

Идентифицируемость математических моделей эпидемиологии: туберкулез, ВИЧ, COVID-19

Криворотько О.И.*^{1,2}, Кабанихин С.И.¹, Петракова В.С.³

¹Институт математики им. С.Л. Соболева СО РАН, Новосибирск, Россия

²Московский физико-технический институт, Долгопрудный, Россия

³Институт вычислительного моделирования СО РАН, Красноярск, Россия

Аннотация. В работе сделан обзор методов анализа идентифицируемости и чувствительности математических моделей эпидемиологии и связанных с ними процессов, описываемых системами дифференциальных уравнений и агентными моделями. Рассмотрены методы анализа чувствительности параметров к вариации данных, а также анализ чувствительности состояний модели по отношению к вариации параметров, на основе подходов линейной и дифференциальной алгебры, байесовском, Монте-Карло. Показано, что для структурной идентифицируемости базовых SIR-моделей (состоящих из системы трех нелинейных обыкновенных дифференциальных уравнений, описывающих динамику восприимчивой (S), инфицированной (I) и удаленной (R) из популяции групп населения) распространения эпидемии и линейных компартментных моделей возможно использование априорной информации о процессе. Продемонстрировано, что модель может быть структурно идентифицируемой, но при этом быть практически неидентифицируемой по причине неполных данных. Показано, что в SEIR-HCD-модели распространения COVID-19, описываемой системой семи обыкновенных дифференциальных уравнений и основанной на законе действующих масс, параметр приобретения гуморального иммунитета является наименее чувствительным к изменениям количества выявленных, критических и умерших в результате COVID-19 случаев. В пространственной SEIR-HCD-модели распространения COVID-19 продемонстрировано увеличение чувствительности параметра длительности частичного иммунитета со временем, а также уменьшение границ изменения параметров заразности и инфицирования. В случае SEIR-HCD-модели среднего поля распространения COVID-19 показана чувствительность системы к индексу самоизоляции и отсутствие чувствительности стохастических параметров системы. В случае агентной модели распространения COVID-19 границы изменения параметра заразности удалось уменьшить более чем в два раза по сравнению с известными оценками методами анализа чувствительности к измерениям выявленных, критических случаев и количества смертей в результате COVID-19. Разработана дифференциальная модель распространения коинфекции туберкулеза и ВИЧ с учетом множественной лекарственной устойчивости и показана ее локальная идентифицируемость.

Ключевые слова: эпидемиология, SIR, анализ чувствительности, идентифицируемость, априорная информация, байесовские методы, метод Соболя.

*krivorotko.olya@mail.ru

ВВЕДЕНИЕ

Важное значение при моделировании процессов в эпидемиологии имеют математические модели иммунологии, фармакокинетики и взаимодействия вирусов. Большинство параметров упомянутых моделей неизвестны для описания конкретного эпидемиологического процесса в рассматриваемом регионе, часть параметров может быть оценена в рамках накопленной статистики, что приводит к необходимости рассматривать обратные задачи для указанных моделей, суть которых состоит в идентификации неизвестных компонент систем или начально-краевых условий по дополнительной информации о процессе [1].

Но прежде чем решать обратную задачу, необходимо найти условия корректности (существование, единственность и/или устойчивость решения). Цель исследования идентифицируемости параметров моделей – выяснить, определимы ли однозначно по экспериментальным данным неизвестные параметры, входящие в модель, или нет [2, 3, 4].

Анализ идентифицируемости условно можно разделить на две группы:

1. *Априорная идентифицируемость* проводится без привлечения конкретных экспериментальных данных, на основе качественного исследования структуры математической модели с учетом возможной схемы эксперимента. Другими словами, анализ проводится на основе предположения об идеальном эксперименте, а получаемые в результате такого исследования выводы носят достаточно общий характер. В связи с этим априорный анализ идентифицируемости относится скорее к прямым задачам, чем к обратным. Здесь также выделяют структурную идентифицируемость, которая означает, что исследование проводится не в какой-либо одной точке (или нескольких точках) параметрического пространства, а для всех точек одновременно (за исключением, возможно, точек множеств меры нуль).
2. *Апостериорная (практическая) идентифицируемость* является противоположностью структурной идентифицируемости и выполняется при условии, что у исследователя имеются конкретные экспериментальные данные, на основе которых предполагается находить оценки параметров в отдельных точках параметрического пространства. Для нахождения оценок неизвестных параметров используются статистические методы, а также методы решения обратных задач.

История развития идентифицируемости математических моделей началась в 80-е годы XX века, когда в публикациях К. Glover и J. Willems [5], R. Bellman и K. Astrom [6], M. Grewal и K. Glover [7] были введены основные понятия и определения. Важные первые результаты опубликованы в статьях 1980 года С. Cobelli и J. DiStefano [8], в 1987 году Т. Tunalı и Т. Tarn [9] и в 1979 году S. Vajda [10], где получены условия идентифицируемости для нелинейных систем дифференциальных уравнений. S. Audoly в 1983 году [11] предложил новый подход для анализа априорной идентифицируемости изолированных систем.

Среди отечественных ученых можно отметить В.Ф. Щербака [12], А.А. Левакова [13], В.В. Карелина [14], которые в 1982–1983 годах получили условия идентифицируемости линейных систем. Т.В. Авдеенко [15] в 2006 году рассмотрела подход, основанный на идее сепараторов параметрического пространства, к исследованию структурной глобальной идентифицируемости линейных динамических моделей, позволяющий избежать трудоемкого аналитического решения системы нелинейных алгебраических уравнений. В обзорной работе 2007 года [2] предлагается вычислительно эффективный подход к анализу структурной идентифицируемости, обусловленный значительным

сокращением размерности матриц идентифицируемости с учетом их структуры. В 2015 году А.А. Ломов [16] рассмотрел задачу параметрической идентифицируемости при наличии возмущений в измерениях и получил необходимые и достаточные условия локальной совместной идентифицируемости коэффициентов уравнений объекта и возмущений. В работе А.А. Якименко [17] получено новое достаточное условие разрешимости задачи идентифицируемости систем одновременных уравнений. Кроме того, рассмотрен метод получения оценок структурной формы модели системы одновременных уравнений по оценкам приведенной формы, выведенных методом наименьших квадратов.

В работах [8, 18] дан обзор понятий методов анализа идентифицируемости, известных к 1980 году, в том числе по функциям чувствительности, по передаточной функции, по информационной матрице Фишера и соавторов. R. Brown [19] рассмотрел свойства и практические аспекты идентификации в области метаболизма, фармакокинетики, экологии и химической кинетики. Априорная глобальная идентифицируемость некоторых нелинейных моделей была проверена в работах [20, 21]. N. Meshkat [22, 23] разработал новый алгоритм разделения области определения параметров моделей для анализа идентифицируемости. В работах E. Carson [24, 25] приведены примеры и алгоритмы создания достоверных математических моделей в физиологии и медицине, включая нелинейные и стохастические модели. A. Raue с соавторами [26] провели анализ структурной идентифицируемости для задач клеточной биологии.

Важным принципом анализа идентифицируемости является анализ чувствительности, который определяет степень влияния параметров и начальных условий модели на результаты моделирования. Методы анализа чувствительности требуют знания начальных параметров модели, относительно которых исследуется чувствительность, расположение и количество измерений. К методам чувствительности относятся методы собственных значений, ортогональный, корреляционный, байесовский, Соболя и другие [3, 29].

Работа организована следующим образом. В первом разделе приведены методы анализа структурной и практической идентифицируемости, а также исследования чувствительности детерминистских математических моделей распространения коинфекции туберкулеза и ВИЧ, а также COVID-19, основанных на SIR-моделях. В результате анализа идентифицируемости выявлены наименее идентифицируемые параметры и условия, необходимые для их идентификации. В следующем разделе приведен алгоритм байесовского метода для анализа чувствительности агентных, стохастических и многопараметрических детерминистских моделей, основная идея которого состоит в построении суррогатной упрощенной модели во избежание вычислительных затрат. В результате работы данного алгоритма удалось уменьшить границы изменения неизвестных параметров на примере агентной и пространственной SEIR-HCD-модели распространения COVID-19. В заключении приведены основные выводы и рассуждения. Логическая структура статьи схематически представлена на рисунке 1.

МЕТОДЫ ИДЕНТИФИЦИРУЕМОСТИ ДЕТЕРМИНИСТСКИХ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ

В данном разделе будут приведены подходы для анализа идентифицируемости моделей, основанных на системах обыкновенных дифференциальных уравнений (ОДУ)

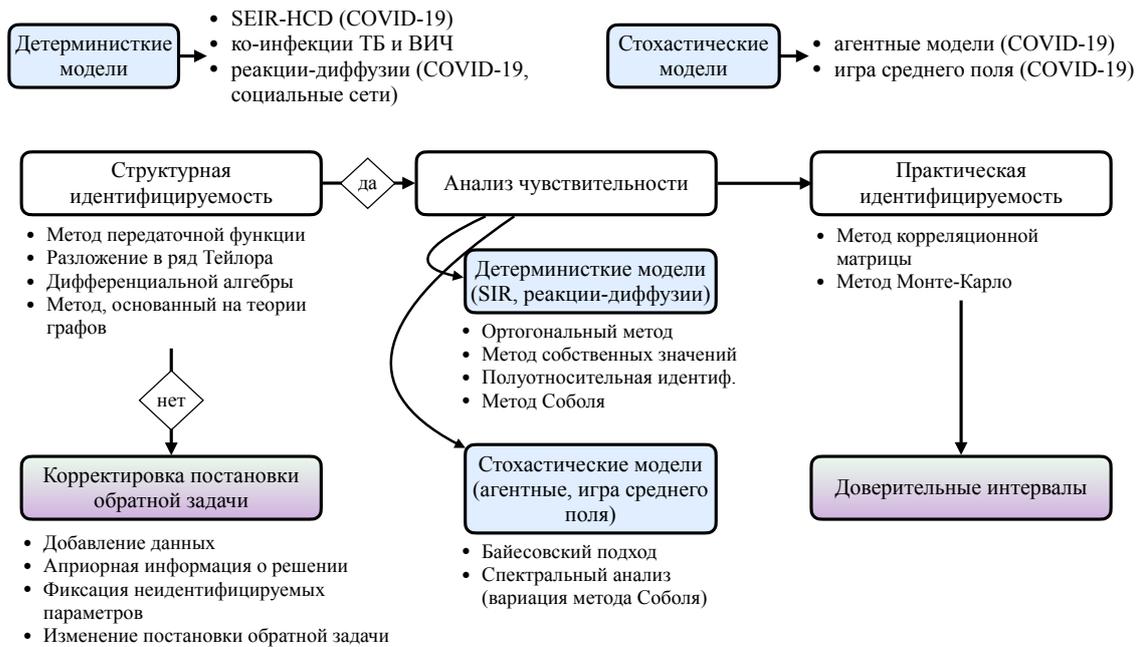


Рис. 1. Логическая структура статьи.

вида

$$\begin{cases} \frac{d\vec{y}}{dt} = \Phi(t, \vec{y}; \vec{q}), & t \in (t_0, T), \\ \vec{y}(0) = \vec{y}_0, \end{cases} \quad (1)$$

а также для моделей типа «реакции-диффузии»

$$\begin{cases} \frac{\partial \vec{y}}{\partial t} = \Phi(t, x, \vec{y}; \vec{q}) + \frac{\partial}{\partial x} \left(N \vec{v} \frac{\partial \vec{y}}{\partial x} \right), & t \in (t_0, T), x \in \Omega, \\ \vec{y}(t_0, x) = \vec{y}_0(x), & x \in \Omega, \\ \left(\gamma_1 \vec{y} + \gamma_2 \frac{\partial \vec{y}}{\partial x} \right) \Big|_{\partial \Omega} = 0, & t \in (t_0, T). \end{cases} \quad (2)$$

Здесь $\vec{y}(t; q, y_0) = (y_1(t), \dots, y_n(t))$ – вектор состояний модели (n – количество уравнений в системе), $\vec{q} \in \mathbb{R}^m$ – вектор коэффициентов модели, $\Phi(t, \vec{y}; \vec{q})$ – вектор-функция правой части, представляющая собой закон действующих масс (в замкнутой системе изменение популяции пропорционально рождаемости и обратно пропорционально общей смертности), $\vec{y}_0 \in \mathbb{R}^n$ – вектор начальных условий. Для модели (2): $\Omega = (l_1, l_2)$ – область в вещественном пространстве \mathbb{R} , $N = N(x)$ – размер популяции в точке x пространства, не зависящий от времени (в силу закона действующих масс $\frac{\partial N(t, x)}{\partial t} = 0$), $\vec{v}(x) \in \mathbb{R}^n$ – вектор скорости каждой группы популяции, который в общем случае зависит от пространственной компоненты, γ_1 и γ_2 – параметры смешанных граничных условий типа Дирихле-Неймана из промежутка $(0, 1)$, $\partial \Omega$ – граница области Ω .

В предположении существования конечных рядов Фурье по пространственной переменной, модели «реакции-диффузии» (2) могут быть сведены к системам ОДУ (1) на коэффициенты Фурье, количество уравнений которых зависит от количества учитываемых коэффициентов Фурье в приближении разложения. В противном случае, анализ чувствительности системы (2) подробнее рассмотрен в разделе .

Предположим, что некоторые компоненты векторов \vec{q} и \vec{y}_0 неизвестны. В обратной

задаче требуется определить q и y_0 в случае, когда решение прямой задачи (1) задано в некоторых точках интервала (t_0, T) :

$$y_i(t_k) = f_{ik}, \quad i \in \mathbb{I} \subset \mathbb{I}_n, \quad k \in \mathbb{I}_K. \quad (3)$$

Здесь и далее $\mathbb{I}_n = \{1, 2, \dots, n\}$.

Для формулировки результатов анализа идентифицируемости систем ОДУ, запишем обратную задачу (1), (3) в более общем виде, опуская знак вектора:

$$\begin{cases} \frac{dy(t)}{dt} = \Phi(y(t), q, u(t)), & t \in (0, T), \\ y(0) = y_0, \\ h(t, y(t), q) = f(t), \end{cases} \quad (4)$$

где $y(t) \in C^1(R^n)$ – вектор функций, $q \in R^m$ – вектор параметров, $f(t) \in C(R^p)$ – функция измерений ($p \leq n$ – количество измеряемых функций), $u(t)$ – входная функция, t – время. В качестве входной функции $u(t)$ в задачах эпидемиологии могут выступать внешние меры контроля и управления эпидемией, в задачах иммунологии – функция «лечения» [31], в задачах фармакокинетики – режим введения лекарственного препарата [24].

Структурная идентифицируемость

На практике у исследователя есть два типа информации: входные данные (функция $u(t)$) и данные на выходе ($f(t)$). Априорный анализ идентифицируемости математической модели отвечает на вопрос: можно ли по информации на входе и на выходе утверждать, что искомый набор параметров определяется единственным образом? Если ответ на данный вопрос положительный, то можно проводить серию измерений для получения экспериментальных данных $f(t)$, не опасаясь иметь, с одной стороны, бесполезные, а с другой – затратные данные.

Определение идентифицируемости [19, 24, 25, 21] формулируется следующим образом:

Определение 1. Модель (4) называется априорно идентифицируемой, если ее параметры $q = [q_1, \dots, q_m]$ можно однозначно определить по входным данным $u(t)$ и данным измерений $f(t) = h(t, y, q)$.

Информация, содержащаяся в измерениях, является функцией от времени и вектора параметров $q = [q_1, \dots, q_n]$, связывающая модель с измерениями.

