_W, \; r_{EF} = \frac{S^{kF}}{N} \cdot n_F, \; r_{ED} = i_I \cdot p_D, \; n_C = i_C \cdot p_C \tag{3}$$

$$r_{IC} = \frac{E_C}{t_{in}}, \; r_{IW} = \frac{E_W}{t_{in}}, \; r_{IF} = \frac{E_F}{t_{in}}, \qquad r_{ID} = \frac{E_D}{t_{in}},$$

$$r_{RC} = \frac{I_C}{t_{il}}, \quad r_{RW} = \frac{I_W}{t_{il}}, \quad r_{RC} = \frac{I_F}{t_{il}}, \qquad r_{RD} = \frac{I_D}{t_{il}}.$$

Note that the waterway of infection is decisive in the spread of cholera infection. The rate of flow of infection by contact depends on the total number of infected. The rate of infection by water does not depend on the total number of infected. It is proportional to the number of positive samples [23]. The most significant impact on the epidemiological process is exerted by the nw coefficient −the intensity of infection by water. It depends on many different factors (virulence of the pathogen, immunity of susceptible, restrictive measures, etc.). The coefficient nw is not constant during the development of the epidemic. Presumably, its dependence on time is decreasing.

Scheme 1 allows to determine empirical values of nw from the collected real statistical data. These values are then used in Scheme 2 for long-term disease prediction.

From (1) and (3) we express the expression

$$r_E = r_{EI} + r_R, \quad r_{EW} = \frac{S^{kw}}{N} \cdot n_W.$$

Believing $r_{EW} \approx r_E$, we will find

$$n_W \approx \frac{r_{EI} + r_R}{S^{kw}} \cdot N. \tag{4}$$

Formula (4) is used to collect RIW_ST statistics on the values of the coefficient n_W. These statistics are collected as a result of the operation of the streaming Scheme 1. The statistics statement is written in the "Action" section of the "Event" element:

```
RIW_ST.add(TotalPopulation*(tableSpeedInfRealFunction(time())+
EI/IverageIllnessDuration)/SusceptibleReal);
```

Scheme 2 is shown in Figure 2. It includes the drives of people in different phases in relation to the disease. It also takes into account various ways of contracting cholera.

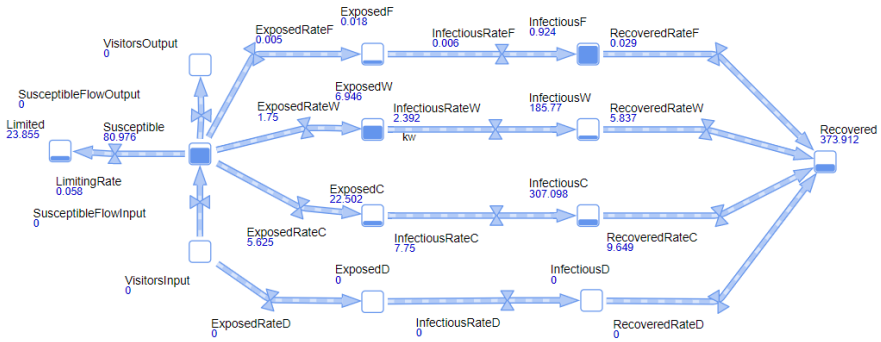

Fig. 2. Scheme 2.

The description of model parameters and variables is given in Table 1. The addition of the symbols C, W, F, D at the end of the model identifiers means relation to contact, water, food-borne and importation modes of infection.

Table 1. Tables, functions, parameters, model variables.

