

Английский текст <https://vavilov.elpub.ru/jour>

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 регуляризирующего 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 contagiousness 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.

Для цитирования: Криворотко О.И., Кабанихин С.И., Сосновская М.И., Андорная Д.В. Анализ чувствительности и идентифицируемости математических моделей распространения эпидемии COVID-19. *Вавиловский журнал генетики и селекции*. 2021;25(1):82-91. DOI 10.18699/VJ21.010

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 ventila-

tor 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

Введение

Многие математические модели в биологии (эпидемиология, иммунология, фармакокинетика, системная биология), медицине (томография), физике и химии (метеорология, химическая кинетика), социологии описываются системами дифференциальных уравнений: обыкновенных (Kermack, McKendrick, 1927), в частных производных (Habtemariam et al., 2008), стохастических (Lee et al., 2020). Коэффициенты этих уравнений характеризуют особенности протекания процессов в конкретных условиях. Для построения адекватных математических моделей необходимо уточнять коэффициенты уравнений по некоторой дополнительной информации о процессе или его известных характеристиках. Так, например, в задачах эпидемиологии скорость передачи инфекции в регионе, вероятности возникновения критических случаев в зависимости от сопутствующих заболеваний, возрастных и демографических характеристик, количество бессимптомных/латентных инфицированных и пр. неизвестны или заданы приближенно из статистической информации. Нередко эти характеристики чувствительны к измерениям, заданным с погрешностью (ошибки округления, погрешность прибора, человеческий фактор), что влечет за собой неустойчивость решения задачи идентификации параметров моделей.

Анализ идентифицируемости систем дифференциальных уравнений, возникающих при моделировании процессов в биологии, медицине, физике, является важным шагом перед разработкой вычислительных алгоритмов их решения (Bellu et al., 2007; Raue et al., 2010, 2014; Miao et al., 2011; Kabanikhin et al., 2016; Voropaeva, Tsgoev, 2019). В статье (Krivorotko et al., 2020a) приведена классификация методов идентифицируемости: структурная, практическая и анализ чувствительности, а также рассмотрены системы обыкновенных дифференциальных уравнений, описывающие процессы в эпидемиологии и иммунологии, на предмет чувствительности параметров к ошибкам в измерениях и практическую идентифицируемость.

Подробный обзор методов и примеров анализа структурной идентифицируемости в задачах биологии, описываемых системами обыкновенных дифференциальных уравнений (ОДУ), можно найти в работах (Miao et al.,

2011; Kabanikhin et al., 2016). Основываясь на структуре модели

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

можно установить единственность решения задачи определения параметров q и начальных условий (или их части) \bar{x}_0 модели по имеющимся измерениям f_{ik} , а также дать рекомендации по добавлению информации или изменению условий задачи поиска параметров для получения единственного решения.

Нами выполнен анализ полуотносительной чувствительности математических моделей эпидемиологии и социальных процессов, предложенный Adams et al. (2004) для анализа систем ОДУ, который позволяет установить степень чувствительности параметров к измерениям и недостающие/излишние измерения относительно некоторого «эталонного» набора для решения поставленной задачи идентификации неизвестных параметров. В качестве примера рассмотрены описываемые системой ОДУ две математические модели распространения новой коронавирусной инфекции, вызванной вирусом SARS-CoV-2. Построен алгоритм регуляризации численного решения задачи идентификации параметров математической модели SEIR-типа и камерной модели на основе состояний агентов в агентно-ориентированной модели по статистическим данным из открытых источников для Новосибирской области. Приведены результаты моделирования, а также сценарий развития распространения заболевания COVID-19 в Новосибирской области.

Анализ чувствительности параметров в системах обыкновенных дифференциальных уравнений

Анализ чувствительности используется для оценки идентифицируемости неизвестных параметров модели для системы ОДУ (1) до построения алгоритмов численного решения задач идентификации параметров. Для методов анализа чувствительности реальные экспериментальные данные не требуются, но количество измерений и момен-

ты времени, в которые выполнены измерения, могут быть необходимы. Исследование чувствительности для математической модели проводится относительно некоторых номинальных параметров q^* , значения которых берутся из литературных источников или доступной статистической информации.

Методы анализа чувствительности основаны на построении матрицы чувствительности. Предположим, что $t_1 \leq t_2 \leq \dots \leq t_K$ – фиксированные моменты времени, в которые проводятся измерения f_{ik} . Тогда коэффициенты матрицы чувствительности для вектора параметров q^* вычисляются по формуле

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

где f_i ($i = 1, \dots, M$) – i -я компонента вектора измеряемых функций, а q_j ($j = 1, \dots, L$) – j -я компонента вектора параметров.

Таким образом, матрица чувствительности определяется следующим образом:

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

Для расчета матрицы чувствительности рассматривается традиционная функция чувствительности:

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

Дифференцируя первое уравнение системы (1) по q_j , получаем, что каждая вектор-функция s_{q_j} удовлетворяет следующей задаче Коши:

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

Численно решая систему дифференциальных уравнений (3), получаем $s_{q_j}(t)$.

