

Original Russian text www.bionet.nsc.ru/vogis/

Sensitivity and identifiability analysis of COVID-19 pandemic models

O.I. Krivorotko^{1,2}✉, S.I. Kabanikhin^{1,2}, M.I. Sosnovskaya², D.V. Andornaya²

¹ Institute of Computational Mathematics and Mathematical Geophysics of Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

² Novosibirsk State University, Novosibirsk, Russia

✉ krivorotko.olya@mail.ru

Abstract. The paper presents the results of sensitivity-based identifiability analysis of the COVID-19 pandemic spread models in the Novosibirsk region using the systems of differential equations and mass balance law. The algorithm is built on the sensitivity matrix analysis using the methods of differential and linear algebra. It allows one to determine the parameters that are the least and most sensitive to data changes to build a regularization for solving an identification problem of the most accurate pandemic spread scenarios in the region. The performed analysis has demonstrated that the virus contagiousness is identifiable from the number of daily confirmed, critical and recovery cases. On the other hand, the predicted proportion of the admitted patients who require a ventilator and the mortality rate are determined much less consistently. It has been shown that building a more realistic forecast requires adding additional information about the process such as the number of daily hospital admissions. In our study, the problems of parameter identification using additional information about the number of daily confirmed, critical and mortality cases in the region were reduced to minimizing the corresponding misfit functions. The minimization problem was solved through the differential evolution method that is widely applied for stochastic global optimization. It has been demonstrated that a more general COVID-19 spread compartmental model consisting of seven ordinary differential equations describes the main trend of the spread and is sensitive to the peaks of confirmed cases but does not qualitatively describe small statistical datasets such as the number of daily critical cases or mortality that can lead to errors in forecasting. A more detailed agent-oriented model has been able to capture statistical data with additional noise to build scenarios of COVID-19 spread in the region.

Key words: parameter sensitivity; identifiability; ordinary differential equations; inverse problems; epidemiology; COVID-19; forecasting; Novosibirsk region.

For citation: Krivorotko O.I., Kabanikhin S.I., Sosnovskaya M.I., Andornaya D.V. Sensitivity and identifiability analysis of COVID-19 pandemic models. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2021;25(1):82-91. DOI 10.18699/VJ21.010

Анализ чувствительности и идентифицируемости математических моделей распространения эпидемии COVID-19

О.И. Криворотко^{1,2}✉, С.И. Кабанихин^{1,2}, М.И. Сосновская², Д.В. Андорная²

¹ Институт вычислительной математики и математической геофизики Сибирского отделения Российской академии наук, Новосибирск, Россия

² Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

✉ krivorotko.olya@mail.ru

Аннотация. Разработан алгоритм анализа чувствительности и идентифицируемости математических моделей распространения эпидемии COVID-19 в Новосибирской области, основанных на системах дифференциальных уравнений и законе действующих масс. Основу алгоритма составляет анализ матрицы чувствительности методами дифференциальной и линейной алгебры, показывающей степень зависимости неизвестных параметров моделей от заданных измерений. В результате работы алгоритма выявляются наименее и наиболее чувствительные к измерениям параметры, что способствует построению регуляризующего алгоритма решения задачи идентификации параметров для построения более точных сценариев развития эпидемии в регионе. Анализ чувствительности математических моделей

распространения коронавирусной инфекции COVID-19 показал, что параметр контагиозности вируса устойчиво определяется по количеству ежедневно выявляемых заболевших, критических и вылечившихся больных. С другой стороны, прогнозируемая доля госпитализированных больных, находящихся в критическом состоянии и требующих подключения аппарата ИВЛ, а также коэффициент смертности определяются гораздо менее устойчиво. Для построения более реалистичного прогноза необходимо добавить дополнительную информацию о процессе (например, о количестве ежедневных случаев госпитализации). Задачи уточнения идентифицируемых параметров по дополнительной информации о количестве выявленных, критических и смертельных случаев в Новосибирской области были сведены к задачам минимизации соответствующих целевых функционалов. Задача минимизации была решена с помощью метода дифференциальной эволюции, широко используемого в задачах стохастической глобальной оптимизации. Показано, что более общая камерная модель, состоящая из семи обыкновенных дифференциальных уравнений, описывает основную тенденцию распространения коронавирусной инфекции, чувствительна к пикам выявленных случаев, однако некачественно описывает небольшие статистики (количество ежедневных критических, смертельных случаев), что может приводить к ошибочным выводам. Более подробная агентно-ориентированная математическая модель, учитывающая поведение отдельных агентов, позволяет улавливать небольшие шумы в данных и строить сценарии развития распространения эпидемии в регионе.

Ключевые слова: чувствительность параметров; идентифицируемость; обыкновенные дифференциальные уравнения; обратные задачи; эпидемиология; COVID-19; прогнозирование; Новосибирская область.

Introduction

Many mathematical models in biology (epidemiology, immunology, pharmacokinetics, systems biology), medicine (tomography), physics and chemistry (meteorology, chemical kinetics), as well as sociology are described by systems of differential equations, whether they be ordinary (Kermack, McKendrick, 1927), partial (Habtemariam et al., 2008), or stochastic differential ones (Lee et al., 2020). Coefficients in these equations characterize specific features of simulated processes under given conditions. To build an adequate mathematical model, one needs to refine the coefficients of the equations based on the known parameters of the process and any additional information available about it. For example, when considering epidemiological problems, the parameters such as the infection transmission rate in the region; the critical case rates depending on comorbidities, age, and other demographic factors; the proportion of asymptomatic carriers/latent infection cases, etc., are unknown or approximately derived based on statistical data. These parameters are often sensitive to the measurements prone to errors (rounding, instrument, and human factor errors), which leads to unstable solutions of parameter identification problems.

Identifiability analysis of the differential equation systems modeling biological, medical, and physical processes is an important step to undertake before developing computational algorithms (Bellu et al., 2007; Raue et al., 2010, 2014; Miao et al., 2011; Kabanikhin et al., 2016; Voropaeva, Tsgoev, 2019). A classification of identifiability types distinguishing between structural identifiability, practical identifiability, and sensitivity analysis is presented in (Krivorotko et al., 2020a). The authors also analyze the systems of ordinary differential equations (ODE) describing epidemiological and immunological processes in terms of practical identifiability and parameter sensitivity to measurement errors.

A detailed review of methods and case studies of structural identifiability analysis in biological problems described by

ODE systems may be found in (Miao et al., 2011; Kabanikhin et al., 2016). The model structure being as follows:

$$\begin{cases} \frac{d\bar{x}}{dt} = g(t, \bar{x}, q), & t \in (0, T), x(t) \in C^1(\mathbb{R}^N), q \in \mathbb{R}^L, \\ \bar{x}(0) = \bar{x}_0, \\ x_i(t_k) = f_{ik}, & i \in \{1, \dots, M\}, k = 1, \dots, K, \end{cases} \quad (1)$$

The analysis based on this model makes it possible to verify the uniqueness of the solution q of the parameter identification problem and the initial conditions of model \bar{x}_0 (or their part) based on available measurements f_{ik} , while also providing recommendations on adding new information and modifying the conditions of the parameter identification problem to ensure the uniqueness of the solution.