Name	Designation On flow diagram	Designation
Database tables and functions		
Table of actual infected persons	db_table_speed_inf_real	
Tabular function on actual infected persons	tableSpeedInfRealFunction	
Parameters		
Total Population	TotalPopulation	N
Probabilities of transmission	Infectivity C|D	p_C, p_D
Contact intensities	RateInfectious C	i_C
Intensity of restrictive measures	RateLimiteded	i_L
Intensities of inbound and outbound flows	RateFlowInput RateFlowOutput	i_I, i_O
Infestation intensity coefficients	RI C|W|F	n_C, n_w,n_F
Duration of the incubation (latency) period	IveragelncubTime	t_{in}
Duration of infection	IveragelllnessDuration	t_{il}
Parameters of infection rates	kW, kF,kC	kW, kF,kC
Levels		
Number of susceptibles in schemes 1 and 2	SusceptibleReal Susceptible	S
Number of immune due to restrictive measures	Limited	L
Number of infected in latent stage	Exposed C|W|F|D	E_C, E_W E_F ,E_D
Number of people infected in the active phase of the disease	Infectious C|W|F|D	I_C, I_W I_F, I_D
Total number of infected people	Infectious	I
Number of people who recovered	Recovered	R

Number of arrivals	VisitorsInput	V_I
Number of people leaving	VisitorsOutput	V_O
Flows		
Rate of flow of commuters	SusceptibleFlowInput	r_{SI}
The speed of the flow of people leaving	SusceptibleFlowOutput	r_{SO}
Infection flow rates	ExposedRate C|W|F|D	r_{EC}, r_{EW} r_{EF}, r_{ED}
Flow rates of the transition to the active phase of the disease	InfectiousRate C|W|F|D	r_{IC}, r_{IW} r_{IF}, r_{ID}
The speed of the recovery flow	RecoveredRate	r_R
Flow rate of restrictive measures	LimitedRate	r_L

Sliders are used to control the model parameters. The sliders allow to adjust the following parameters: transmission probabilities, infection intensity coefficients, intensity of restrictive measures, and others. Initially, all parameters have default values. When the model is started, a pause is provided to adjust the parameters using the sliders. If the user does not adjust the parameters using sliders, the model is run with default parameter values. As a result of the model run, a single graph can show the curves of change over time of the number of actually infected (Scheme 1) and infected according to the model (Scheme 2). In addition, the parameters can be changed during the simulation using sliders. Through the collection of RIW_ST statistics, empirical values of the coefficient and its modelled values are compared. The empirical values can be transferred to the modelling system. In case of a large discrepancy between the curves of the number of infected persons and those infected according to the model, the model parameters should be adjusted, in particular, the model value of the coefficient should be set close to its empirical value.

3 Results

Three epidemic scenarios are considered: with a blurred peak, with a pronounced peak, and with a plateau and several peaks. The paper presents the results of modelling for 250 days. Statistical data corresponding to each scenario are recorded in the database db_table_speed_inf_real. The table functions tableSpeedInfRealFunction are formed on them, which are used to calibrate the model parameters.

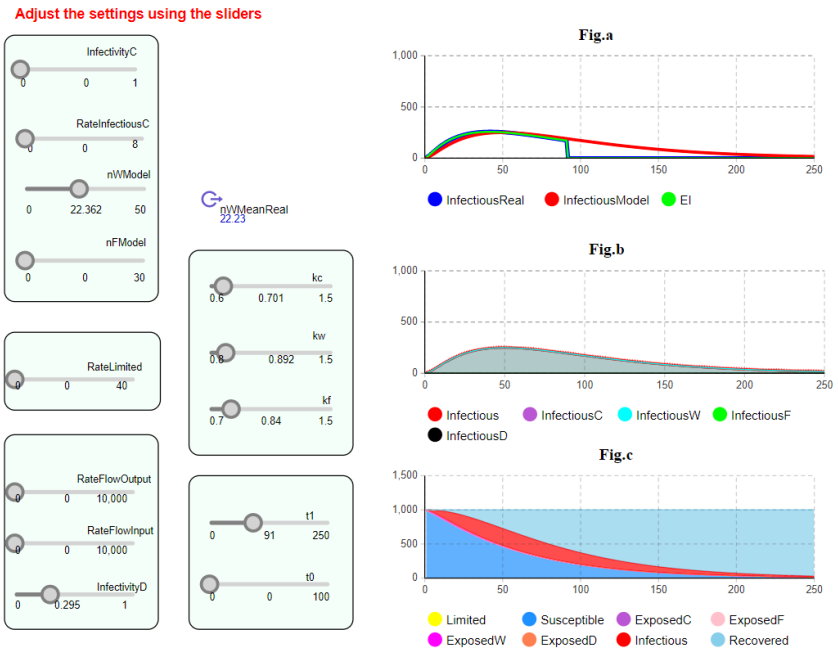

Fig. 3. Scenario of an epidemic with a blurred peak.