В качестве первого шага оцениваются те параметры q , к которым решение модели наиболее чувствительно. Они, в свою очередь, определяются с помощью расчета полуотносительной чувствительности. Представим информацию о чувствительности как функцию времени на интересующем интервале. Мы хотим иметь некоторую общую меру чувствительности решения к параметрам, поэтому для каждой комбинации состояния/параметр берем норму (в пространстве L_2) по времени, а затем ранжируем полученные скаляры, чтобы определить наиболее чувствительные параметры. Чем меньше значение

$\left\| \frac{\partial f_i(t)}{\partial q_k} q_k^* \right\|_2$, тем меньше влияние параметра q_k на переменную f_i . Данную общую меру назовем полуотносительной чувствительностью.

Далее для анализа чувствительности в работе использован ортогональный метод (Yao et al., 2003). Его основная идея состоит в том, чтобы исследовать линейные зависимости столбцов матрицы чувствительности S . Таким образом, одновременно можно оценить как чувствительность параметров к входным данным, так и зависимость между параметрами.

Анализ чувствительности математических моделей распространения COVID-19

Особенность разработанных к настоящему времени математических моделей распространения COVID-19 состоит в анализе поведения бессимптомного протекания болезни и влияния инкубационного периода заболевания на характер эпидемиологической ситуации в регионах. Разработан ряд пакетов с открытым кодом (Gomez et al., 2020; Tuomisto et al., 2020; Wolfram, 2020), а также web-сервисов для моделирования сценариев развития COVID-19:

- в странах мира: <https://covid19-scenarios.org/> (Базельский университет, Швейцария);
- в Москве, Новосибирской области и ряде европейских стран: <https://covid19.biouml.org/> (Федеральный исследовательский центр Институт вычислительных технологий СО РАН, Новосибирск);
- в Алматы, Казахстан: <http://covid19.mmay.info/almaty/?fbclid=IwAR20yx7F4MdWRqwUDzripUK29IWAvoyCSkDPafgpj25ummay23e7oFHBdXg>.

Основные подходы в моделировании распространения эпидемий можно разделить на две группы.

1. *Камерный подход* (моделирование «сверху вниз»). Взаимодействие агентов в популяции, распределенных в характерные группы со схожими признаками (восприимчивые, (бес)симптомные инфицированные, госпитализированные агенты, критические случаи и т.п.), строится на основе закона действующих масс в рамках камерного моделирования, впервые предложенного в 1927 г. (Kermack, McKendrick, 1927). Распределение агентов во времени происходит в зависимости от заданных вероятностей перехода между группами: вероятность инфицирования, параметр контагиозности вируса, смертность и др.
2. *Агентно-ориентированный подход* (моделирование «снизу вверх»), базирующийся на исследовании взаимодействий отдельных индивидуумов и их влияния на глобальные показатели (параметр контагиозности, смертность, вероятность тяжелого протекания болезни и др.). Агентно-ориентированные модели характеризуются случайными графами, в которых длинам ребер соответствуют вероятности перемещения агента в различные состояния.

Часто параметры перехода между группами и состояниями агентов неизвестны или заданы в широком диапазоне (например, инкубационный период заболевания составляет 2–14 дней по оценкам ВОЗ), что затрудняет анализ модели и построение качественных сценариев развития заболевания.

Рассмотрим два варианта разделения популяции конкретного возраста (20–29 лет) на группы, в которых агенты переходят в различные состояния в ходе прогрессирования