На практике точные экспериментальные данные $f(t)$ недоступны из-за ошибок измерений (например, настроек прибора по измерению данных). Обычно в наборе дискретных точек по времени $[t_1, \dots, t_N]$ экспериментальные данные имеют погрешность, о которой известна некоторая статическая информация.

Один из часто используемых методов анализа структурной идентифицируемости основан на теории дифференциальной алгебры, в рамках которой система (4) рассматривается как система алгебраических уравнений в переменных состояния $y(t)$, измерений $f(t)$, входных данных $u(t)$, а также соответствующих производных этих функций. Ниже приведена краткая схема анализа идентифицируемости:

1. Введем формальное упорядочение: будем считать, что $u < \dot{u} < \ddot{u} < \dots < f < \dot{f} < \ddot{f} < \dots < y < \dot{y} < \ddot{y} < \dots$. Если дифференциальные полиномы являются рациональными функциями, то необходимо привести их к общему знаменателю.
2. Проводится упорядочение всех уравнений системы (4) в соответствии со старшим

слагаемым в каждом дифференциальном полиноме.

3. Проводится специальная процедура псевдоделения, исключаящая из рассматриваемых уравнений функции $y, \dot{y}, \ddot{y}, \dots$. Подробно данная процедура изложена в работах [24, 21, 22]. В результате такого алгоритма остается столько дифференциальных многочленов относительно $f(t), u(t)$, их производных, а также всевозможных комбинаций, сколько в исходной системе (4) выходных данных $f_i(t)$.
4. Составляем нелинейную систему алгебраических уравнений, приравнивая коэффициенты при каждом слагаемом дифференциальных полиномов к наблюдаемым параметрам. Анализ разрешимости системы проводится с использованием теории базисов Гребнера.
5. Вопрос идентифицируемости системы (4) сводится к вопросу разрешимости нелинейной алгебраической системы.

Данный метод привлекателен тем, что структура алгоритма позволяет автоматизировать его и проводить анализ идентифицируемости нелинейных динамических систем с использованием ЭВМ [21, 23].

В работе [3] приведен обзор методов анализа структурной идентифицируемости, основанных на методах линейной и дифференциальной алгебры и теории графов, большая часть которых внедрена в программные комплексы. Наиболее используемый открытый программный пакет – DAISY [32].

Основываясь на этом анализе можно показать, что рассмотренные в конце раздела математические модели распространения COVID-19 и коинфекции ВИЧ и туберкулез (обычный и с учетом множественной лекарственной устойчивости, МЛУ) являются структурно идентифицируемыми (т.е. решение задачи существует и единственно при идеальных условиях проведения эксперимента). Однако, необходимо установить чувствительность неизвестных параметров \vec{q} к измерениям f с целью контроля устойчивости полученного решения обратной задачи и качества прогнозирования [33, 34, 26].

Пример анализа SIR-модели

Рассмотрим базовую SIR-модель Кермака и МакКендрика [35]

$$\begin{cases} \frac{dS}{dt} = -\alpha \frac{SI}{N}, & t > 0, \\ \frac{dI}{dt} = \alpha \frac{SI}{N} - \beta I, \\ \frac{dR}{dt} = \beta I. \end{cases} \quad (5)$$

Здесь популяция из $N = \text{const}$ особей разделена на три группы (камеры): S – восприимчивые, I – инфицированные ($I \ll N$) и R – вылеченные и умершие, связанные между собой вероятностными переходами $\alpha, \beta \in (0, 1)$. В начале эпидемии, в момент времени $t = 0$, заражено определенное количество людей $S(0) = N - I_0, I(0) = I_0$ и $R(0) = 0$, предполагая $0 < I_0 < N$.

Ввиду закона действующих масс $R(t) = N - S(t) - I(t)$ SIR-модель может быть

переписана в виде:

$$\begin{cases} \frac{dS}{dt} = -\alpha \frac{SI}{N}, & t > 0, \\ \frac{dI}{dt} = \alpha \frac{SI}{N} - \beta I, \\ f(t) = kI(t) \end{cases} \quad (6)$$

с учетом наблюдений $f(t) = kI(t)$, описывающие долю инфицированных индивидуумов (например, выявленные случаи не включая бессимптомных носителей инфекции). Такие измерения используются, например, в работе [36].

Лемма 1 (о локальной идентифицируемости SIR-модели). *Предположим, что необходимо определить параметры S , I , N , α , β и k . Тогда система (6) не идентифицируема.*

Функции $kI(t)$, $\frac{\alpha}{kN}$, $\frac{\alpha S}{N}$, β модели (6) являются идентифицируемы. Или, что эквивалентно, функции $kS(t)$, $kI(t)$, β , $\frac{\alpha}{kN}$ модели (6) – идентифицируемы.

В частности, если N задано и $k = 1$ или $k = \beta$, то система (6) – идентифицируема.

Если $k = \beta$, тогда S , I , β , $\frac{\alpha}{N}$ – идентифицируемы.

Доказательство леммы основано на работах [37, 4]. Покажем, что параметры и состояния системы (6) могут быть выражены через наблюдения и их производные [38]. Рассмотрим вектор-функцию

$$\mathcal{U} = \{(S, I) \mid S > 0, I > 0, S + I < N\}.$$

Открытое множество \mathcal{U} является положительным инвариантом. Так как состояние равновесия $(S_0, 0)$ не наблюдаемо [4], предположим $f \neq 0$ и по тем же причинам $S \neq 0$. Используя метод дифференциальной алгебры [39, 40], выразим все производные измерения $f(t)$ через конечное число рациональных функций. Имеем (при $S \neq 0$ и $I \neq 0$)

$$\begin{aligned} f &= kI, & \frac{\dot{f}}{f} &= \frac{\alpha S}{N} - \beta = y_1, & -\frac{\dot{y}_1}{f} &= \frac{\alpha^2}{kN^2} S = y_2, \\ -\frac{\dot{y}_2}{f} &= \frac{\alpha^3}{k^2 N^3} S = y_3, & \frac{y_3}{y_2} &= \frac{\alpha}{kN}, & \frac{y_2^2}{y_3} &= \frac{\alpha S}{N}, \\ y_1 - \frac{y_2^2}{y_3} &= \beta. \end{aligned}$$

Все производные функции f могут быть выражены через выражения kI , $\frac{\alpha}{kN}$, $\frac{\alpha S}{N}$, β . Поэтому именно эти соотношения на параметры и функции являются идентифицируемыми.

Если N – задано и $k = 1$ или $k = \beta$, тогда параметры α , β , S , I , R являются полиномиальными функциями от (f, \dot{f}, \ddot{f}) , таким образом, система (6) является идентифицируемой.

Аналогичные рассуждения можно привести для параметров $kS(t)$, $kI(t)$, β , $\frac{\alpha}{kN}$. \square

Пример анализа структурной идентифицируемости модели фармакокинетики

Рассмотрим модель контроля глюкозы-инсулина [25]. Следует отметить, что в рамках камерного подхода к моделированию фармакокинетических процессов в качестве камер

могут выступать не только органы, но и различные вещества внутри камеры (например, глюкозы и инсулина в крови):

$$\begin{cases} \frac{dy_1(t)}{dt} = -q_1y_1(t) - q_2y_2(t) + u(t), & t > 0 \\ \frac{dy_2(t)}{dt} = -q_3y_2(t) - q_4y_1(t), \\ f_1(t) = \frac{y_1(t)}{V_1}, \\ y_1(0) = y_2(0) = 0, \end{cases} \quad (7)$$

где $y_1(t)$, $y_2(t)$ – массы глюкозы (назовем глюкозу – 1-я камера) и инсулина (2-я камера), $u(t) = D \cdot \delta(t)$ – внутривенная инфузия, соответствующая дозе D , $q = [q_1, q_2, q_3, q_4, V_1]$ – вектор неизвестных параметров. Пусть V_2 – объем второй камеры. Предположим, что в качестве данных измерений нам известна только концентрация глюкозы $f_1(t)$. Следует отметить, что на практике начальные данные системы (7) могут быть не равны нулю. В этом примере такие значения выбраны для удобства и никак не влияют на схему анализа.

Для определения идентифицируемости данной модели воспользуемся методом передаточной функции [27]. Запишем передаточную функцию в виде: $H_1(s) = \frac{F_1(s)}{U(s)}$. Применяя преобразования Лапласа для уравнений системы (7) получим:

$$\begin{cases} sY_1(s) - y_1(0) = -q_1Y_1(s) - q_2Y_2(s) + D, \\ sY_2(s) - y_2(0) = -q_3Y_2(s) - q_4Y_1(s), \\ F_1(s) = \frac{Y_1(s)}{V_1}. \end{cases}$$

Прописными буквами обозначены преобразования Лапласа для соответствующих функций. Находим выражение для $F_1(s)$ путем выражения $Y_1(s)$ из первых двух уравнений в последней системе. Получим следующее выражение:

$$F_1(s) = \frac{Y_1(s)}{V_1} = \frac{s^2 + (q_1 + q_3)s + q_1q_3 + q_2q_4}{DV_1(s + q_3)}.$$

Теперь представляем значения $F_1(s)$ и $U(s)$ в формулу передаточной функции:

$$H_1(s) = \frac{F_1(s)}{U(s)} = \frac{\frac{s + q_3}{V_1}}{s^2 + (q_1 + q_3)s + q_1q_3 + q_2q_4} = \frac{\varphi_2s + \varphi_1}{s^2 + \varphi_4s + \varphi_3}.$$

Здесь $H_1(s)$ – рациональный многочлен с коэффициентами, зависящими от неизвестных параметров. Записываем систему алгебраических уравнений (8), приравнявая коэффициенты при мономах к наблюдаемым параметрам φ_i :

$$\frac{q_3}{V} = \varphi_1, \quad \frac{1}{V_1} = \varphi_2, \quad q_1q_3 + q_2q_4 = \varphi_3, \quad q_1 + q_3 = \varphi_4. \quad (8)$$

Система (8) не имеет единственного решения, следовательно, модель априорно неидентифицируема. Если рассмотреть вместо параметров q_2 и q_4 – их произведение, считая их одним параметром, то система разрешится, и набор параметров, определяемых

единственным образом, имеет вид: $q = [q_1, q_3, V_1, q_2q_4]$. Предположим, что в качестве данных измерений известна также концентрации инсулина, т.е. $f_2(t) = \frac{y_2(t)}{V_2}$. Для таких данных измерений также выпишем передаточную функцию. Применяя преобразование Лапласа для соответствующих уравнений системы, получим выражение для $F_2(s)$, передаточная функция $H_2(s)$ имеет следующий вид:

$$H_2(s) = \frac{F_2(s)}{U(s)} = \frac{\frac{q_4}{V_2}}{s^2 + (q_1 + q_3)s + q_1q_3 + q_2q_4}.$$

К системе алгебраических уравнений (8) добавляется еще одно уравнение, связывающее наблюдаемые данные с данными измерений: $\frac{q_4}{V_2} = \varphi_5$. Тогда набор идентифицируемых параметров имеет вид: $q = [q_1, q_3, V_1, q_2V_2, \frac{q_4}{V_2}]$.

Следует отметить, что если известно значение V_2 (или, например, что $V_1 = V_2$), модель станет идентифицируемой. Таким образом, можно заключить, что добавление информации об исследуемом процессе делает вычисление параметров более надежным, что согласуется с интуитивным представлением.

Связь идентифицируемости и понятия r -решения

Понятие идентифицируемости тесно связано с существованием r -решения компактного оператора или системы линейных алгебраических уравнений [41, 42]. Действительно, применив к исследуемым обратным задачам (1), (3) процедуру линеаризации относительно параметров q (для модели внутриклеточной динамики ВИЧ инфекции она подробно описана в работе [43]), получим систему линейных алгебраических уравнений вида

$$Aq = f.$$

Здесь $f = \{f_{ik}\}_{\substack{i \in \mathbb{I} \\ k \in \mathbb{I}_k}}$ – вектор измерений, A – матрица линеаризованной обратной задачи.

Применяя метод сингулярного разложения к произвольным прямоугольным матрицам $A = U\Sigma V^*$, где U и V – унитарные матрицы, получаем

$$U\Sigma V^*q = f \rightarrow \Sigma V^*q = U^*f,$$

где $\Sigma = \text{diag} \{\sigma_1, \dots, \sigma_j\}$, σ_j – сингулярные числа матрицы A . Обозначим $g = U^*f$, $z = V^*q$, тогда решение системы алгебраических уравнений с точностью до унитарного преобразования можно выписать в следующем виде:

$$z_1 = \frac{g_1}{\sigma_1}, z_2 = \frac{g_2}{\sigma_2}, \dots, z_r = \frac{g_r}{\sigma_r}, z_i = 0, \text{ если } \sigma_i < \delta, i = r + 1, \dots, p.$$

Здесь r – натуральное число обрыва последовательности малых сингулярных чисел, т.е. тех σ_i , которые меньше некоторого заданного δ (в [41, 42] выбор δ связан с уровнем ошибки в измерениях f).

Полученная последовательность $q = Vz$ решения задачи $Aq = f$, в которой r ненулевых компонент, аналогично последовательности идентифицируемых параметров для линеаризованных моделей.

Практическая идентифицируемость

Анализ практической идентифицируемости позволяет оценить параметры модели с приемлемой точностью по зашумленным экспериментальным данным.

Результат применения методов практической идентифицируемости для анализа математических моделей зависит от качества и количества экспериментальных данных. Модель может быть структурно идентифицируемой, но при этом быть практически неидентифицируемой по причине плохого качества данных, например, плохого соотношения сигнал/шум, ошибок в измерениях или разреженной выборки [30]. Еще одна причина анализа практической идентифицируемости является недостаток данных. Эта ситуация соответствует наблюдению вблизи состояния равновесия системы (4).

Пусть измерения заданы с ошибкой:

$$f_k = h(t, y(t_k), q) + \varepsilon_k, \quad k = 1, \dots, K, \quad (9)$$

где ε_k – независимые и одинаково распределенные случайные величины с нулевым средним $\mathbb{E}[\varepsilon_k] = 0$ и постоянной дисперсией σ^2 .

Далее мы рассмотрим такие методы практической идентифицируемости, как методы Монте-Карло и корреляционной матрицы.

Метод корреляционной матрицы

Практическая идентифицируемость часто ассоциируется с понятием доверительных интервалов [45]. Доверительные интервалы могут быть получены из информационной матрицы Фишера FIM (Fisher Information Matrix). Диагональные элементы корреляционной матрицы, обратной к матрице Фишера, соответствуют дисперсии оценок параметров. Их квадратные корни (стандартные отклонения) дают доверительные интервалы для параметров, тем самым предоставляя информацию о практической идентифицируемости.

В работе [46] предложен подход, который изучает корреляции между параметрами модели. Если корреляция близка к единице, то эти два параметра зависят друг от друга и неразличимы для модели, т.е. не могут быть оценены по отдельности. Ниже приведен алгоритм метода корреляционной матрицы:

1. Задаём точные значения параметров q^0 , которые могут быть известны из статистической информации или из литературных источников.
2. Вычисляем матрицу $\frac{\partial f(t_k, q^0)}{\partial q}$, которую назовем матрицей чувствительности относительно заданного набора параметров q^0 (см. Раздел).
3. Строим матрицу Фишера по формуле:

$$FIM = \sum_{k=1}^K \left(\frac{\partial f(t_k, q^0)}{\partial q} \right)^T V^{-1} \left(\frac{\partial f(t_k, q^0)}{\partial q} \right) = S^T V^{-1} S,$$

где K – количество измерений, V – известная положительно определённая матрица весов.

4. Вычисляем матрицу $C = FIM^{-1}$.
5. Вычисляем коэффициенты корреляции R_{ij} по формуле

$$R_{ij} = \begin{cases} \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}}, & i \neq j, \\ 1, & i = j. \end{cases}$$

Здесь C_{ij} – коэффициенты матрицы C .

