Mathematical Modeling and Forecasting of COVID-19 in Moscow and Novosibirsk Region

O. I. Krivorot'ko^{1,2,3*}, S. I. Kabanikhin^{1,2,3**}, N. Yu. Zyat'kov^{1***}, A. Yu. Prikhod'ko^{1,2,3****}, N. M. Prokhoshin^{2,3****}, and M. A. Shishlenin^{1,2,3*****}

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

²Mathematical Center in Akademgorodok, Novosibirsk, Russia

³Novosibirsk State University, Novosibirsk, Russia Received July 20, 2020; in final form, July 30, 2020; accepted July 30, 2020

Abstract—We investigate inverse problems of finding unknown parameters of mathematical models SEIR-HCD and SEIR-D of COVID-19 spread with additional information about the number of detected cases, mortality, self-isolation coefficient, and tests performed for the city of Moscow and Novosibirsk region since 23.03.2020. In SEIR-HCD the population is divided into seven groups, and in SEIR-D into five groups with similar characteristics and transition probabilities depending on the specific region of interest. An identifiability analysis of SEIR-HCD is made to reveal the least sensitive unknown parameters as related to the additional information. The parameters are corrected by minimizing some objective functionals which is made by stochastic methods (simulated annealing, differential evolution, and genetic algorithm). Prognostic scenarios for COVID-19 spread in Moscow and in Novosibirsk region are developed, and the applicability of the models is analyzed.

DOI: 10.1134/S1995423920040047

Keywords: mathematical modeling, pandemic, COVID-19, SIER-HCD, SIER-D, scenarios, inverse problem, identifiability, optimization, differential evolution, simulated annealing, genetic algorithm, Moscow, Novosibirsk region.

INTRODUCTION

In December 2019, a pneumonia outbreak took place in Wuhan (China), as a result of which a COVID-19 strain was first detected in a nucleic acid analysis in a patient with the pneumonia. By the end of June 2020 the pandemic captured 188 countries, with more than 10 million of detected infection cases, and 505,500 people have died. According to the cases detected, the Russian Federation is the third in ranking of infected countries, behind the United States and Brazil, with 484,630 infection cases by June 9, 2020 (see Table 1). Although a peak of the disease in the world seems to have been passed, no decrease in the number of detected cases has been observed for a sufficiently long period [1].

To create proper scenarios of disease development in Moscow (the largest number of detected cases in Russia) and Novosibirsk region (moderate detection rate) is an important step in taking appropriate measures to curb the epidemic in the regions.

Papers [2] and [3] are the first to describe some scenarios of the coronavirus epidemic development in Moscow.

^{*}E-mail: **krivorotko.olya@mail.ru**

^{**}E-mail: ksi52@mail.ru

^{***}E-mail: nikolay.zyatkov@gmail.com

^{*****}E-mail: a.prikhodko@g.nsu.ru

E-mail: nikita.prokhoshin@gmail.com

E-mail: mshishlenin@ngs.ru

MATHEMATICAL MODELING AND FORECASTING OF COVID-19

Country	Infe	ected	Deaths	Cured
	Total	New cases		
USA	1,961,185	17,250	111,007	524,855
Brazil	707,412	15,654	37,134	396,737
Russia	485,253	8 595	6141	241,917
Great Britain	288,834	1213	40,680	1 257

Table 1. COVID-19 statistics for five leading countries in the number of cases detected by June 9, 2020

The COVID-19 epidemic spread in China, taking into account the incubation period, passenger traffic, and decisions of the Health Care Commission is studied with SEIR mathematical models in [4-9]. A review of the mathematical models and methods for their solution can be found in [10], and a review of the literature on estimating contagiousness parameters for various countries, in [11].

A model called QSEIR that takes into account the quarantine measures in China to curb the epidemic is proposed in [12]. The model parameters are estimated by a statistical method. Some forecasts of the epidemic based on some strategies of quarantine measures in China have been made. A forecast of the pandemic development in China based on an approximation of the data with some basic functions has also been made in [13].

In paper [14], a model called SIR-D is used to estimate the contagiousness parameter, daily mortality, and recovery rate. Some parameters are found from Chinese data, and a forecast is made for three weeks up to February 29. It was shown in [15] that the Chinese epidemic would have a peak by the end of February, and gradually decline by the end of April. It was shown with machine learning methods that quarantine lifting in Hubei would cause a second peak of the epidemic in this province in mid-March and continue the epidemic to the end of April.

A mathematical model with some control measures called SEIR-HD is proposed in [16].

In paper [17], a generalized SEIR model is applied to SARS-CoV-2 spread in Italy. A particle swarm method is used to find the model parameters for increasing the reliability of 30-day forecasts. The results obtained are compared with data and forecasts for Spain and South Korea. The model is used to take into account the mobility of people according to Google COVID-19 Community Mobility Reports. In [18, 19], modified SEIR models are used to obtain scenarios of the epidemic in Italy.

In [20], a SIR-D model is used to obtain scenarios of the epidemic in Brazil taking into account school closures and isolation of elderly people.

In [21], a SEIR-D model with time-dependent parameters and calculation results are presented for some Indian districts, as well as for Lombardy (Italy) and Moscow (Russia).

The purpose of the present paper is to study the reliability of forecasts by the SEIR models versus the available information on the COVID-19 pandemic spread.

The SIR (Susceptible, Infected, Recovered) model is used as a basis for describing the spread of infectious diseases; it was proposed in the 1920s by Scottish epidemiologists Anderson Kermack and William McKendrick. According to the SIR, the population is divided into three groups: susceptible (S), infected (I), and recovered (R):

- *S*—susceptible (uninfected) (age three and older);
- *I*—infected with symptoms;
- *R*—cured.

With time, transitions $S \rightarrow I$ (infection) and $I \rightarrow R$ (recovery or death) are possible. The SIR-model does not take into account many factors (for instance, population densities in different regions or different transmission routes). Therefore, the SIR has been refined many times [22–24]. Today there is a whole family of models (and even open codes) developed on the basis of the SIR model [25–28]:

- SIRS: to describe the dynamics of diseases with temporary immunity.
- SEIR (Exposed): to describe the spread of diseases taking into account incubation period.
- SIS: to describe the spread of diseases with no immunity.
- MSEIR (Maternally derived immunity): to take into account maternally derived immunity.