In the present paper, the analysis of the semi-relative sensitivity of the mathematical models to describe epidemiological and social processes is presented. This approach, proposed in (Adams et al., 2004) for analyzing ODE systems, shows the degree of parameter sensitivity to measurements and identifies lacking/excessive measurements based on a certain reference set of parameters for solving the stated parameter identification problem. Two mathematical models of the spread of the new coronavirus infection caused by the SARS-CoV-2 virus described by ODE systems are used as examples. A regularization algorithm for numerical solution of the parameter identification problem is developed for SEIR compartmental model and agent-based model using the statistical data from public sources. The modeling results and the scenario of COVID-19 spread in the Novosibirsk region are presented.

Parameter sensitivity analysis in systems of ordinary differential equations

Sensitivity analysis is used for identifiability assessment of the unknown parameters of the model represented by ODE system (1) before developing a numerical solution algorithm

for the parameter identification problem. These methods do not require real experimental data, but the number of measurements and their time may be a necessity. Sensitivity analysis for a mathematical model is performed with regard to a set of nominal parameters q^* , whose values are taken from the literature or statistical data available.

Sensitivity analysis methods are based on a sensitivity matrix. Assume that $t_1 \leq t_2 \leq \dots \leq t_K$ are the fixed times of measurements f_{ik} . Then, the sensitivity matrix coefficients for parameter vector q^* are calculated as:

$$s_{ij}(t) = \frac{\partial f_i(t, q^*)}{\partial q_j} \cdot q_j^*, \quad (2)$$

where, f_i , $i = 1, \dots, M$, is the i th entry of the measurement function vector, and q_j , $j = 1, \dots, L$, is the j th entry of the parameter vector.

Thus, the sensitivity matrix is determined as follows:

$$S_{M \cdot K \times L} = \begin{pmatrix} s_{11}(t_1) & \dots & s_{1L}(t_1) \\ \vdots & \ddots & \vdots \\ s_{M1}(t_1) & \dots & s_{ML}(t_1) \\ \vdots & \ddots & \vdots \\ s_{11}(t_K) & \dots & s_{1L}(t_K) \\ \vdots & \ddots & \vdots \\ s_{M1}(t_K) & \dots & s_{ML}(t_K) \end{pmatrix}.$$

The sensitivity matrix is calculated using the conventional sensitivity function:

$$s_{q_j}(t) = \frac{\partial x}{\partial q_j}(t), \quad j = 1, \dots, L.$$

When the first equation from (1) is differentiated with respect to q_j , each vector function s_{q_j} should satisfy the Cauchy problem as follows:

$$\begin{cases} \dot{s}_{q_j}(t) = \frac{\partial g}{\partial \bar{x}}(t, \bar{x}(t; q), q) \cdot s_{q_j}(t) + \frac{\partial g}{\partial q_j}(t, \bar{x}(t; q), q), \\ s_{q_j}(t_0) = \frac{\partial \bar{x}_0}{\partial q_j}. \end{cases} \quad (3)$$

So, $s_{q_j}(t)$ is obtained by numerically solving the Cauchy problem.

First, the assessment is performed for the parameters q , to which the model's solution is most sensitive. These parameters, in turn, are defined by calculating semi-relative sensitivity. Here, sensitivity is considered a time function on the interval of interest. To obtain a general measure of parameter sensitivity of the solution, a time norm (over space L_2) is derived for each state/parameter combination and the obtained scalar quantities are ranked to identify the most sensitive pa-

rameters. The lower the value $\left\| \frac{\partial f_i(t)}{\partial q_k} q_k^* \right\|_2$, the less the effect

of q_k on f_i . This general measure will be referred to as semi-relative sensitivity.

The orthogonal method is then used for sensitivity analysis. The idea of the method suggested in (Yao et al., 2003) is to investigate linear dependencies of the columns of sensitivity matrix S . In such a way it will be possible to assess parameter

sensitivity to the input data and parameter interdependence at the same time.

Sensitivity analysis of COVID-19 spread models

The feature of currently developed COVID-19 spread models is that they analyze the behavior of asymptomatic cases and the effect of the incubation period on the epidemiological situation in the regions. Several open-source suites (Gomez et al., 2020; Tuomisto et al., 2020; Wolfram, 2020) and web services have been developed for modeling COVID-19 spread scenarios:

- on a global scale: <https://covid19-scenarios.org/> (University of Basel, Switzerland);
- in Moscow, the Novosibirsk region, and some European countries: <https://covid19.biouml.org/> (Institute of Computational Technologies, SB RAS, Novosibirsk);
- in Almaty, the Republic of Kazakhstan: <http://covid19.mmay.info/almaty/?fbclid=IwAR20yx7F4MdWRqwUDzripUK291WAvoyCSkDPafgpj25ummay23e7oFHBdXg>.

Two fundamental approaches to epidemic propagation modeling may be distinguished:

1. *Compartmental approach* (top-down modeling). Here, the interaction between the agents within the population grouped by similar attributes (susceptible group, (a)symptomatic carriers, hospital admission cases, critical cases, etc.) is described using the mass balance law within the compartmental model first suggested in 1927 (Kermack, McKendrick, 1927). Agents are distributed in time depending on the assigned transition probabilities between groups such as infection probability, virus contagiousness, mortality rate, etc.
2. *Agent-based approach* (bottom-up modeling) is based on studying the interactions between individuals and their effect on global parameters (e.g. virus contagiousness, mortality rate, severe case probability, etc.). Agent-based models are characterized by random graphs whose arc lengths describe the probabilities of transition to different agent states.

The parameters of transition between groups or agent states are often unknown or broadly defined. For instance, the incubation period of the disease according to WHO data varies from 2 to 14 days, which complicates the analysis of the model and the building of adequate disease spread scenarios.

Let us consider two breakdowns of the population of a particular age (e.g., ages 20–29) into groups. The transition into different agent states in the course of the disease caused by the SARS-CoV-2 virus is presented in Fig. 1. These models do not take into account such factors as sex segregation, annual birth and mortality rates (since the modeling interval of less than a year is analyzed), vaccination, passenger traffic, and comorbidities, which affect the probabilities of transition to different agent states. Our goal here is to demonstrate the correlation between dependences of similar parameters on the same measurements and recommend what parameters can be determined consistently and based on what measurements.