6. Строим корреляционную матрицу

$$R = \begin{pmatrix} R_{11}(q_1, q_1) & \dots & R_{1m}(q_1, q_m) \\ \vdots & \ddots & \vdots \\ R_{m1}(q_m, q_1) & \dots & R_{mm}(q_m, q_m) \end{pmatrix}.$$

Метод корреляционной матрицы позволяет проверить различимость какой-либо пары параметров. Следует заметить также, что метод выявляет неразличимую пару и не указывает на тот параметр, который более неидентифицируемый.

Метод Монте-Карло

Метод Монте-Карло [47] является методом выборки с использованием случайных чисел и распределений вероятностей. Метод позволяет моделировать различные сценарии с различным количеством наблюдений при разных уровнях шума или погрешностей измерений для разных экспериментальных конструкций, когда такие проекты могут оказаться невозможными для практических экспериментов. Алгоритм метода Монте-Карло приведен ниже:

1. Задаём точные значения параметров q^0 , которые могут быть известны из статистической информации или из литературных источников.
2. Решаем прямую задачу (1) и получаем измерения f_{ik} вида (9) в фиксированные моменты времени t_k .
3. Генерируем N наборов (например, $N = 1000$) данных измерений f с заданным уровнем шума ϵ .
4. Решаем обратную задачу для каждого из N смоделированных наборов данных для получения вектора параметров q^i , $i = 1, \dots, N$.
5. Вычисляем среднюю относительную ошибку (Average relative error) ARE для каждого элемента вектора q по формуле

$$ARE_l = \frac{1}{N} \sum_{i=1}^N \frac{|q_l^0 - q_l^i|}{|q_l^0|} \cdot 100\%, \quad l = 1, \dots, m, \quad (10)$$

где q_l^0 – l -параметр вектора q^0 , а q_l^i – l -й параметр вектора q^i .

Для очень небольшой погрешности измерения параметры должны быть близки к истинным значениям и ARE должно быть близко к нулю. Когда ошибка измерения увеличивается, значение ARE также будет возрастать. Если значение ARE для некоторого параметра недопустимо высокое, можно утверждать, что этот параметр не является идентифицируемым. Кроме того, нет чёткого правила, насколько высоко должно быть значение ARE , прежде чем заявить о неидентифицируемости параметра для конкретной задачи.

Практическая идентифицируемость и вычисление доверительных интервалов

Понятия практическая идентифицируемость, вычисление доверительных интервалов и метод наименьших квадратов тесно связаны.

Рассмотрим наблюдения для системы ОДУ (n уравнений) в виде (9). Тогда матрица Фишера запишется в виде

$$FIM(q, \sigma) = \frac{1}{\sigma^2} \sum_{k=1}^K \left(\frac{\partial f(t_k, q^0)}{\partial q} \right)^T \left(\frac{\partial f(t_k, q^0)}{\partial q} \right).$$

Применяя метод наименьших квадратов (Ordinary Least Squares, OLS), получим оценку вектора параметров (m параметров)

$$\hat{q}_{OLS} = \arg \min_q \sum_{k=1}^K (f_k - h(t_k, y_k, q))^2.$$

Даже если распределение ошибки не задано, асимптотическая теория может быть использована для аппроксимации среднего и дисперсии оцениваемого вектора параметров q (случайная величина). Аппроксимация для смещений (с $n + m$ параметров) имеет вид:

$$\hat{\sigma}_{OLS}^2 = \frac{1}{K - (n + m)} \sum_{k=1}^K [f_k - h(t_k, y_k, \hat{q}_{OLS})]^2.$$

Аппроксимации вектора параметров и дисперсии, полученные выше, используются для аппроксимации матрицы ковариации C :

$$\hat{C} := [FIM(\hat{q}_{OLS}, \hat{\sigma}_{OLS}^2)]^{-1}.$$

Стандартная ошибка (Standard error, SE) для вектора параметров \hat{q}_{OLS} вычисляется как квадратные корни из диагональных элементов матрицы ковариаций:

$$SE(\hat{q}_{OLS}(i)) = \sqrt{\hat{C}_{ii}}, \quad i = 1, \dots, n + m.$$

Наконец, для вычисления 95% доверительного интервала для i -й компоненты вектора параметров \hat{q}_{OLS} с $n + m$ можно использовать распределение Стьюдента с $K - (n + m)$ степенями свободы: пусть

$$\zeta(i) = t_{0.025}^{K-(n+m)} \times SE(\hat{q}_{OLS}(i)),$$

тогда доверительный интервал для параметров $\hat{q}_{OLS}(i)$ определяется следующим образом:

$$\hat{q}_{OLS}(i) - \zeta(i) < \hat{q}_{OLS}(i) < \hat{q}_{OLS}(i) + \zeta(i).$$

Из этих формул следует, что обусловленность информационной матрицы Фишера играет существенную роль. Большие доверительные интервалы дают представление о практической идентифицируемости.

Анализ чувствительности

Методы анализа чувствительности не требуют фактических экспериментальных данных, хотя может потребоваться знание количества измерений и моментов времени, когда эти измерения были проведены [29, 48]. Для исследования математической модели методами анализа чувствительности необходимо знать значения параметров, относительно которых исследуется чувствительность. Эти параметры могут быть известны из литературных источников или из статистической информации. Чем

чувствительнее параметр к вариации данных, тем он более идентифицируем. И наоборот, чем меньше чувствительность параметра к изменениям данным, тем он менее идентифицируем в модели, что требует применение априорной информации относительно параметра.

В основе методов анализа чувствительности исследуются матрицы чувствительности. Предположим, что $t_1 \leq t_2 \leq \dots \leq t_K$ – моменты времени, в которые будет измеряться вектор-функция $f(t)$. Тогда коэффициенты матрицы чувствительности для заданного вектора параметров q^* вычисляются по формуле $s_{ij} = \frac{\partial f_i(t, q^*)}{\partial q_j}$, где f_i – i -я компонента вектора измерений f , $i = 1, \dots, p$ и q_j – j -я компонента вектора параметров, $j = 1, \dots, m$.

Таким образом, матрица чувствительности определяется следующим образом:

$$S_{P \cdot K \times L} = \begin{pmatrix} s_{11}(t_1) & \dots & s_{1m}(t_1) \\ \vdots & \ddots & \vdots \\ s_{p1}(t_1) & \dots & s_{pm}(t_1) \\ \dots & \dots & \dots \\ s_{11}(t_K) & \dots & s_{1m}(t_K) \\ \vdots & \ddots & \vdots \\ s_{p1}(t_K) & \dots & s_{pm}(t_K) \end{pmatrix}. \quad (11)$$

Для вычисления матрицы чувствительности рассмотрим традиционную функцию чувствительности:

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

Путём дифференцирования по q_j первого уравнения системы (4) получаем, что каждая вектор-функция s_{q_j} , $j = 1, \dots, m$, удовлетворяет линеаризованной задаче Коши

$$\begin{cases} \dot{s}_{q_j}(t) = \frac{\partial \Phi}{\partial y}(t, y(t; q), q) s_{q_j}(t) + \frac{\partial \Phi}{\partial q_j}(t, y(t; q), q), \\ s_{q_j}(t_0) = \vec{0}. \end{cases} \quad (12)$$

Решая систему дифференциальных уравнений (12), получаем $s_{q_j}(t)$. Затем, дифференцируя третье уравнение системы (4) по q_j , получаем, что коэффициенты матрицы чувствительности вычисляются по формуле

$$s_{ij} = \frac{\partial h_i}{\partial y} s_{q_j} + \frac{\partial h_i}{\partial q_j}, \quad i = 1, \dots, p, \quad j = 1, \dots, m.$$

Далее мы рассмотрим ортогональный метод и метод собственных значений анализа свойств матрицы чувствительности S для систем ОДУ, а также свойства матриц на основе анализа полуотносительной чувствительности.

Ортогональный метод

Основная идея ортогонального метода, предложенного в работе [48], состоит в том, чтобы исследовать линейные зависимости столбцов матрицы чувствительности S , определённой формулой (11). Таким образом, одновременно можно оценить как чувствительность параметров к входным данным, так и зависимость между параметрами. Основным преимуществом использования данного метода является возможность

упорядочения параметров от наиболее до наименее чувствительного. В литературе существуют несколько вариантов определения линейно зависимых столбцов матрицы, два из которых мы посчитали основными и вынесли в два алгоритма, остальные варианты являются вариацией имеющихся.

Алгоритм 1 ортогонального метода [49]:

1. Задаём критерий остановки δ_1 , массив номеров идентифицируемых параметров $I = \emptyset$ и массив номеров неидентифицируемых параметров $U = 1, \dots, m$. Строим матрицу чувствительности S вида (11).
2. Выбираем столбец l с наибольшей суммой квадратов элементов, добавляем его в матрицу E как первый столбец и удаляем его из S . Элемент l добавляем в массив I и удаляем из U .
3. Если U пусто, то останавливаем алгоритм: для данной модели все параметры идентифицируемы. Иначе переходим на шаг 4.
4. Для каждого столбца $S_h, h = 1, \dots, n$, матрицы S , где n — количество оставшихся столбцов из S , вычисляем перпендикуляры:

$$S_h^\perp = S_h - S_h^{proj}, \quad S_h^{proj} = \sum_{k=1}^{m-n} \frac{(S_h, E_k)}{E_k, E_k} E_k,$$

$$E = (E_1, \dots, E_{m-n}), \quad h = 1, \dots, n.$$

5. Из полученной матрицы перпендикуляров S^\perp выбираем столбец l с наибольшей суммой квадратов элементов. Если $\|S_h^\perp\| \leq c$, то останавливаем алгоритм. Все параметры из I идентифицируемы, иначе переходим на шаг 6.
6. Добавляем элемент l в I , удаляем его из U , соответствующий столбец удаляем из S и добавляем в E , переходим на шаг 3.

В работе [48] процесс выбора идентифицируемых параметров останавливают, как только максимальная величина будет ниже значения отсечки $\delta_1 \approx 0$, выбор которого довольно произволен. Например, в [50] уставлено $\delta_1 = 10^{-4}$. В данной работе мы не применяем критерий остановки, а оцениваем все параметры от наиболее идентифицируемых до наименее идентифицируемых параметров и анализируем значение δ_1 в целом.

Алгоритм 2 ортогонального метода [48]:

1. Для каждого столбца матрицы чувствительности S вычисляем сумму квадратов элементов.
2. В качестве первого оцениваемого параметра выбираем параметр, соответствующий столбцу матрицы S с наибольшим значением суммы квадратов элементов.
3. Столбец с наибольшей суммой квадратов элементов обозначаем как X_m ($m = 1$ для первой итерации).
4. Вычисляем \hat{S}_m по формуле $\hat{S}_m = X_m(X_m^T X_m)^{-1} X_m^T S$.
5. Вычисляем остаточную матрицу $R_m: R_m = S - \hat{S}_m$.
6. Для каждого столбца остаточной матрицы R_m вычисляем сумму квадратов элементов. Столбец с наибольшей суммой соответствует следующему оцениваемому параметру.

7. Выбираем соответствующий столбец в S и расширяем матрицу X_m путём добавления нового столбца. Обозначаем расширенную матрицу как X_{m+1} .
8. Устанавливаем $m = m + 1$ и повторяем шаги 4–7 до тех пор, пока столбец наибольшим значением суммы квадратов элементов в остаточной матрице не станет меньше параметра остановки δ_1 .

Метод собственных значений

Метод собственных значений, предложенный в [10], основан на свойствах собственных значений и собственных векторов матрицы Гесса:

$$H_{ml} = \sum_{i,k} \frac{\partial f_i(t_k, q^*)}{\partial q_l} \cdot \frac{\partial f_i(t_k, q^*)}{\partial q_m} = (S^T S)_{ml}, \quad (13)$$

где S – матрица чувствительности, определённая формулой (11), q^* – заданный вектор параметров модели, относительно значений которого исследуется чувствительность.

Алгоритм метода собственных значений:

1. Задаём критерий остановки δ_2 , массив идентифицируемых параметров $I = 1, \dots, m$ и массив неидентифицируемых параметров $U = \emptyset$. Строим матрицу чувствительности S вида (11).
2. Если I пусто, то останавливаем алгоритм. Для данной модели все параметры неидентифицируемы. Иначе переходим на шаг 3.
3. Вычисляем матрицу H по формуле (13). Для неё находим собственные значения λ^l ($\lambda^1 \leq \lambda^2 \leq \dots \leq \lambda^m$) и собственные векторы u^l .
4. Если $\lambda^1 \geq \delta^2$, то останавливаем алгоритм. Параметры, находящиеся в I , идентифицируемы. Иначе переходим на шаг 5.
5. Выбираем l : $|u_l^1| = \max(|u_2^1|, |u_3^1|, \dots, |u_m^1|)$ (максимальный элемент в собственном векторе, который соответствует минимальному собственному значению). Добавляем l в множество неидентифицируемых параметров U , удаляем из I , удаляем соответствующий столбец из матрицы S и переходим на шаг 2.

Как и для предыдущего метода, мы оцениваем все параметры, не выбирая значения отсечки для $\delta_2 \approx 0$ от наиболее идентифицируемых до наименее идентифицируемых.

Полуотносительная чувствительность

Еще один метод упорядочивания параметров по степени чувствительности основан на свойствах полуотносительной нормы к исследуемым параметрам. Чтобы описать эту методологию, предположим, что мы хотим определить относительную чувствительность наблюдаемых модельных величин f к конкретным параметрам q_k .

Рассмотрим систему ОДУ в векторно-матричном виде (1). Мы хотим иметь некоторую общую меру чувствительности решения к параметрам, поэтому для каждой комбинации состояние/параметр берем норму (в пространстве \mathbb{L}_2) по времени, а затем ранжируем полученные скаляры, чтобы определить наиболее чувствительные параметры. Полуотносительная чувствительность модельного решения к параметру q_k определяется выражением $\frac{\partial f_i(t; q^*)}{\partial q_k} \cdot q_k^*$ и вычисляется путем формального дифференцирования системы ОДУ (1) по q_k с изменением порядка дифференцирования по времени и параметрам. Таким образом, мы получаем $(n \times m)$ -мерную систему дифференциальных

уравнений для функции чувствительности $f_q(t; q^*) = \partial f(t; q^*) / \partial q$ (где n – это количество уравнений в системе, m – количество неизвестных параметров модели), удовлетворяющей системе (12). Здесь для удобства обозначений используется $f_q(t)$ вместо функции $s_q(t)$.

Данный алгоритм дает информацию о чувствительности как функцию от времени на интересующем интервале. 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 .

Метод Соболя

Еще один метод исследования чувствительности модели к неизвестным параметрам является метод на основе анализа дисперсии модели [51, 52]. В основе данных методов лежит рассмотрение распределения неизвестных параметров с помощью методов Монте-Карло, главной задачей которого является подсчет индексов чувствительности для каждого из исследуемых параметров. Такой метод не подойдет к анализу стохастических и агентных моделей ввиду его трудоемкости. Поэтому чаще используется для систем дифференциальных уравнений с небольшими размерностями пространства неизвестных параметров.