In the present paper, the following two models are used: SEIR-HCD and SEIR-D, to which we add

- *E*—infected or in incubation period;
- *H*—hospitalized, that is, with a severe form of the disease;
- *C*—individuals in critical condition under artificial lung ventilation (ALV);
- *D*—dead.

The choice of the models is based on the SEIR, since COVID-19 has a rather long incubation period (5–14 days), during which there may be carriers who do not show any symptoms of the disease [29].

In the present paper, first the coefficients of transition from one group to another are specified by solving an inverse problem with data on the number of infected individuals I(t) and E(t), and the number of the dead D(t). Then we solve a direct problem, that is, calculate a scenario of the epidemic.

The following additional information about the disease is used: duration of the incubation period, duration of the latent period (from the time of infection to the time when the infected individual becomes a source of infection), and a contagiousness parameter.

With the generalized model, we can calculate scenarios of the disease in some cities of Russia: For instance, predict a disease peak in Moscow and Novosibirsk region, the time of reaching a plateau, and a decline in the epidemic in each specific case. In these models we plan to take into account social distancing, climatic conditions, age characteristics, peculiarities of human immunity, passenger intercity transportation in the Russian Federation, etc.





1. MATHEMATICAL MODEL SEIR-HCD

Mathematical model SEIR-HCD of COVID-19 spread in a specific region of the Russian Federation was first proposed in [30]. The model is based on a system of seven nonlinear ordinary differential equations in an interval $t \in [t_0, T]$:

$$\begin{cases} \frac{dS}{dt} = -\frac{5-a(t-\tau)}{5} \left(\frac{\alpha_I S(t)I(t)}{N} + \frac{\alpha_E S(t)E(t)}{N} \right) + \gamma R(t), \\ \frac{dE}{dt} = \frac{5-a(t-\tau)}{5} \left(\frac{\alpha_I S(t)I(t)}{N} + \frac{\alpha_E S(t)E(t)}{N} \right) - (\kappa + \rho)E(t), \\ \frac{dI}{dt} = \kappa E(t) - (\beta + \nu)I(t), \\ \frac{dR}{dt} = \beta I(t) + \rho E(t) - \gamma R(t) + \varepsilon_{HR}H(t), \\ \frac{dH}{dt} = \nu I(t) + \varepsilon_{CH}C(t) - (\varepsilon_{HC} + \varepsilon_{HR})H(t), \\ \frac{dC}{dt} = \varepsilon_{HC}H(t) - (\varepsilon_{CH} + \mu)C(t), \\ \frac{dD}{dt} = \mu C(t) \end{cases}$$
(1)

with initial conditions

$$S(t_0) = S_0, \quad E(t_0) = E_0, \quad I(t_0) = I_0, \quad R(t_0) = R_0,$$

$$H(t_0) = H_0, \quad C(t_0) = C_0, \quad D(t_0) = D_0.$$
(2)

The entire population N = S + E + I + R + H + C + D consists of seven groups (see a schematic diagram of the model in Fig. 1, left).

The parameters and their average values are given in Table 2.

1.1. Inverse Problem

Assume that additional measurements about the following three functions for fixed times are available:

$$E(t_k) = b_k f_k, \quad I(t_k) = (1 - b_k) f_k, \quad \Delta D(t_k) = g_k, \quad t_k \in (t_0, T), \quad k = 1, \dots, K,$$
(3)

where $b(t) \in [0, 1]$ is the fraction of asymptomatic patients in identified cases (the data are taken from stopkoronavirus.rf), f_k is the number of detected cases per day k, g_k is the number of deaths per day k, $\Delta D(t_k) = D(t_k) - D(t_{k-1})$, and K is the number of days in the statistics.

Unknown parameters: $q = (\alpha_E, \alpha_I, \kappa, \beta, \nu, \varepsilon_{CH}, \mu, E_0) \in \mathbb{R}^8$.

The inverse problem (1)-(3) is in determining the vector of parameters q using the additional measurements (3).

The inverse problem is reduced to minimizing the following objective functional:

$$J(q) = \sum_{k=1}^{K} \left[(c_k^{\text{test}} E(t_k; q) - b_k f_k)^2 + (c_k^{\text{test}} I(t_k; q) - (1 - b_k) f_k)^2 + (\Delta D(t_k; q) - g_k)^2 \right].$$
(4)

Here $c_k^{\text{test}} \in [0, 1]$ is the ratio of the number of tests to the number of healthy individuals in the region under study per day k. The functional (4) is based on the following assumptions: $100b_k \%$ of the identified cases do not show any symptoms (group E(t)). Hence, $100b_k \%$ and $100(1 - b_k)\%$ of the detected cases are proportional to the tested (testing coefficient c_k^{test}) asymptomatic and symptomatic cases, respectively.

Symbol	Description	Limits	
		SEIR-HCD	SEIR-D
a(t)	Self-isolation index according to Yandex	(0, 5)
τ	Latent period (characterizes delays in the release of virions or contagiousness)	2 day	/S
α_I	Infection transmission between infected and susceptible individuals associated with virus contagiousness and social factors	(0, 1)
α_E	Infection transmission between asymptomatic and susceptible population groups ($\alpha_E \gg \alpha_I$)	(0, 1)
κ	Frequency of symptoms in open cases leading to transition from asymptomatic to infected population	(0, 1)
ρ	Recovery rate of detected cases (recovering without any symptoms)	0	(0, 1)
β	Recovery rate of infected cases	(0, 1)
γ	Reinfection rate. This parameter is reciprocal of virus immunity level (0—stable immunity, 0.001—probability of reinfection)	0	
ν	Hospitalized cases with severe disease	(0, 1)	—
ε_{HR}	Probability of recovery in severe condition	0.225	_
ε_{HC}	Fraction of hospitalized cases with severe disease under artificial lung ventilation (ALV)	0.025	
ε_{CH}	Probability for a patient to be taken off the ALV	(0, 1)	—
$c^{\rm isol}$	Coefficient of impact of self-isolation index on infection	_	(0, 1)
μ	COVID-19 mortality	$(10^{-4}, 10^{-1})$	(0, 0.1)
E_0	Initial number of asymptomatically infected	(1,800)	(0,600)
R_0	Initial number of cured	—	(0,600)

