

# Simulation of COVID-19 Spread Scenarios in the Republic of Kazakhstan Based on Regularization of the Agent-Based Model

O. I. Krivorotko<sup>1,2,3\*</sup>, S. I. Kabanikhin<sup>2,3\*\*</sup>, M. A. Bektemesov<sup>4\*\*\*</sup>,  
M. I. Sosnovskaya<sup>3\*\*\*\*</sup>, and A. V. Neverov<sup>1,3\*\*\*\*\*</sup>

<sup>1</sup>*Institute of Computational Mathematics and Mathematical Geophysics, Siberian Branch,  
Russian Academy of Sciences, Novosibirsk, 630090 Russia*

<sup>2</sup>*Sobolev Institute of Mathematics, Siberian Branch, Russian Academy of Sciences,  
Novosibirsk, 630090 Russia*

<sup>3</sup>*Novosibirsk State University, Novosibirsk, 630090 Russia*

<sup>4</sup>*Abai Kazakh National Pedagogical University, Almaty, 050010 Kazakhstan*  
*e-mail: \*krivorotko.olya@mail.ru, \*\*ksi52@mail.ru, \*\*\*maktagali@mail.ru,*  
*\*\*\*\*m.sosnovskaya@alumni.nsu.ru, \*\*\*\*\*a.neverov@g.nsu.ru*

Received July 4, 2022; revised September 27, 2022; accepted September 28, 2022

**Abstract**—We propose an algorithm for modeling scenarios for newly diagnosed cases of COVID-19 in the Republic of Kazakhstan. The algorithm is based on treating incomplete epidemiological data and solving the inverse problem of reconstructing the parameters of the agent-based model (ABM) using the set of available epidemiological data. The main tool for constructing the ABM is the *Covasim* open library. In the event of a drastic change in the situation (appearance of a new strain, removal or introduction of restrictive measures, etc.), the model parameters are updated taking into account additional information for the previous month (online data assimilation). The inverse problem is solved by stochastic global optimization (of tree-structured Parzen estimators). As an example, we give two scenarios of COVID-19 propagation calculated on December 12, 2021 for the period up to January 20, 2022. The scenario that took into account the New Year holidays (published on December 12, 2021 on <http://covid19-modeling.ru>) almost coincided with what happened in reality (the error was 0.2%).

**Keywords:** *agent-based model, COVID-19, inverse problem, optimization, regularization, scenario, basic reproduction number*

**DOI:** 10.1134/S1990478923010118

## INTRODUCTION

From December 2019 to September 2022, more than 604 million cases of COVID-19 infection were registered in the world, of which approximately 6.5 million were fatal. As of September 5, 2022, the Republic of Kazakhstan (RK) ranks 51st among 216 countries in terms of the number of detected cases of COVID-19 and 119th in terms of the share of the population vaccinated against COVID-19. Agent-based models of the spread of infectious diseases are used to monitor the epidemiological situation in regions and countries as well as to analyze the efficiency of containment measures.

The agent-based model (ABM) of the spread of COVID-19 allows one

- To take into account demographic information for a specific country (the number and age structure of the population).
- To build realistic transmission networks in various social strata, including households, schools, organizations, and public places.

- To take into account the age-related features of the development of the disease.
- To determine the viral burden of the agent, including the rate of transmission of the infection.
- To take into account physical distancing and wearing of masks, vaccination, testing (including asymptomatic), isolation, contact monitoring, quarantine in regions with 100 to 20 000 000 people [1–3].

In the agent-based model, each acting agent is equipped with attributes (age, social status, susceptibility to disease, etc.).

#### *Brief Survey of Papers on COVID-19 ABM*

Here is a brief survey of the publications on agent-based modeling of COVID-19, which are largely based on earlier work on the modeling of infectious diseases in 2013 [4] and 2020 [5].

Aleta et al. [6] used ABM to describe the spread of COVID-19 in Boston. A response system based on advanced testing and contact tracing has been shown to play an important role in easing social distancing restrictions in the absence of herd immunity against SARS-CoV-2.

Lau et al. [7] calculated, using the example of the state of Georgia in the USA, that infected people under 60 years of age can be 2.78 times more contagious than older people and tend to be the main driving force of superspreading.

In [8–10], antiepidemic programs in various regions of France and the UK were analyzed based on the ABM of the spread of COVID-19. It was shown that to control a new wave of the COVID-19 outbreak, it suffices to use contact and isolation data over the most recent three months.

Nielsen and Sneppen [11] constructed an ABM in which superspreaders acted as infection sources, and it was shown that superspreading dramatically increases the importance of restrictions on personal contacts.

In 2021, a group of American scientists [3] developed the *Covasim* software package [12], which is based on an agent-based approach to modeling an epidemic taking into account the characteristics of the disease as well as pharmaceutical (vaccination) and social (restricted visits, wearing masks) measures. This software package has been used to build COVID-19 epidemic scenarios, study pandemic dynamics, and support administrative decision-making in more than a dozen countries in Africa, Asia-Pacific, Europe, and North America.

#### *Open-Source Software Packages*

Here is a brief overview of the software systems on the spread of COVID-19.

1. <http://covid19-scenarios.org>, Basel University, Switzerland. An age-structured compartmental model of the spread of coronavirus infection is being implemented, based on 9 differential equations with the possibility of varying model parameters [13].
2. <http://covid19.biouml.org>, Institute of Computational Technologies, Siberian Branch of the Russian Academy of Sciences, Novosibirsk. The spread of COVID-19 in Moscow, the Novosibirsk oblast, Germany, France, and Italy is modeled. The package uses extended compartmental and agent-based models whose parameters are identified on the basis of published statistics. Forecasts are made not only of the number of registered cases, cured and dead, but also of the number of free beds, artificial lung ventilation (ALV) devices, and other characteristics necessary for effective management of the situation in an epidemic.
3. <http://anylogic.com/healthcare>, Bogota, Colombia. The package relies on an agent-based model taking into account the geographical features of the city (location of schools, medical institutions, public places) and the social distance between the agents with the possibility of varying parameters.
4. <http://github.com/kausaltech/reina-model>, Helsinki, Finland. Based on the agent-based model, different ages of agents and seven possible states of disease progression are taken into account with a random structure of contacts also considered; i.e., the agents interact randomly [14].
5. <http://github.com/institutefordiseasemodeling/covasim>, USA. The package is built around an agent-based model with random structure with a possibility of identifying param-

eters for a particular region [3]. It served as the basis for creating our own software package (see item 6) and was also adapted to obtain results for the Republic of Kazakhstan in the framework of the present paper.

6. <http://covid19-modeling.ru>, Institute of Computational Mathematics and Mathematical Geophysics, Siberian Branch of the Russian Academy of Sciences, Novosibirsk. The complex is based on combining compartmental and agent-based models and provides the ability to construct development scenarios by solving inverse problems [15, 16].