In Fig. 3. the sliders show the parameter values used in the modelling. Fig.a shows three graphs corresponding to the data of actual infected people and the data obtained by streaming Scheme 1 and Scheme 2. Fig .b displays plots of the variation in the number of infected corresponding to the different modes of infection. Fig. c shows plots of the change in the number of people in different states in relation to the disease.
Calibration of the model in the first case showed that such a scenario is characteristic if one mode of infection , waterborne, is significantly predominant, with the coefficient of intensity of waterborne infection $n_w \approx 22$. In Fig. 3.(Fig.a) the time interval t1=90 days is used for parameter selection and further prediction of the infection dynamics.
The introduction of restrictive measures (Fig. 4) increases the rate of infection attenuation. The model allows determining the required intensity of their implementation. For the case under consideration, the value of of the variable i_L=20.

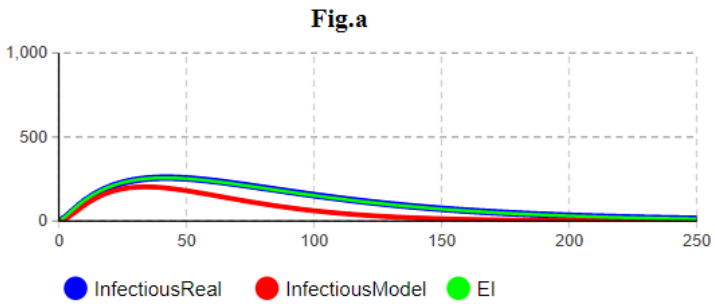

Fig. 4. Epidemic development during the introduction of restrictive measures.

Calibration of the model in the second case (Fig. 5) showed that this scenario is characteristic if infection occurs through several pathways (water, contact, food). The values of the model parameters found from the model runs are presented on the sliders.

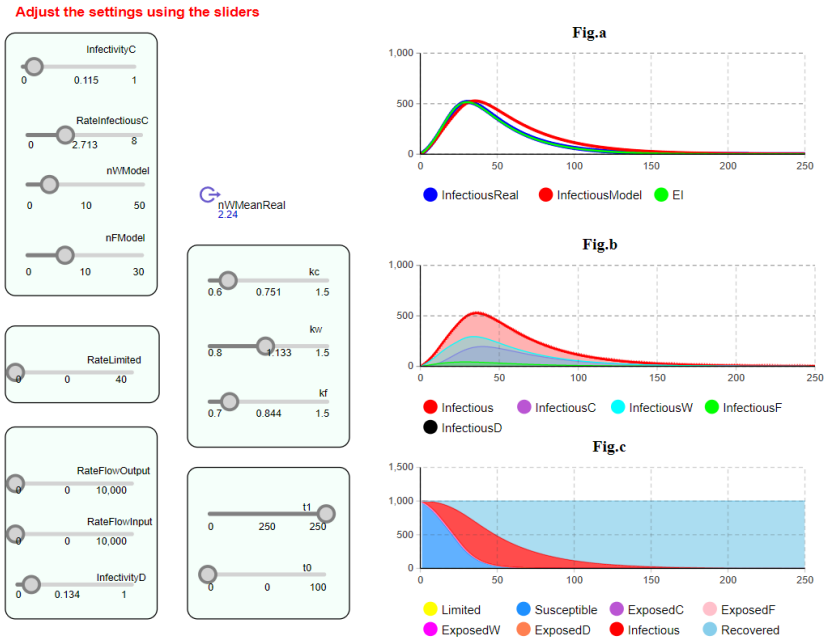

Fig. 5. Epidemic scenario with a pronounced peak.

