

СИБИРСКИЕ ЭЛЕКТРОННЫЕ
МАТЕМАТИЧЕСКИЕ ИЗВЕСТИЯ

Siberian Electronic Mathematical Reports

<http://semr.math.nsc.ru>

Том 16, стр. 144–144 (2019)

DOI 10.33048/semi.2019.16.xxx

УДК 519.688

MSC 65M32

МАТЕМАТИЧЕСКОМ МОДЕЛИРОВАНИИ COVID-19

О.И. КРИВОРОТЬКО, С.И. КАБАНИХИН

АБСТРАКТ. The mathematical models for analysis and forecasting of COVID-19 pandemic based on time-series models, differential equations (SIR models based on ordinary, partial and stochastic differential equations), agent-based models, mean field games and its combinations are considered. Inverse problems for mathematical models in epidemiology of COVID-19 are formulated in the variational form. The numerical results of modeling and scenarios of COVID-19 propagation in Novosibirsk region are demonstrated and discussed. The epidemiology parameters of COVID-19 propagation in Novosibirsk region (contagiosity, hospitalization and mortality rates, asymptomatic cases) are identified. The combination of differential and agent-based models increases the quality of forecast scenarios.

Keywords: epidemiology, COVID-19, time-series models, SIR, agent-based models, mean field games, inverse problems, forecasting.

1. ВВЕДЕНИЕ

В работе изложены математические основы программного комплекса COVID-19, созданного в СО РАН совместно с коллегами из РФЯЦ-ВНИИТФ им. академ. Е.И. Забабахина (Снежинск), ФИЦ КНТ (Красноярск), МФТИ и МГУ (Москва), а также коллегами из Болгарии, Великобритании, Казахстана, Китая, США. В основе комплекса программ лежит комбинация трех основных типов моделей: SIR, АОМ и ИСП.

В силу новизны и сложности заболевания COVID-19 параметры большинства математических моделей, как правило, неизвестны, и это приводит к необходимости рассматривать и решать обратные задачи. Основные проблемы моделирования распространения COVID-19:

KRIVOROTKO, O.I., KABANIKHIN S.I. ON MATHEMATICAL MODELS OF COVID-19 PANDEMIC.

© 2022 Криворотько О.И., Кабанихин С.И.

Работа поддержана РНФ (грант 18-71-10044-П).

Поступила 1 декабря 2022 г., опубликована 31 декабря 2022 г.

1. Данные для решения обратной задачи являются неполными и зашумленными, а также представляют собой большие данные (ежедневные сводки о заболевших, заразившихся, вакцинированных и т.д.).
2. Параметры меняются со временем: контагиозность $\alpha(t)$, вероятность появления тяжелых случаев $1 - \beta(t)$, смертность $\mu(t)$, процент бессимптомных случаев $\alpha_E(t)$ и т.д.
3. Процесс распространения COVID-19 существенно изменяется при введении или отмене ограничительных мер (ношение масок, социальная дистанция, перевод на удаленный рабочий режим, закрытие школ, предприятий, районов и городов), появление каждого нового штамма (альфа, бета, гамма, дельта, ...).

С учетом 1.-3. решение обратных задач требует новых подходов и методов, включая байесовский, стохастическую оптимизацию, природоподобные алгоритмы (генетический, имитации отжига, глубокие нейронные сети) и другие методы машинного обучения и искусственного интеллекта. В работе будет приведен краткий обзор и анализ работ по указанным новым подходам.

Статья организована следующим образом. В разделе 2 изложены методы анализа временных рядов, которые составляют основу статистических данных в эпидемиологии, а именно: статистические методы (подраздел 2.2), методы на основе машинного обучения (подраздел 2.3) и на базе фильтрации (подраздел 2.4). В разделе 3 приведен обзор математических моделей моделирования вспышки COVID-19, в основе которых лежат системы обыкновенных дифференциальных уравнений (подраздел 3.2) и уравнений в частных производных (подраздел 3.3). Ключевой характеристикой описания распространения эпидемии является базовый индекс репродукции вируса \mathcal{R}_0 , показывающий количество возможных инфицированных в результате контакта одного инфицированного с восприимчивой популяцией. Алгоритм вывода \mathcal{R}_0 в общем случае, а также для SEIR-HCD модели, представлен в разделе 3.5. В разделе 4 приведено описание имитационных моделей, а именно: в Разделе 4.2 описаны математические модели, основанные на теории клеточных автоматов, в Разделе 4.3 – агентно-ориентированные модели (АОМ). Подробное построение агентной модели для описания распространения COVID-19 приведено в подразделе 4.3.1. Раздел 5 посвящен моделям игр среднего поля (ИСП) с управляющим уравнением первого порядка. В разделе 5.2 описано уравнение Колмогорова-Фоккера-Планка (КФП), определяющее распределение агентов в моделях эпидемиологии, а в разделе 5.3 – уравнение Гамильтона-Якоби-Беллмана (ГЯБ), отвечающее за оптимальное управление. В разделе 6 приведены численные алгоритмы решения прямых (подраздел 6.1) и обратных (подраздел 6.2) задач эпидемиологии. Результаты численных расчетов проанализированы в разделе 7.2. Также в подразделе 7.4 сформулированы выводы и в 7.5 – направления дальнейших исследований.

В приложении А изложено описание эпидемиологических данных распространения COVID-19 в Новосибирской области, используемые в численных расчетах. В приложении В приведено описание комплекса программ и характеристики численных расчетов.