The ODE system (1) describing COVID-19 spread in the population is divided into 10 groups (Kerr et al., 2020) based on the mass balance law and is expressed as follows:

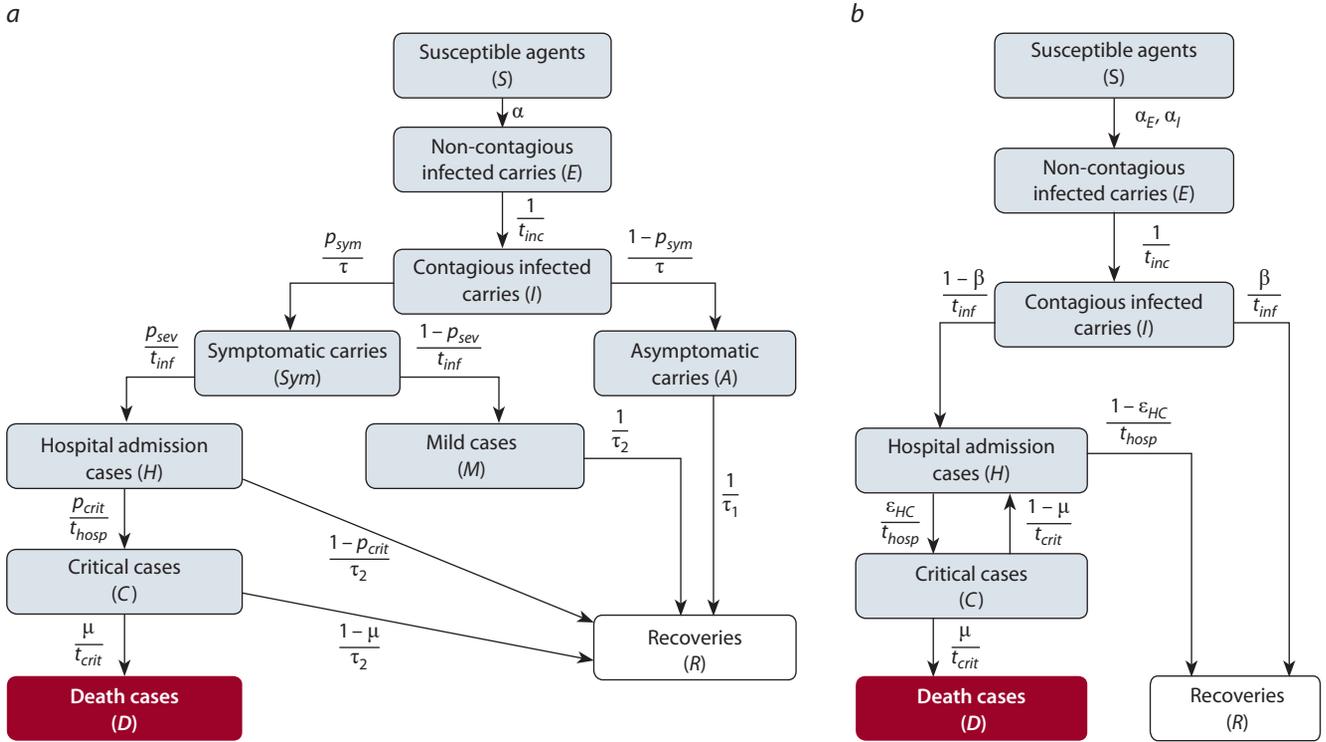


Fig. 1. Agent-state diagram in (a) the COVASIM package (Kerr et al., 2020) and (b) the SEIR-HCD model (Unlu et al., 2020).

$$\begin{cases}
 \frac{dS}{dt} = -\alpha S(t), \\
 \frac{dE}{dt} = \alpha S(t) - \frac{1}{t_{inc}} E(t), \\
 \frac{dI}{dt} = \frac{1}{t_{inc}} E(t) - \frac{1}{\tau} I(t), \\
 \frac{dA}{dt} = \frac{1-p_{sym}}{\tau} I(t) - \frac{1}{\tau_1} A(t), \\
 \frac{dSym}{dt} = \frac{p_{sym}}{\tau} I(t) - \frac{p_{sev}}{t_{inf}} Sym(t) - \frac{1-p_{sev}}{\tau} Sym(t), \\
 \frac{dR}{dt} = \frac{1}{\tau_1} A(t) + \frac{1-p_{crit}}{\tau_2} H(t) + \frac{1}{\tau_1} M(t) + \frac{1-\mu}{\tau_2} C(t), \\
 \frac{dH}{dt} = \frac{p_{sev}}{t_{inf}} Sym(t) - \frac{p_{crit}}{t_{hosp}} H(t) - \frac{1-p_{crit}}{\tau_2} H(t), \\
 \frac{dM}{dt} = \frac{1-p_{sev}}{\tau} Sym(t) - \frac{1}{\tau_2} M(t), \\
 \frac{dC}{dt} = \frac{p_{crit}}{t_{hosp}} H(t) - \frac{\mu}{t_{crit}} C(t) - \frac{1-\mu}{\tau_2} C(t), \\
 \frac{dD}{dt} = \frac{\mu}{t_{crit}} C(t),
 \end{cases}
 \quad (4)$$

with the initial conditions:

$$\begin{aligned}
 S(0) &= S_0, E(0) = E_0, I(0) = I_0, A(0) = A_0, Sym(0) = Sym_0, \\
 R(0) &= R_0, H(0) = H_0, M(0) = M_0, C(0) = C_0, D(0) = D_0.
 \end{aligned}$$

Model (4) characterizes a class of agent states for an age group within the agent-based model (see Fig. 1, a).

The equation system for the SEIR-HCD model, where the population is divided into 7 groups (Krivorotko et al., 2020b; Unlu et al., 2020), is composed in a similar fashion:

$$\begin{cases}
 \frac{dS}{dt} = -\frac{5-a(t-\tau)}{5} \left[\frac{\alpha_I S(t)I(t)}{N(t)} + \frac{\alpha_E S(t)E(t)}{N(t)} \right], \\
 \frac{dE}{dt} = \frac{5-a(t-\tau)}{5} \left[\frac{\alpha_I S(t)I(t)}{N(t)} + \frac{\alpha_E S(t)E(t)}{N(t)} \right] - \frac{1}{t_{inc}} E(t), \\
 \frac{dI}{dt} = \frac{1}{t_{inc}} E(t) - \frac{1}{t_{inf}} I(t), \\
 \frac{dR}{dt} = \frac{\beta}{t_{inf}} I(t) + \frac{1-\varepsilon_{HC}}{t_{hosp}} H(t), \\
 \frac{dH}{dt} = \frac{1-\beta}{t_{inf}} I(t) + \frac{1-\mu}{t_{crit}} C(t) - \frac{1}{t_{hosp}} H(t), \\
 \frac{dC}{dt} = \frac{\varepsilon_{HC}}{t_{hosp}} H(t) - \frac{1}{t_{crit}} C(t), \\
 \frac{dD}{dt} = \frac{\mu}{t_{crit}} C(t),
 \end{cases}
 \quad (5)$$

with the initial conditions:

$$\begin{aligned}
 S(0) &= S_0, E(0) = E_0, I(0) = I_0, R(0) = R_0, \\
 H(0) &= H_0, C(0) = C_0, D(0) = D_0.
 \end{aligned}$$

Here, $S(t)$ is a susceptible agent at time t , $E(t)$ – a noncontagious infected (not transmitting the virus), $I(t)$ – a contagious infected (transmitting the virus), $A(t)$ – an asymptomatic case, $Sym(t)$ – a symptomatic case, $H(t)$ – a severe case, $C(t)$ – a critical case (requiring a ventilator), $M(t)$ – a mild case, $R(t)$ – a recovered case, and $D(t)$ – a mortality case. The averaged parameters of models (4) and (5) for the Novosibirsk region are presented in Table 1 (Lauer et al., 2020; Verity et al., 2020; Wölfel et al., 2020).