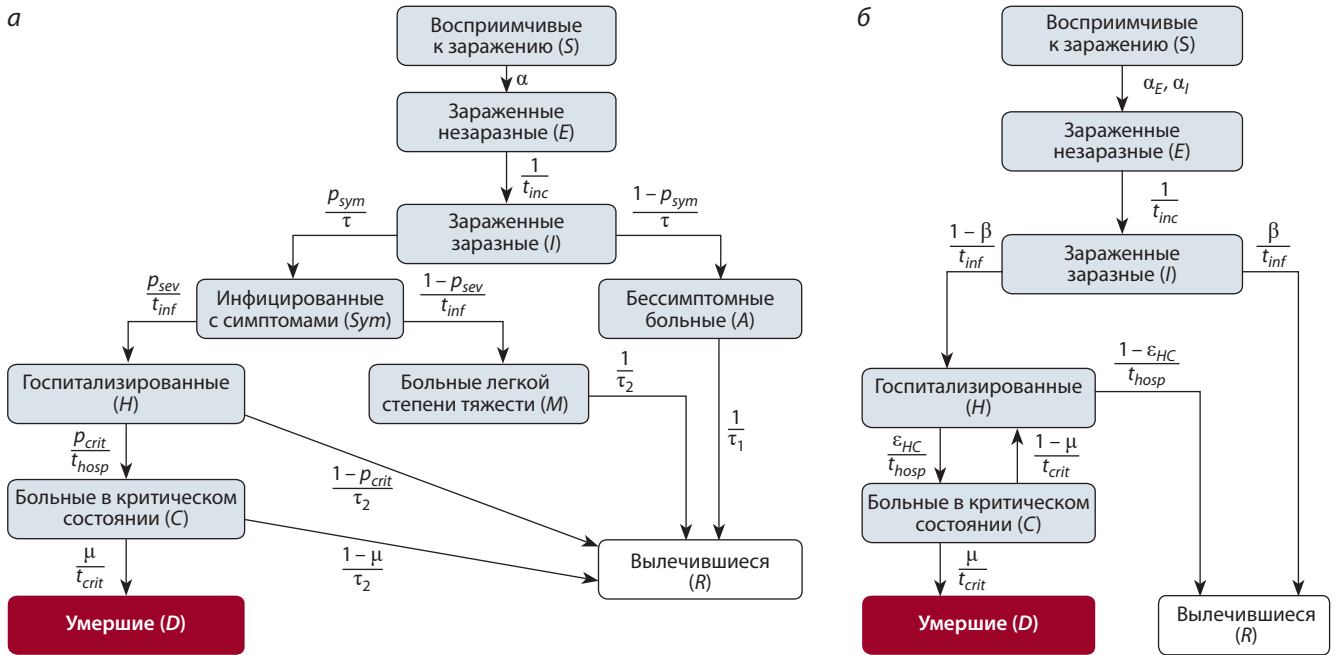


Рис. 1. Диаграмма состояний агентов: а – в пакете COVASIM (Kerr et al., 2020); б – в рамках SEIR-HCD модели (Unlu et al., 2020).

заболевания, вызванного вирусом SARS-CoV-2 (рис. 1). В моделях не учитываются: разделение по половому признаку; ежегодные рождаемость и смертность (поскольку промежуток моделирования составляет менее одного года); вакцинация; пассажиропотоки; сопутствующие заболевания агентов, влияющие на вероятности перехода между группами. Цель анализа двух математических моделей – продемонстрировать корреляцию зависимостей схожих параметров к одним и тем же измерениям, а также дать рекомендации, какие параметры по каким измерениям удастся определить устойчиво.

Система ОДУ типа (1), описывающая распространение COVID-19 в популяции, разделенной на десять характерных групп (Kerr et al., 2020), записывается на основе баланса масс следующим образом:

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

с начальными условиями:

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

Модель (4) характеризует один из классов состояний агентов одной возрастной группы в агентно-ориентированной модели (см. рис. 1, а).

Аналогично записывается система уравнений SEIR-HCD модели, в которой популяция разделена на семь групп (Krivorotko et al., 2020b; Unlu et al., 2020):

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

с начальными условиями:

$$S(0) = S_0, E(0) = E_0, I(0) = I_0, R(0) = R_0, \\
 H(0) = H_0, C(0) = C_0, D(0) = D_0.$$

Здесь $S(t)$ – восприимчивый агент в момент времени t ; $E(t)$ – зараженный незаразный (не передающий вирус); $I(t)$ – зараженный заразный (передающий вирус); $A(t)$ – больной без симптомов; $Sym(t)$ – больной с симптомами;

Таблица 1. Параметры моделей (4), (5) и их усредненные значения для Новосибирской области (Kerr et al., 2020; Unlu et al., 2020)

Параметр	Описание	Значение/ интервал
τ_1	Количество дней до выздоровления бессимптомных и легких больных	6–11
τ_2	Количество дней до выздоровления зараженных пациентов в критическом и тяжелом состоянии	12–17
p_{sym}	Вероятность проявления симптомов после заражения	0.6
p_{sev}	Вероятность перехода больного с симптомами в тяжелое состояние (нуждается в госпитализации)	0.0072
p_{crit}	Вероятность перехода больного из тяжелого состояния в критическое (нуждается в ИВЛ)	0.00036
α	Вероятность передачи вируса для агентов, имевших контакт	(0.01, 0.025)
$E(0)$	Начальное количество инфицированных людей в популяции	(1, 100)
$a(t)$	Индекс самоизоляции от Яндекс, который описывает степень изоляции населения по шкале от 0 (отсутствие изоляции) до 5 (полная изоляция)	(0, 5)
α_E	Вероятность заражения между бессимптомной и восприимчивой группами населения ($\alpha_E \gg \alpha_i$)	(0, 1)
α_i	Вероятность заражения между инфицированным и восприимчивым населением, которая связана с contagiозностью вируса и социальными факторами	(0, 1)
β	Вероятность выздоровления зараженных пациентов, которые перенесли болезнь без осложнений	(0, 1)
ε_{HC}	Доля госпитализированных больных, находящихся в критическом состоянии и требующих подключения аппарата ИВЛ	(0, 1)
μ	Вероятность смерти в результате COVID-19	(0, 0.5)
τ	Продолжительность латентного периода (характеризует запаздывание выделения вирионов), дней	2
t_{inc}	Продолжительность инкубационного периода, дней	2–14
t_{inf}	Продолжительность периода заражения, дней	2.5–14
t_{hosp}	Продолжительность периода госпитализации, дней	4–5
t_{crit}	Продолжительность использования аппарата ИВЛ, дней	10–20

$H(t)$ – тяжелобольной; $C(t)$ – критически больной (требующий подключения аппарата ИВЛ); $M(t)$ – больной легкой степени тяжести; $R(t)$ – выздоровевший; $D(t)$ – умерший. Усредненные параметры моделей (4) и (5) для Новосибирской области приведены в табл. 1 (Lauer et al., 2020; Verity et al., 2020; Wölfel et al., 2020).