Пусть модель задана в виде:

$$Y = f(q_1, \dots, q_m), \quad (14)$$

где q_1, \dots, q_m – набор неизвестных параметров. Построим матрицу \mathbf{Q} размерности $N \times (m + 2)$, которая является случайным набором неизвестных параметров $\vec{q} = \{q_i, i = 1, \dots, m\}$ в заданных неуточненных границах. Тогда основанный на дисперсии эффект первого порядка на модель (14) для параметра q_i можно записать как:

$$V_{q_i}(\mathbb{E}_{\mathbf{Q}_i}(Y|q_i)).$$

Здесь q_i – i -й параметр, \mathbf{Q}_i – матрица неизвестных параметров без q_i . Смысл оператора математического ожидания $\mathbb{E}_{\mathbf{Q}_i}(Y|q_i)$ заключается в том, что среднее значение для Y берется по всем возможным значениям \mathbf{Q}_i при фиксированном q_i . В то же время дисперсия V_{q_i} берется по всем возможным значениям q_i . Соответствующая мера чувствительности параметра q_i (коэффициент чувствительности первого порядка) записывается следующим образом:

$$S_i = \frac{V_{q_i}(\mathbb{E}_{\mathbf{Q}_i}(Y|q_i))}{V(Y)}, \quad i = 1, \dots, m. \quad (15)$$

Здесь $V(Y)$ – дисперсия по всем строкам матрицы Y . Таким образом, индексы чувствительности представляют собой ожидаемое сокращение дисперсии, которое было бы получено, если бы значение параметра q_i можно было зафиксировать, нормированное на общую дисперсию. Чем больше значение S_i , тем более чувствительный параметр q_i к анализируемым данным. Считается, что при $S_i \in (0.5, 1)$ исследуемый параметр можно считать чувствительным к измерениям [52].

Подробно с программной реализацией и исходным кодом можно ознакомиться в документации <https://salib.readthedocs.io/en/latest/basics.html>, а также в работе [53].

Тестирование амплитудной чувствительности Фурье (FAST)

Распространенным методом глобального анализа чувствительности моделей является расширенный тест амплитудной чувствительности Фурье (eFAST) [70]. Также как и метод Соболя метод eFAST позволяет разбить общую дисперсию выхода

модели на составляющие, т.е. количественно определить, какая доля дисперсии обусловлена каждым входным параметром и их взаимодействием посредством расчетов индексов чувствительности. Однако, по сравнению с другими методами анализа глобальной чувствительности [74], eFAST требует относительно небольшой выборки, что значительно ускоряет расчеты для сложных моделей.

Классический метод FAST использует спектральный анализ для разложения дисперсии $V(Y)$ после предварительного исследования пространства входных параметров с использованием синусоидальных функций разных частот для каждого входного параметра. Модифицированная версия FAST – расширенный метод FAST (eFAST), предложенный в [70], повышает вычислительную эффективность оригинального метода за счет выбора кривой поиска, сканирующей пространство входных параметров. Аналогично методу Соболя представим интересующую нас модель в виде (14), где множество значений, которые могут принимать входные параметры q_1, \dots, q_m , распределены (по возможности равномерно) в единичном гиперкубе

$$K^m = \{x \mid 0 \leq x_i \leq 1; i = 1, \dots, m\}.$$

Вычисление индексов чувствительности (15) сопряжено с вычислением многомерных интегралов [70], что и делает метод Соболя затратным по времени. В [70] авторы показали, что, используя одномерное разложение Фурье вдоль некоторой кривой поиска, исследующей пространство K , можно разложить дисперсию выхода Y по входным параметрам. Для расширенного метода FAST кривая определяется следующей системой выражений:

$$q_i(s) = \frac{1}{2} + \frac{1}{\pi} \arcsin(\sin(\omega_i s + \varphi_i)) \quad i = 1, 2, \dots, m, \quad (16)$$

где ω_i – целочисленная константа (частота), сопряженная с параметром q_i , s – переменная, изменяющаяся в пределах от $-\pi$ до π , φ_i – стартовая точка кривой.

В таких обозначениях симулятор $f(x)$ может быть представлен в виде отображения:

$$f(s) = f(G_1(\sin(\omega_1 s)); \dots, G_n(\sin(\omega_n s))),$$

где G_i – функция замены переменных (16) каждого входного параметра. Тогда $V(Y)$ может быть представлена как

$$V(Y) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f^2(s) ds - \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} f(s) ds \right)^2 \approx \sum_{j=-\infty}^{+\infty} (A_j^2 + B_j^2) - (A_0^2 + B_0^2) \approx 2 \sum_{j=1}^{N_s} (A_j^2 + B_j^2),$$

где N_s – достаточно большое число, а A_j, B_j – коэффициенты Фурье:

$$A_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s) \cos(js) ds, \quad B_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s) \sin(js) ds. \quad (17)$$

Вклад изменения каждого входного параметра q_i в общую дисперсию выхода $V_{q_i}(\mathbb{E}_{\mathbf{Q}_i}(Y|q_i))$ может быть получен путем вычисления коэффициентов Фурье вдоль

основной частоты ω_i и её гармоник $p\omega_i$ [70]:

$$V_{q_i}(\mathbb{E}_{Q_i}(Y|q_i)) = \sum_{p \in \mathbb{Z} \setminus \{0\}} (A_{p\omega_i}^2 + B_{p\omega_i}^2).$$

Таким образом, если параметр q_i является чувствительным, то колебания f на соответствующей частоте ω_i и других его гармониках $p\omega_i$ будут иметь большую амплитуду. Такие амплитуды количественно оцениваются по спектру $\Lambda^2(u) = A_u^2 + B_u^2$ на тех частотах, которые составляют основу для расчета чувствительности модели.

На практике коэффициенты Фурье (17) рассчитываются по квадратурным формулам на множестве равноотстоящих N_s точек на кривой поиска (16), где N_s обычно выбирают в соответствии с теоремой Найквиста [70]:

$$N_s = N_r \left(2M_p \max_{i=1, \dots, n} \{w_i\} + 1 \right),$$

где M_p – число гармоник ($p = 1, \dots, M_p$), обычно выбираемое равным 4 или 6, а N_p – количество смещений по s при изменении начальной точки $\varphi_i, i = 1, \dots, N_r$. Подробнее численная реализация метода eFAST описана в работах [70, 72].

Отметим, что существует две особенности, затрудняющие использование метода eFAST. Во-первых, частоты $w_i, i = 1, \dots, n$, должны быть выбраны таким образом, чтобы никакая гармоника $p\omega_i$ не являлась линейной комбинацией любого другого подмножества гармоник любых других частот, чтобы избежать двусмысленности расчетов, то есть должно выполняться условие

$$\sum_{i=1}^n r_i \omega_i \neq 0 \quad -\infty < r_i < +\infty. \quad (18)$$

Поиск последовательности чисел, удовлетворяющих условию (18) сам по себе является трудоёмкой задачей. Отметим, что для $m < 20$ последовательности, удовлетворяющие (18), выписаны в [73].

Во-вторых, индекс чувствительности S_i первого порядка (15) отображает только вклад отдельно взятого входа на общий выход, т.е. не учитывает взаимодействие входов между собой. Классический метод FAST позволяет вычислять только индексы чувствительности первого порядка в отличие от метод Соболя. Отметим, однако, что расширенная версия метода (eFAST) позволяет вычислять также общие индексы чувствительности [70]:

$$ST_i = 1 - V_{-q_i}/V(Y), \quad (19)$$

где V_{-q_i} выражает дисперсию при изменении всех параметров за исключением q_i . Таким образом, индекс ST_i отображает влияние на выход всех подгрупп входных параметров, куда входит параметр q_i .

Примеры моделей

SEIR-HCD-модель распространения COVID-19

Для описания распространения новой коронавирусной инфекции COVID-19, вызванной вирусом SARS-CoV-2, была выбрана модель SEIR-HCD, характеризующаяся

системой из семи обыкновенных дифференциальных уравнений [54, 55, 56]

$$\left\{ \begin{array}{l} \frac{dS}{dt} = -\tilde{a}(t) \left(\frac{\alpha_I(t)S(t)I(t)}{N} + \frac{\alpha_E(t)S(t)E(t)}{N} \right) + \frac{1}{t_{imm}}R(t), \\ \frac{dE}{dt} = \tilde{a}(t) \left(\frac{\alpha_I(t)S(t)I(t)}{N} + \frac{\alpha_E(t)S(t)E(t)}{N} \right) - \frac{1}{t_{inc}}E(t), \\ \frac{dI}{dt} = \frac{1}{t_{inc}}E(t) - \frac{1}{t_{inf}}I(t), \\ \frac{dR}{dt} = \frac{\beta(t)}{t_{inf}}I(t) + \frac{1 - \varepsilon_{HC}(t)}{t_{hosp}}H(t) - \frac{1}{t_{imm}}R(t), \\ \frac{dH}{dt} = \frac{1 - \beta(t)}{t_{inf}}I(t) + \frac{1 - \mu(t)}{t_{crit}}C(t) - \frac{1}{t_{hosp}}H(t), \\ \frac{dC}{dt} = \frac{\varepsilon_{HC}(t)}{t_{hosp}}H(t) - \frac{1}{t_{crit}}C(t), \\ \frac{dD}{dt} = \frac{\mu(t)}{t_{crit}}C(t) \end{array} \right. \quad (20)$$

с начальными условиями

$$\begin{aligned} S(t_0) &= N - E_0 - I_0 - R_0 - H_0 - C_0 - D_0, & E(t_0) &= E_0, \\ I(t_0) &= I_0, & R(t_0) &= R_0, & H(t_0) &= H_0, & C(t_0) &= C_0, & D(t_0) &= D_0. \end{aligned} \quad (21)$$

Схема модели (20) приведена на рисунке 2а, где вся популяция $N = S(t) + E(t) + I(t) + R(t) + H(t) + C(t) + D(t)$ разделена на семь групп:

- S – восприимчивое неиммунизированное население,
- E – бессимптомные носители инфекции,
- I – симптомные инфицированные индивидуумы,
- R – вылеченная группа населения от COVID-19,
- H – госпитализированные индивидуумы,
- C – критические случаи, требующие подключения к аппарату искусственной вентиляции легких (ИВЛ),
- D – умершие случаи в результате COVID-19.

Модель (20) допускает единственное состояние равновесие без инфекции, которое при отсутствии иммунизации населения (слагаемого $t_{imm}^{-1}R(t)$) устойчиво по Ляпунову при условии $\mathcal{R}_0 < 1$, где $\mathcal{R}_0 = \tilde{a}(t_{inc}\alpha_E + t_{inf}\alpha_I)$ – базовый индекс репродукции вируса.

В обратной задаче для модели (20) распространения COVID-19 необходимо определить коэффициенты

$$\vec{q}(t) = (\alpha_I(t), \alpha_E(t), \beta(t), \varepsilon_{HC}(t), \mu(t))$$

и начальные условия модели E_0 и R_0 по дополнительной информации о количестве симптомных выявленных $(1 - b_k)f_k$, критических C_k (подключенных к аппарату ИВЛ в день k) и умерших g_k в результате инфекции в фиксированные моменты времени

$$I(t_k) = (1 - b_k)f_k, \quad C(t_k) = C_k, \quad D(t_k) = g_k, \quad k = 1, \dots, K, \quad (22)$$

а также статистических данных ВОЗ о течении заболевания: среднее время инкубационного периода t_{inc} , инфицирования t_{inf} , госпитализации t_{hosp} , пребывания

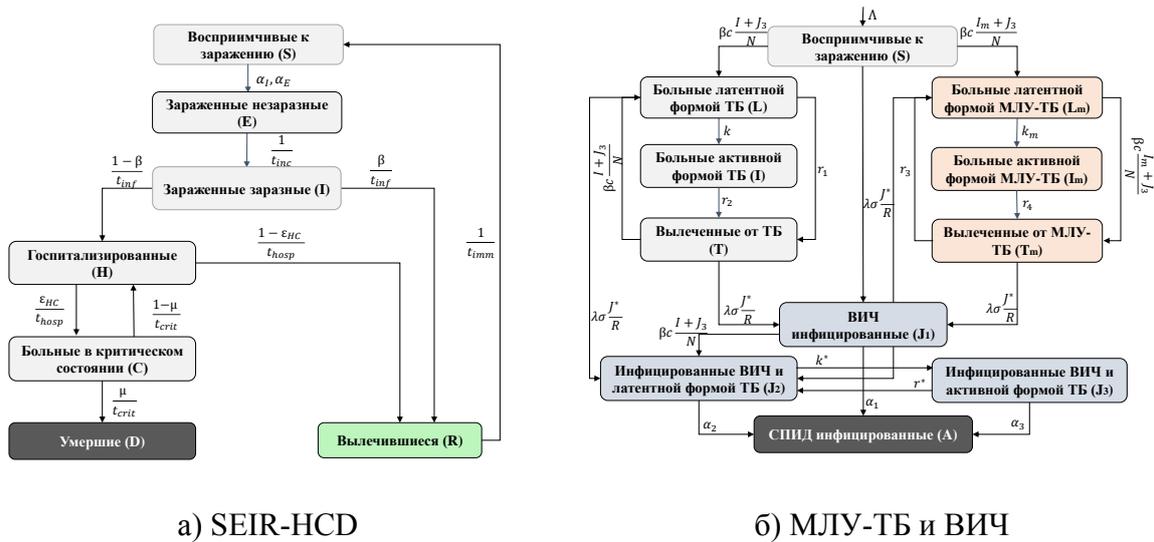


Рис. 2. Блок-схемы дифференциальных моделей эпидемиологии.

в критическом состоянии t_{crit} и продолжительности гуморального иммунитета после перенесенного заболевания t_{imm} . Здесь b_k – доля бессимптомных случаев из выявленных больных методов ПЦР-тестирования в день k , f_k – количество выявленных инфицированных COVID-19 в день k .

Результаты численных расчетов полуотносительной чувствительности параметров $q = (\alpha_E, \alpha_I, \beta, \epsilon_{HC}, \mu, E_0)$ к измерениям $I(t_k)$, $C(t_k)$ и $D(t_k)$ (22) в модели распространения COVID-19 SEIR-HCD типа (20) приведены в Таблице 1. Отметим, что в случае анализа чувствительности части начальных данных в начальных условиях системы на функцию чувствительности (12) будет не $\vec{0}$, а частные производные по анализируемому начальным условиям $\frac{\partial y_0}{\partial q}$.