Table 1. Averaged parameters used in models (4), (5) for the Novosibirsk region (Kerr et al., 2020; Unlu et al., 2020)

Parameter	Description	Value/interval
τ_1	Days before recovery for asymptomatic and mild cases	6–11
τ_2	Days before recovery for severe and critical cases	12–17
p_{sym}	Probability of symptom expression after infection	0.6
p_{sev}	Probability of a symptomatic case becoming a severe one (requiring hospital admission)	0.0072
p_{crit}	Probability of a severe case becoming a critical one (requiring a ventilator)	0.00036
α	Probability of virus transmission to contact agents	(0.01, 0.025)
$E(0)$	Initial quantity of infected agents in the population	(1, 100)
$a(t)$	Yandex self-isolation index describing the compliance with self-isolation principles in the population on a scale from 0 (lack of isolation) to 5 (complete isolation)	(0, 5)
α_E	Contagion probability between the asymptomatic and susceptible groups ($\alpha_E \gg \alpha_I$)	(0, 1)
α_I	Contagion probability between the infected and susceptible groups linked with virus contagiousness and social factors	(0, 1)
β	Recovery probability for infected cases without complications	(0, 1)
ϵ_{HC}	Proportion of admitted critical-state patients requiring a ventilator	(0, 1)
μ	Probability of COVID-19 related mortality	(0, 0.5)
τ	Latent period in days (delay in virion isolation)	2
t_{inc}	Incubation period in days	2–14
t_{inf}	Infection period in days	2.5–14
t_{hosp}	Hospitalization period in days	4–5
t_{crit}	Days on a ventilator	10–20

Note that the coefficients t_{inc}^{-1} , t_{inf}^{-1} , t_{hosp}^{-1} , t_{crit}^{-1} , τ^{-1} , τ_1^{-1} , τ_2^{-1} at the respective agent states in models (4) and (5) describe the delay in transition between states (Likhoshvai et al., 2004). Consider the following equation (5):

$$\frac{dI}{dt} = \frac{1}{t_{inc}} E(t) - \frac{1}{t_{inf}} I(t),$$

where coefficient t_{inc}^{-1} (in the linear approximation) indicates the delay of t_{inc} days before the transition from non-contagious infected group $E(t)$ to contagious infected group $I(t)$, and coefficient t_{inf}^{-1} – that the agent stays in the contagious infected group for the infection period of t_{inf} days.

Mathematical model 1 (see the diagram in Fig. 1, a). Assume that additional information on recoveries and deaths on fixed days is available for the mathematical model (4):

$$R(t_k) = R_k, D(t_k) = D_k, k = 1, \dots, 225. \quad (6)$$

Here, R_k is the number of recovered agents on day k , D_k is the number of disease-related deaths on day k .

The model analyzes the semi-relative sensitivity of two unknown parameters, i. e. contagiousness α and initial quantity of asymptomatic cases $E(0)$ in the model (4) to the measurements (6). It will allow us to determine the possibility of consistent identification of the unknown parameters based on the data available for an adequate representation of epidemiological situation in the region. The sensitivities of parameters (q_k) = (α , $E(0)$), $k = 1, 2$, to measurements (f_i) = (R , D), $i = 1, 2$,

represented by norm $\left\| \frac{\partial f_i(t)}{\partial q_k} q_k^* \right\|_2$ and sorted in descending

Table 2. Semi-relative sensitivities of various model (4) states to the parameters sorted in descending order

Variable, f_i	Parameter, q_k	$\left\ \frac{\partial f_i(t)}{\partial q_k} q_k^* \right\ _2$
$R(t)$	$E(0)$	$8.9 \cdot 10^6$
	α	$7.6 \cdot 10^{14}$
$D(t)$	$E(0)$	$6.7 \cdot 10^{-14}$
	α	$6.07 \cdot 10^{-6}$

order are presented in Table 2. The lower the value $\left\| \frac{\partial f_i(t)}{\partial q_k} q_k^* \right\|_2$, the less the effect of q_k on f_i .

Figure 2 demonstrates how sensitive function $\frac{\partial f_i(t)}{\partial q_k} q_k^*$ changes in time depending on the parameter. Thus, parameters α and $E(0)$ within the model (4) are less sensitive to variable $D(t)$ and are therefore not identifiable by the mortality data alone. On the other hand, these parameters are sensitive to function $R(t)$, and, as a result, are recovered more consistently based on the recovery data.

Mathematical model 2 (see the diagram in Fig. 1, b). Let us now investigate the SEIR-HCD mathematical model (5). Assume that additional information on diagnoses, critical cases, and mortality on fixed days is available:

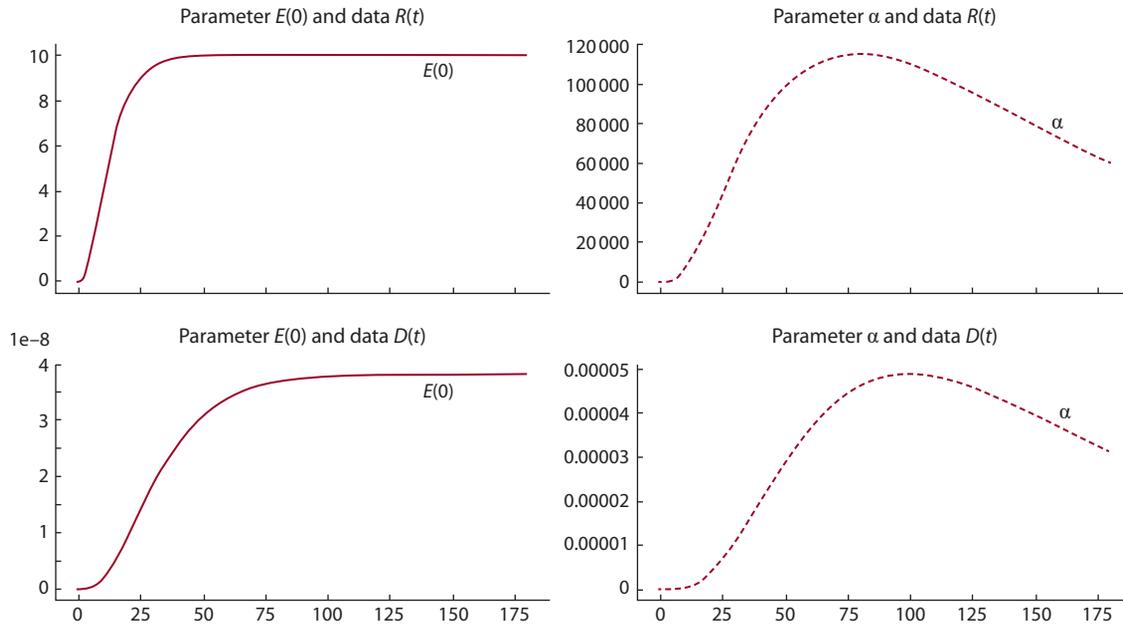


Fig. 2. Sensitive function $\frac{\partial f_i(t)}{\partial q_k} q_k$ for model (4) for the period from 12.03.2020 to 09.09.2020 (182 days).