### *Inverse Problems for ABM*

The same virus can spread and affect people differently in different regions. Due to the novelty and complexity of the COVID-19 disease (frequent changes in the infectiousness of the virus and the average age of the severe course of the disease, the duration of the incubation period, etc.), the parameters of most mathematical models are usually unknown; this makes it difficult to adapt existing software solutions to analyze the situation in a particular region taking into account the introduction of restrictions during different periods and pharmaceutical interventions. The main challenges in modeling the spread of COVID-19 are as follows.

1. The data for solving the inverse problem is incomplete and noisy and also represents big data (daily reports of sick, infected, and vaccinated people and so on).
2. Parameters such as virus contagiousness  $\beta(t)$ , probability of severe cases  $p_{\text{sev}}(t)$ , mortality  $p_{\text{death}}(t)$ , etc. change over time.
3. The spread of COVID-19 changes substantially when restrictive measures are introduced or lifted (masks, social distancing, switching to a remote working mode, closing schools, enterprises, districts and cities).

These problems lead to the need to consider and solve inverse problems of identifying unknown epidemiological parameters in a particular region taking into account the mutation of the virus and various (administrative, pharmaceutical) interventions based on additional information on the number of PCR tests performed, detected cases, hospitalized and critically ill patients, and deaths from COVID-19. Due to the ill-posedness of inverse problems (the solution may be nonunique and/or unstable), regularization is used taking into account restrictions on the desired parameters obtained from the data of the World Health Organization (WHO) and the Ministry of Health of the region under study (in our case, the Republic of Kazakhstan).

### *Specific Features of the ABM in the Republic of Kazakhstan*

The specific feature of the present work lies in the adaptation of the agent-based model to the Republic of Kazakhstan taking into account the distribution of the population (by age and region) as well as restrictive measures and the type and quality of statistical (incomplete and inaccurate) data. As of 2021, 40.8% of the population of the Republic of Kazakhstan was rural, and its density is comparable to that of the urban population.

We assumed that the urban and rural population of the Republic of Kazakhstan is on an equal footing in terms of polymerase chain reaction (PCR) testing and the spread of COVID-19. The paper proposes an algorithm for step-by-step refinement of the epidemiological parameters of the ABM of spread of COVID-19 in the Republic of Kazakhstan, as well as an algorithm for constructing scenarios for the development of the epidemic in a region based on a combination of machine learning methods and solving inverse problems.

### *Paper Structure*

For reliable modeling of certain factors in the spread of COVID-19, the calculations should be based on adequate initial data—from the population and its distribution over social groups to the workload of various modes of transport or shops [1]. Due to incomplete and noisy statistical data, the paper uses methods of machine learning and regression analysis and of processing and analysis of time series. The time series (the number of PCR tests performed) is extrapolated based on statistical models so as to construct scenarios for the spread of COVID-19 used in the agent-based approach (Sec. 1). Based on the characteristics of the data for the Republic of Kazakhstan, in

Sec. 2 we construct an agent-based model whose main purpose is to compile development scenarios and assess the impact of various interventions on the epidemic and formulate the direct and inverse problems for the ABM. Section 3 describes an algorithm for solving inverse problem based on tree-structured Parzen estimators for minimizing the absolute objective functional and data assimilation. In Sec. 4, scenarios for the spread of COVID-19 in the Republic of Kazakhstan are constructed and analyzed taking into account restrictive measures from December 13, 2021 to January 20, 2022. It is shown that an increase in the concentration of agents on New Year's holidays in public places (shops, theaters, parks) increases the number of detected cases of COVID-19 (by January 15, 2022, it increased by 3.5 times compared to January 12, 2021). The main conclusions are given in the last section.

## 1. Analysis of Data for the Republic of Kazakhstan

Let us present the main demographic and epidemiological data for the Republic of Kazakhstan that were used in the subsequent construction of the agent-based model of the spread of COVID-19 described in Sec. 2.

### 1.1. Demographics

Statistical data on the population by age categories (age groups are divided into periods of 10 years) were provided by M. A. Bektemesov (Table 1). This kind of data is used when initiating an artificial ABM population (see Sec. 2.2). The total population of the Republic of Kazakhstan is 18 879 552 people.

### 1.2. Epidemiological Data

The data sources are

- <http://ourworldindata.org> (OWD).
- <http://worldhealthorg.shinyapps.io/covid> (WHO).
- <http://coronavirus2020.kz/ru> (CV2020).
- <http://kt.kz> (KT).

A program written in the Python programming language was created to collect and process epidemiological data. The scheme of the program is as follows.

1. Specify the URL of the resource from which the data is collected.
2. For news sites, carry out an article search using the site's tools and single out the necessary statistical data.
3. In the html markup of the page, select elements containing the following data using preselected keys:

**Table 1.** Distribution of the population by age in the Republic of Kazakhstan as of October 1, 2021

Age category (years)	Number of people
0–9	3 890 241
10–19	2 855 522
20–29	2 537 252
30–39	2 987 296
40–49	2 292 980
50–59	1 994 801
60–69	1 439 848
70–79	588 375
80+	293 237

- For news sites, find data by specified keywords in sentences.
- For sites with statistics, the data does not require additional search and is selected in a predetermined order.

The following epidemiological data on the prevalence of COVID-19 on day  $t$  were collected and processed:

- The number  $\mathcal{T}(t)$  of PCR tests performed (Fig. 1)—OWD.
- The number  $f(t)$  of diagnosed COVID-19 cases (PCR-positive patients) (Fig. 2)—OWD, WHO, CV2020, KT.
- The number of vaccinated people—OWD, WHO, KT.
- The number  $H(t)$  of hospitalizations patients with COVID-19—OWD (number of beds only), KT, CV2020.
- The number  $C(t)$  of patients connected to the ventilator—KT, CV2020.
- The number  $D(t)$  of deaths from COVID-19—OWD, WHO, CV2020, KT.

Some of the data posted on the indicated sites is incomplete (some days, weeks, and months are missing). Missing data of intermediate values in the time series were interpolated using cubic splines using the *interpolate* method of the `pandas` library [17].

### 1.3. Regression Model for Processing and Extrapolating Seasonal Time Series

The statistical data  $\mathcal{T}(t)$  on the number of PCR tests performed in the Republic of Kazakhstan have gaps since June 8, 2021, as well as weekly seasonality (periodicity); therefore, to build ABM-based scenarios for the spread of COVID-19, we used the time series  $\mathcal{T}(t)$  by the seasonal autoregressive model SARIMA, which is a modification of the model ARIMA (AutoRegressive Integrated Moving Average) [18], describing one-dimensional time series with a seasonal component.