Table 2. Parameters of models (1), (2), (5), and (6) and their limits

2. MATHEMATICAL MODEL SEIR-D

Model SEIR-D of COVID-19 spread is described by a system of five nonlinear ordinary differential equations in an interval $t \in [t_0, T]$ [31] (a schematic diagram of the model is shown in Fig. 1 right):

$$\begin{pmatrix}
\frac{dS}{dt} = -c(t-\tau) \left(\frac{\alpha_I S(t) I(t)}{N} + \frac{\alpha_E S(t) E(t)}{N} \right) + \gamma R(t), \\
\frac{dE}{dt} = c(t-\tau) \left(\frac{\alpha_I S(t) I(t)}{N} + \frac{\alpha_E S(t) E(t)}{N} \right) - (\kappa + \rho) E(t), \\
\frac{dI}{dt} = \kappa E(t) - \beta I(t) - \mu I(t), \\
\frac{dR}{dt} = \beta I(t) + \rho E(t) - \gamma R(t), \\
\frac{dD}{dt} = \mu I(t)
\end{cases}$$
(5)

with initial data

$$S(t_0) = S_0, \quad E(t_0) = E_0, \quad I(t_0) = I_0, \quad R(t_0) = R_0, \quad D(t_0) = D_0.$$
 (6)

Here N = S + E + I + R + D is the entire population.

A function that uses some restrictions on social mobility is

$$c(t) = 1 + c^{\text{isol}}\left(1 - \frac{2}{5}a(t)\right), \quad c(t) \in (0, 2).$$

2.1. Statement of the Inverse Problem

Additional information:

- infected per day $f_k, k = 1, \ldots, K$,
- deaths per day $g_k, k = 1, \ldots, K$.

In the model (5) this means

$$0.58f_k = \kappa E_{k-1}, \quad g_k = D(t_k) - D(t_{k-1}), \quad t_k \in (t_0, T), \ k = 1, \dots, K.$$
(7)

There is very little information about the real number of asymptomatically infected persons. However, it seems that almost all infected individuals with symptoms are detected during the day. On average, 58% of the infected individuals detected per day (f_k) are with symptoms (I). The infected individuals detected on day k with symptoms $(0.58f_k)$ are a fraction of the infected individuals without symptoms on day k - 1 ($\kappa E(t_{k-1})$).

Unknown parameters of the model are $q = (\alpha_E, \alpha_I, \kappa, \rho, \beta, \mu, c^{\text{isol}}, E_0, R_0) \in \mathbb{R}^9$.

The inverse problem (5)–(7) is in determining the vector of parameters q from the additional information (7).

The inverse problem (5)-(7) is reduced to minimizing the functional

$$J(q) = \sum_{k=1}^{K} \left[w_1(\kappa E(t_{k-1};q) - 0.58f_k)^2 + w_2(D(t_k;q) - D(t_{k-1};q) - g_k)^2 \right].$$
 (8)

The weights in the functional (8) are chosen as follows:

$$w_1 = \sum_{k=1}^{K} \frac{f_k}{K}, \qquad w_2 = \sum_{k=1}^{K} \frac{g_k}{K}.$$

3. NUMERICAL EXPERIMENTS AND FORECAST RESULTS

This section presents a comparison of numerical results where the inverse problems (1)–(3) are used for model SEIR-HCD, and (5)–(7) for model SEIR-D in Moscow and Novosibirsk region. The criteria of the comparison are as follows::

- 1. Accuracy of epidemic peak prediction in Moscow (date and number of detected infected individuals).
- 2. Accuracy of disease spread prediction in Moscow according to daily detected cases f_k and mortality $g_k, k = 1, \ldots, K$.
- 3. Accuracy of processing of the historical data of daily detected cases f_k and mortality g_k in Moscow.
- 4. Required computing resources (runtime of algorithms, computing power).
- 5. Comparison of the necessary a priori information on the limits of the desired parameters (see Table 2, columns 3 and 4).

3.1. Data Processing

The self-isolation index data a(t) in Moscow and Novosibirsk region are taken from Yandex maps. The testing coefficient was calculated using statistical information from open sources by the following formula:

$$c_k^{\text{test}} = \frac{T_k}{Z_k},$$

where T_k is the number of tests made in the region under study on day k, which was obtained by multiplying the number of tests done in the country on day k by the proportion of the region's population in the country's population, Z_k is the number of uninfected individuals in the region by the day kcalculated as the difference between the region's population and the total number of sick and dead individuals over the entire observation period up to day k (no data on recovered individuals were used). T_k and Z_k were extrapolated for the future period of forecasting.

3.2. Algorithms for Solving the Inverse Problems

3.2.1. Model SEIR-HCD

To numerically solve the inverse problem (1)-(3), the following sequence of steps was used:

- 1. Processing of data f_k , g_k , c_k^{test} , $a(t_k)$, b_k , k = 1, ..., K, for time periods (a)-(d) (see Section 3.3).
- 2. Determining the parameter limits for the sought-for vector q (see Table 2, column 3).
- 3. Using a method of differential evolution with module *scipy.optimize.differential_evolution* of the *SciPy* library to determine the minimum of the functional (4) and obtain the optimal vector of parameters *q*^{*} for the mathematical model (1), (2):
 - 3.1. Solving the direct problem (1), (2) and calculating the functional (4) for each iteration of the method. To solve the direct problem (1), (2) module *scipy.integrate.odeint* of the *SciPy* library was used.
 - 3.2. Determining the optimal vector of parameters q^* .
- 4. Solving the direct problem (1), (2) for time periods (a)–(d) using the thus found optimal parameters from q^* and making forecasts for corresponding time periods (a)–(d). The number of infected individuals per day \hat{f}_k is calculated by the formula $\hat{f}_k = c_k^{\text{test}}(I(t_k) + E(t_k))$.

3.2.2. Model SEIR-D

The calculations were made using the Python programming language. To numerically solve the inverse problem (5)—(7), the following sequence of steps was used:

- 1. Processing of data f_k , g_k , $a(t_k)$, k = 1, ..., K, for time periods (a)-(d).
- 2. Determining the parameter limits for the sought-for vector q (see Table 2, column 4).
- 3. Determining the minimum of the functional (8) and obtaining the optimal vectors of the parameters q^* of the mathematical model (5), (6) using the following four methods:

1) differential evolution method (module *scipy.optimize.differential_evolution*),

- 2) simulated annealing method (module *scipy.optimize.dual_annealing*),
- 3) genetic algorithm (https://pypi.org/project/geneticalgorithm/ library),
- 4) particle swarm method (https://pythonhosted.org/pyswarm/ library).