Отметим, что коэффициенты t_{inc}^{-1} , t_{inf}^{-1} , t_{hosp}^{-1} , t_{crit}^{-1} , τ^{-1} , τ_1^{-1} , τ_2^{-1} перед соответствующими состояниями агентов в моделях (4) и (5) описывают запаздывание перехода между состояниями (Лихошвай и др., 2004). Например, в уравнении (модель (5))

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

коэффициент t_{inc}^{-1} означает (в линейном приближении) запаздывание в t_{inc} дней перехода из группы зараженных незаразных индивидуумов $E(t)$ в группу инфицированных заразных $I(t)$, а коэффициент $-t_{inf}^{-1}$ – задержку агента в группе инфицированных заразных индивидуумов в течение периода инфицирования t_{inf} , дней.

Математическая модель 1 (схема приведена на рис. 1, а). Предположим, что известна дополнительная информация о количестве вылеченных и умерших больных в фиксированные дни в случае математической модели (4):

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

Здесь R_k – количество выздоровевших агентов в день k ; D_k – количество умерших в результате заболевания в день k .

Проанализируем полутносительную чувствительность двух неизвестных параметров заразности α и началь-

Таблица 2. Полуотносительные чувствительности различных состояний модели (4) к параметрам, отсортированные по убыванию

Переменная f_i	Параметр 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	$6.07 \cdot 10^{-6}$

го количества бессимптомных больных $E(0)$ в модели (4) к измерениям (6). Это позволит установить возможность устойчивого определения неизвестных параметров по имеющимся данным для построения адекватной эпидемиологической картины в регионе. В табл. 2 приведены отсортированные по убыванию значения функции чувствительности исследуемых параметров $q_1 = \alpha$, $q_2 = E(0)$ на измерения $(f_i) = (R, D)$, $i = 1, 2$, представленной в виде нормы $\left\| \frac{\partial f_i(t)}{\partial q_k} q_k^* \right\|_2$. Чем меньше значение $\left\| \frac{\partial f_i(t)}{\partial q_k} q_k^* \right\|_2$, тем меньше влияние параметра q_k на измерения f_i .

На рис. 2 представлены графики изменения функции чувствительности $\frac{\partial f_i(t)}{\partial q_k} q_k^*$ от времени в зависимости от варьируемого параметра. Таким образом, в модели (4)

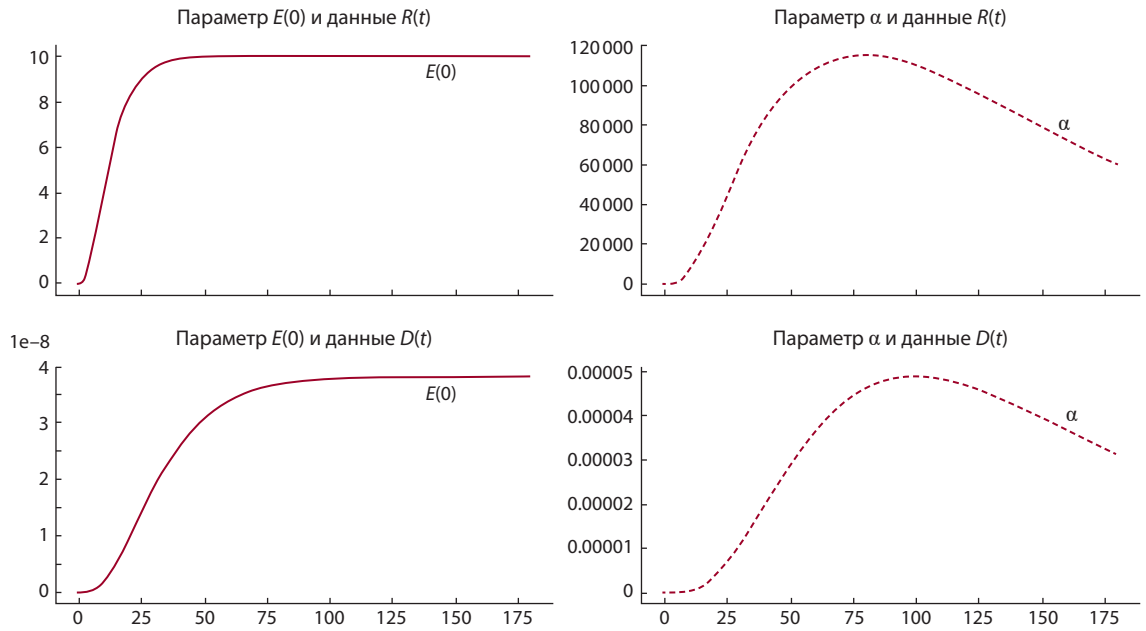


Рис. 2. Функция чувствительности $\frac{\partial f_i(t)}{\partial q_k}$ для модели (4) для периода моделирования 182 дня (с 12.03.2020 по 09.09.2020).