Таблица 1. Полуотносительные чувствительности состояний модели к параметрам (значения норм $\|\frac{\partial f_i(t)}{\partial q_k} \cdot q_k^*\|_2$) для модели (20)

Параметр q	I	C	D
α_E	$2.865 \cdot 10^{14}$	$2.386 \cdot 10^{12}$	$7.11 \cdot 10^{11}$
α_I	$2.396 \cdot 10^{14}$	$1.996 \cdot 10^{12}$	$5.947 \cdot 10^{11}$
β	$2.03 \cdot 10^{-6}$	$3.428 \cdot 10^4$	$2.164 \cdot 10^4$
ϵ_{HC}	$1.6 \cdot 10^{-7}$	$4.833 \cdot 10^4$	$3.041 \cdot 10^4$
μ	$3 \cdot 10^{-8}$	$3.695 \cdot 10^2$	$2.982 \cdot 10^4$
E_0	$1.854 \cdot 10^{14}$	$1.544 \cdot 10^{12}$	$4.601 \cdot 10^{11}$

Качество определения параметров β , ϵ_{CH} 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 рисунке (рис. 3) представлены графики изменения от времени функции чувствительности $\frac{\partial f_i(t)}{\partial q_k} \cdot q_k^*$ в зависимости от варьируемого параметра. Чем более

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

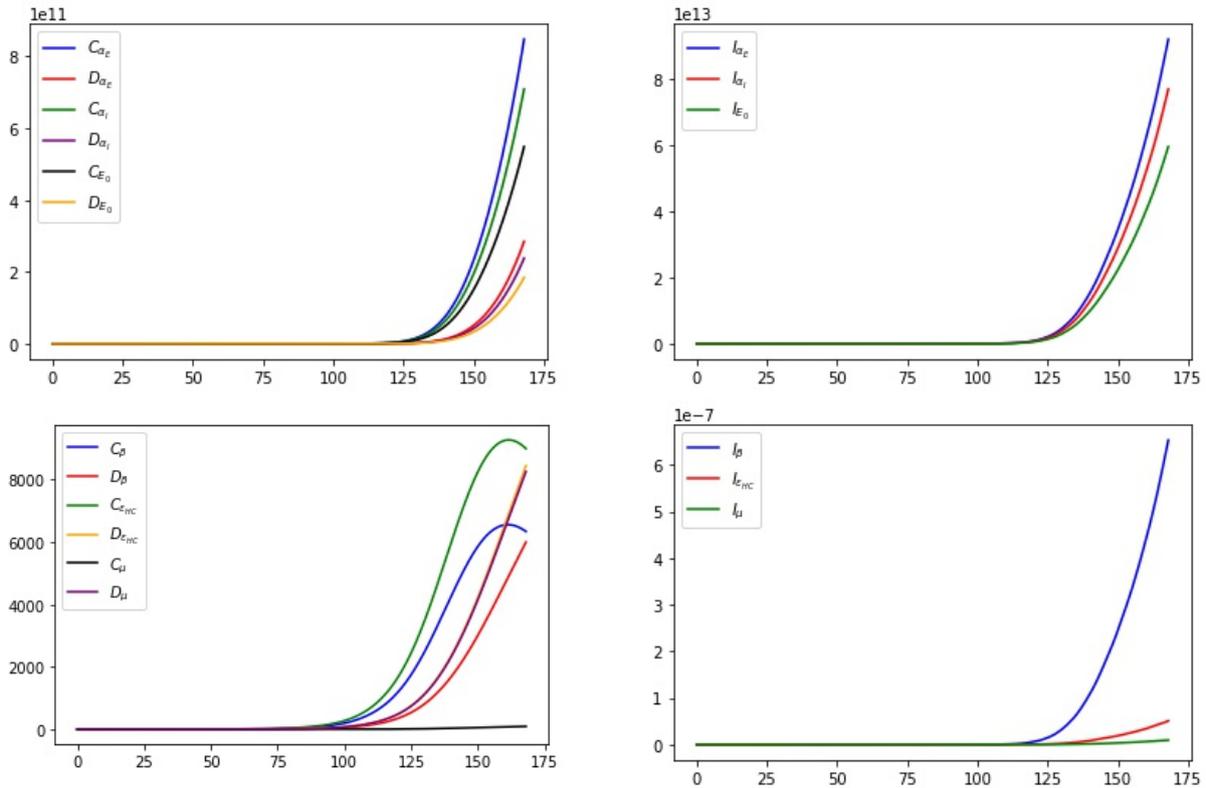


Рис. 3. Функция полуотносительной чувствительности $\frac{\partial f_i(t)}{\partial q_k} \cdot q_k^*$ для временного промежутка с 15.04.2020 по 01.10.2020 (по оси Ox $T = 170$ дней).

На рисунке (рис. 4) приведены результаты анализа чувствительности параметров математической модели (20) при различных итерациях ортогонального алгоритма (описание алгоритма приведено выше). По горизонтальной оси отложены итерации ортогонального алгоритма, количество которых на 1 меньше размерности вектора неизвестных параметров (количества столбцов матрицы чувствительности), а по вертикальной – значения норм перпендикуляров для полученных преобразований матриц чувствительности.

После анализа идентифицируемости можно заключить, что наименее чувствительными (более идентифицируемыми) параметрами модели к вариациям в данных (погрешностям) являются параметры заражения между бессимптомной и восприимчивой группами населения α_E , заражения между инфицированным и восприимчивым населением, который связан с контагиозностью вируса и социальными факторами α_I , а также начальное значение зараженных или находящихся в инкубационном периоде индивидуумов E_0 . Другими словами, эти параметры более устойчиво определяются при решении обратной задачи (20), (22).

Наименее чувствительным (менее идентифицируемым) к ошибкам в измерениях является параметр β , характеризующий скорость получения иммунитета после заболевания (с наименьшими значениями норм перпендикуляров в матрице чувствительности). Поэтому в качестве априорной информации для параметра β была использована информация о доле популяции в рассматриваемом регионе с наличием поздних антител IgG к вирусу SARS-CoV-2, полученную с сайта Инвитро.

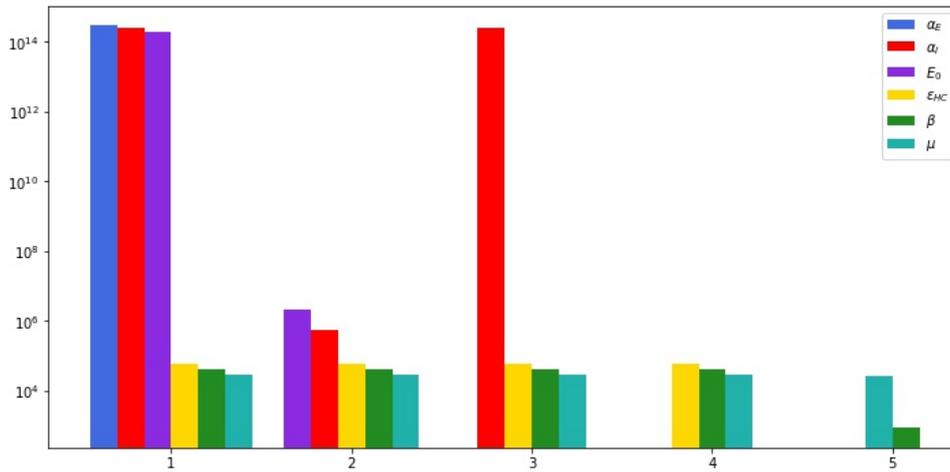


Рис. 4. Величины норм перпендикуляров (вертикальная ось) для каждого параметра (соответствует цвету) на различных итерациях (горизонтальная ось) ортогонального алгоритма чувствительности параметров математической модели (20).

Модель коинфекции туберкулеза и ВИЧ

За основу описания коинфекции туберкулеза (ТБ) и вируса иммунодефицита человека (ВИЧ) была взята SIR-модель, характеризующаяся системой из 8 дифференциальных уравнений [57], к которым добавлены три группы с туберкулезом с множественной лекарственной устойчивостью (МЛУ-ТБ) (с индексом m)

$$\left\{ \begin{array}{l}
 \frac{dS}{dt} = \Lambda - \beta c S \frac{I + J_3}{N} - \beta c S \frac{I_m + J_3}{N} - \lambda \sigma S \frac{J^*}{R} - \mu S; \\
 \frac{dL}{dt} = \beta c (S + T) \frac{I + J_3}{N} - \lambda \sigma L \frac{J^*}{R} - (\mu + k + r_1)L; \\
 \frac{dI}{dt} = kL - (\mu + d + r_2)I; \\
 \frac{dT}{dt} = r_1L + r_2I - \beta c T \frac{I + J_3}{N} - \lambda \sigma T \frac{J^*}{R} - \mu T; \\
 \frac{dJ_1}{dt} = \lambda \sigma T_m \frac{J^*}{R} + \lambda \sigma (S + T) \frac{J^*}{R} - \beta c J_1 \frac{I + J_3}{N} - (\alpha_1 + \mu)J_1; \\
 \frac{dJ_2}{dt} = \lambda \sigma L_m \frac{J^*}{R} + \lambda \sigma L \frac{J^*}{R} + \beta c J_1 \frac{I + J_3}{N} - (\alpha_2 + \mu + k^*)J_2 + r^*J_3; \\
 \frac{dJ_3}{dt} = k^*J_2 - (\alpha_3 + \mu + d^* + r^*)J_3; \\
 \frac{dA}{dt} = \alpha_1J_1 + \alpha_2J_2 + \alpha_3J_3 - (\mu + f)A; \\
 \frac{dL_m}{dt} = \beta c (S + T) \frac{I_m + J_3}{N} - \lambda \sigma L_m \frac{J^*}{R} - (\mu + k_m + r_3)L_m; \\
 \frac{dI_m}{dt} = k_mL_m - (\mu + d_m + r_4)I_m; \\
 \frac{dT_m}{dt} = r_3L_m + r_4I_m - \beta c T_m \frac{I_m + J_3}{N} - \lambda \sigma T_m \frac{J^*}{R} - \mu T_m
 \end{array} \right. \quad (23)$$

с начальными условиями

$$\begin{aligned} S(0) &= S_0, L(0) = L_0, I(0) = I_0, T(0) = T_0, \\ J_1(0) &= J_{1,0}, J_2(0) = J_{2,0}, J_3(0) = J_{3,0}, A(0) = A_0, \\ L_m(0) &= L_{m,0}, I_m(0) = I_{m,0}, T_3(0) = T_{m,0}. \end{aligned} \quad (24)$$

Здесь $N = \text{const}$ – размер исследуемой популяции, $R = N - I(t) - I_m(t) - J_3(t) - A(t) = S(t) + L(t) + T(t) + J_1(t) + J_2(t) + L_m(t) + T_m(t)$ – популяция без активной фазы туберкулеза (обычного и МЛУ) и СПИД, $J^* = J_1(t) + J_2(t) + J_3(t)$ – ВИЧ инфицированные, у которых еще не развилась стадия СПИДа. Подробная схема математической модели коинфекции ВИЧ и МЛУ туберкулеза представлена на рисунке 26. Подробный анализ модели в окрестностях состояния равновесия приведен в работе [57].

Результаты численных расчетов полутносительной чувствительности параметров $q = (\beta c, \lambda \sigma, k, k^*, r_1, r_2, r^*, \alpha_1, \alpha_2, \alpha_3, \theta, r_3, r_4, k_m)$ для модели (23) коинфекции ВИЧ и ТБ представлены в Таблице 2.

Таблица 2. Полутносительные чувствительности измеряемых состояний модели (23) к неизвестным параметрам (значения норм $\|\frac{\partial f_i(t)}{\partial q_k} \cdot q_k^*\|_2$)

Параметр q	I	J_3	A	I_m
βc	0.131	$1 \cdot 10^{-5}$	$4 \cdot 10^{-5}$	0.067
$\lambda \sigma$	$1 \cdot 10^{-6}$	0.0001	0.151	$2.7 \cdot 10^{-8}$
k	1410453.27	$5 \cdot 10^{-6}$	$2 \cdot 10^{-5}$	5.04
k^*	0.00052	5510.63	369.42	0.009
r_1	1205447.68	0.00012	0.00061	4.648
r_2	1031500.99	$5 \cdot 10^{-6}$	$2 \cdot 10^{-5}$	4.75
r^*	0.00038	3971.85	265.29	0.0068
α_1	$1 \cdot 10^{-12}$	$1 \cdot 10^{-7}$	602.93	0
α_2	0.00019	1779.423	52430.77	0.003
α_3	0.000029	279.81	1485.35	0.0005
θ	4.103	0.0	0.0	7.23
r_3	0.0	$1.5 \cdot 10^{-7}$	$7.7 \cdot 10^{-7}$	9235.84
r_4	0.0	0.0	0.0	4505.68
k_m	0.0	0.0	0.0	34.71

Параметры $\beta c, k, r_1, r_2$ не могут быть устойчиво идентифицированы по измерениям J_3 и A , однако чувствительны к измерению I . Параметры α_1 и $\lambda \sigma$ чувствительны только к измерению A . В свою очередь параметры $k^*, r^*, \alpha_2, \alpha_3$ могут быть определены по данным измерений J_3 и A , и не чувствительны к данным I .

Параметры r_3, r_4, k_m , связанные с МЛУ формой ТБ, имеют большое значение нормы полутносительной чувствительности для данных I_m и могут быть восстановлены по измерениям о количестве зараженных лекарственно-устойчивым туберкулезом. В то

же время данные параметры не чувствительны к остальным измерениям модели, т.к. они отражают вероятности перехода только между камерами, связанными с большими лекарственно-устойчивыми штаммами туберкулеза.

Пространственная SEIR-HCD-модель распространения COVID-19

Модель SEIR-HCD распространения COVID-19 (20) была преобразована в модель типа «реакции-диффузии» [58]

$$\left\{ \begin{array}{l} \partial_t s = \nabla(n \bar{v}_s \nabla s) - \tilde{a}(\bar{\alpha}_i s i + \bar{\alpha}_e s e) + t_{imm}^{-1} r, \\ \partial_t e = \nabla(n \bar{v}_e \nabla e) + \tilde{a}(\bar{\alpha}_i s i + \bar{\alpha}_e s e) - t_{inc}^{-1} e, \\ \partial_t i = \nabla(n \bar{v}_i \nabla i) + t_{inc}^{-1} e - t_{inf}^{-1} i, \\ \partial_t r = \nabla(n \bar{v}_r \nabla r) + \beta(t) t_{inf}^{-1} i + (1 - \bar{\epsilon}_{HC}) t_{hosp}^{-1} h - t_{imm}^{-1} r, \\ \partial_t h = (1 - \beta(t)) t_{inf}^{-1} i + (1 - \bar{\mu}) t_{crit}^{-1} c - t_{hosp}^{-1} h, \\ \partial_t c = \bar{\epsilon}_{HC} t_{hosp}^{-1} h - t_{crit}^{-1} c, \\ \partial_t d = \bar{\mu} t_{crit}^{-1} c, \end{array} \right. \quad (25)$$

со следующими начальными

$$\begin{aligned} s(0, x) &= e^{-(x+1)^4} + e^{-\frac{(x-0.35)^2}{10^{-2}}} + \frac{1}{8} \left(e^{-\frac{(x-0.62)^4}{10^{-5}}} + e^{-\frac{(x-0.52)^4}{10^{-5}}} + e^{-\frac{(x-0.42)^4}{10^{-5}}} \right) + \\ &\quad + \frac{1}{4} e^{-\frac{(x-0.735)^4}{10^{-5}}}, \\ e(0, x) &= \frac{1}{20} e^{-\frac{(x-0.75)^4}{10^{-5}}}, \quad i(0, x) = I_0, \quad r(0, x) = R_0, \\ h(0, x) &= H_0, \quad c(0, x) = C_0, \quad d(0, x) = D_0 \end{aligned} \quad (26)$$

и граничными условиями

$$\partial_x y(t, 0) = 0, \quad y(t, 1) = 0. \quad (27)$$

Здесь $\bar{v} = (\bar{v}_s, \bar{v}_e, \bar{v}_i, \bar{v}_r)$ – скорости перемещения в пространстве групп восприимчивых, бессимптомных и инфицированных носителей COVID-19, а также выздоровевших от COVID-19, $q = (\bar{\alpha}_i, \bar{\alpha}_e, \bar{\epsilon}_{HC}, \bar{\mu})$ – вектор постоянных параметров модели, характеризующие распространение COVID-19 в конкретном регионе, $y(t, x) = (s, e, i, r, h, c, d)(t, x)$. В работе предполагается, что госпитализированная группа населения, критическая и умершие не могут перемещаться в пространстве.

Обозначим вектор параметров модели

$$\vec{q} = (\bar{\alpha}_i, \bar{\alpha}_e, t_{inc}, t_{inf}, \beta(t), \bar{\epsilon}_{HC}, t_{hosp}, t_{imm}, \bar{\mu}, t_{crit}, v_s, v_e, v_i, v_r) \in \mathbb{R}^{14},$$

и дополнительную информацию о процессе при K днях измерений интегрально по всем потенциальным источникам x , т.е.

$$\int_{\Omega} i(t_k, x) dx = (1 - b_k) f_k, \quad \int_{\Omega} c(t_k, x) dx = C_k, \quad \int_{\Omega} d(t_k, x) dx = g_k, \quad (28)$$

$$k = 1, \dots, K.$$

С помощью метода Соболя в реализации библиотеки Salib на языке Python <https://github.com/SALib/SALib> оценим значения индексов чувствительности, построенными по формуле (15).