- From the vectors q* obtained by methods 1)-4), the best one is chosen in terms of the functional (8):
 - 1) Solving the direct problem (5), (6) and calculating the functional (8) for each iteration of the method. To solve the direct problem (5), (6) module *scipy.integrate.ode_ivp* of the *SciPy* library, method *RK45*, was used.
 - 2) Determining the optimal vector of parameters q^* of the mathematical model (5), (6).
- 5. Solving the direct problem (5), (6) for time periods (a)–(d) using the thus found optimal parameters from q^* and making forecasts for corresponding time periods (a)–(d). The number of infected individuals detected per day \hat{f}_k is calculated by the formula $\hat{f}_k = \frac{\kappa E(t_{k-1})}{0.58}$.

3.3. Numerical Calculations for Moscow

As initial data on the COVID-19 spread in Moscow ($N_0 = 11514330$) for SEIR-HCD model (1), we used the following statistical information for March 23, 2020:

 $S_0 = 11\ 514\ 241 - q_8, \ E_0 = q_8, \ I_0 = 71, \ R_0 = 9, \ H_0 = 8, \ C_0 = 1, \ D_0 = 0.$

In the SEIR-D (5), the following initial data were used:

$$S_0 = 11514330 - I_0 - q_8 - q_9, E_0 = q_8, I_0 = 71, R_0 = q_9, D_0 = 0.$$

Data for the inverse problems were obtained from open sources for Moscow in the following time periods:

- (a) 23.03.2020-21.05.2020 (K = 60 days), (b) 23.03.2020-11.05.2020 (K = 50 days),
- (c) 23.03.2020-01.05.2020 (K = 40 days), (d) 23.03.2020-21.04.2020 (K = 30 days).

The calculations were made using the Python programming language. The forecast was made for the following time periods according to above data (a)-(d):

(a)	22.05.2020–21.06.2020 (30 days),	(b) 12.05.2020–11.06.2020 (30 days),
(c)	02.05.2020-01.06.2020 (30 days),	(d) 22.04.2020-21.05.2020 (30 days).

3.3.1. Comparison of the Results

A comparison of the results obtained by using two mathematical models, SEIR-HCD and SEIR-D, was made. The parameters of these models were specified for Moscow by optimization methods (differential evolution, simulated annealing, genetic algorithm, and particle swarm).

Figure 2 shows the number of infected individuals detected for the two models verified by using real data up to May 27, 2020 for time periods (a)–(d). Note that even with the scarce statistical data the epidemic spread forecast is more accurate with SEIR-HCD (curve 3). This means that the forecast of the number of detected cases in Moscow is rather accurate (see the ratio of the SEIR-HCD simulation result (curve 3) to the statistical data not used to solve the inverse problem (curve 2)). SEIR-D has a larger error in predicting the number of detected cases, but it describes historical data with better accuracy (see the ratio of the SEIR-D simulation result (curve 4) to the statistical data used to solve the inverse problem (curve 1)).

The parameters reconstructed for SEIR-HCD and SEIR-D are presented in Tables 3–6. An analysis of the behavior of the parameters for SEIR-HCD with the various data sets has shown that the parameters ε_{CH} and β have larger variation ranges, in contrast to the other parameters, as noted in Section 4 where a sensitivity analysis is performed. Thus, in solving the inverse problem some form of regularization and a priori information [32] (for instance, in the form of statistical constraints) were used in the calculations.

Tables 7–10 present comparisons of the simulation results for \hat{y} with real data y for (a)–(d) according to the following metrics taking into account the time period length M:



Fig. 2. The number of detected cases in Moscow for measurements from 23.03.2020 to simulation periods of 90, 80, 70, 60, and 90 days (shown in the plots), respectively: Curve 1—real data, f_k , used in solving the inverse problems, curve 2—real data up to 27.05.2020 not used in solving the inverse problems, curve 3—solution for SEIR-HCD, curve 4—solution for SEIR-D. The reconstructed parameters for the mathematical models are presented in Tables 3–6.

Model	α_E	α_I	κ	ρ	β	ν	ε_{CH}	μ	c^{isol}	E_0	R_0
SEIR-HCD	0.995	0.499	0.280	_	0.146	0.06	0.003	0.0076	-	795	_
SEIR-D	0.510	0.618	0.002	0.187	0.230	_	—	0.0052	0.9	783	187
Table 4. Reconstructed parameters for 23.03.2020–11.05.2020 measurement period											
Model	α_E	α_I	κ	ρ	β	ν	ε_{CH}	μ	$c^{\rm isol}$	E_0	R_0
SEIR-HCD	0.999	0.5	0.290	_	0.16	0.037	0.001	0.011	_	798	_
SEIR-D	0.874	0.669	0.037	0.66	0.97	_	—	0.0163	0.346	598	48
Table 5. Reconstructed parameters for 23.03.2020–01.05.2020 measurement period											
Model	α_E	α_I	κ	ρ	β	ν	ε_{CH}	μ	$c^{\rm isol}$	E_0	R_0
SEIR-HCD	0.994	0.491	0.28	—	0.165	0.04	0.0065	0.011	_	794	_
SEIR-D	0.849	0.716	0.011	0.58	0.36	—	_	0.0089	0.56	586	146
Table	6. Recor	istructed	d parame	eters for	23.03.2	2020-2	1.04.202	0 measur	ement	period	
Model	α_E	α_I	κ	ρ	β	ν	ε_{CH}	μ	c^{isol}	E_0	R_0
SEIR-HCD	0.998	0.493	0.268	_	0.063	0.00	7 0.171	0.0473	3 –	798	_
SEIR-D	0.078	0.787	0.643	0.703	0.173	-	_	0.0039	0.63	17	332