The model SARIMA( $p, d, q$ )( $P, D, Q$ ) $_s$  for the nonstationary time series  $\mathcal{T}(n)$  has the form [19]

$$\Phi(L^s)\varphi(L)\Delta^d\Delta_s^D\mathcal{T}(n) = \theta_0 + \Theta(L^s)\theta(L)\varepsilon(n).$$

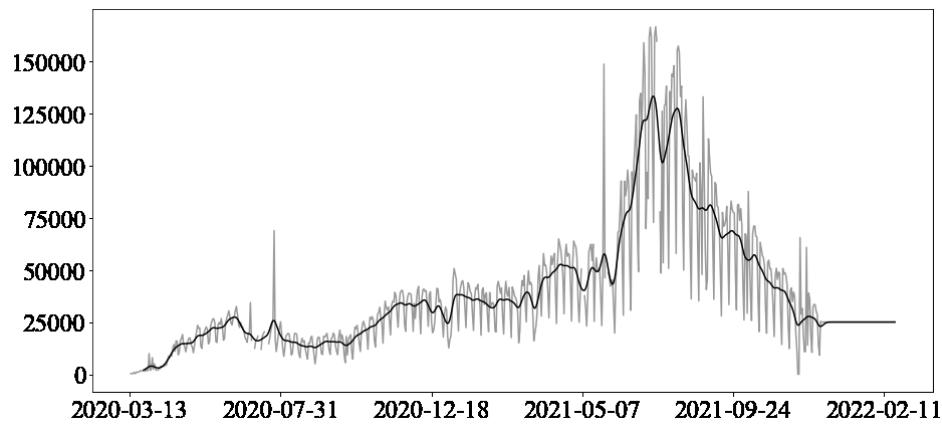
Here the parameters  $p, d, q$  correspond to the nonseasonal part of the time series, and  $P, D, Q$  correspond to the seasonal components of the series,  $s = 7$  is the season length,  $\varepsilon(n)$  is stationary time series of white noise,  $\Delta^d$  is the operator of the difference of the time series of order  $d$  that guarantees the series being stationary (successively taking the first-order differences  $d$  times, first of the time series and then of the resulting differences of the first order, then of the second order, and so on),  $\Delta_s^D$  is the operator of the difference of the time series of order  $D$  for the seasonal component,  $\varphi$  and  $\Phi$  are the parameters of the autoregression for nonseasonal and seasonal components of the series,  $\theta$  and  $\Theta$  are the parameters of the nonseasonal and seasonal sliding average, respectively, and  $n$  is a temporal parameter (days).

The time series extrapolation algorithm is as follows.

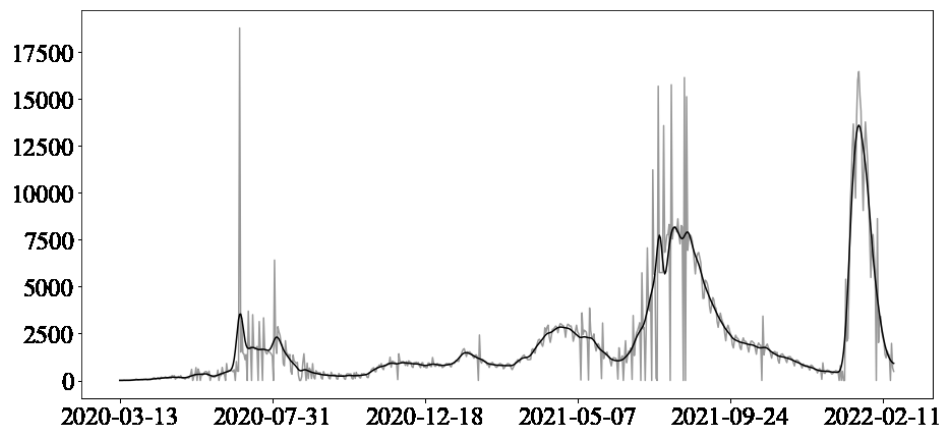
- STEP 1. Apply the Box–Cox transform [18] to reduce the variance.
- STEP 2. Calculate the seasonal difference (shifted by  $s = 7$  days) of the first order.
- STEP 3. Calculate the second difference (shifted by 1 day) of the series obtained at STEP 2.
- STEP 4. Check whether the series obtained at STEP 3 is stationary using the Dickey–Fuller criterion [20].
- STEP 5. Pass the parameters corresponding to STEPS 1–4 to the SARIMA(1, 1, 2)(0, 1, 1) $_7$  model and select the rest based on minimizing the Akaike information criterion. The series from STEP 1 is passed as data.
- STEP 6. Apply the inverse Box–Cox transform to the resulting model with the tuned hyperparameters.

### 1.4. Time Series Extrapolation Result

A forecast of the time series  $\mathcal{T}(t)$  of the number of daily PCR tests in the Republic of Kazakhstan from June 8, 2021 to January 20, 2022 is presented in Fig. 1.



**Fig. 1.** Statistical data on the number  $\mathcal{T}(t)$  of PCR tests performed in the Republic of Kazakhstan, used when constructing the model.



**Fig. 2.** Statistical data on diagnosed COVID-19 cases  $f(t)$  in the Republic of Kazakhstan used when constructing the model.

In view of high fluctuations due to seasonality, as well as jumps in the time series as a result of incorrect data collection, it was decided to presmooth the series using the Gaussian filter before passing it to the model [21]. In this way, we get rid of the possible instability and uninterpretability of the model results due to sharp jumps in the data obtained from the OWD open source (gray line in Figs. 1, 2). The smoothed data that has been passed to the model is represented by the black line in Figs. 1 and 2.

## 2. THE AGENT-BASED MODEL

Agent-based modeling relies on the study of the dynamics of the development of the disease by studying the interaction between individuals, while global changes in the system arise as a result of the activity of many agents (bottom-up modeling). A general description of the ABM of the spread of COVID-19 in the region is given in Sec. 2.1. It includes population initiation (Sec. 2.2), disease spread rules (Sec. 2.3), and agent testing (Sec. 2.4). In Sec. 2.5, the statement of the direct problem for the ABM is given, and Sec. 2.6 provides the statement of the inverse problem for the ABM.

### 2.1. General Description of the Model

Within the framework of this study, a stochastic ABM was implemented for the Republic of Kazakhstan. The main tool for creating the model was the `Covasim` library [3], implemented in the Python language (the open source code is available at [12]) and created to study COVID-19 agent-based models with nontrivial structures. The general algorithm is as follows: all the necessary

parameters and statistical data are loaded, and an artificial population is created taking into account the age distribution in the region. Next, the agents are connected into contact networks, which are complete graphs. Then a time cycle begins: at each step (the time interval is equal to one day), the epidemiological status of the agent is updated as a superposition of probabilities taking into account its contact structure and the restrictive measures introduced (self-isolation, closing public places, wearing masks, etc.).

### 2.2. Population Initiation

The artificial population is initiated on the basis of statistical data in the region and depends on the following parameters of the agents:

- Age ( $t^*$ ). All agents are divided into age groups 10 years each (0–9, 10–19, ..., 80+ years) according to the statistical data of the Republic of Kazakhstan (Tab. 1).
- Social status (worker, student, child, retired), depending on the agent's age  $t^*$ .

Households are filled with agents according to statistical data on the average family size in the region. Depending on their age, agents contact each other in contact networks, which are complete graphs, the degree of which is determined by a Poisson random variable with parameter  $\lambda$ ,

- For households  $\lambda = 3.496$  is the average family size (persons) [22].
- For organizations  $\lambda = 8$ .
- For public places and educational institutions  $\lambda = 20$ .