$$I(t_k) = (1 - b_k) f_k, \quad C(t_k) = C_k, \quad D(t_k) = D_k, \quad (7)$$

$$t_k \in (t_0, T), \quad k = 1, \dots, 205,$$

where $b(t) \in [0, 1]$ is the proportion of asymptomatic carriers in the diagnoses, f_k – the daily number of diagnoses on day k , C_k – the number of critical cases on day k .

Parameters $q = (\alpha_E, \alpha_I, \beta, \varepsilon_{HC}, \mu, E_0)^T \in \mathbb{R}^6$ are considered unknown. To analyze the semi-relative sensitivity of the parameter vector q to measurements (7) within the mathematical model (5), we derive $\frac{\partial f_i(t)}{\partial q_k} q_k^*$, $(f_i) = (I, C, D)$, $i = 1, 2, 3$,

and analyze the values $\left\| \frac{\partial f_i(t)}{\partial q_k} q_k^* \right\|_2$ (Table 3). Consistency of

identifying parameters β , ε_{HC} , and μ as the result of solving the inverse problem barely depends on the available measurements of the number of infected carriers $I(t)$; however, it is not the case for the more sensitive coefficients α_E, α_I, E_0 .

Figure 3 shows how sensitive function $\frac{\partial f_i(t)}{\partial q_k} q_k^*$ changes in time depending on the parameter. The more the parameter changes in time, the higher its sensitivity to the measurements analyzed, and the more consistently it is identified.

In Fig. 4, the results of parameter sensitivity analysis for the model (5) at various iterations of the orthogonal algorithm are presented (see the description of the algorithm in (Krivorotko et al., 2020a)). Iterations of the orthogonal algorithm, whose total number is one less than the dimension of the unknown parameter vector (i. e. the number of columns in the sensitivity matrix), are plotted along the horizontal axis, and the norms of perpendiculars for the obtained transformations of sensitivity matrices – along the vertical axis. It was shown that the contagion between the asymptomatic and susceptible groups α_E ,

the contagion between the infected and susceptible groups α_I linked with virus contagiousness and social factors, and the initial quantity of infected carriers and the agents in incubation period E_0 turned out to be the more identifiable parameters. The ranking of the parameters obtained via sensitivity analysis

Table 3. Semi-relative sensitivities of various states of the model (5) to parameters, sorted in descending order

Variable, f_i	Parameter, q_k	$\left\ \frac{\partial f_i(t)}{\partial q_k} q_k^* \right\ _2$
I	α_E	$2.865 \cdot 10^{14}$
I	α_I	$2.396 \cdot 10^{14}$
I	E_0	$1.854 \cdot 10^{14}$
C	α_E	$2.386 \cdot 10^{12}$
C	α_I	$1.996 \cdot 10^{12}$
C	E_0	$1.544 \cdot 10^{12}$
D	α_E	$7.110 \cdot 10^{11}$
D	α_I	$5.947 \cdot 10^{11}$
D	E_0	$4.601 \cdot 10^{11}$
C	ε_{HC}	$4.833 \cdot 10^4$
C	β	$3.428 \cdot 10^4$
D	ε_{HC}	$3.041 \cdot 10^4$
D	μ	$2.982 \cdot 10^4$
D	β	$2.164 \cdot 10^4$
C	μ	$3.695 \cdot 10^2$
I	β	$2.03 \cdot 10^{-6}$
I	ε_{HC}	$1.6 \cdot 10^{-7}$
I	μ	$3 \cdot 10^{-8}$

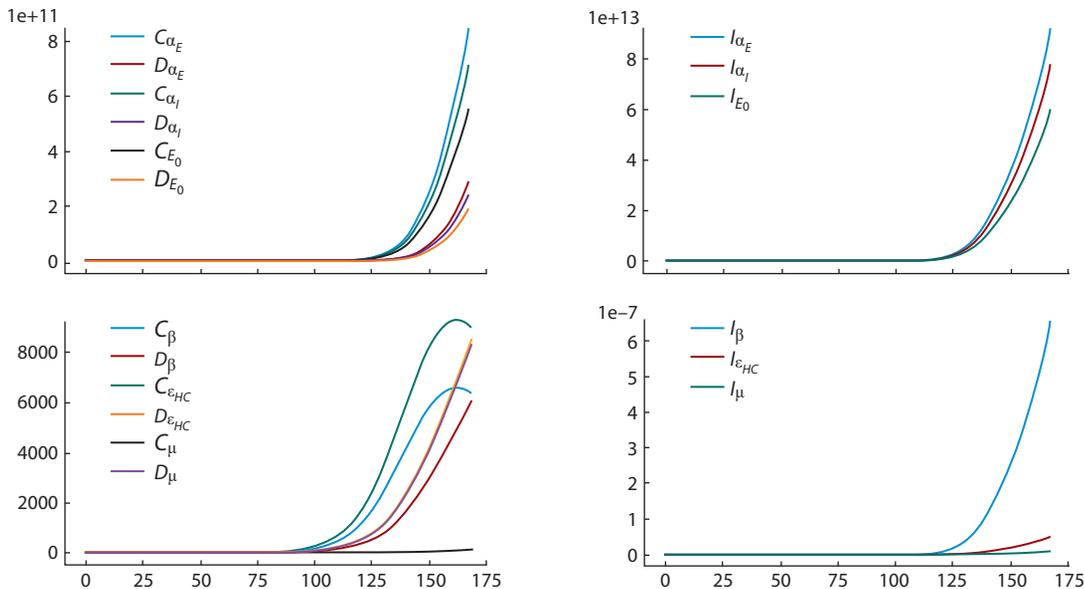


Fig. 3. Semi-relative sensitivity function $\frac{\partial f_i(t)}{\partial q_k} q_k^*$ for the time interval from 15.04.2020 to 01.10.2020 (170 days).