Table 3. Reconstructed parameters for 23.03.2020-21.05.2020 measurement period

MATHEMATICAL MODELING AND FORECASTING OF COVID-19

Model	Criterion 1		Criter	rion 2	Criterion 3		
	δ_t	δ_f	$Err_{\rm fore}^{\rm MAE}$	$Err_{\rm fore}^{\rm MSE}$	$Err_{\rm desc}^{\rm MAE}$	$Err_{\rm desc}^{\rm MSE}$	
SEIR-HCD	2	254	338	157690	769	931618	
SEIR-D	3	930	198	54977	454	365087	

Table 7. Comparison of the models for data of 23.03.2020-21.05.2020, K = 60, N = 6

Table 8. Comparison of the models for data of 23.03.2020-11.05.2020, K = 50, N = 16

Model	Crite	erion 1	Crite	erion 2	Criterion 3	
	δ_t	δ_f	$Err_{\rm fore}^{\rm MAE}$	$Err_{\rm fore}^{\rm MSE}$	$Err_{\rm desc}^{\rm MAE}$	$Err_{\rm desc}^{\rm MSE}$
SEIR-HCD	2	174	614	556795	780	974866
SEIR-D	12	817	3346	12745883	338	247688

Table 9. Comparison of the models for data of 23.03.2020-01.05.2020, K = 40, N = 26

Model	Crit	terion 1	Crite	rion 2	Criterion 3	
	δ_t	δ_{f}	$Err_{\rm fore}^{\rm MAE}$	$Err_{\rm fore}^{\rm MSE}$	$Err_{\rm desc}^{\rm MAE}$	$Err_{\rm desc}^{\rm MSE}$
SEIR-HCD	2	207	779	842300	714	889008
SEIR-D	3	2898	1546	3531484	233	105670

Table 10. Comparison of the models for data of 23.03.2020-21.04.2020, K = 30, N = 36

Model	Cr	iterion 1	Crit	terion 2	Criter	rion 3	
	δ_t	δ_{f}	$Err_{\rm fore}^{\rm MAE}$	$Err_{\rm fore}^{\rm MSE}$	$Err_{\rm desc}^{\rm MAE}$	$Err_{\rm desc}^{\rm MSE}$	
SEIR-HCD	5	7537	5242	30788912	342	183419	
SEIR-D	31	348120	48996	6350947086	148	69867	

• mean-square error:
$$MSE(y, \hat{y}) = \sum_{k=1}^{M} \frac{(y_k - \hat{y}_k)^2}{M};$$

• mean average error:
$$MAE(y, \hat{y}) = \sum_{k=1}^{M} \frac{|y_k - \hat{y}_k|}{M}$$
.

Let us introduce the following notation: $Err_{\text{fore}}^{\text{MAE}} = \text{MAE}(f, \hat{f})$ and $Err_{\text{fore}}^{\text{MSE}} = \text{MSE}(f, \hat{f})$ are errors in the corresponding test period N whose data were not used in solving the inverse problems (in Fig. 2 the data are denoted by curve 2), $Err_{\text{desc}}^{\text{MAE}} = \text{MAE}(f, \hat{f})$ and $Err_{\text{desc}}^{\text{MSE}} = \text{MSE}(f, \hat{f})$ are errors in the corresponding training period K whose data f_k and g_k were used in solving the inverse problems (in Fig. 2 the data are denoted by curve 1).

To check criterion 1, let us introduce δ_t — deviation of the epidemic peak date in the model from the real peak date on May 7, 2020 (in days), δ_f — deviation of the epidemic peak magnitude in the model from the real peak magnitude, which was 6703 detected cases.

3.4. Numerical Calculations for Novosibirsk Region

As initial data on the COVID-19 spread in Novosibirsk region ($N_0 = 2,798,170$) we used the following statistical information on March 23, 2020 for SEIR-HCD:

$$S_0 = 2798170 - q_8, E_0 = q_8, I_0 = 0, R_0 = 0, H_0 = 0, C_0 = 0, D_0 = 0, C_0 =$$

and for SEIR-D:

$$S_0 = 2798170 - q_8 - q_9, E_0 = q_8, I_0 = 0, R_0 = q_9, D_0 = 0.$$

Fig. 3 presents the results of simulation and forecast of the number of detected cases in Novosibirsk region with measurements from March 23 to May 31, 2020 (Fig. 3a) and up to June 15, 2020 (Fig. 3b) obtained using the above algorithms for these two mathematical models. Note that the simulation of real data on the detected cases obtained with SEIR-D (curve 2) is more accurate than the simulation with SEIR-HCD (curve 1).

It is difficult to analyze the simulation results of mortality in Novosibirsk region for both models, since the available data are insufficient for a good qualitative forecast of mortality in this region. Note that the self-isolation index coefficient has a zero influence on contagiousness c^{isol} in Novosibirsk region for SEIR-D. This, in fact, means that self-isolation does not affect the contagiousness in the region (see Tables 11 and 12)

4. SEIR-HCD SENSITIVITY ANALYSIS

By using a software package called DAISY [33] on the structural identifiability analysis of the model (1), (2), it has been shown that the mathematical model of the spread of coronavirus in the



Fig. 3. The number of detected cases of COVID-19 in Novosibirsk region from 23.03.2020 to 21.06.2020 (90 days) with measurements from 23.03.2020 to (a) 31.05.2020 and (b) 15.06.2020. Solution for SEIR-HCD (curve 1), solution for SEIR-D (curve 2), real data f_k used in solving the inverse problems (curve 3). Tables 11 and 12 present the parameters reconstructed for these two mathematical models.

Table 11. Reconstructed parameters for 23.03.2020-31.05.2020 measurement	period	l, Novosibirsł	c region
--	--------	----------------	----------

Model	α_E	α_I	κ	ρ	β	ν	ε_{CH}	μ	c^{isol}	E_0	R_0
SEIR-HCD	0.001	0.224	0.108	_	0.013	0.006	0.055	0.072	_	1001	_
SEIR-D	0.999	0.999	0.042	0.952	0.999	_	_	0.0188	0	99	24

Table 12. Reconstructed parameters for 23.03.2020-15.06.2020 measurement period, Novosibirsk region

Model	α_E	α_I	κ	ρ	β	ν	ε_{CH}	μ	$c^{\rm isol}$	E_0	R_0
SEIR-HCD	0.219	0.0001	0.118	_	0.046	0.001	0.001	0.1	_	6412	-
SEIR-D	0.271	0.999	$1.9\cdot 10^{-5}$	10^{-9}	0.007	—	—	0.0009	0	93	96

population is identifiable. It is important to estimate the sensitivity of the parameters q to the model functions in order to control the stability of the inverse problem solving and the quality of forecasting [34–36].