All agents have contacts in households and in public places, agents aged 6–21 years old can also contact in educational institutions with agents of their own age, and agents aged 22–65 years old, at work. Contact networks are built on the basis of the *SynthPops* open source algorithm [23], which is able to generate realistic contact networks for populations. This method is based on previously published models and empirical studies that allow one to determine the characteristic number of contacts for specific age groups in the model.

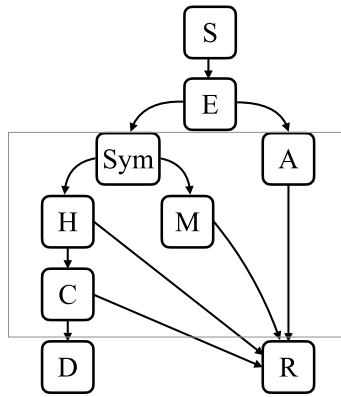
In the case of educational institutions, the algorithm selects a pupil or a student and, depending on the age, forms an age mixing matrix in the educational institution to determine the probable age in the contact network. The students are selected from an ordered list of households so that they reproduce an approximation to the neighborhood dynamics of children attending district educational institutions together. Teachers and other support staff are selected from the adult population of the workforce and are distributed as needed by schools, reflecting the average student-to-teacher to student-to-staff ratios. In large educational institutions, close contacts are modeled by a random set of  $n$  contacts from among those possible in the institution, where  $n$  is defined as a random variable with a Poisson distribution with parameter  $\lambda = 20$ , corresponding to the average class or group size.

The workforce is calculated using age-disaggregated employment rates, and unemployed persons are assigned to jobs using organization size data. The primary reference worker is selected from the labor force, and his/her colleagues are inferred based on age patterns of workforce mixing. All workers (including teachers) are randomly selected from the population to reflect the general mix of adults from different areas at work. Close contacts are modeled by a random set of  $n$  contacts in the organization, where  $n$  is a Poisson random variable with parameter  $\lambda = 8$ , equal to the estimated maximum number of close contacts in the workplace.

To model contacts in public places for each person,  $n$  random contacts in a population are used, where  $n$  has a Poisson distribution with parameter  $\lambda = 20$ . On this level, connections reflect the nature of contacts in parks and rest areas, shopping centers, public transport, etc. All connections between individuals are considered undirected to reflect the ability of any individual in a pair to infect the other.

### 2.3. Spread of the Disease in the ABM

Within the framework of the model, it is assumed that the virus is transmitted between agents connected by a graph edge. Infection by close contact is described by a piecewise constant parameter  $\beta(t)$ , which, depending on the contact structure, is multiplied by the corresponding constant  $w_\beta$  (for households  $w_\beta = 3$ , for educational institutions and places of work  $w_\beta = 0.6$ , and for public



**Fig. 3.** Diagram of agent states in Covasim.

places  $w_\beta = 0.3$ ). Thus, the probability of transmission of the virus is different for each contact network.

Each agent can be in one of 9 disease stages:  $S$  susceptible to infection,  $E$  infected noncontagious,  $A$  asymptomatic,  $Sym$  symptomatic patients,  $M$  mild patients,  $H$  hospitalized,  $C$  critically ill patients (needing resuscitation),  $R$  cured,  $D$  deceased (Fig. 3). Framed are those states in which the agent has the opportunity to get a positive test for COVID-19. Transition from one disease stage to another is controlled by age-dependent parameters (i.e., the older the agent, the more vulnerable he/she is):  $p_{sym}$  is the probability of showing symptoms after infection,  $p_{sev}$  is the probability of a symptomatic patient going into a critical condition (needs hospitalization),  $p_{crit}$  is the probability of a patient going from a severe condition to critical (needs resuscitation),  $p_{death}$  is the probability of death for a patient in intensive care. The numerical values of the parameters for each age group are listed in Table 2 [3].

The duration of each stage of the disease is a random log-normal value with different mean and variance parameters consistent with WHO statistical estimates (means and variances of the distributions are presented in Table 3).

Thus, an agent susceptible to infection ( $S$ ), upon contact with infected agents connected by a graph edge, passes into the stage of an infected noncontagious agent ( $E$ ) with probability  $\beta$  at time  $t$ . Then the agent in the infected noncontagious state ( $E$ ) can switch to the infected state with symptoms ( $Sym$ ) with probability  $p_{sym}$  after  $t_{sym}$  days or remain asymptomatic ( $A$ )  $t_{inc}$  days after infection with probability  $1 - p_{sym}$ . Asymptomatic patients are cured after  $t_{rec1}$  days and transferred to group ( $R$ ). Those infected with symptoms ( $Sym$ ) may develop severe disease and be hospitalized ( $H$ ) with probability  $p_{sev}$  or remain mildly ill ( $M$ ) with probability  $1 - p_{sev}$  in  $t_{inf}$  days after falling into group ( $Sym$ ). Mild patients are cured after  $t_{rec2}$  days and transferred to group ( $R$ ). Hospitalized patients ( $H$ ) may go on to develop a critical condition ( $C$ ), i.e., need a ventilator, with probability  $p_{crit}$  in  $t_{hosp}$  days after hospitalization or recover with probability  $1 - p_{crit}$  in  $t_{rec2}$  days. Critically ill patients ( $C$ ) die with probability  $p_{death}$  in  $t_{crit}$  days or recover with probability  $1 - p_{death}$

**Table 2.** Probabilities of transition between disease stages depending on the age group

Parameter	Age								
	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80+
$p_{sym}$	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9
$p_{sev}$	$5 \times 10^{-4}$	0.0016	0.0072	0.0208	0.0343	0.0765	0.133	0.207	0.246
$p_{crit}$	$3 \times 10^{-5}$	$8 \times 10^{-5}$	$4 \times 10^{-4}$	0.001	0.0022	0.0093	0.0364	0.089	0.174
$p_{death}$	$2 \times 10^{-5}$	$2 \times 10^{-5}$	$10^{-4}$	$3 \times 10^{-4}$	0.0098	0.0026	0.008	0.024	0.082



**Table 3.** Duration of disease stages in each epidemiological status

Parameter	Description	Distribution
$t_{inc}$	Number of days from exposure to the moment the agent becomes contagious	LogN(4.6, 4.8) [24]
$t_{sym}$	Number of days since the agent became contagious to the onset of symptoms	LogN(1, 0.9) [24]
$t_{rec1}$	Duration of illness for asymptomatic and mild cases	LogN(8, 2) [25]
$t_{rec2}$	Duration of illness for severe and critical cases	LogN(14, 2.4) [26]
$t_{inf}$	Number of days it takes for an agent to go from a mild to severe status	LogN(6.6, 4.9) [24]
$t_{hosp}$	Number of days it takes for an agent to go from a severe to critical status	LogN(3, 7.4) [27]
$t_{crit}$	Duration of an agent's stay in a critical status	LogN(6.2, 1.7) [26]

in  $t_{rec2}$  days. All the listed probabilities can be written in the following form:

$$\begin{aligned}
 p(S \rightarrow E) &= \beta, & p(E \rightarrow \text{Sym}) &= \frac{p_{sym}}{t_{sym}}, & p(E \rightarrow A) &= \frac{1 - p_{sym}}{t_{inc}}, \\
 p(\text{Sym} \rightarrow H) &= \frac{p_{sev}}{t_{inf}}, & p(\text{Sym} \rightarrow M) &= \frac{1 - p_{sev}}{t_{inf}}, & p(H \rightarrow C) &= \frac{p_{crit}}{t_{hosp}}, \\
 p(H \rightarrow R) &= \frac{1 - p_{crit}}{t_{rec2}}, & p(M \rightarrow R) &= \frac{1}{t_{rec1}}, & p(A \rightarrow R) &= \frac{1}{t_{rec1}}, \\
 p(C \rightarrow D) &= \frac{p_{death}}{t_{crit}}, & p(C \rightarrow R) &= \frac{1 - p_{death}}{t_{rec2}}.
 \end{aligned}$$