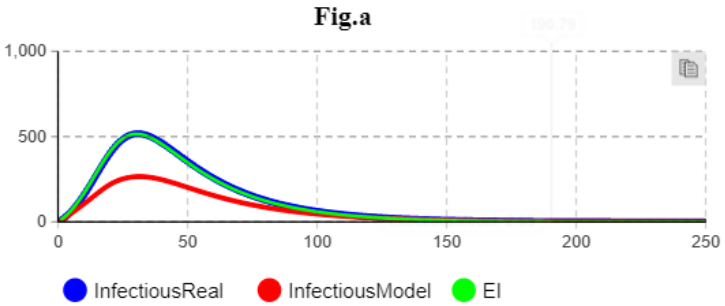

Fig. 6. Epidemic scenario with a pronounced peak.

The introduction of restrictive measures with an intensity of i_L=30 halves the peak of infection development.

Fig. 7 shows graphs of infection development with a plateau and two peaks. For its realisation in Scheme 2, the rates of infection flows were equal to zero during a given time interval (t1,t2). The values of t1,t2 were selected based on the data of the actual infected in Scenario 3. The following code was used for the input flow rate of the ExposedW drive :

time()<t0 || (time()<t3 & time()>t2) ? 0 : pow(Susceptible,kw)/TotalPopulation*RIW.

For other drives, the code has the same structure.

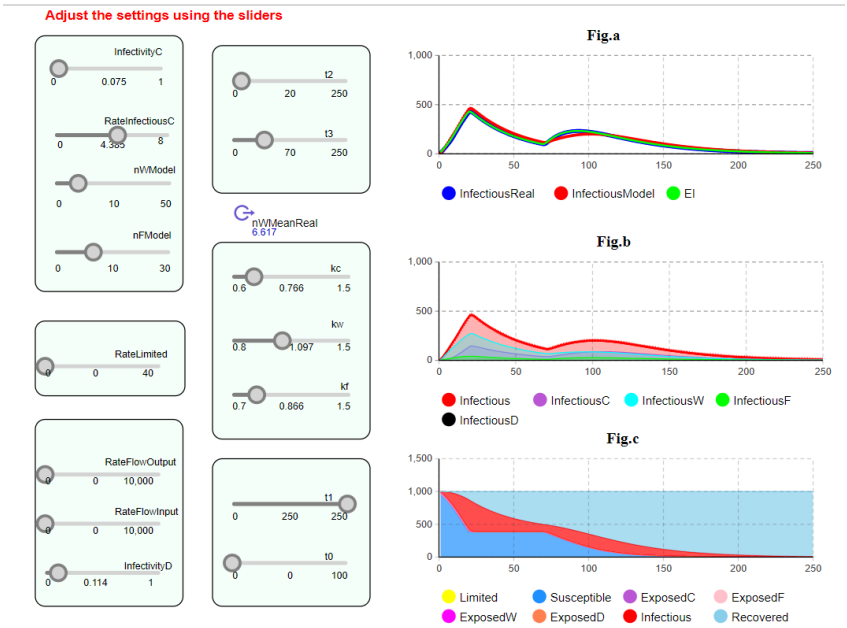

Fig. 7. Epidemic scenario with a plateau and two peaks.

4 Conclusion

Modern methods and approaches to study the epidemiological process are reviewed. A system-dynamic model of cholera spreading has been developed, taking into account different ways of infection and restrictive measures. The model allows calibrating a large number of system parameters based on the data on the number of infected real observations. Observational data are pre-recorded in database tables, which is connected during the modelling process. Three scenarios of epidemic development (blurred peak, pronounced peak, plateau and two peaks) are investigated. Parameter values corresponding to the considered scenarios are found. The developed model can be modified for other scenarios and epidemics. It does not require large computational and human resources for its implementation. It can be used to predict the dynamics of the epidemic development in advance, as well as to organise effective measures to prevent pandemic-related threats.

References

1. M. Kondrat'ev. Computer research and modelling, **5(5)**, 863-882 (2013) https://doi.org/10.20537/2076-7633-2013-5-5-863-882.
2. V. Zakharov, Yu. Balykina. Informatics and tomatization, **20(5)**, 1034–1064 (2021) https://doi.org/10.15622 /20.5.2.
3. O. Krivorotko, N. Zyatkov, S. Kabanikhin. Computational Mathematics and Mathematical Physics**, 63(10)**,1929-1941 (2023) https://doi.org/10.1134/s096554252310007x.
4. A. Muhammad, J. Shumaila, A. Ruhul, et al. Bioengineering (Basel),**10(7)**, 850 (2023) https://doi.org/10.3390/bioengineering10070850.
5. Z. Tarek, M. Shams, S. Tovfek, H. Alqahtani, et al. Bioengineering (Basel), **8(7)**, 552 (2023) https://doi.org/10.3390/biomimetics8070552.