параметры α и $E(0)$ наименее чувствительны к переменной $D(t)$ и не могут быть определены только по данным о смертельных случаях. В свою очередь эти параметры чувствительны к функции $R(t)$, а следовательно, по измерениям о случаях выздоровления восстанавливаются более устойчиво.

Математическая модель 2 (схема приведена на рис. 1, б). Теперь исследуем математическую модель SEIR-HCD (5). Предположим, что известны дополнительные измерения о количестве выявленных, критических и смертельных случаев в фиксированные моменты времени:

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

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

где $b(t) \in [0, 1]$ – доля бессимптомных больных в выявленных случаях; f_k – количество выявленных больных в день k ; C_k – количество пациентов в критическом состоянии в день k .

Пусть параметры $q = (\alpha_E, \alpha_I, \beta, \varepsilon_{HC}, \mu, E_0)^T \in \mathbb{R}^6$ неизвестны. Проанализируем полуотносительную чувствительность вектора параметров q к измерениям (7) для математической модели (5). Для этого построим $\frac{\partial f_i(t)}{\partial q_k} q_k^*$,

$(f_i) = (I, C, D)$, $i = 1, 2, 3$, и проанализируем значение $\left\| \frac{\partial f_i(t)}{\partial q_k} q_k^* \right\|_2$ (табл. 3). Качество определения параметров

β, ε_{HC} и μ при решении обратной задачи практически не зависит от имеющихся измерений количества инфицированных $I(t)$, в отличие, например, от более чувствительных к этим данным коэффициентов α_E, α_I, E_0 .

На рис. 3 представлены графики изменения от времени функции чувствительности $\frac{\partial f_i(t)}{\partial q_k} q_k^*$ в зависимости от варьируемого параметра. Чем более изменчив параметр

в динамике, тем чувствительность к данным измерениям выше, а значит, определяться он будет более устойчиво.

Результаты анализа чувствительности параметров математической модели (5) при различных итерациях ортогонального алгоритма (описание алгоритма см. (Krivorotko

Таблица 3. Полуотносительные чувствительности различных состояний модели (5) к параметрам, отсортированные по убыванию

Переменная f_i	Параметр 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	$3 \cdot 10^{-8}$

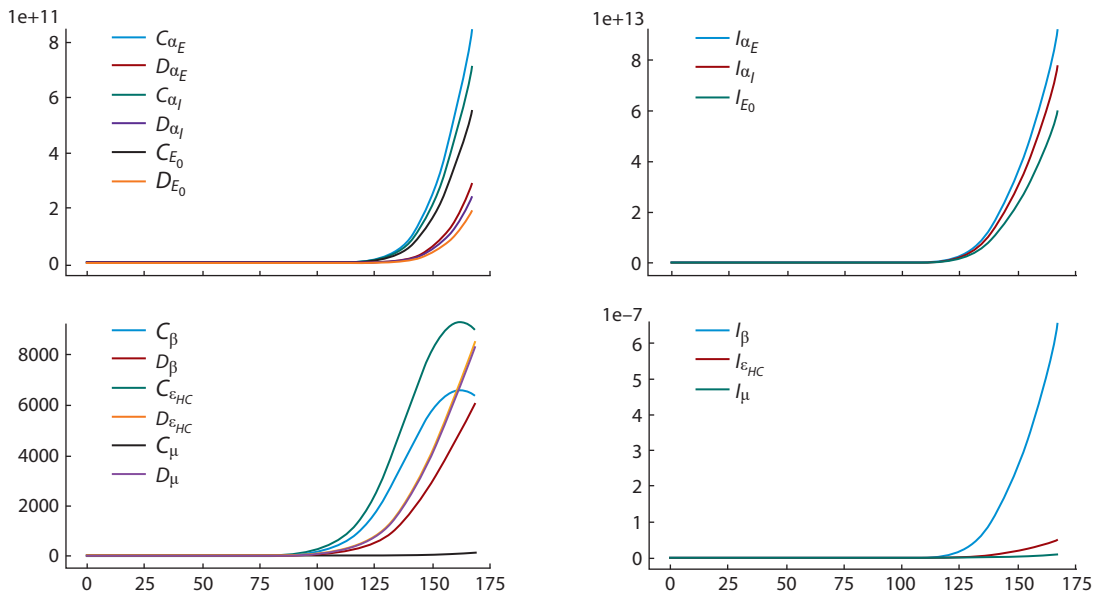


Рис. 3. Функция полуотносительной чувствительности $\frac{\partial f_i(t)}{\partial q_k} q_k^*$ для временного интервала 170 дней (с 15.04.2020 по 01.10.2020).