На рисунке 5 представлены значения индексов чувствительности $S_i, i = 1, \dots, 14$, для всех неизвестных параметров в различные временные срезы (ось Ox), т.е. исследовалась дисперсия системы в пяти временных точках решения прямой задачи: на 40-ой, 80-ый, 120-ый, 160-ый и 200-ый дни моделирования, которые на рисунке соединены линиями.

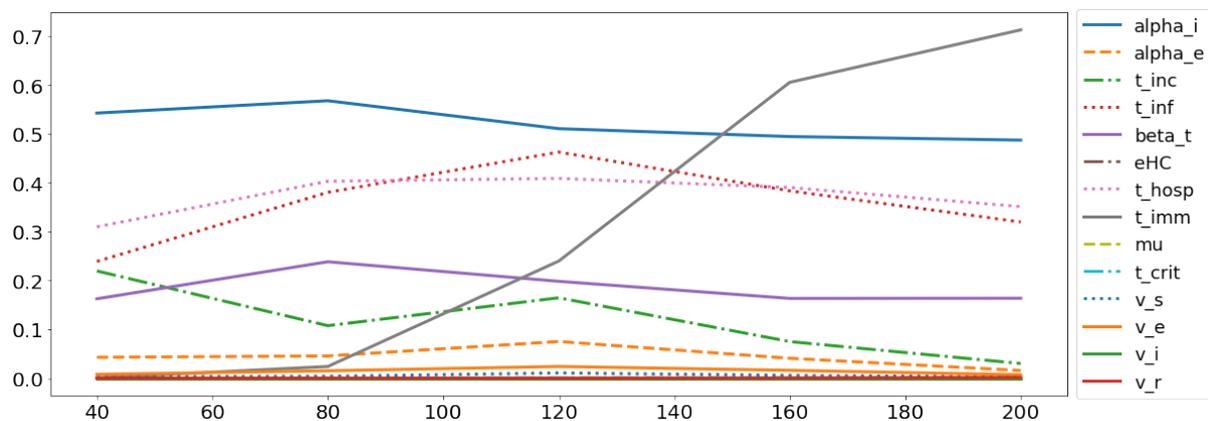


Рис. 5. Индексы чувствительности S_i для неизвестных параметров $\vec{q} \in \mathbb{R}^{14}$ в зависимости от времени (с 40 по 200 день).

Отметим, что параметр t_{imm} со временем становится более влиятельным для системы, что естественно, поскольку популяционный иммунитет возрастает со временем. В то же время индексы чувствительности параметров $\bar{\alpha}_i, t_{inc}, t_{inf}, t_{hosp}$ наоборот постепенно уменьшаются. Это может быть вызвано специфичностью модели – группы, отвечающие за распространение заболевания (I, E, H), в уравнениях которых участвуют последние параметры, имеют волнообразную особенность и со временем затухают. Диффузионные коэффициенты v_s, v_e, v_i, v_r не чувствительны к измерениям обратной задачи (28).

На рисунке 6 представлены столбчатые диаграммы значений индексов чувствительности $S_i, i = 1, \dots, 14$, в зависимости от времени измерений обратной задачи (40, 80, 120 и 160 дни с момента моделирования). По диаграмме можно сделать заключение о том, что $\bar{\alpha}_i$ является самым чувствительным параметром, в то время как диффузионные параметры v_s, v_i, v_r, v_e являются наименее чувствительными к данным (28).

Как уже упоминалось ранее, метод Соболя является достаточно трудоемким даже для анализа 14 параметров к вариации 3х типов данных пространственной SEIR-HCD-модели распространения COVID-19 (25)-(26). Ввиду этого к анализу данной модели также был применен байесовский подход, описанный в следующем разделе, который был предложен авторами в 2015 году [59] и адаптирован в данной работе к исследуемым АОМ и пространственной SEIR-HCD распространения COVID-19 в регионе.

SEIR-HCD-модель «среднего поля»

В настоящее время использование теории «игр среднего поля» (Mean Field Games в оригинале) для прогнозирования заболеваемости COVID-19 успешно развивается, см., например, работы [66, 67, 68, 75]. Наибольшее распространение получил подход, при котором модель «среднего поля» строится на основе дифференциальной модели SIR типа [67, 68, 66], при этом на среднесрочном периоде моделирования прогнозы модели «среднего поля» (МСП), как правило, оказываются более успешными, чем для систем на основе ОДУ. В работах показано, что МСП могут быть более гибкими в учете влияний внешних факторов на эпидемиологический процесс, но являются вычислительно более трудоемкими, чем базовые дифференциальные модели, и, также как и базовые

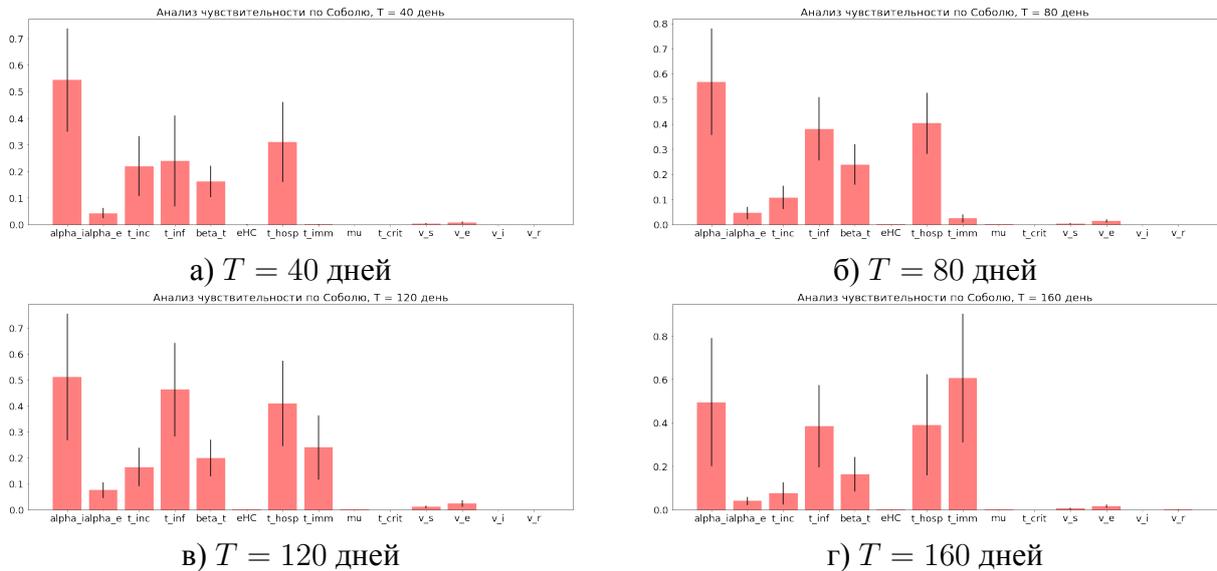


Рис. 6. Значения индексов чувствительности S_i для параметров $\vec{q} \in \mathbb{R}^{14}$ при различных измерениях данных. Красными столбцами представлены значения индексов S_i , черными линиями – доверительный интервал для S_i , $i = 1, \dots, 14$.

дифференциальные модели, являются чувствительными по отношению к используемым параметрам.

Для постановки эпидемиологической МСП введем в рассмотрение плотность распределения популяции внутри семи эпидемиологических групп согласно системе ОДУ (20) $m_i(t, x) : [0, T] \times [0, 1] \rightarrow \mathbb{R}$, $i \in \{S, E, I, R, H, C, D\}$, где m_i принимается вместо соответствующих значений i , используемых в (20) и $\int_0^1 m(t, x) dx$ представляет собой долю населения в каждой эпидемиологической группе в момент времени t . Предположим, что переменная состояния x колеблется в пределах $\Omega = [0, 1]$ и обозначает склонность населения к соблюдению карантинных мер, таких как физическое дистанцирование и самоизоляция. Здесь «0» означает, что представитель популяции лоялен к введенным ограничительным мерам и склонен их соблюдать, а «1» означает обратное. Введем также функцию $v(t, x) : [0, T] \times [0, 1] \rightarrow \mathbb{R}$, обозначающую стратегию изоляции в пределах всей массы населения. На основе описания эпидемиологического процесса (20) и принципов теории «среднего поля» функции $m_i(t, x)$ являются решением следующей системы уравнений в частных производных [66, 67, 68]:

$$\left\{ \begin{array}{l} \frac{\partial m_S}{\partial t} + \frac{\partial (m_S v)}{\partial x} - \frac{\sigma^2}{2} \frac{\partial^2 m_S}{\partial x^2} + \frac{5-a}{5} (\alpha_I m_S m_I + \alpha_E m_S m_E) - t_{imm}^{-1} m_R = 0, \\ \frac{\partial m_E}{\partial t} + \frac{\partial (m_E v)}{\partial x} - \frac{\sigma^2}{2} \frac{\partial^2 m_E}{\partial x^2} - \frac{5-a}{5} (\alpha_I m_S m_I + \alpha_E m_S m_E) + t_{inc}^{-1} m_E = 0, \\ \frac{\partial m_I}{\partial t} + \frac{\partial (m_I v)}{\partial x} - \frac{\sigma^2}{2} \frac{\partial^2 m_I}{\partial x^2} - t_{inc}^{-1} m_E + t_{inf}^{-1} m_I = 0, \\ \frac{\partial m_R}{\partial t} + \frac{\partial (m_R v)}{\partial x} - \frac{\sigma^2}{2} \frac{\partial^2 m_R}{\partial x^2} - \beta \cdot \omega_{inf} m_I - (1 - \varepsilon_{HC}) \cdot t_{hosp}^{-1} m_H + t_{imm}^{-1} m_R = 0, \\ \frac{\partial m_H}{\partial t} - (1 - \beta) \cdot t_{inf}^{-1} m_I - (1 - \mu) \cdot t_{crit}^{-1} m_C + t_{hosp}^{-1} m_H = 0, \\ \frac{\partial m_C}{\partial t} - \varepsilon_{HC} t_{hosp}^{-1} m_H + t_{crit}^{-1} m_C = 0, \\ \frac{\partial m_D}{\partial t} - \mu t_{crit}^{-1} m_C = 0 \end{array} \right. \quad (29)$$

с начальными

$$m_i(0, x) = m_{0i}(x) \quad \forall i \in S, E, I, R, H, C, D \quad (30)$$

и граничными условиями

$$\frac{\partial m_i}{\partial x} = 0, \quad \forall t \in [0, T] \text{ и } x = 0, 1. \quad (31)$$

Здесь σ – неотрицательный параметр, отображающий стохастичность процесса [69].

Поставим задачу оптимизации [69]: предположим, что члены популяции рациональны и склонны выбирать стратегию поведения, максимизирующую их собственную выгоду (минимизирующую затраты) в соответствии с функционалом стоимости

$$\min_v J = \min_v \int_0^T \int_0^1 \left(\frac{1}{2} (m_S(t, x) + m_E(t, x) + m_I(t, x) + m_R(t, x)) v^2(t, x) + \right. \\ \left. + (m_E^2(t, x) + m_I^2(t, x) + m_D^2(t, x)) \right) dx dt. \quad (32)$$

Функционал вида (32) определяет оптимальную стратегию как функцию, минимизирующую число заболевших и умерших в рассматриваемом регионе. Подробнее о различных эпидемиологических сценариях, выраженных через функционал стоимости, а также сравнение прогнозов МСП с прогнозами систем ОДУ, на базе которых строится эпидемиологическая МСП, представлено в работах [67, 68]. Там же предложен численный метод решения таких задач.

Таким образом, подход «среднего поля» в применении к эпидемиологии приводит к задаче оптимизации, требующей поиска оптимальной стратегии $v(t, x)$ путем минимизации функционала типа (32) с ограничением в виде системы уравнений в частных производных (29)-(31).

Для анализа чувствительности в качестве выхода $Y = (y_1, \dots, y_7)$ будем считать долю населения в каждой эпидемиологической группе в выбранный момент времени \tilde{t} :

$$y_1 = \int_0^1 m_S(\tilde{t}, x) dx, \quad y_2 = \int_0^1 m_E(\tilde{t}, x) dx, \quad \dots, \quad y_7 = \int_0^1 m_D(\tilde{t}, x) dx. \quad (33)$$

Анализ чувствительности проводился методом eFAST, описанным в разделе на нескольких выборках входных параметров, определяемых комбинациями M_p и N_r :

$$\begin{aligned} M_p &= 4, \quad N_r = 1, 2; \\ M_p &= 6, \quad N_r = 1, 2, \end{aligned} \quad (34)$$

что позволяет рассчитать усредненные индексы чувствительности S_i для каждого выхода и их стандартное отклонение по общей выборке. В каждом случае начальная точка кривой поиска выбиралась как $\varphi_r = r \cdot \pi/4, = 0, \dots, N_r - 1$.

На рисунке 7 представлены расчеты индексов чувствительности для вектора входных параметров $\vec{q} = (a, \alpha_I, \alpha_E, \beta, \varepsilon_{HC}, \mu, t_{inc}, t_{inf}, t_{hosp}, t_{crit}, t_{imm}, \sigma^2)^T \in \mathbb{R}^{12}$ относительно каждого выхода, выраженного согласно (33) на 40/80/120/160 день. Результаты показывают, что более всего модель «среднего поля» чувствительна к параметрам $\{a, \alpha_I, \alpha_E, \beta, \varepsilon_{HC}, \mu, t_{inc}, t_{inf}\}$ при этом абсолютно все выходы чувствительны к параметру a , характеризующему соблюдение агентами самоизоляции.

Для описания взаимосвязи параметров модели на рисунке 8 представлено сравнение индексов общей чувствительности (ST_i – зеленый столбец) и индексов чувствительности первого порядка (S_i – голубой столбец), полученных методом eFAST для выхода



Рис. 7. Индексы чувствительности, полученные методом eFAST для всех выходов (33), полученные на 40/80/120/160-й день для модели «среднего поля».

y_1 (33) на 40/80/120/160-й день. Чем больше разница между столбцами, тем более сложно идентифицируем параметр. Из рисунка видно, что, несмотря на то, что система

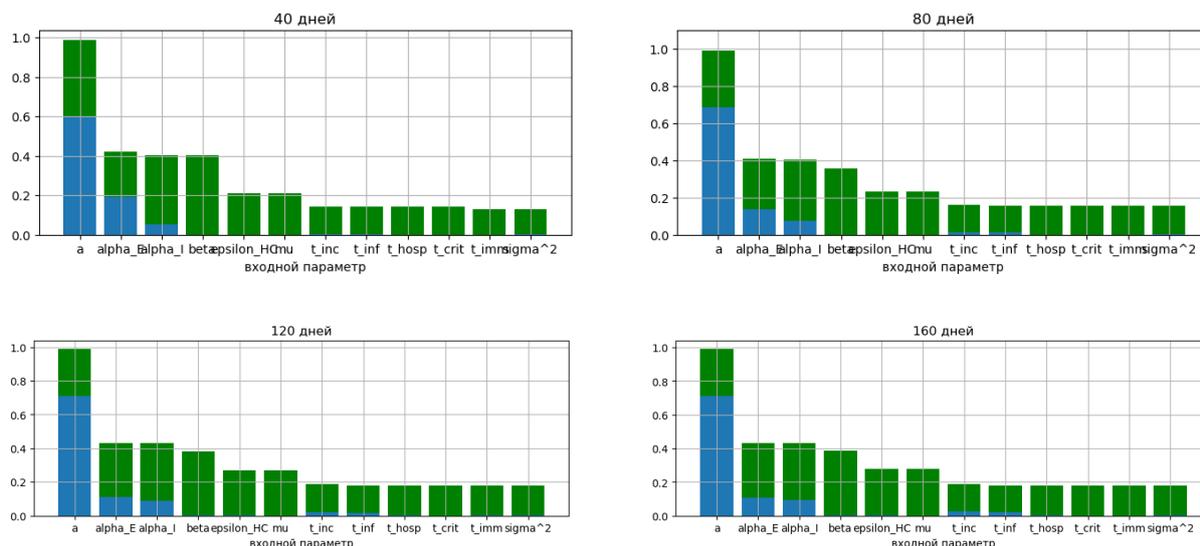


Рис. 8. Индексы общей чувствительности (ST_i – зеленый столбец) и индексы чувствительности первого порядка (S_i – голубой столбец), полученные методом eFAST для выхода y_1 (33) на 40/80/120/160-й день для модели «среднего поля».