First we will estimate the parameters to which the solution of the model is most sensitive. They are determined by the calculation of semi-relative sensitivity.

To describe this method, assume that we wish to determine the relative sensitivity of the observed model quantities $x = (E, I, D)^{\top}$ (and, hence, the model solution $\bar{x} = (S, E, I, R, H, C, D)^{\top}$) to the parameters q_k , k = 1, ..., 8. The semi-relative sensitivity of the model solution to the parameter q_k is defined by the expression $\frac{\partial x_i(t;q)}{\partial q_k}q_k$ and calculated by formal differentiation of the ODE model

$$\begin{cases} \frac{d\bar{x}}{dt} = f(t,\bar{x};q), \\ \bar{x}(0) = \bar{x}_0 \end{cases}$$
(9)

with respect to q_k and change in the order of differentiation with respect to time and the parameters.

Thus, we obtain a (7×8) -dimensional system of differential equations for the sensitivity function $\bar{x}_q(t;q) = \partial \bar{x}(t;q)/\partial q$:

$$\frac{d}{dt}\left(\frac{\partial \bar{x}(t)}{\partial q}\right) = \frac{\partial f}{\partial \bar{x}}\frac{\partial \bar{x}(t)}{\partial q} + \frac{\partial f}{\partial q}$$

with initial conditions

$$\frac{\partial \bar{x}(0)}{\partial q} = \frac{\partial \bar{x}_0}{\partial q} = 0$$

This means that the zero matrix for any initial condition is not a function of all estimated parameters. Here $\partial f/\partial \bar{x}$ is the Jacobian of the ODE system, and $\partial f/\partial q$ is the derivative of the right-hand side with respect to these parameters.

This process provides some information about the sensitivity as a function of time in the interval of interest. It would be desirable to have some general measure of sensitivity of the solution to the parameters. Therefore, for each state/parameter combination we take a norm (in the space L_2) in time, and then rank the thus obtained scalars to determine the most sensitive parameters (see Table 13).



Fig. 4. Sensitivity function $c^{\text{test}}(t) \frac{\partial x_i(t)}{\partial q_k} q_k$ for the time period from 23.03.2020 to 21.04.2020 (30 days along the horizontal line).



(a) Norms of perpendiculars for each parameter at various iterations of orthogonal algorithm.



(b) Elements of the eigenvector corresponding to the minimum eigenvalue of the Hessian.

Fig. 5. Sensitivity analysis for the mathematical model (1) by orthogonal method (a) and eigenvalue method (b) [35] (logarithmic scale).

The smaller $\|\frac{\partial x_i(t)}{\partial q_k}\|_2$, the less the influence of the parameter q_k on the variable x_i . For instance, the quality of determining the parameters ε_{CH} and μ in solving the inverse problem does not depend on the available measurements of the number of infected individuals I(t) + E(t) in contrast, for example, to the coefficients α_E , ρ , and κ , which are more sensitive to these data.

Figure 4 presents the time variation of the sensitivity function $c^{\text{test}}(t) \frac{\partial x_i(t)}{\partial q_k} q_k$ versus the parameters. The more variable is the dynamics of a parameter, the higher is the sensitivity to the measurement data and, therefore, it will be determined with higher stability. In the inverse problem (1)–(3) α_E , ρ , and κ are the most identifiable parameters, and μ , ε_{CH} , and α_I are the parameters that are least sensitive to the data.

Figure 5 shows the results of the sensitivity analysis of the parameters of the mathematical model (1) for various iterations of an orthogonal algorithm and an eigenvalue method, respectively [35]. One can

Variable x_i	Parameter q_k	$\left\ c_i(t)\frac{\partial x_i(t)}{\partial q_k}q_k\right\ _2$	Variable x_i	Parameter q_k	$\left\ c_i(t)\frac{\partial x_i(t)}{\partial q_k}q_k\right\ _2$
D	α_E	225.81	D	eta	13.60
I + E	$lpha_E$	156.69	D	ε_{CH}	12.64
D	μ	89.06	I + E	ν	3.38
D	ρ	63.27	I + E	eta	1.21
I + E	ρ	57.73	D	α_I	0.22
D	ν	53.78	I + E	α_I	0.19
D	κ	47.44	I + E	μ	0.00
I + E	κ	30.42	I + E	ε_{CH}	0.00

Table 13. Semi-relative sensitivity of various model states to parameters arranged in descending order

Table 14. Sequences of parameters obtained by using sensitivity analysis methods for mathematical model (1): from most to least sensitive parameter

Orthogonal method	Eigenvalue method
$\alpha_I, \beta, \varepsilon_{CH}, \nu, \rho, \mu, \kappa, \alpha_E$	$\alpha_I, \beta, \kappa, \varepsilon_{CH}, \rho, \mu, \nu, \alpha_E$

see that the most identifiable are the parameters of infection between the asymptomatic and susceptible population groups α_E , mortality μ , and the recovery rate of identified cases ρ .

Table 14 presents sequences of parameters obtained by the two methods. The identifiability analysis shows that the model parameters that are least sensitive (most identifiable) to the data variations (errors) are α_E , κ , and μ . In other words, these parameters are more stably determined in solving the inverse problem (1)–(3). The parameters α_I , ε_{CH} , and β are most sensitive (less identifiable) to the measurement errors. Therefore, it is necessary to develop a regularization algorithm to control the quality of determining the sensitive parameters.

5. CONCLUSIONS

Two mathematical models, SEIR-HCD and SEIR-D, based on the SEIR structure have been analyzed. For each of these models, inverse problems of specifying their parameters for Moscow and Novosibirsk region have been formulated. Identifiability and sensitivity to statistical data errors for these parameters have been analyzed. Algorithms for solving the inverse problems have been developed, and scenarios of the COVID-19 epidemic spread in Moscow and Novosibirsk region for various measurement and forecast time periods have been constructed.