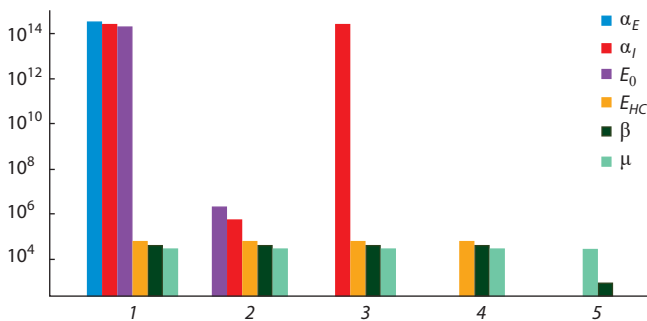


Рис. 4. Величины норм перпендикуляров для каждого параметра на различных итерациях (1–5) ортогонального алгоритма чувствительности параметров математической модели (5).

et al., 2020a) приведены на рис. 4. По горизонтальной оси отложены итерации ортогонального алгоритма, количество которых на единицу меньше размерности вектора неизвестных параметров (количества столбцов матрицы чувствительности), а по вертикальной – значения норм перпендикуляров для полученных преобразований матриц чувствительности. Показано, что наиболее идентифицируемыми также оказались параметры заражения между бессимптомной и восприимчивой группами населения α_E , между инфицированным и восприимчивым населением α_I (связан с контагиозностью вируса и социальными факторами), а также начальное значение зараженных или находящихся в инкубационном периоде индивидуумов E_0 . С помощью метода анализа чувствительности для математической модели (5) получена последовательность параметров (от наиболее до наименее чувствительного): $\alpha_E, E_0, \alpha_I, \epsilon_{HC}, \mu, \beta$.

После анализа идентифицируемости можно заключить, что наименее чувствительными (более идентифицируе-

мыми) параметрами модели к вариациям в данных (погрешностям) являются α_E, E_0 и α_I , иными словами, эти параметры более устойчиво определяются при решении обратной задачи (5), (7). Наиболее чувствительными (менее идентифицируемыми) к ошибкам в измерениях являются параметры ϵ_{HC}, μ и β (с наименьшими значениями норм перпендикуляров в матрице чувствительности), т. е. необходимо разработать алгоритм регуляризации, позволяющий контролировать качество определения чувствительных параметров.

Математическое моделирование распространения COVID-19 в Новосибирской области

Для математического моделирования COVID-19 в Новосибирской области были использованы данные открытых источников:

- количество тестированных (в том числе выявленных f и процента $b(t)$ бессимптомных из них), вылеченных (R) и умерших (D) от COVID-19;
- промежутки длительности инкубационного t_{inc} , латентного τ периодов болезни, периодов инфицирования t_{inf} , госпитализации t_{hosp} , длительности использования аппарата ИВЛ t_{crit} ;
- скорости выздоровления легких τ_1 и тяжелых τ_2 случаев;
- демографические показатели (возрастные распределения в регионе, размер популяции);
- сведения о среднем размере семьи (2.6 человека) в Российской Федерации по данным ООН в 2019 г. (<https://population.un.org/Household/#/countries/840>).
Исследуемая информация периодически обновлялась на сайтах:
 - Минздрава Новосибирской области: <https://zdrav.nso.ru/> (данные пункта (г));

Таблица 4. Меры по сдерживанию коронавирусной инфекции, принятые в Новосибирской области в 2020 г., которые учитываются в моделях (4) и (5)

Дата	Мера
18 марта	Начало удаленных занятий в школах и университетах Новосибирской области
28 марта	Приостановлены все массовые, развлекательные, общественные мероприятия на территории региона
27 апреля	Указ губернатора об обязательном ношении масок в магазинах
6 июля	Открытие летних веранд в кафе и ресторанах
1 сентября	Начало очных занятий в школах и университетах
28 сентября	Введение обязательного ношения масок в любых помещениях, ужесточение мер в образовательных учреждениях

- Федеральной службы государственной статистики, Новосибирская область: <https://novosibstat.gks.ru/folder/31729> (данные пункта (в));
- Стопкоронавирус: <https://стопкоронавирус.рф> (данные пункта (а));
- Всемирной организации здравоохранения: <https://www.who.int> (данные пункта (б)).

Моделирование проводилось с учетом введенных в Новосибирской области сдерживающих мер в отношении коронавирусной инфекции COVID-19 (табл. 4).

Решения обратных задач (4), (6) и (5), (7) сводятся к решению задачи минимизации целевого функционала (Kabanikhin, 2008):

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

Здесь s – статистики, по которым сравнивались данные (кумулятивные выявленные, критические, умершие); w_s – весовой коэффициент; $c_d^{i,s}, c_m^{i,s}$ значения данных (с индексом d) и модели (с индексом m); T – количество дней моделирования; q – вектор неизвестных параметров: $q_1 = (\beta, E_0)^T$ в случае обратной задачи (4), (6) и $q_2 = (\alpha_E(t), \alpha_I(t), \beta, \varepsilon_{HC}, \mu, E_0)^T$ в случае обратной задачи (5), (7). В вычислительных экспериментах использовалась абсолютная норма

$$G_1 = \frac{|c_d^{i,s} - c_m^{i,s}|}{M}, \text{ где } M = \max_i \{c_d^{i,s}\} - \text{нормирующий член,}$$

и квадратичное отклонение в виде $G_2 = (c_d^{i,s} - c_m^{i,s})^2 / T$.