наиболее чувствительна к параметру a , он может быть определен независимо. Параметры α_I , α_E , β , ϵ_{HC} , μ имеют большое общее влияние на систему, при этом зависимы от вариаций других параметров.

ИДЕНТИФИЦИРУЕМОСТЬ АГЕНТНО-ОРИЕНТИРОВАННЫХ И СТОХАСТИЧЕСКИХ МОДЕЛЕЙ

Для решения обратных задач в случае агентно-ориентированных моделей (АОМ) и стохастических моделей применение стохастических алгоритмов нецелесообразно, а детерминированные алгоритмы носят локальный характер. Для увеличения точности решения обратной задачи необходимо восстанавливать параметры в заранее уточненных границах. Существует ряд классических алгоритмов для исследования чувствительности и идентифицируемости параметров для дифференциальных моделей [34, 3, 60, 61]. Однако они неприменимы для агентных моделей, так как в них нет явных зависимостей, выраженных в системах ОДУ и УЧПР. Для такого рода стохастических систем разработан метод на основе байесовского подхода [59], позволяющий разрешить следующие проблемы:

1. Задача идентификации параметров \vec{q} по экспериментальным данным f_k , $k = 1, \dots, K$, может быть переопределена.
2. Если параметры \vec{q} не идентифицируемы для некоторых измерений, то они могут быть идентифицируемы для другого набора данных.
3. Если имеется дополнительная статистическая информация о \vec{q} , которая идет вразрез с полученными в рамках модели оценками и некорректными распределениями.

Воспользуемся реализацией данного подхода для исследования чувствительности параметров агентной модели распространения COVID-19 в Новосибирской области, подробное описание которой приведено в работе [62].

Построение эмулятора

Основная идея байесовского подхода состоит в построении гибридной модели (эмулятора) для исследуемой агентной модели (симулятора). Вычисление на основе

эмулятора требует меньше машинного времени, следовательно он может быть запущен многократно для более точной оценки результата. Общий вид эмулятора представляется в виде:

$$g(\vec{q}) = \sum_{i=1}^p h_i(\vec{q}) \cdot \beta_i + u(\vec{q}), \quad (35)$$

где $h_i(\vec{q})$ – известные базисные функции (например, мономы), β_i – регрессионные коэффициенты, $u(\vec{q})$ – гауссовский процесс с заранее заданной функцией ядра.

Алгоритм для построения эмулятора и исследования идентифицируемости параметров скорости распространения инфекции α , начального количества инфицированных $E(0)$ и шанса быть протестированным p агентной модели представлен схематично на рисунке 9 и имеет следующий вид:

1. В заданных неуточненных границах a_i, b_i ($i = 1, \dots, m$, где m – размерность вектора \vec{q}) выбираются точки $\widehat{q}_1, \dots, \widehat{q}_n$ с помощью латинского гиперкуба n -го порядка [63], который дает равномерное распределение точек в пространстве параметров.
2. Решается прямая задача с помощью симулятора (вычисляется решение агентной модели на основе комплекса Covasim) с использованием сгенерированных n точек, и сохраняются результаты модели – выходные данные $\vec{y}(t)$ в последний день симуляции.
3. Создается эмулятор с нулевой функцией среднего и экспоненциальной корреляционной функцией гауссовского процесса:

$$c(\vec{q}, \vec{q}') = \sigma^2 \cdot \exp \left[- \sum_{i=1}^p \frac{(q_i - q'_i)^2}{\delta_i^2} \right].$$

4. Реализуется подбор параметров $\beta_i, \sigma, \delta_i$ эмулятора с базисными функциями $H\beta = \beta_0 + \beta_1 q + \beta_2 q^2 + \beta_3 q^3$ ($p = 4$) для сопоставления выходным данным симулятора (АОМ), которые были получены в пункте 2. Для этого решается обратная задача с помощью квазиньютоновского метода оптимизации L-BFGS-B [64].
5. Проводится валидация эмулятора следующим образом: генерируются новые точки $\widehat{q}_1, \dots, \widehat{q}_m$, запускается симулятор со значениями параметров в этих точках и затем сравнивается со выводом эмулятора.
6. Последним шагом в анализе является определение пространства идентифицируемости параметров, которое проводится с помощью метода сопоставления истории. Для этого генерируется большое количество наборов параметров ($n = 50000$), а затем выходные данные эмулятора на этих точках сравниваются с реальными наблюдениями для определения границ правдоподобности с учетом данных. Сопоставление истории включает в себя вычисление метрики неправдоподобности, которая устанавливает, насколько качественно конкретный набор входных данных описывает реальные наблюдения:

$$I(\vec{q}) = \frac{|f - \mathbb{E}(g(\vec{q}))|}{\sqrt{\text{Var}[f - \mathbb{E}(g(\vec{q}))]}}. \quad (36)$$

Здесь f – реальные наблюдения (статистические данные), \vec{q} – набор параметров, $g(\vec{q})$ – значения эмулятора в точке \vec{q} .

После вычисления метрики неправдоподобности пространство параметров разделяется на две части на основе порогового значения (по правилу 3σ [65]). Таким образом, нас интересуют точки с $I(\vec{q}) < 3$, которые образуют границы правдоподобного пространства.

7. На последнем шаге алгоритма полученные множества для всех выходных данных (количества выявленных, смертей и пациентов, находящихся в отделении интенсивной терапии) пересекаются.

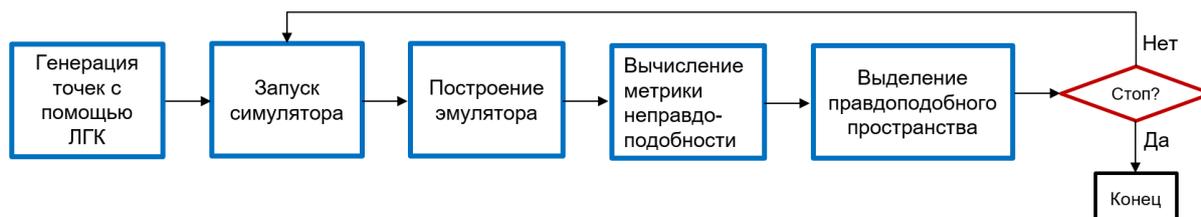


Рис. 9. Алгоритм построения пространства правдоподобности на основе байесовского подхода.

Анализ идентифицируемости для Новосибирской области

Была исследована идентифицируемость агентной модели с тремя неизвестными параметрами: параметр заразности α , начальное количество инфицированных $E(0)$ и параметр тестирования $p(y)$ с заданными границами (Таблица 3) по статистическим данным о количестве выявленных, умерших g_k и пациентов, находящихся в отделении интенсивной терапии C_k (измерения обратной задачи).

Таблица 3. Параметры АОМ распространения COVID-19 и их границы

Параметр	Неуточненные границы	Уточненные границы
α	[0.005; 0.05]	[0.005; 0.022]
$E(0)$	[1; 700]	[1; 700]
$p(y)$	[1; 500]	[1; 500]

Результат исследования представлен на рисунке (рис. 10). С помощью описанного выше подхода удалось уменьшить границы поиска параметра α более чем в 2 раза (на рисунках слева и посередине концентрация возможных значений параметра заразности смещена к одной границе), в то время как границы параметров $p(y)$ и $E(0)$ остались неизменными. Отметим, что был проведен анализ чувствительности АОМ в случае меньшего числа измерений: количество ежедневно выявленных и умерших случаев в результате COVID-19. Для такого набора данных границы параметра заразности α были уменьшены не так значительно, как в случае данных измерений по 3м эпидемиологическим состояниям.

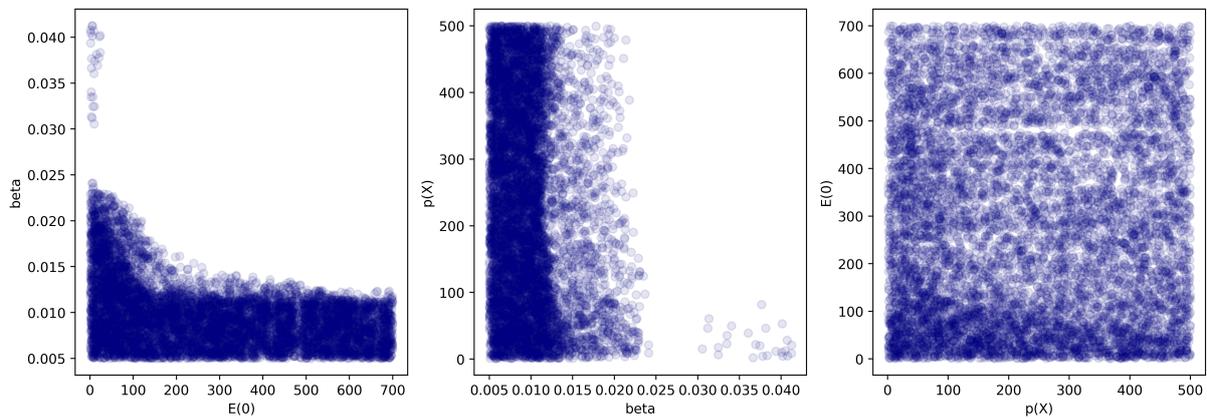


Рис. 10. Пространство правдоподобности $I(\vec{q})$ изменения параметра заразности α , начального количества инфицированных $E(0)$ и шанса быть протестированным $p(y)$ АОМ в зависимости от количества выявленных, критических и умерших случаев в результате COVID-19.

ЗАКЛЮЧЕНИЕ

В данной главе приведен краткий обзор и описание методов анализа идентифицируемости и чувствительности математических моделей биологии, описываемых системами дифференциальных уравнений (обыкновенные и стохастические дифференциальные уравнения, уравнения в частных производных) и агентных моделей. Методы анализа идентифицируемости можно разделить на анализ структурной и практической идентифицируемости. Показано, что для базовой SIR-модели необходимо проводить анализ структурной идентифицируемости для получения условий существования единственного решения в окрестности заданных параметров. И даже линейные модели (на примере модели фармакокинетики) могут быть структурно неидентифицируемы, в результате чего необходимо переформулировать задачу или задать дополнительные условия на параметры (априорную информацию). Показано, что понятие идентифицируемости тесно связано с понятием r -решения для систем алгебраических дифференциальных уравнений.

Модель может быть структурно идентифицируемой, но при этом быть практически неидентифицируемой по причине неполных и зашумленных данных. Анализ практической идентифицируемости позволяет оценить диапазон доверительных интервалов параметров моделей с помощью метода наименьших квадратов и Монте-Карло.

Важным этапом является анализ чувствительности параметров к вариации в данных, так как позволяет выявить наименее чувствительные параметры модели, а также сократить границы их изменения для применения локальных быстрых методов идентификации параметров по данным обратных задач. Для исследуемых моделей применялись ортогональный метод, метод собственных значений, анализ полуотносительной чувствительности, метод Соболя, метод eFAST и байесовский метод. В результате анализа чувствительности получены следующие выводы по моделям:

1. В модели SEIR-HCD (20) параметр β приобретения гуморального иммунитета является наименее чувствительным к измерениям количества выявленных, критических и умерших в результате COVID-19 случаев. Было принято решение зафиксировать этот параметр и в качестве априорной информации принять долю населения региона с наличием поздних антител IgG к вирусу SARS-CoV-2 (данные с сайта Инвитро).

2. Модели распространения коинфекции ТБ и ВИЧ с учетом МЛУ-ТБ (23) являются локально идентифицируемы в случае измерения количества больных ТБ, ТБ и ВИЧ, СПИД больных и инфицированных МЛУ-ТБ.
3. В пространственной SEIR-HCD-модели (25) показано, что влияние параметра длительности частичного иммунитета растет с увеличением времени моделирования. Границы изменения параметра скорости передачи инфекции удалось уменьшить в более чем 2 раза, среднее время инфицирования – до 6 дней, время пребывания в критическом состоянии – до 12 дней, а длительность частичного иммунитета может варьироваться от 2х до 4х месяцев.
4. Для модели игры «среднего поля» на основе SEIR-HCD (29)-(32) показано, что система наиболее чувствительна к вариации параметра a , характеризующего соблюдение популяцией изоляционных мероприятий. В свою очередь параметр a может быть определен независимо от других параметров из измерений. Стохастический параметр системы не оказывает влияния на модель при долгосрочном моделировании, а параметры, отвечающие за скорости распространения инфекции, влиятельны при вариации других подмножеств параметров (предположительно, параметра a).
5. Для агентно-ориентированной модели показано, что использование достоверных статистических данных позволяет уменьшить границы изменения чувствительного параметра. Так, в случае измерения количества выявленных, критических и умерших в результате COVID-19 случаев границы параметра заразности удалось уменьшить более чем в 2 раза.

Работа выполнена при поддержке Министерства науки и высшего образования Российской Федерации (госзадание) № 075-00337-20-03, номер проекта 0714-2020-0005.