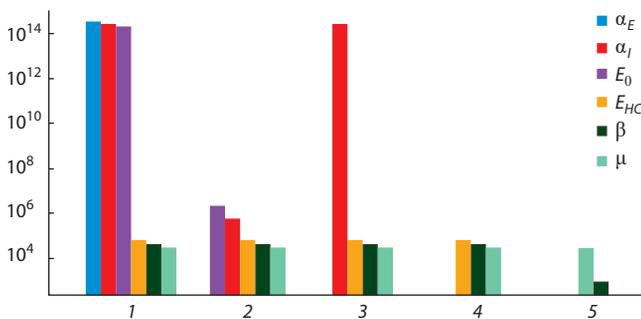


Fig. 4. Normalised perpendiculars for each parameter (different colors) at the different iterations (1–5) of the sensitivity-based orthogonal algorithm (5).

for the model (5) from the most sensitive to the least sensitive is as follows: $\alpha_E, E_0, \alpha_I, \epsilon_{HC}, \mu, \beta$.

As a result of identifiability analysis, a conclusion can be made that model parameters α_E, E_0 , and α_I are the least sensitive to data variations (errors), i. e. are more identifiable. In other words, these parameters are identified more consistently as a result of solving the inverse problem (5), (7). In turn, parameters ϵ_{HC}, μ , and β are the most sensitive to measurement errors, i. e., less identifiable (and have the lowest values of the norms of perpendiculars in the sensitivity matrix). Hence, the regularization algorithm should be developed to ensure the consistent identification of sensitive parameters.

Mathematical modeling of COVID-19 spread in the Novosibirsk region

To build a COVID-19 spread model for the Novosibirsk region, the following publicly reported data were used:

- (a) Number of people tested (including the number of diagnoses f and proportion of asymptomatic carriers $b(t)$), recovered cases (R), and COVID-19 related deaths (D);

- (b) Duration of incubation period t_{inc} , latent period τ , infection period t_{inf} , hospitalization period t_{hosp} , and duration of ventilation t_{crit} ;
(c) Recovery time for mild τ_1 and severe τ_2 cases;
(d) Demographic profiles (population size and its age distribution in the region);
(e) Average household size (2.6 people) in the Russian Federation in 2019, according to UN data (<https://population.un.org/Household/#/countries/840>).

Additional information was regularly obtained from the following websites:

- Ministry of Health of the Novosibirsk region: <https://zdrav.nso.ru/> (d).
- Federal State Statistics Service of the Novosibirsk region: <https://novosibstat.gks.ru/folder/31729> (c).
- Stopcoronavirus website: <https://стопкоронавирус.рф> (a).
- World Health Organization: <https://www.who.int> (b).

The modeling was performed taking into account the measures to contain the COVID-19 spread (Table 4).

Solutions of inverse problems (4), (6) and (5), (7) were reduced to misfit function minimization (Kabanikhin, 2008):

$$J(q) = \sum_s \sum_{i=1}^T w_s \cdot G(c_d^{i,s}, c_m^{i,s}(q)).$$

Here, s is the statistics used for data comparison (cumulative diagnoses, critical cases, and mortality), w_s – the weight coefficient, $c_d^{i,s}, c_m^{i,s}$ – data (with subscript d) and model values (with subscript m), T – the modeling interval in days, q – the unknown parameter vector: $q_1 = (\beta, E_0)^T$ for the inverse problem (4), (6) and $q_2 = (\alpha_E(t), \alpha_I(t), \beta, \epsilon_{HC}, \mu, E_0)^T$ for the inverse problem (5), (7). The absolute norm for computational experiments was set as follows:

$$G_1 = \frac{|c_d^{i,s} - c_m^{i,s}|}{M}, \text{ where } M = \max_t \{c_d^{i,s}\} \text{ was the normalization item; and the standard deviation was } G_2 = (c_d^{i,s} - c_m^{i,s})^2 / T.$$

Table 4. COVID-19 containment measures in the Novosibirsk region to be used in the models (4), (5)

Date	Containment measures
March 18, 2020	Introduction of distance learning in the schools and higher education institutions of the Novosibirsk region
March 28, 2020	Suspension of all mass gathering and public entertainment events in the region
April 27, 2020	Implementation of the face mask regime for shops by the Governor of the Novosibirsk region
July 6, 2020	Opening of outdoor terraces in cafes and restaurants
September 1, 2020	Reintroduction of contact learning in schools and higher education institutions
September 28, 2020	Implementation of the face mask regime for all premises and restricted containment measures in educational institutions

Minimization of misfit function $J(q)$ was implemented using the differential evolution method from the SciPy.Optimize Python library. The general algorithm of global minimum search was as follows:

1. Creation of the initial generation $\{\vec{q}_i\} \in B, i = 1 \dots N_q$.

2. Creation of a new generation:

• Mutation:

For all $\vec{q}_i \in B$ three random vectors were selected as follows: $\vec{v}_1, \vec{v}_2, \vec{v}_3 \in B, (\vec{v}_j \neq \vec{q}_i, j = 1, 2, 3)$.

Mutant vector: $\vec{v} = \vec{v}_1 + F(\vec{v}_2 - \vec{v}_3), F \in [0, 2]$.

• Crossover: trial vector \vec{u} was calculated as follows:

$$u_k = \begin{cases} v_k, & \text{if } rand < p, \\ q_k, & \text{if } rand \geq p, \end{cases} \quad k = 1 \dots N_q.$$

3. Selection:

$$\vec{q}_i = \begin{cases} \vec{q}_i, & \text{if } J(\vec{x}_i) < J(\vec{u}_i), \\ \vec{u}_i, & \text{else.} \end{cases}$$

The results of COVID-19 spread modeling in the Novosibirsk region with the forecast up to December 10, 2020, are presented in Fig. 5. The model was built using the agent-based approach relying on the investigation of interactions between individuals and their effect on global parameters. The modeling was performed using Covasim, a simulator for developing stochastic agent-based models. A detailed discussion of the model structure may be found in (Kerr et al., 2020). We also used the statistical data on diagnoses and deaths from March 12 to October 23, 2020. The following misfit function was minimized taking into account the identifiability analysis results for the model (4), (6):

$$J(q_1) = \frac{1}{T} \sum_{i=1}^T (f_d^i - f_m^i)^2 + 100 \cdot (D_d^i - D_m^i)^2.$$

Here, f_d^i, f_m^i are cumulative diagnoses, and D_d^i, D_m^i are cumulative deaths.