6. A. Ageeva. Modeling, optimization and information technology, **3(30)**, 3-4 (2020) https://doi.org/10.26102/2310-6018/2020.30.3.030.
7. O. Krivorotko, S. Kabanikhin, N. Zyatkov. Numer. Analys. Appl. **13,** 332–348 (2020) https://doi.org/10.1134/S1995423920040047.
8. O, Krivorotko , S. Kabanikhin, M. Sosnovskaya, D. Andornaya. *Sensitivity and identifiability analysis of COVID-19 pandemic models.* Vavilov Journal of Genetics and Breeding, 82-91 (2021) https://doi.org/10.18699/VJ21.010.
9. T. Vonareva, O. Krivorotko. Computational Mathematics and Mathematical Physics, **62(4)**,674-684 (2022) https://doi.org/10.1134/s0965542522040145
10. O. Krivorotko, M. Sosnovskaia, S. Kabanikhin. Journal of Inverse and Ill-Posed Problems, **3**, 1-17 (2023) https://doi.org/10.1515/jiip-2021-0038.
11. A. Neverov, O . Krivorotko. *Numerical Modelling of Mean-Field Game Epidemic. Optimization and Applications* : 14th International Conference, OPTIMA 2023, Petrovac, Montenegro, September 18–22, 2023, Revised Selected Papers, Springer, 207-217 (2023) https://doi.org/10.1007/978-3-031-47859-8_15.
12. O. Krivorotko, S. Kabanikhin,V. Petrakova. Mathematical biology and bioinformatics, **18(1),** 177-214 (2023) https://doi.org/10.17537/2023.18.177.
13. I. Kiselev, I. Akberdin, F. Kolpakov. *A delay differential equation approach to model the COVID-19 pandemic.* MedRxiv (2021) https://doi.org/10.1101/2021.09.01.21263002
14. C. Silva, C. Cruz, D. Torres, et al. Scientific Reports, **11**, 3451 (2021) https://doi.org/10.1038/s41598-021-83075-6.
15. V.I. Baran, E.P. Baran. Bulletin of the Russian University of Cooperation, **3(45),** 9-13(2021) https://doi.org/10.52623/2227-4383-3-45-2.
16. A. Shabunin. Applied nonlinear dynamics. Bulletin of the Russian University of Cooperation, **27 (2)** ,5–20 (2019) https://doi.org/10.18500/0869-6632-2019-27-2-5-20 .
17. A. Noskov, V. Kruglikov, E. Moskvitina, et al. Problems of Particularly Dangerous Infections,**1**, 56-66 (2023) https://doi.org/10.21055/0370-1069-2023-1-56-66.
18. Noskov, V. Kruglikov, E. Moskvitina, et al. Problems of Particularly Dangerous Infections, **1**, 24-34 (2022) https://doi.org/10.21055/0370-1069-2022-1-24-34.
19. E. Monakhova, A. Ghosh, A. Mutreja, F. Weill. Problems of especially dangerous infections, **3,** 17-26 (2020) https://doi.org/10.21055/0370-1069-2020-3-17-26
20. A.Boyko, S. Evdokimova, T. Novikova. *Simulation modeling: types and applications New aspects of modeling systems and processes:* Materials of the International Scientific and Practical Conference, 18-26 (2023) https://doi.org/10.58168/NAMSP_18-26.
21. M. Bashabsheh. Naukovedenie Internet journal, **6**, 127-135 (2013)
22. N. Baturina, G. Anisimova. E3S Web of Conferences **224**,03024 (2020) https://doi.org/10.1051/e3sconf/202022403024.
23. N. Baturina, G. Anisimova. E3S Web of Conferences **371**,01053 (2023) https://doi.org/10.1051/e3sconf/202337101053.