#### 2.4. Testing Agents in the ABM

Testing of agents is carried out in the amount corresponding to the daily statistical data in the Republic of Kazakhstan (see Sec. 1.2). The chance of being tested for COVID-19 depends on the epidemiological state of the agent (susceptible, infected with symptoms, hospitalized, etc.). At each modeling step, the tests are distributed among the entire population (excluding the deceased). A positive result can be obtained by agents whose status is circled in Fig. 3 (infected asymptomatic and with symptoms, hospitalized, mild cases, and critical cases). In the case of a positive test for COVID-19, the agents are included in the daily detected statistics. The model assumes that symptomatic agents are more likely to be tested than asymptomatic ones. This ratio of chances is controlled by a parameter  $p$  reconstructed when solving the inverse problem (see Sec. 2.6).

#### 2.5. Statement of the Direct Problem for the ABM

The direct problem of agent-based modeling is to determine the number of infected (including  $f(t)$  cases detected as a result of PCR testing), hospitalized, deceased, and other states of the agent that are taken into account in the ABM. In the direct problem, all input parameters of the model are assumed to be known. In this case, the agent-based model allows one to calculate the values of the vector

$$\vec{X}(t) = (S(t), E(t), A(t), \text{Sym}(t), M(t), H(t), C(t), R(t), D(t))$$

on the next day, i.e.,  $\vec{X}(t + 1)$ . Since many parameters of the vector

$$\vec{q}(t) = (E(0), \beta, p, \beta_d(i), \beta_c(i)), \quad i = 1, \dots, N,$$

of the ABM of the spread of COVID-19 are unknown, it is necessary to state and solve the inverse problem using additional information on the current day  $t$ . Here  $E(0)$  is the initial number of infected,  $\beta$  is a virus transmission parameter,  $p$  is a testing parameter, the  $\beta_d(i)$  are the days of change in the parameter  $\beta$ , the  $\beta_c(i)$  are the values by which the parameter  $\beta$  changes on days  $\beta_d(i)$ , and  $i$  corresponds to the month of change of the contagiousness parameter  $\beta$ ; i.e.,  $i + 1 = t + 30$ .

### 2.6. Statement of the Inverse Problem for the ABM

In this paper, its own inverse problem is solved at each time stage (equal to 30 days; see Sec. 3.2 for details). The inverse problem 1 consists in reconstructing the parameter vector

$$\vec{q}(0) = (E(0), \beta, p, \beta_d(1), \beta_c(1))$$

based on auxiliary information about the number  $f(t)$  of daily detected cases,  $t$  is measured in days.

The model assumes that the virus variability (the emergence of new strains, pharmaceutical and social measures) occurs no more than once a month. In view of this, the inverse problem 2 is solved, which consists in reconstructing the parameter vector monthly,

$$\vec{q}(t + 30) = (\beta_d(i), \beta_c(i)), \quad i + 1 = t + 30,$$

based on auxiliary information about the number  $f(t)$  of daily detected cases. Here  $i$  corresponds to the month of simulation.

Solution of the inverse problems of reconstructing the vector  $\vec{q}(t)$  was reduced to solving the problem of minimizing the objective functional

$$J(\vec{q}) = \sum_{t_i=1}^T \frac{|f_d(t_i) - f_m(t_i, \vec{q})|}{M_{\text{diag}}}. \quad (1)$$

Here  $f_d(t_i)$  and  $f_m(t_i, \vec{q})$  are the smoothed data and the result of simulation of daily detected cases, respectively,  $T$  is the number of simulation days, and  $M_{\text{diag}} = \max_{t_i} \{f_d(t_i)\}$  is a normalization term.

In the papers [28, 29], the authors analyzed the sensitivity of unknown parameters to measurements for the model under study using the methods of differential algebra and the Bayesian approach. It is shown that the parameter  $\beta$  responsible for virus transmission is most sensitive to measurements. With the help of sensitivity analysis methods, the range of the parameter  $\beta$  was reduced by 2 times owing to the addition of auxiliary information about the epidemic (namely, information on critical cases was added to the measurements of the number of detected cases and deaths).

## 3. ALGORITHM FOR SOLVING THE INVERSE PROBLEM

In the course of solving the inverse problem, the vector of unknown parameters  $\vec{q}$  was reconstructed using the **Optuna** package [30], which is based on the method of tree-structured Parzen estimators (or TPE for short) as well as the data assimilation approach for stage-by-stage reconstruction of the agent-based model parameters (Sec. 3.2).

### 3.1. Method of Tree-Structured Parzen Estimators

The idea of the method is as follows: the probabilities  $p(\vec{q} | J(\vec{q}))$  and  $p(J(\vec{q}))$  are calculated to determine parameter domain for minimizing the functional  $J$ . To this end, the space  $\mathcal{D}_K = \{q_k, J(q_k) | k = 1, \dots, K\}$  of parameter values is divided into 2 subsets  $\mathcal{D}_{K_l}^l$  and  $\mathcal{D}_{K_g}^g$  such that  $\mathcal{D}_{K_l}^l$  contains the level quantile  $\gamma$  of the smallest values of the functional at points from  $\mathcal{D}_K(J_\gamma)$ ; i.e.,  $P(J < J_\gamma) = \gamma$ . The subset  $\mathcal{D}_{K_g}^g$  contains all other points from  $\mathcal{D}_K$ . Next, using the Parzen window method, the distribution densities  $l(x)$  and  $g(x)$  obtained from  $\mathcal{D}_{K_l}^l$  and  $\mathcal{D}_{K_g}^g$ , respectively, are estimated. Thus, using  $l(x)$ , one can obtain the domain of points at which the functional reaches its least values. Thus, the probability  $p(\vec{q}_{K+1} | J(\vec{q}))$  is determined as

$$p(\vec{q}_{K+1} | J(\vec{q})) = \begin{cases} l(\vec{q}), & J(\vec{q}_{K+1}) < J^\gamma \\ g(\vec{q}), & J(\vec{q}_{K+1}) \geq J^\gamma. \end{cases}$$

Then a set of vectors is generated according to the density  $l(x)$ . From these vectors, we select a vector  $\vec{q}^*$  on which the maximum expected improvement  $EI(x)$  expressed by the following formula is achieved:

$$EI(\vec{q}) = \left( \gamma + \frac{g(\vec{q})}{l(\vec{q})}(1 - \gamma) \right)^{-1}.$$

Convergence in probability of statistical methods is reflected in the general theorem in [31].