СПИСОК ЛИТЕРАТУРЫ

1. Kabanikhin S.I. Definitions and examples of inverse and ill-posed problems. *Journal of Inverse and Ill-posed Problems*. 2008. V. 16. № 4. P. 317–357.
2. Авдеенко Т.В., Горский В.Г. Построение динамических моделей в пространстве состояний: Анализ структурной идентифицируемости: монография. Новосибирск: Изд-во НГТУ, 2007. 292 с. (Серия «Монографии НГТУ»).
3. Miao H., Xia X., Perelson A.S., Wu H. On identifiability of nonlinear ODE models and applications in viral dynamics. *SIAM Review*. 2011. V. 53. № 1. P. 3–39.
4. Hamelin F., Iggidr A., Rapaport A., Sallet G. Observability, Identifiability and Epidemiology. A survey. *Arxiv*. URL: <https://arxiv.org/abs/2011.12202> (data accessed: 07.10.2021).
5. Glover K., Willems J. Parametrization of linear dynamical systems: canonical forms and identifiability. *IEEE Trans on Automatic Control*. 1974. V. 19. P. 640–646.
6. Bellman R., Astrom K. On structural identifiability. *Math. Biosci.* 1970. V. 30. № 4. P. 65–74.
7. Grewal M., Glover K. Identifiability of linear and nonlinear dynamical systems. *IEEE Trans on Automatic Control*. 1976. V. 21. № 6. P. 833–837.
8. Cobelli C., DiStefano J. Parameter and structural identifiability concepts and ambiguities: a Critical review and analysis. *Amer. J. Physiology-Regulatory, Integrative and Comparative Physiology*. 1980. V. 3. P. 369–380.
9. Tunali T., Tarn T. New Results for Identifiability of Nonlinear Systems. *IEEE Transactions on Automatic Control*. 1987. V. 15. P. 45–51.
10. Vajda S. Identifiability of first order reaction systems. *Reaction Kinetics and Catalysis Letters*. 1979. V. 11. № 1. P. 39–43.
11. Audoly S., D'angiò L. On the identifiability of linear compartmental system: a revisited transfer function approach based on topological properties. *Mathematical Biosciences*. 1983. V. 66. No. 2. P. 201–208.
12. Щербак В.Ф. Условия идентифицируемости динамических систем. Математическая физика. *IEEE Transactions on Automatic Control*. 1983. V. 34. P. 105–108.
13. Леваков А.А. Идентифицируемость нелинейных систем. *Дифференциальные уравнения*. 1983. Т. 19. № 6. С. 1074–1078.
14. Карелин В.В. Алгоритм для оценки вектора параметров линейных динамических систем с дискретно-измеряемыми функциями. *Вопросы механики и процессов управления*. 1982. Т. 359. № 5.
15. Авдеенко Т.В., Каргин С.А. Анализ идентифицируемости линейных динамических моделей с использованием сепараторов параметрического пространства. *Сиб. журн. индустр. матем.* 2006. Т. 9. № 3. С. 3–16.
16. Ломов А.А. Совместная идентифицируемость коэффициентов линейных разностных уравнений объекта и аддитивных возмущений. *Вестник НГУ. Серия: математика, механика, информатика*. 2015. Т. 15. № 4. С. 46–57.
17. Якименко А.А. К вопросу об идентифицируемости систем одновременных уравнений. *Труды БГТУ*. 2022. Т. 3. № 2. С. 10–13.
18. Saccomani M., Cobeli C. Qualitative Experiment Design in Physiological System Identification. *IEEE Control System*. 1992. V. 3. № 12. P. 18–23.
19. Brown R. Identifiability: role in design of pharmacokinetic experiments. *IEEE Transactions on Biomedical Engineering*. 1982. V. 14. № 3. P. 31–41.
20. Saccomani M., Audoly S., Angio L. An Effective Automatic Procedure for Testing Parameter Identifiability of HIV/AIDS Models. *Bull Math Biol*. 1978. V. 73. P. 1734–1753.
21. Bellu G., Saccomani M., Audoly S. Comput Methods Programs. *Biomed. Mathematical Biosciences*. 2007. V. 5. P. 67–75.

22. Meshkat N., Anderson C., III J.D. Alternative to Ritt's Pseudodivision for finding the input-output equations of multi-output models. *Mathematical Biosciences*. 2012. V. 239. № 1. P. 117–123.
23. Meshkat N., Er-zhen Kuo C., DiStefano J. Finding and Using Identifiable Parameter Combinations in Nonlinear Dynamic Systems Biology Models and COMBOS: A Novel Web Implementation. *Plos One*. 2014. V. 9. № 10. Article No. e110261.
24. Carson E., Cobelli C. *Modelling Methodology for Physiology and Medicine*. New-York: Academic Press, 2001.
25. Carson E., Cobelli C. *Introduction to Modelling in Physiology and Medicine*. New-York: Academic Press, 2008.
26. Raue A., Karlsson J., Saccomani M.P., Jirstrand M., Timmer J. Comparison of approaches for parameter identifiability analysis of biological systems. *Bioinformatics*. 2014. V. 30. № 10. P. 1440–1448.
27. Кабанихин С.И., Воронов Д.А., Гродзь А.А., Криворотько О.И. Идентифицируемость математических моделей медицинской биологии. *Вавиловский журнал генетики и селекции*. 2015. Т. 19. № 6. С. 738–744.
28. Villaverde A.F., Barreiroc A. Identifiability of large nonlinear biochemical networks. *MATCH Commun. Math. Comput. Chem*. 2016. V. 76. P. 259–296.
29. Saltelli A., Chan K., Scott M. *Sensitivity Analysis*. Wiley, Hoboken, 2000.
30. Raue A., Kreutz C., Maiwald T. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*. 2009. V. 25. № 15. P. 1923–1929.
31. Adams B.M., Banks H.T., Davidian M., Hee-Dae Kwon, Tran H.T., Wynne S.N., Rosenberg E.S. On Hiv dynamics: modeling, data analysis, and optimal treatment protocols. *Journal of Computational and Applied Mathematics*. 2005. V. 184. № 1. P. 10–49.
32. Bellu G., Saccomani M.P., Audoly S., D'Angiò L. DAISY: A new software tool to test global identifiability of biological and physiological systems. *Computer Methods and Programs in Biomedicine*. 2007. V. 88. № 1. P. 52–61.
33. Криворотько О.И., Андорная Д.В., Кабанихин С.И. Анализ чувствительности и практическая идентифицируемость математических моделей биологии. *Сибирский журнал индустриальной математики*. 2020. Т. 23. № 1. С. 107–125.
34. Raue A., Becker V., Klingmüller U., Timmer J. Identifiability and observability analysis for experimental design in nonlinear dynamical models. *Chaos*. 2010. V. 20. № 4. P. 045105.
35. Kermack W.O., McKendrick A.G. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*. 1927. V. 115. № 772. P. 700–721.
36. Magal P., Webb G. The parameter identification problem for SIR epidemic models: identifying unreported cases. *Journal of Mathematical Biology*. 2018. V. 77. № 6–7. P. 1629–1648.
37. Evans N.D., White L.J., Chapman M.J., Godfrey K.R., Chappell M.J. The structural identifiability of the susceptible infected recovered model with seasonal forcing. *Mathematical Biosciences*. 2005. V. 194. № 2. P. 175–197.
38. Diop S., Fliess M. Nonlinear observability, identifiability, and persistent trajectories. In: *[1991] Proceedings of the 30th IEEE Conference on Decision and Control*. 1991. V. 1. P. 714–719.
39. Kolchin E.R. *Differential Algebra, Algebraic Groups*. Cambridge: Academic Press, 1973.
40. Ljung L., Glad T. On global identifiability for arbitrary model parametrizations. *Automatica*. 1994. V. 30. № 2. P. 265–276.
41. Годунов С.К., Антонов А.Г., Кирилюк О.П., Костин В.И. *Гарантированная*

- точность решения систем линейных уравнений в евклидовых пространствах. Изд. 2. Новосибирск: Наука, 1992.
42. Чеверда В.А., Костин В.И. r -псевдообратный для компактного оператора. *Сибирские электронные математические известия*. 2010. Т. 7. С. 258–282.
 43. Banks H.T., Kabanikhin S.I., Krivorotko O.I., Yermolenko D.V. A numerical algorithm for constructing an individual mathematical model of HIV dynamics at cellular level. *Journal of Inverse and Ill-posed Problems*. 2018. V. 26. № 6. P. 859–873.
 44. Kabanikhin S.I., Krivorotko O.I., Ermolenko D.V., Kashtanova V.N., Latyshenko V.A. Inverse problems of immunology and epidemiology. *Eurasian Journal of Mathematical and Computer Applications*. 2017. V. 5. № 2. P. 14–35.
 45. Wieland F.-G., Hauber A.L., Rosenblatt M., Tönsing C., Timmer J. On structural and practical identifiability. *Current Opinion in Systems Biology*. 2021. V. 25. P. 60–69.
 46. Rodriguez-Fernandez M., Egea J.A., Banga J. R. Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems. *BMC Bioinformatics*. 2006. V. 7. № 483.
 47. Metropolis N., Ulam S. The Monte Carlo method. *Journal of the American Statistical Association*. 1949. V. 44. № 247. P. 335–341.
 48. Cacuci D.G. *Sensitivity and uncertainty analysis: theory*. New York: Chapman & Hall/CRC, 2003. V. 1. 304 p.
 49. Latyshenko V., Krivorotko O.I., Kabanikhin S.I., Zhang S., Kashtanova V.N., Yermolenko D.V. Identifiability analysis of inverse problems in biology. In: *Proceedings of the 2nd International Conference on Computational Modeling, Simulation and Applied Mathematics (CMSAM2017)*. 2017. P. 567–571.
 50. Quaiser T., Mönnigmann M. Systematic identifiability testing for unambiguous mechanistic modeling – application to JAK-STAT, MAP kinase, and NF- κ B signaling pathway models. *BMC Systems Biology*. 2009. V. 3. № 1.
 51. Cukier R.I., Fortuin C.M., Schuler K.E., Petschek A.G., Schaibly J.H. Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. *The Journal of Chemical Physics*. 1973. V. 59. P. 3873–3878.
 52. Соболев И.М. Об оценке чувствительности нелинейных математических моделей. *Математическое моделирование*. 1990. Т. 2. № 1. С. 112–118.
 53. Saltelli A., Annoni P., Azzini I., Campolongo F., Ratto M., Tarantola S. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*. 2010. V. 180. P. 259–270.
 54. Unlu E., Leger H., Motornyi O., Rukubayihunga A., Ishacian T., Chouiten M. Epidemic analysis of COVID-19 Outbreak and Counter-Measures in France. *Medrxiv*. URL: <https://www.medrxiv.org/content/10.1101/2020.04.27.20079962v1> (data accessed: 01.05.2020).
 55. Криворотько О.И., Кабанихин С.И., Зятков Н.Ю., Приходько А.Ю., Прохошин Н.М., Шишленин М.А. Математическое моделирование и прогнозирование COVID-19 в Москве и Новосибирской области. *Сибирский журнал вычислительной математики*. 2020. Т. 23. № 4. С. 395–414.
 56. Krivorotko O.I., Zyatkov N.Y. Data-driven regularization of inverse problem for SEIR-HCD model of COVID-19 propagation in Novosibirsk region. *Eurasian Journal of Mathematical and Computer Applications*. 2022. V. 10. № 1. P. 51–68.
 57. Roeger L.I., Feng Z., Castillo-Chavez C. Modeling TB and HIV co-infections. *Mathematical Biosciences and Engineering*. 2009. V. 6. № 4. P. 815–837.
 58. Колмогоров А.Н., Петровский И.Г., Пискунов Н.С. Исследование уравнения диффузии, соединенной с возрастанием количества вещества, и его применение

- к одной биологической проблеме. *Бюллетень Московского государственного университета*. 1937. Т. 1. № 6. С. 1–26.
59. Andrianakis I., Vernon I.R., McCreesh N., McKinley T.J., Oakley J.E., Nsubuga R.N., Goldstein M., White R.G. Bayesian history matching of complex infectious disease models using emulation: a tutorial and a case study on HIV in Uganda. *PLOS Computational Biology*. 2015. V. 11. № 1. Article No. e1003968.
 60. Raue A., Karlsson J., Saccomani M.P., Jirstrand M., Timmer J. Comparison of approaches for parameter identifiability analysis of biological systems. *Bioinformatics*. 2014. V. 30. № 10. P. 1440–1448.
 61. Криворотко О.И., Кабанихин С.И., Сосновская М.И., Андорная Д.В. Анализ чувствительности и практическая идентифицируемость математических моделей биологии. *Вавиловский журнал генетики и селекции*. 2021. Т. 25. № 1. С. 82–91.
 62. Krivorotko O., Sosnovskaia M., Vashchenko I., Kerr C., Lesnic D. Agent-based modeling of COVID-19 outbreaks for New York state and UK: Parameter identification algorithm. *Infect. Dis. Model.* 2022. V. 7. № 1. P. 30–44.
 63. McKay M.D., Beckman R.J., Conover W.J. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*. 1979. V. 21. № 2. P. 239–254.
 64. Malouf R. A comparison of algorithms for maximum entropy parameter estimation. In: *Proceedings of the Sixth Conference on Natural Language Learning*. 2002. V. 20. P. 49–55.
 65. Pukelsheim F. The three sigma rule. *The American Statistician*. 1994. V. 48. № 2. P. 88–91.
 66. Lee W., Liu S., Tembine H., Li W., Osher S. Controlling Propagation of epidemics via mean-field control. *SIAM J. Appl. Math.* 2021. V. 81. № 1. P. 190–207.
 67. Petrakova V., Krivorotko O. Mean field game for modeling of covid-19 spread. *Journal of Mathematical Analysis and Applications*. 2022. V. 514. № 1. Article No. 126271.
 68. Petrakova V., Shaydurov V. MFG SIRV-D Model with Managed Rates of Epidemic Spread. (in print).
 69. Bensoussan A., Frehse J., Yam Ph. *Mean Field Games and Mean Field Type Control Theory*. New York: Springer, 2013. 128 p.
 70. Saltelli A., Tarantola S., Chan K.-S. A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics*. 1999. V. 41. P. 39–56.
 71. Cukier R.I., Fortuin C.M., Shuler K.E., Petschek A.G., Schaibly J.H. Study of the Sensitivity of Coupled Reaction Systems to Uncertainties in Rate Coefficients. I. Theory. *The Journal of Chemical Physics*. 1973. V. 59. P. 3873–3878.
 72. Saltelli A., Tarantola S., Campolongo F. Sensitivity Analysis as an Ingredient of Modeling. *Statistical Science*. 2000. V. 15. № 4. P. 377–395.
 73. Schaibly J.H., Shuler K.E. Study of the Sensitivity of Coupled Reaction Systems to Uncertainties in Rate Coefficients. Part II, Applications. *Journal of Chemical Physics*. 1973. V. 59. P. 3879–3888.
 74. Frey H., Patil S. Identification and Review of Sensitivity Analysis Methods. *Risk Analysis*. 2002. V. 22. № 3. P. 553–578.
 75. Tembine H. COVID-19: Data-Driven Mean-Field-Type Game Perspective. *Games*. 2020. V. 11. № 151. P. 107.

Рукопись поступила в редакцию 18.01.2023.

Переработанный вариант поступил 01.05.2023.

Дата опубликования 20.06.2023.

The identifiability of mathematical models in epidemiology: Tuberculosis, HIV, COVID-19

Olga Krivorotko^{*1,2}, Sergey Kabanikhin^{†1}, Victoriya Petrakova^{‡3}

¹*Sobolev Institute of Mathematics of SB RAS, Novosibirsk, Russia*

²*The Moscow Institute of Physics and Technology, Dolgoprudny, Russia*

³*Institute of Computational Modeling of SB RAS, Krasnoyarsk, Russia*

Abstract. The paper presents a short review of sensitivity-based identifiability approaches for analyzing mathematical models of epidemiology and related processes described by systems of differential equations and agent-based models. It is shown that for structural identifiability of basic SIR models (describe the dynamic of Susceptible, Infected and Removed groups based on nonlinear ordinary differential equations) of epidemic spread and linear compartmental models it is possible to use a priori information about the process. It is demonstrated that a model can be structurally identifiable but be practically non-identifiable due to incomplete data. The paper uses methods for analyzing the sensitivity of parameters to data variation, as well as analyzing the sensitivity of model states to parameter variation, based on linear and differential algebra, Bayesian, and Monte Carlo approaches. It was shown that in the SEIR-HCD model of COVID-19 propagation, described by a system of seven ordinary differential equations and based on the mass balance law, the parameter of humoral immunity acquisition is the least sensitive to changes in the number of diagnosed, critical and mortality cases of COVID-19. The spatial SEIR-HCD model of COVID-19 propagation demonstrated an increase the sensitivity of the partial immunity duration parameter over time, as well as a decrease in the limits of change in the infectivity and infection parameters. In the case of the SEIR-HCD mean-field model of COVID-19 propagation, the sensitivity of the system to the self-isolation index and the lack of sensitivity of the stochastic parameters of the system are shown. In the case of the agent-based COVID-19 propagation model, the change in the infectivity parameter was reduced by more than a factor of 2 compared to the statistics. A differential model of co-infection HIV and tuberculosis spread with multiple drug resistance was developed and its local identifiability was shown.

Key words: *epidemiology, SIR, sensitivity-based identifiability analysis, identifiability, a priori information, Bayesian methods, Sobol method.*