Modeling results f_m^i and statistics f_d^i of cumulative and daily diagnoses are presented in Fig. 5, a, b. Modeling results D_m^i and statistics D_d^i of cumulative COVID-19 related deaths in the Novosibirsk region are presented in Fig. 5, c. Note that the second wave of the epidemic may be observed in the Novosibirsk region in mid-September in both the statistical data and modeling results. Its growth will be insignificant (i.e. it will not exceed 215 new daily diagnoses by mid-December,

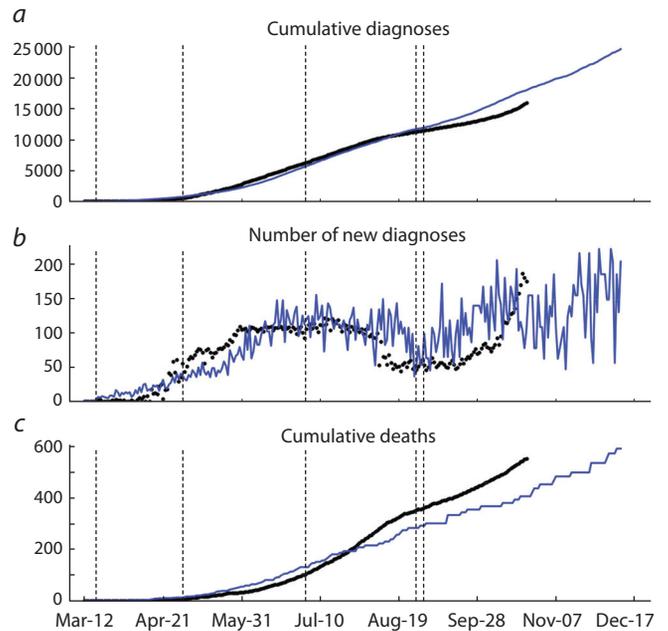


Fig. 5. COVID-19 spread model for the Novosibirsk region (solid blue line) using the agent-based approach and statistical data (black dots) with containment measures (vertical dashed lines).

2020) due to the introduction of stricter containment measures from October 28.

The inverse problem (5), (7) was reduced to the minimization of the following misfit function (Krivorotko et al., 2020b):

$$J(q_2) = \sum_{k=1}^K (w_1 |t_{inc}^{-1} E(t_{k-1}; q_2) - (1 - b_k) f_k| + w_2 |C(t_k; q_2) - C_k| + w_3 |D(t_k; q_2) - D_k|).$$

Infection rate parameters $\alpha_E(t)$ and $\alpha_I(t)$ linked to virus contagiousness and varying in time were represented as piecewise constant functions depending on the interventions (see Table 4).

Based on the identifiability analysis results for the model (5), (7), more rigid restrictions were imposed for poorly identifiable parameters (see Table 1). The result of solving the inverse problem (5), (7) for the SEIR-HCD model for the period from April 15, 2020, to October 3, 2020, is presented in Fig. 6.

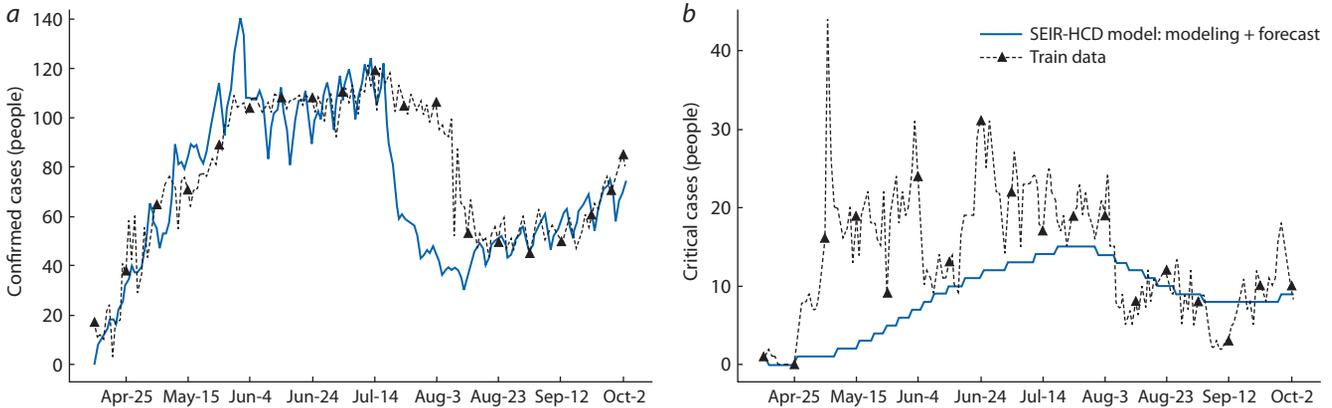


Fig. 6. Modeling COVID-19 spread in the Novosibirsk region (solid blue line) from 15.04.2020 to 03.10.2020 and the statistical data (dashed black line) for (a) daily confirmed cases f_k and (b) the critical cases C_k requiring a ventilator.

Note that although the rough mathematical model (with ODE system of 7 equations) captures the general trend based on the number of diagnoses (the peak of confirmed cases in the region, see Fig. 6, a), it is still unable to qualitatively describe highly variable statistics (critical cases requiring a ventilator, see Fig. 6, b). Nonsmooth solutions in Fig. 6 result from the use of the Yandex self-isolation index characterized by weekly seasonality. In this case, smoothing would undermine the use of the tool. A more detailed analysis of modeling and forecasting of COVID-19 spread in the Moscow and Novosibirsk regions is presented in (Krivorotko et al., 2020b). This case requires the agent-based approach capable of detailed description of small statistical datasets.

Conclusions

In the present study, sensitivity-based identifiability analysis has been performed for the COVID-19 pandemic spread models based on systems of differential equations. The algorithm is based on the analysis of the sensitivity matrix using the differential and linear algebra apparatus, which shows the degree of dependence of the unknown model parameters on the given measurements.

The analysis has shown that the virus contagiousness is consistently identified based on the number of new daily diagnoses, critical cases, and recoveries. On the other hand, the predicted proportion of admitted critical state patients requiring a ventilator and the mortality rate are identified much less consistently. It has been demonstrated that developing a more realistic forecast will require additional information about the process such as the number of daily hospital admissions.

The identifiable parameters refinement problems have been reduced to the minimization of the respective misfit functions describing the proximity of the modeling data to the statistics of the diagnoses, critical cases, and mortality in the Novosibirsk region. The use of absolute and quadratic norms as measures of deviation between the data and the modeling results in the minimization procedures has not yielded any significant differences in terms of analyzing the modeling results. It has been shown that a rough compartmental model

of seven ODEs describes the general trend of the coronavirus infection propagation, as it is sensitive to peaks of confirmed cases; however, it is unable to qualitatively describe smaller statistics (daily numbers of critical cases t_k and deaths), which may lead to improper conclusions. A more detailed mathematical model using the agent-based approach, where a class of agent states is expressed by a system of ten ODEs, will make it possible to describe noisy statistical datasets in more detail and build adequate scenarios of COVID-19 pandemic spread.