The stopping criterion in the presented algorithm is the limitation on the number of iterations  $\text{max\_iter} = 100$ . The scheme of the algorithm is as follows.

---

**Algorithm 1** (tree-structured Parzen estimators algorithm).

---

**Require:** the values of parameters  $\gamma$ ,  $n_{\text{samp}}$ , and  $\text{max\_iter}$

- 1:     **Initialize:** the space  $\mathcal{D}_{\text{init}} = \{\vec{q}_k, J(\vec{q}_k), k = 1, \dots, n_{\text{init}}\}$  of values of unknown parameters
  - 2:     **for**  $m = 0, \dots, \text{max\_iter}$  **do**
  - 3:         Divide  $\mathcal{D}_{n_{\text{init}}+m}$  to generate spaces  $\mathcal{D}_{m_g}^g, \mathcal{D}_{m_l}^l$
  - 4:         Produce an estimate of density  $l(\vec{q})$  for tuples of parameters from  $\mathcal{D}_{n_{\text{init}}+m_l}^l$
  - 5:         Produce an estimate of density  $g(\vec{q})$  for tuples of parameters from  $\mathcal{D}_{n_{\text{init}}+m_g}^g$
  - 6:         Generate  $\vec{q}^s = \{\vec{q}_k^s \mid k = 1, \dots, n_{\text{samp}}\}$ , where  $\vec{q}_k^s \sim l(\vec{q})$
  - 7:         Choose  $\vec{q}_{m+1} = \underset{\vec{q} \subset \vec{q}^s}{\text{argmax}} EI(\vec{q})$
  - 8:         Calculate  $J(\vec{q}_{m+1})$
  - 9:          $\mathcal{D}_{n_{\text{init}}+m} \leftarrow \mathcal{D}_{n_{\text{init}}+m+1}$
  - 10:     **end for**
- 

More details about the method of tree-structured Parzen estimators can be found in [32].

### 3.2. Step-by-Step Reconstruction of Unknown Parameters

The parameter  $\beta$  is assumed to be piecewise constant. Accordingly, the longer the simulation period under consideration, the greater the number of unknown parameters. However, each of the parameters  $\beta_d(i)$  and  $\beta_c(i)$ ,  $i = 1, \dots, N$ , where  $N = 18$  is the number of simulation months, depends only on the data on a specific simulation subperiod. Each interval was calibrated sequentially one after the other (data assimilation method), and the parameters reconstructed at the previous step were used in the subsequent run of the optimization algorithm (the algorithm is described in [33]). Thus, on the first interval from March 13, 2020 to April 12, 2020 ( $i = 1$ ), we used the vector of unknown parameters (the initial conditions at  $t = 0$  correspond to the date March 3, 2020)

$$\vec{q}(0) = (E(0), \beta, p, \beta_d(1), \beta_c(1)),$$

and on all subsequent intervals,

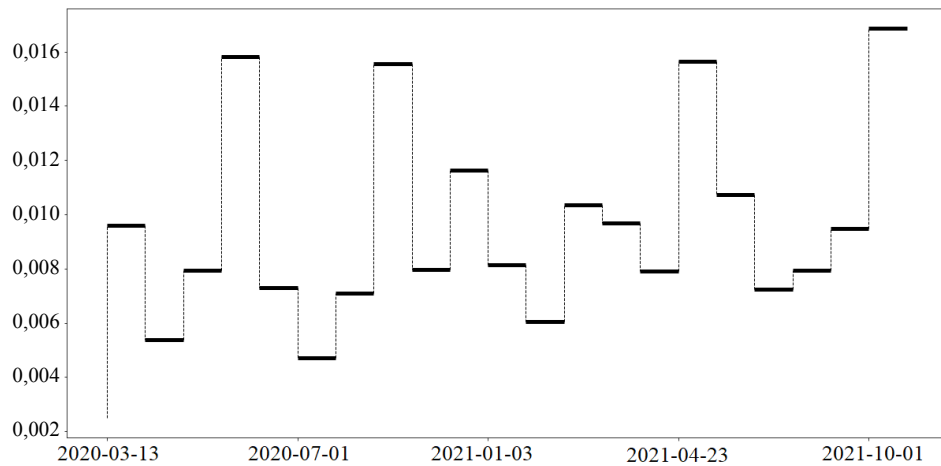
$$\vec{q}(i) = (\beta_d(i), \beta_c(i)), \quad i = 2, \dots, N,$$

where  $N$  is the number of months of simulation.

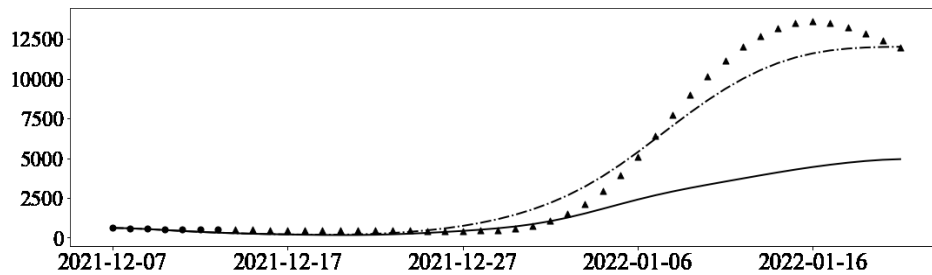
The regularization of the solution of the inverse problem consists in the use of constraints on the required parameters obtained in the sensitivity analysis based on the measurements used [28, 29].

## 4. NUMERICAL RESULTS

Let us present the results of mathematical modeling of the spread of COVID-19 in the Republic of Kazakhstan. As shown in Sec. 3.2, at the first stage, we reconstruct the number of asymptomatic infected patients  $E(0)$ , the rate of transmission of the virus from the infected to susceptible agent  $\beta$ ,



**Fig. 4.** Graph of the change in the contagiousness parameter  $\beta_c(i)$  and days of its change  $\beta_d(i)$  in the Republic of Kazakhstan in the period from March 13, 2020 to December 12, 2021.



**Fig. 5.** Simulation of two scenarios for the spread of daily detected cases as a result of PCR testing in the Republic of Kazakhstan.

and the value  $\beta_c(1)$  by which the parameter  $\beta$  will change on day  $\beta_d(1)$ , as well as the chance  $p$  of being tested. Monthly, based on the solution of the inverse problem, the parameters of the day  $\beta_d(i)$  and the values  $\beta_c(i)$  of the change in the rate of transmission of the virus in the Republic of Kazakhstan were updated (the result of reconstructing  $\beta_d(i)$  and  $\beta_c(i)$  is shown in Fig. 4).

When constructing scenarios for the spread of daily detected cases of COVID-19, it is assumed that the average level of testing of the population  $\mathcal{T}(t)$  is preserved in the region. The number of PCR tests on day  $t$  is calculated using the regression model SARIMA (see Sec. 1).