Поиск минимума функционала $J(q)$ реализуется с помощью метода дифференциальной эволюции библиотеки SciPy.Optimize Python. Общая схема алгоритма нахождения глобального минимума записывается следующим образом:

1. Генерация начального поколения $\{\vec{q}_i\} \in B, i = 1 \dots N$.
2. Генерация нового поколения.

- Мутация:

$\forall \vec{q}_i \in B$ выбираются три случайных вектора

$$\vec{v}_1, \vec{v}_2, \vec{v}_3 \in B (\vec{v}_j \neq \vec{q}_i, j = 1, 2, 3).$$

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

- Скрещивание: пробный вектор \vec{u} вычисляется следующим образом:

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

3. Отбор:

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

Результат моделирования распространения коронавирусной инфекции в Новосибирской области с прогнозом до 10 декабря 2020 г. представлен на рис. 5. При конструировании данной модели мы использовали агентно-ориентированный подход, основанный на исследовании взаимодействий отдельных индивидуумов и их влияния на глобальные показатели. Моделирование производили с помощью пакета Covasim – инструмента для создания стохастических агентных моделей. Подробнее со структурой модели можно ознакомиться в работе (Kerr et al., 2020). Использовались также статистические данные по выявленным случаям и смертям в период с 12 марта по 23 октября 2020 г. Минимизировался следующий функционал с учетом проведенного анализа идентифицируемости моделей (4), (6):

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

Здесь f_d^i, f_m^i – кумулятивные выявленные случаи; D_d^i, D_m^i – кумулятивные смерти.

На рис. 5, а, б представлены результаты моделирования f_m^i и статистики f_d^i по выявленным случаям – кумулятивным и ежедневным соответственно, на рис. 5, в – результаты моделирования D_m^i и статистики D_d^i по кумулятивной смертности в результате COVID-19 в Новосибирской области. Отметим, что в статистических данных и в результатах моделирования наблюдается вторая волна эпидемии в середине сентября, которая после введения с 28 октября более жестких мер растет незначительно (не более 215 ежедневных выявленных случаев к середине декабря 2020 г.).

Обратная задача (5), (7) была сведена к минимизации целевого функционала (Krivorotko et al., 2020b):

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

Параметры скорости распространения инфекции $\alpha_E(t), \alpha_I(t)$, связанные с контагиозностью вируса и зависящие от времени, представлялись в виде кусочно-постоянных функций в зависимости от введенных карантинных мер (см. табл. 4).

Учитывая анализ идентифицируемости математической модели (5), (7), были введены более жесткие ограничения на поиск слабоидентифицируемых параметров (см. табл. 1). Результат моделирования решения обратной задачи (5), (7) для SEIR-HCD модели в Новосибирской

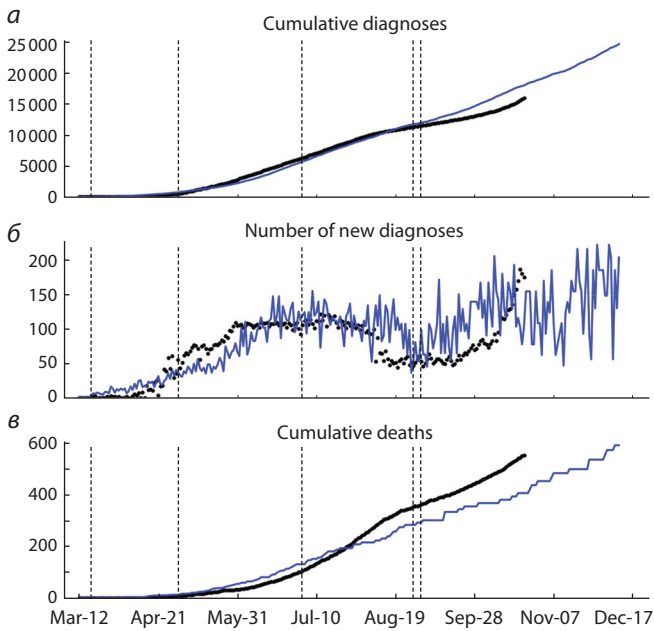


Рис. 5. Моделирование распространения COVID-19 в Новосибирской области (синяя линия) с помощью агентного подхода и статистические данные (черные точки) с указанием введенных мер, влияющих на результаты моделирования и статистику (вертикальные штриховые линии).

области на период с 15 апреля по 3 октября 2020 г. приведен на рис. 6.

Отметим, что более грубая математическая модель (семь уравнений в системе ОДУ) улавливает основную тенденцию по выявленным случаям (пик заболевания в регионе, см. рис. 6, а), однако в некоторых моментах не согласуется с общей статистикой (критические случаи, требующие аппаратов ИВЛ, см. рис. 6, б). Негладкие решения на рис. 6 являются результатом применения индекса самоизоляции от Яндекс, характеризующегося недельной сезонностью, сглаживание которого нарушит суть использования в математическом моделировании. Более подробный анализ моделирования и прогнозирования распространения коронавирусной инфекции в Москве и

Новосибирской области приведен в работе (Krivorotko et al., 2020b). В этом случае необходимо применять агентно-ориентированную модель, подробно описывающую небольшие статистики.