References

- Adams B.M., Banks H.T., Davidiana M., Kwona H.D., Trana H.T., Wynnea S.N., Rosenbergb E.S. HIV dynamics: modeling, data analysis, and optimal treatment protocols. *J. Comput. Appl. Math.* 2004; 184:10-49. DOI 10.1016/j.cam.2005.02.004.
- Bellu G., Saccomani M.P., Audoly S., D'Angiò L. DAISY: a new software tool to test global identifiability of biological and physiological systems. *Comput. Methods Programs Biomed.* 2007;88(1):52-61. DOI 10.1016/j.cmpb.2007.07.002.
- Gomez J., Prieto J., Leon E., Rodriguez A. INFEKTA: a general agent-based model for transmission of infectious diseases: studying the COVID-19 propagation in Bogotá – Colombia. *MedRxiv.* 2020. DOI 10.1101/2020.04.06.20056119.
- Habtemariam T., Tameru B., Nganwa D., Beyene G., Ayanwale L., Robnett V. Epidemiologic modeling of HIV/AIDS: use of computational models to study the population dynamics of the disease to assess effective intervention strategies for decision-making. *Adv. Syst. Sci. Appl.* 2008;8(1):35-39.
- Kabanikhin S.I. Definitions and examples of inverse and ill-posed problems. *J. Inverse Ill-Posed Probl.* 2008;16(4):317-357. DOI 10.1515/JIIP.2008.019.
- Kabanikhin S.I., Voronov D.A., Grodz A.A., Krivorotko O.I. Identifiability of mathematical models in medical biology. *Russ. J. Genet. Appl. Res.* 2016;6(8):838-844. DOI 10.1134/S2079059716070054.
- Kermack W.O., McKendrick A.G. A contribution of the mathematical theory of epidemics. *Proc. R. Soc. Lond. A.* 1927;115:700-721. DOI 10.1098/rspa.1927.0118.
- Kerr C., Stuart R., Mistry D., Abesuriya R., Hart G., Rosenfeld K., Selvaraj P., Nunez R., Hagedorn B., George L., Izzo A., Palmer A., Delpont D., Bennette C., Wagner B., Chang S., Cohen J., Panovska-Griffiths J., Jastrzebski M., Oron A., Wenger E., Famulare M., Klein D. Covasim: an agent-based model of COVID-19 dynamics and interventions. *MedRxiv.* 2020. DOI 10.1101/2020.05.10.20097469.

- Krivorotko O.I., Andornaya D.V., Kabanikhin S.I. Sensitivity analysis and practical identifiability of some mathematical models in biology. *J. Appl. Ind. Math.* 2020a;14:115-130. DOI 10.1134/S1990478920010123.
- Krivorotko O.I., Kabanikhin S.I., Zyat'kov N.Yu., Prikhod'ko A.Yu., Prokshoshin N.M., Shishlenin M.A. Mathematical modeling and forecasting of COVID-19 in Moscow and Novosibirsk region. *Numer. Analysis Applications.* 2020b;13(4):332-348. DOI 10.1134/S1995423920040047.
- Lauer S.A., Grantz K.H., Bi Q., Jones F.K., Zheng Q., Meredith H., Azman A.S., Reich N.G., Lessler J. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann. Intern. Med.* 2020;172:577-582. DOI 10.7326/m20-0504.
- Lee W., Liu S., Tembine H., Li W., Osher S. Controlling propagation of epidemics via mean-field games. *ArXiv.* 2020;arXiv:2006.01249.
- Likhoshvai V.A., Fadeev S.I., Demidenko G.V., Matushkin Yu.G. Modeling nonbranching multistage synthesis by an equation with retarded argument. *Sibirskiy Zhurnal Industrialnoy Matematiki = Journal of Applied and Industrial Mathematics.* 2004;7(1):73-94. (in Russian)
- Miao H., Xia X., Perelson A.S., Wu H. On identifiability of nonlinear ODE models and applications in viral dynamics. *SIAM Rev.* 2011;53(1):3-39. DOI 10.1137/090757009.
- Raue A., Becker V., Klingmüller U., Timmer J. Identifiability and observability analysis for experimental design in nonlinear dynamical models. *Chaos.* 2010;20(4):045105. DOI 10.1063/1.3528102.
- Raue A., Karlsson J., Saccomani M.P., Jirstrand M., Timmer J. Comparison of approaches for parameter identifiability analysis of biological systems. *Bioinformatics.* 2014;30(10):1440-1448. DOI 10.1093/bioinformatics/btu006.
- Tuomisto J.T., Yrjölä J., Kolehmainen M., Bonsdorff J., Pekkanen J., Tikkanen T. An agent-based epidemic model REINA for COVID-19 to identify destructive policies. *MedRxiv.* 2020. DOI 10.1101/2020.04.09.20047498.
- Unlu E., Leger H., Motornyi O., Rukubayihunga A., Ishacian T., Chouiten M. Epidemic analysis of COVID-19 outbreak and counter-measures in France. *MedRxiv.* 2020. DOI 10.1101/2020.04.27.20079962.
- Verity R., Okell L., Dorigatti I., Winskill P., Whittaker C., Imai N., Cuomo-Dannenburg G., Thompson H., Walker P., Fu H., Dighe A., Griffin J., Baguelin M., Bhatia S., Boonyasiri S., Cori A., Cucunubá Z., FitzJohn R., Gaythorpe K., Green W., Hamlet A., Hinsley W., Laydon D., Nedjati-Gilani G., Riley S., Elstrand S., Volz E., Wang H., Wang Y., Xi X., Donnelly C., Ghani A., Ferguson N.M. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect. Dis.* 2020;20(6):669-677. DOI 10.1016/S1473-3099(20)30243-7.
- Voropaeva O.F., Tsgoev Ch.A. A numerical model of inflammation dynamics in the core of myocardial infarction. *J. Appl. Ind. Math.* 2019;13(2):372-383. DOI 10.1134/S1990478919020182.
- Wolfram C. An agent-based model of COVID-19. *Complex Syst.* 2020;29(1):87-105. DOI 10.25088/ComplexSystems.29.1.87.
- Wölfel R., Corman V.M., Guggemos W., Seilmaier M., Zange S., Müller M.A., Niemeyer D., Jones T.C., Vollmar P.V., Rothe C., Hoelscher M., Bleicker T., Brünink S., Schneider J., Ehmann R., Zwirgmaier K., Drosten C., Wendtner C. Virological assessment of hospitalized patients with COVID-2019. *Nature.* 2020;581:465-469. DOI 10.1038/s41586-020-2196-x.
- Yao K.Z., Shaw B.M., Kou B., McAuley K.B., Bacon D.W. Modeling ethylene/butene copolymerization with multi-site catalysts: parameter estimability and experimental design. *Polymer Reaction Engineer.* 2003;11(3):563-588. DOI 10.1081/PRE-120024426.

ORCID ID

O.I. Krivorotko orcid.org/0000-0003-0125-4988

Acknowledgements. The work was supported by the Russian Foundation for Basic Research (project no. 18-31-20019) and the Council for Grants of the President of the Russian Federation (project no. 075-15-2019-1078 (MK-814.2019.1)).

Conflict of interest. The authors declare no conflict of interest.

Received October 25, 2020. Revised December 17, 2020. Accepted December 18, 2020.