An epidemic peak in Moscow with an error of 2 days and with 174 detected cases less than the real number was obtained by model SEIR-HCD with data on detected cases and mortality in Moscow from March 23, 2020 to May 1, 2020. A forecast of the epidemic development (the number of new detected cases per day) with the smallest error was obtained by model SEIR-HCD for all data (30, 40, and 50 days of the statistics), except for the longest period (60 days of the statistics), in which SEIR-D had the smallest error. Evidently, the use of a coarser mathematical model (with a smaller number of homogeneous groups) is justified in the case of a large amount of statistical data and a short forecasting period. In fact, SEIR-D gives the best approximation of the historical data on detected cases and mortality in Moscow.

Simulation and forecast of the development of coronavirus infection in Novosibirsk region were made for two sets of measurements: from 23.03.2020 to 31.05.2020 and up to 15.06.2020, due to the scarce statistics up to April 15, 2020. It has been shown that no epidemic peak has yet been reached. For the above measurement intervals, SEIR-D provides the most accurate simulation results for the coronavirus infection spread in Novosibirsk region.

In the case of a large simulation interval it is necessary to take into account the dynamics of measures taken to combat the coronavirus infection (mandatory use of masks in closed rooms, reopening of borders for passenger traffic, reopening of cinemas, etc.), which affects real data, quantitative and qualitative data on the reconstructed parameters, scenarios, and properties of the models. In this case the unknown parameters q of the models should be considered as functions of time (piecewise constant, piecewise linear ones, etc.). The time variation of the reconstructed parameters will help in explaining the disease development in the regions and how the introduction of some measures (or their slackening) affects the number of infected cases. In further work it is planned to consider $\alpha_E(t)$, $\alpha_I(t)$, $c^{isol}(t)$, and others as piecewise constant functions of time.

FUNDING

This work was supported by the Russian Science Foundation under project no. 18-71-10044 (problem statement, identifiability analysis, and numerical solution of the forecasting problem for mathematical model SEIR-HCD, Sections 1, 3, and 4), and the Mathematical Center in Akademgorodok under agreement no. 075-15-2019-1675 with the Ministry of Science and Higher Education of the Russian Federation (formulation and numerical solution of the inverse problem for mathematical model SEIR-D, Sections 2 and 3).

ACKNOWLEDGMENTS

The authors would like to thank members of COVID-19 Simulation Working Group of the Russian Academy of Sciences for valuable recommendations and discussion. We are also grateful to A.S. Kozelkov and O.A. Kuzenkov for helpful discussions and comments.

REFERENCES

- Coronavirus COVID-19 Global Cases by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University, March 21, 2020; https://gisanddata.maps.arcgis.com/apps/opsdashboard/ index.html#/bda7594740fd40299423467b48e9ecf6
- Tamm, M.V., COVID-19 in Moscow: Prognoses and Scenarios, FARMAKOEKONOMIKA. Sovr. Farmacoeconom. Farmacoepidemiology, 2020, vol 13, no. 1, pp. 43–51; DOI: 10.17749/2070-4909.2020. 13.1.43-51.
- 3. Koltsova, E.M., Kurkina, E.S., and Vasetsky, A.M., Mathematical Modeling of the Spread of COVID-19 in Moscow and Russian Regions, 2020; arXiv:2004.10118 [q-bio.PE].
- 4. Zlojutro, A., Rey, D., and Gardner, L., Optimizing Border Control Policies for Global Out-Break Mitigation, *Sci. Rep.*, 2019, vol. 9, p. 2216; https://rdcu.be/bniOs.
- 5. Chen, Y., Cheng, J., Jiang, Y., and Liu, K., A Time Delay Dynamical Model for Outbreak of 2019-nCoV and the Parameter Identification, *J. Inv. Ill-Posed Probl.*, 2020, vol. 28, iss. 2, pp. 243–250.
- Tang, B., Wang, X., Li, Q., Bragazzi, N.L., Tang, S., Xiao, Y., and Wu, J., Estimation of the Transmission Risk of 2019-nCoV and Its Implication for Public Health Interventions, SSRN; https://ssrn.com/abstract=3525558.
- Shi Pengpeng, Cao Shengli, and Feng Peihua, SEIR Transmission Dynamics Model of 2019 nCoV Coronavirus with Considering the Weak Infectious Ability and Changes in Latency Duration, medRxiv, 2020; DOI: 10.1101/2020.02.16.20023655.
- 8. Prem, K., Liu, Yang, Russell, T.W., Kucharski, A.J., Eggo, R.M., Davies, N., Jit, M., and Klepac, P., The Effect of Control Strategies to Reduce Social Mixing on Outcomes of the COVID-19 Epidemic in Wuhan, China: A Modelling Study, *The Lancet Public Health*, 2020, vol. 5, no. 5, pp. E261–E270.
- 9. Fan Ru-guo, Wang Yi-bo, Luo Ming, Zhang Ying-qing, and Zhu Chao-ping, SEIR-Based COVID-19 Transmission Model and Inflection Point Prediction Analysis, *J. Univ. Electronic Sci. Technol. China*, 2020, vol. 49, no. 3, pp. 369–374.
- 10. Kabanikhin, S.I. and Krivorot'ko, O.I., Mathematical Simulation of the Wuhan COVID-2019 Epidemic and Inverse Problems, *Zh. Vych. Mat. Mat. Fiz.*, 2020, vol. 11 (in press).
- Liu, Ying, Gayle, A.A., Wilder-Smith, A., and Rocklöv, J., The Reproductive Number of COVID-19 is Higher Compared to SARS Coronavirus, J. Travel Med., 2020, vol. 27, iss. 2, taaa021; DOI: org/10.1093/ jtm/taaa021.