#### 4.1. Scenarios for the Spread of COVID-19 in the Republic of Kazakhstan

Figure 5 shows the result of simulating the average number of daily detected cases as a result of PCR testing in the Republic of Kazakhstan with a 45-day forecast (dots represent real data from March 13, 2020 to December 12, 2021 that are involved in the solution of the inverse problem, and triangles show data from December 13, 2021 to January 1, 2022 that were used to check the forecast). Two types of scenarios were taken into account when constructing the forecast.

- The base scenario (solid line), which did not take into account the increased congestion of people during the New Year holidays but only the increase in the mobility of citizens in the period of preparation for the holidays (from December 12, 2021 to December 12, 2021).
- The increased mobility of citizens for the New Year holidays (dashed line). The increased transmission of the virus in public places in the period from January 1, 2022 to January 6, 2022 was taken into account. It is characterized by an increase in the value of  $\beta_c(t)$  of virus transmission by 2.5 times for the period of pre-New Year holidays December 20–30, 2021—it reaches the value of 0.039, then increases to 0.548 in the period January 1–10, 2022, after which it drops to 0.03.

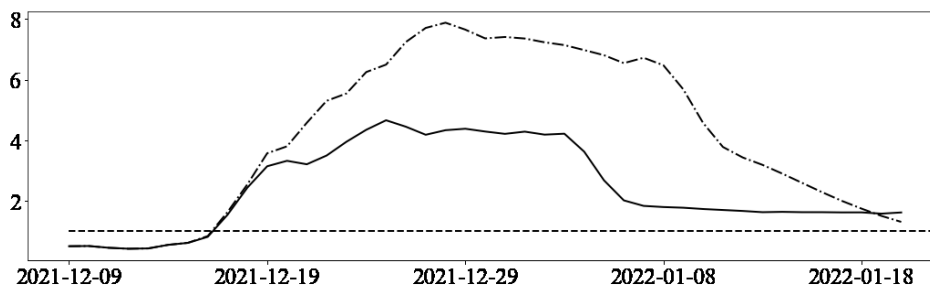


Fig. 6. Basic reproduction number  $\mathcal{R}(t)$ .

Based on the results of model validation, it can be concluded that taking into account the increased mobility of citizens demonstrates a more accurate correspondence to real data. The values of the parameter  $\beta$  in the second scenario (dotted line) indicate that during the New Year holidays, the transmission of the virus increased by 3.5 times compared to early December. So, as of January 20, 2022, the number of detected cases of COVID-19 was 12 032 people, while the number of expected detected cases in the Republic of Kazakhstan according to the ABM base scenario (solid line) calculated on December 12, 2021 was 4 939 people (the error 59%) and when taking into account the increased mobility of citizens in public places, it was 12 007 people (the error 0.2%).

#### 4.2. Base Virus Reproduction Index in the Republic of Kazakhstan

The main indicator of the spread of an epidemic is the basic reproduction number  $\mathcal{R}(t)$ , which characterizes the average number of people who become infected from those actively infected in a completely nonimmunized environment in the absence of special epidemiological measures. In this paper, we used the expression for the base reproduction index proposed by Kerr et al. [3],

$$\mathcal{R}(t) = \frac{I_N(t) \cdot d}{I_C(t)}. \quad (2)$$

Here  $I_N(t)$  is the number of newly infected on day  $t$ ,  $I_C(t)$  is the current number of infected on day  $t$ , and  $d$  is the average duration of the disease in days. If  $\mathcal{R}(t) < 1$ , then the epidemic is considered to stop spreading; otherwise, it grows. Figure 6 shows the graph of  $\mathcal{R}(t)$  for the Republic of Kazakhstan for the two scenarios considered above. The results show an increase in the number of detected cases of COVID-19 in the Republic of Kazakhstan and a high burden on the healthcare system from December 19, 2021 to January 18, 2022 for the scenario with increased mobility of citizens in public places due to the New Year holidays (dot-and-dash line), after which the number of newly detected cases is reduced to the volume of the base development scenario (solid line). The horizontal dashed line denotes the threshold value  $\mathcal{R}(t) = 1$ .

## CONCLUSIONS

An agent-based model of the spread of COVID-19 in the Republic of Kazakhstan has been developed. It is based on the *Covasim* software package and implemented in the Python language and includes the initiation of the population based on the demographic data of the country and the rules for the spread of the disease and the testing of agents depending on age and epidemiological status. The first stage involves collection, processing, and analysis of incomplete data using regression analysis and machine learning methods. At the second stage, the epidemiological parameters of the agent-based model (rates of infection transmission and testing, initial number of infected agents) are refined using additional information on the number of detected cases of COVID-19 in the Republic of Kazakhstan. For this purpose, a data assimilation algorithm has been developed within the framework of which unknown parameters are identified monthly with additional information about the number of daily detected cases of COVID-19 based on the global optimization method of tree-structured Parzen estimators. At the third stage, nonpharmaceutical interventions in the epidemic's

spreading process are taken into account in order to construct the most realistic scenarios for the spread of COVID-19.

When implementing the model, data on the number of detected cases from March 13, 2020 to December 12, 2021 in the Republic of Kazakhstan were used. It has been shown that an increase in the density of agents during the New Year holidays in public places (shops, theaters, parks) increases the number of detected cases of COVID-19 (by January 1, 2022, it increased by 3.5 times compared to December 1, 2021). As an example, two scenarios are given for the spread of COVID-19 calculated on December 12, 2021 for the period up to January 20, 2022. The scenario that took into account the New Year holidays (published on December 12, 2021 on the website <http://covid19-modeling.ru>) almost coincided with what happened in reality (the error was 0.2%, or 25 people). Thus, mathematical modeling makes it possible to obtain a qualitative and quantitative agreement between the forecast of the epidemiological situation and reality.

A specific feature of modeling the spread of coronavirus infection in the Republic of Kazakhstan is that it is necessary to take into account the concentration of the population in large cities, including Almaty (1 993 067 people), Astana (1 199 083 people), Shymkent (1 090 160 people), and others, inhabited by more than 11 million people, and in the villages, more than 7 million people. For more detailed modeling of COVID-19 spread scenarios, it is necessary to take into account traffic flows between the largest cities, as well as traffic flows at the city–region level. To obtain more detailed scenarios for the spread of COVID-19, it is necessary to combine agent and SIR models, as is done in the Novosibirsk oblast [2, 15, 16].

The contribution of each author to the work is as follows:

- O.I. Krivorotko and S.I. Kabanikhin: formulation of direct and inverse problems, formulation of solution algorithms and analysis of calculation results, and coordination of work.
- M.A. Bektemesov: provision of data on the Republic of Kazakhstan.
- M.I. Sosnovskaya: implementation of the agent-based model and algorithms for solving the direct and inverse problems.
- A.V. Neverov: data processing and development of a program for computing on the SSCC SB RAN cluster.

#### FUNDING

This research was supported by the Council for Grants under the President of the Russian Federation, project no. MK–4994.2021.1.1, the Russian Foundation for Basic Research, project no. 21–51–10003, and the Ministry of Education and Science of the Republic of Kazakhstan, project no. AP09260317.