Заключение

В работе проведен анализ чувствительности и идентифицируемости математических моделей распространения эпидемии COVID-19, основанных на системах дифференциальных уравнений. Основу алгоритма составляет анализ матрицы чувствительности методами дифференциальной и линейной алгебры, показывающей степень зависимости неизвестных параметров моделей от заданных измерений.

Анализ чувствительности математической модели распространения коронавирусной инфекции COVID-19 показал, что параметр контагиозности вируса устойчиво определяется по количеству ежедневно выявляемых заболевших, критических больных и вылечившихся. С другой стороны, прогнозируемая доля госпитализированных пациентов, больных, находящихся в критическом состоянии и требующих подключения аппарата ИВЛ, а также коэффициент смертности определяются гораздо менее устойчиво. Для построения более реалистичного прогноза необходимо добавлять дополнительную информацию о процессе (например, о количестве ежедневных случаев госпитализации).

Задачи уточнения идентифицируемых параметров были сведены к задачам минимизации соответствующих целевых функционалов, описывающих близость данных моделирования к статистикам по выявленным, критическим и смертельным случаям в Новосибирской области. Использование абсолютной и квадратичной нормы отклонения данных от результатов моделирования при минимизации целевых функционалов не показало существенных различий при анализе результатов моделирования. Более общая камерная модель, состоящая из семи ОДУ, описывает основную тенденцию распространения коронавирусной инфекции, чувствительна к пикам выявленных случаев, однако некачественно описывает небольшие статистики (количество критических случаев в день t_k , смертельных

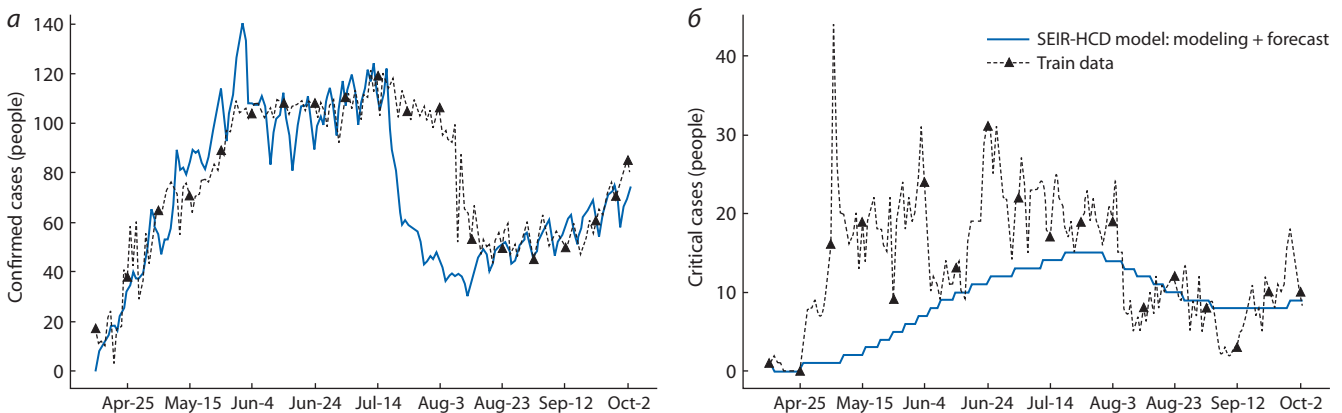


Рис. 6. Моделирование распространения COVID-19 в Новосибирской области (синяя линия) с 15 апреля по 3 октября 2020 г. и статистические данные (штриховая черная линия):

а – ежедневные выявленные случаи f_k ; б – критические случаи C_k , требующие подключения ИВЛ.

случаев), что может приводить к ошибочным выводам. Более подробная агентно-ориентированная математическая модель, в которой один из классов состояний агентов описывается системой из десяти ОДУ, позволяет детальнее улавливать небольшие изменения в статистике данных и строить сценарии развития распространения эпидемии.

Список литературы/ References

- Лихошвай В.А., Фадеев С.И., Демиденко Г.В., Матушкин Ю.Г. Моделирование уравнением с запаздывающим аргументом многостадийного синтеза без ветвления. *Сиб. журн. индустр. математики*. 2004;7(1):73-94.
- [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)]
- Adams B.M., Banks H.T., Davidiana M., Kwona H.D., Trana H.T., Wynnea S.N., Rosenbergs 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., Abeysuriya 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., Prokhoshin 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.
- 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., Motorny 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 Engineering*. 2003;11(3):563-588. DOI 10.1081/PRE-120024426.

ORCID ID

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

Благодарности. Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (проект № 18-31-20019) и Совета по грантам Президента Российской Федерации (соглашение № 075-15-2019-1078, МК-814.2019.1).

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 25.10.2020. После доработки 17.12.2020. Принята к публикации 18.12.2020.