- 12. Liu Xiuli, Hewings Geoffrey, J.D., Wang Shouyang, Qin Minghui, Xiang Xin, Zheng Shan, and Li Xuefeng, Modelling the Situation of COVID-19 and Effects of Different Containment Strategies in China with Dynamic Differential Equations and Parameters Estimation, medRxiv, 2020; DOI:10.1101/2020. 03.09.20033498.
- 13. Lijun Pei, Prediction of Numbers of the Accumulative Confirmed Patients (NACP) and the Plateau Phase of 2019-nCoV in China, *Cognit. Neurodyn.*, 2020, vol. 14, pp. 411–424.
- 14. Anastassopoulou, C., Russo, L., Tsakris, A., and Siettos, C.I., Data-Based Analysis, Modelling and Forecasting of the Novel Coronavirus (2019-nCoV) Outbreak, *Plos One*, 2020; DOI: 10.1101/2020. 02.11.20022186.
- Yang, Z., Zeng, Z., Wang, K., Wong, S.S., Liang, W., Zanin, M., Liu, P., Cao, X., Gao, Z., Mai, Z., Liang, J., Liu, X., Li, S., Li, Y., Ye, F., Guan, W., Yang, Y., Li, F., Luo, S., Xie, Y., Liu, B., Wang, Z., Zhang, S., Wang, Y., Zhong, N., and He, J., Modified SEIR and AI Prediction of the Epidemics Trend of COVID-19 in China under Public Health Interventions, *J. Thorac Dis.*, 2020, vol. 12, no. 3, pp. 165–174; DOI: 10.21037/jtd.2020.02.64.
- Ivorra, B., Ferrández, M.R., Vela-Pérez, M., and Ramos, A.M., Mathematical Modeling of the Spread of the Coronavirus Disease 2019 (COVID-19) Taking into Account the Undetected Infections. The Case of China [published online ahead of print, 2020 Apr. 30], *Comm. Nonlin. Sci. Numer. Simul.*, 2020, vol. 88:105303; DOI: 10.1016/j.cnsns.2020.105303.
- Godio, A., Pace, F., and Vergnano, A., SEIR Modeling of the Italian Epidemic of SARS-CoV-2 Using Computational Swarm Intelligence, *Int. J. Envir. Res. Public Health*, 2020, vol. 17, p. 3535; DOI: 10.3390/ijerph17103535.
- 18. Carcione, J.M., Santos, J.E., Bagaini, C., and Ba, J., A Simulation of a COVID-19 Epidemic Based on a Deterministic SEIR Model., *Front. Public Health*, 2020, vol. 8:230; DOI: 10.3389/fpubh.2020.00230.
- 19. Ding, Y. and Gao, L., An Evaluation of COVID-19 in Italy: A Data-Driven Modeling Analysis, *Infect. Disease Model.*, 2020; DOI: 10.1016/j.idm.2020.06.007.
- Canabarro, A., Tenorio, E., Martins, R., Martins, L., Brito, S., and Chaves, R., Data-Driven Study of the COVID-19 Pandemic via Age-Structured Modelling and Prediction of the Health System Failure in Brazil Amid Diverse Intervention Strategies, medRxiv, 2020; DOI: 10.1101/2020.04.03.20052498.
- Taarak Rapolu, Brahmani Nutakki, Sobha Rani, T., and Durga Bhavani, S., A Time-Dependent SEIRD Model for Forecasting the COVID-19 Transmission Dynamics, medRxiv, 2020; DOI: 10.1101/2020.05.29. 20113571.
- 22. Klepac, P., Pomeroy, L.W., Bjørnstad, O.N., Kuiken, T., Osterhaus, A.D., and Rijks, J.M., Stage-Structured Transmission of Phocine Distemper Virus in the Dutch 2002 Outbreak, *Proc. Biol. Sci.*, 2009, vol. 276, pp. 2469–2476.
- 23. Klepac, P. and Caswell, H., The Stage-Structured Epidemic: Linking Disease and Demography with a Multi-State Matrix Approach Model, *Theor. Ecol.*, 2011, vol. 4, pp. 301–319.
- 24. Siettos, C.I. and Russo, L., Mathematical Modeling of Infectious Disease Dynamics, *Virulence*, 2013, vol. 4, no. 4, pp. 295–306.
- 25. Jenness, S.M. and Goodreau, S.M., Martina Morris EpiModel: An R Package for Mathematical Modeling of Infectious Disease over Networks, *J. Stat. Soft.*, 2018, vol. 84, no. 8; DOI: 10.18637/jss.v084.i08.
- 26. Noll, N.B., Aksamentov, I., Druelle, V., Badenhorst, A., Ronzani, B., Jefferies, G., Albert, J., and Neher, R., COVID-19 Scenarios: An Interactive Tool to Explore the Spread and Associated Morbidity and Mortality of SARS-CoV-2, medRxiv, 2020; DOI: 10.1101/2020.05.05.20091363.
- 27. https://covid19.biouml.org/COVID-19 Statistics and Forecast at the Regional Level.
- 28. https://covid19-projections.com/COVID-19 Projections Using Machine Learning.
- 29. Coronavirus Disease 2019 (COVID-19), Situation Report, May 31, 2020; https://covid19.who.int/.
- Unlu, E., Leger, H., Motornyi, O., Rukubayihunga, A., Ishacian, T., and Chouiten, M., Epidemic Analysis of COVID-19 Outbreak and Counter-Measures in France, medRxiv, 2020; DOI: 10.1101/2020. 04.27.20079962.
- 31. Sameni, R., Mathematical Modeling of Epidemic Diseases. A Case Study of the COVID-19 Coronavirus, 2020, arXiv:2003.11371.
- 32. Kabanikhin, S.I. and Shishlenin, M.A., Quasi-Solution in Inverse Coefficient Problems, *J. Inv. Ill-Posed Probl.*, 2008, vol. 16, no. 7, pp. 707–715.

- Bellu, G., Saccomani, M.P., Audoly, S., and D'Angi'o, L., DAISY: A New Software Tool to Test Global Identifiability of Biological and Physiological Systems, *Comp. Meth. Progr. Biomed.*, 2007, vol. 88, no. 1, pp. 52–61.
- 34. Raue, A., Becker, V., Klingmüller, U., and Timmer, J., Identifiability and Observability Analysis for Experimental Design in Nonlinear Dynamical Models, *Chaos*, 2010, vol. 20:045105; DOI: 10.1063/1.3528102.
- 35. Krivorotko, O.I., Andornaya, D.V., and Kabanikhin, S.I., Sensitivity Analysis and Practical Identifiability of Some Mathematical Models in Biology, *Sib. Zh. Ind. Mat.*, 2020, vol. 23, no. 1, pp. 107–125.
- Raue, A., Karlsson, J., Saccomani, M.P., Jirstrand, M., and Timmer, J., Comparison of Approaches for Parameter Identifiability Analysis of Biological Systems, *Bioinform.*, 2014, vol. 30, no. 10, pp. 1440–1448; DOI: 10.1093/bioinformatics/btu006.