#### REFERENCES

1. V. A. Adarchenko, S. A. Baban', A. A. Bragin, et al., "Modeling the development of the coronavirus epidemic using differential and statistical models" (Snezhinsk, 2020), *Preprint RFIATC-VNIITF*, no. 264 [in Russian].
2. O. I. Krivorotko and S. I. Kabanikhin, "Mathematical models of the spread of COVID-19" (Novosibirsk, 2022), *Preprint Inst. Mat. im. S.L. Soboleva*, no. 300 [in Russian].
3. C. C. Kerr, R. M. Stuart, D. Mistry, R. G. Abeysuriya, K. Rosenfeld, G. R. Hart, et al., "Covasim: An agent-based model of COVID-19 dynamics and interventions," *PLOS Comput. Biol.* **17** (7), e1009149 (2021).
4. R. Laubenbacher, F. Hinkelmann, and M. Oremland, "Agent-based models and optimal control in biology: A discrete approach," in *Mathematical Concepts and Methods in Modern Biology: Using Modern Discrete Models* (Academic Press, San Diego, CA, 2013), ch. 5, 143–178.
5. A. I. Vlad, T. E. Sannikova, and A. A. Romaniuha, "Modeling the spread of respiratory viral infections in the city: a multi-agent approach," *Mat. Biol. Bioinf.* **15** (2), 338–356 (2020) [in Russian].
6. A. Aleta, D. Martin-Corral, Y. Pastore, A. Piontti, et al., "Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19," *Nat. Hum. Behav.* **4** (9), 964–971 (2020).

7. M. S. Y. Lau, B. Grenfell, M. Thomas, M. Bryan, et al., “Characterizing superspreading events and age-specific infectiousness of SARS-CoV-2 transmission in Georgia, USA,” *Proc. Natl. Acad. Sci. USA*. **117** (36), 22430–22435 (2020).
8. A. J. Kucharski, P. Klepac, A. J. K. Conlan, S. M. Kissler, et al., “Effectiveness of isolation, testing, contact tracing, and physical distancing on reducing transmission of SARS-CoV-2 in different settings: A mathematical modelling study,” *Lancet Infect. Dis.* **20** (10), 1151–1160 (2020).
9. N. Hoertel, M. Blachier, C. Blanco, M. Olsson, et al., “A stochastic agent-based model of the SARS-CoV-2 epidemic in France,” *Nat. Med.* **26** (9), 1417–1421 (2020).
10. J. Hellewell, S. Abbott, A. Gimma, N. I. Bosse, et al., “Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts,” *Lancet Glob. Health* **8** (4), e488–e496 (2020).
11. B. F. Nielsen and K. Sneppen, “COVID-19 superspreading suggests mitigation by social network modulation,” *Phys. Rev. Lett.* **126** (11), 118301 (2021).
12. COVID-19 Agent-based Simulator, 2022. Available at <https://github.com/InstituteForDiseaseModeling/covasim>.
13. N. B. Noll, I. Aksamentov, V. Druelle, A. Badenhorst, et al., COVID-19 scenarios: An interactive tool to explore the spread and associated morbidity and mortality of SARS-CoV-2 (2020) (e-Print Archive; medrxiv:10.1101/2020.05.05.20091363v2). <https://doi.org/10.1101/2020.05.05.20091363>
14. J. T. Tuomisto, J. Yrjölä, M. Kolehmainen, J. Bonsdorff, et al., An agent-based epidemic model REINA for COVID-19 to identify destructive policies (2020) (e-Print Archive; medrxiv:10.1101/2020.04.09.20047498v2) <https://doi.org/doi.org/10.1101/2020.04.09.20047498>
15. O. I. Krivorotko, S. I. Kabanikhin, N. Yu. Ziat'kov, A. Yu. Prikhod'ko, et al., “Mathematical modeling and prediction of COVID-19 in Moscow and in Novosibirsk oblast,” *Sib. Zh. Vychisl. Mat.* **23** (4), 395–414 (2020) [in Russian].
16. O. Krivorotko, M. Sosnovskaia, I. Vashchenko, C. Kerr, and D. Lesnic, “Agent-based modeling of COVID-19 outbreaks for New York state and UK: Parameter identification algorithm,” *Infect. Dis. Model.* **7**, 30–44 (2022).
17. Interpolation function. Available at <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.interpolate.html>.
18. G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control. Issue 1* (Holden-Day, San Francisco, 1970; Mir, Moscow, 1974).
19. P. P. Dabral and M. Z. Murry, “Modelling and forecasting of rainfall time series using SARIMA,” *Environ. Process.* **4**, 399–419 (2017).
20. D. A. Dickey and W. A. Fuller, “Distribution of the estimators for autoregressive time series with a unit root,” *J. Am. Stat. Assoc.* **74**, 427–431 (1979).
21. Gaussian filter. Available at [https://docs.scipy.org/doc/scipy/reference/generated/scipy.ndimage.gaussian\\_filter.html](https://docs.scipy.org/doc/scipy/reference/generated/scipy.ndimage.gaussian_filter.html).
22. Household Size, 2019, UN. Available at <https://population.un.org/Household/#/countries/840/>.
23. SynthPops algorithm. Available at <https://github.com/institutefordiseasemodeling/synthpops/>.
24. S. A. Lauer, K. H. Grantz, Q. Bi, et al., “The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application,” *Ann. Int. Med.* **172** (9), 577–582 (2020).
25. R. Wölfel, V. M. Corman, W. Guggemos, et al., “Virological assessment of hospitalized patients with COVID-2019,” *Nature* **581**, 465–469 (2020).
26. R. Verity, L. C. Okell, I. Dorigatti, et al., “Estimates of the severity of coronavirus disease 2019: a model-based analysis,” *Lancet Infect. Dis.* **20** (6), 669–677 (2020).

27. D. Wang, B. Hu, C. Hu, et al., “Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China,” *JAMA* **323** (11), 1061–1069 (2020).
28. O. I. Krivorotko, S. I. Kabanikhin, M. I. Sosnovskaya, D. V. Andornaya, “Analysis of sensitivity and identifiability of mathematical models of the spread of the COVID-19 epidemic,” *Vavilov. Zh. Genet. Sel.* **25** (1), 82–91 (2021) [in Russian].
29. O. Krivorotko, M. Sosnovskaia, and I. Vashchenko, “Agent-based mathematical model of COVID-19 spread in Novosibirsk region: Identifiability, optimization and forecasting,” *J. Inverse Ill-Posed Probl.* (2022) (in print).
30. OPTUNA: hyperparameter optimization framework. Available at <https://optuna.org/>.
31. A. A. Zhiglavskii and A. G. Zhilinskias, *Methods for Finding the Global Extremum* (Nauka, Moscow, 1991) [in Russian].
32. J. Bergstra, R. Bardenet, B. Kégl, and Y. Bengio, “Algorithms for hyper-parameter optimization,” *Adv. Neural Inf. Process. Syst.* **24**, 2546–2554 (2011).
33. M. L. Daza-Torres, M. A. Capistrán, A. Capella, and J. A. Christen, “Bayesian sequential data assimilation for COVID-19 forecasting,” *Epidemics* **39**, 100564 (2022).