Рис. 1. Диаграмма моделей прогнозирования временных рядов.

В качестве иллюстрации в работе приведены результаты расчетов по трем моделям: SEIR-HCD – камерная модель, основанная на системе из 7 обыкновенных дифференциальных уравнений (Раздел 3.5.2), агентная (Раздел 4.3.1) и модель игры среднего поля (Раздел 5).

2. МОДЕЛИ, ОСНОВАННЫЕ НА АНАЛИЗЕ ВРЕМЕННЫХ РЯДОВ

В разделе будут рассмотрены регрессионные и сетевые модели, методы фильтрации и их взаимосвязи [1], на основе анализа наиболее достоверных статистических данных – количество ежедневных ПЦР-тестов $\mathcal{T}(t)$ и индекс самоизоляции $a(t)$. Время t во всех моделях измеряется в днях.

2.1. Введение. Особенностью функций $\mathcal{T}(t)$ и $a(t)$ является повторяемость подъемов и спадов по времени с изменяющейся амплитудой. Например, $\mathcal{T}(t)$ возрастает по вторникам и снижается к понедельнику, а $a(t)$ снижается в период выходных, праздников и отпусков.

Методы математической статистики и машинного обучения помогают обрабатывать, анализировать эпидемиологические данные и проводить краткосрочное прогнозирование поведения $\mathcal{T}(t)$ и $a(t)$ при отсутствии резких изменений ситуации (введение ограничительных мер, мутации вируса) [2]. В разделе будут приведены регрессионные, сетевые модели, а также методы фильтрации и их взаимосвязи, диаграмма которых приведена на рис. 1.

Отметим, что в данном разделе мы рассматриваем модели, которые используют только значения наблюдаемых показателей $\mathcal{T}(t)$ и $a(t)$.

2.2. Регрессионные модели. Регрессионные модели подразделяются на неадаптивные модели, для оценки параметров которых используются все имеющиеся данные, и адаптивные, значения параметров которых рассчитываются на основе скользящего окна наблюдений [3].

Вид регрессионной зависимости для неадаптивной модели выбирается исходя из свойств анализируемого временного ряда. В отдельных случаях можно применять полиномиальные или степенные функции, но чаще модель должна учитывать сезонный характер заболеваемости (например, модель Серфлинга [4]). Неадаптивные регрессионные модели учитывают всю предысторию заболеваемости, но игнорируют локальные колебания эпидемических показателей, поэтому появление нового штамма вируса и введение ограничительных

мер в регионе снижают значимость данных предыдущего периода (устаревшие данные).

Адаптивные регрессионные модели используют ограниченный отрезок временного ряда и поэтому более чувствительны к изменению ситуации. Важную роль при использовании адаптивных моделей играет ширина скользящего окна (в пределах нескольких месяцев) – число последних наблюдений, на основе которых оценивают параметры модели.

В регрессионных моделях предполагается, что невязка модели (ошибка предсказания) – независимая случайная величина, имеющая нормальный закон распределения с нулевым математическим ожиданием и постоянной дисперсией [1]. Одной из проблем прогнозирования является наличие существенной автокорреляции невязок. Регрессионная модель может быть дополнена, а прогноз – уточнен, например, с помощью авторегрессионных моделей.

2.2.1. *Авторегрессионная модель прогнозирования временного ряда.* Для построения прогнозов сезонных рядов, например, количества проведенных ПЦР-тестов в регионе, использовалась авторегрессионная модель SARIMA являющейся модификацией модели ARIMA (AutoRegressive Integrated Moving Average) [5], которая описывает одномерные временные ряды с сезонной компонентой [6]. ARIMA является расширением моделей типа ARMA для нестационарных временных рядов, которые можно сделать стационарными взятием разностей некоторого порядка от исходного временного ряда (так называемые интегрированные или разностно-стационарные временные ряды).

Модель $ARIMA(p, d, q)$ для нестационарного временного ряда $\mathcal{T}(n)$ имеет вид:

$$\Delta^d \mathcal{T}(n) = c + \sum_{i=1}^p a_i \Delta^d \mathcal{T}(n-i) + \sum_{j=1}^q b_j \varepsilon(n-j) + \varepsilon(n).$$

Здесь $\varepsilon(n)$ – стационарный временной ряд белого шума, c, a_i, b_j – параметры модели, Δ^d – оператор разности временного ряда порядка d , гарантирующий стационарность ряда (последовательное взятие d раз разностей первого порядка – сначала от временного ряда, затем от полученных разностей первого порядка, затем от второго порядка и т.д.). Обычно, при построении модели ARIMA порядок разностей ограничивается числом $d = 2$.

В нашем же случае при использовании сезонности ряда строится модель $SARIMA(p, d, q)(P, D, Q)_s$ для нестационарного временного ряда $\mathcal{T}(n)$ [6]:

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

Здесь параметры p, d, q отвечают за несезонную часть временного ряда, а P, D, Q соответствуют сезонным компонентам ряда, s – длина сезона, Δ_s^D – оператор разности временного ряда порядка D для сезонной компоненты, φ и Φ – параметры авторегрессии для несезонных и сезонных компонент ряда, θ и Θ – параметры несезонного и сезонного скользящего среднего соответственно, L – оператор сдвига, n – временной параметр (день). В общем случае выражения

имеют вид:

$$\begin{aligned}\varphi(L) &= 1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_p L^p, \\ \Phi(L^s) &= 1 - \Phi_1 L^s - \Phi_2 L^{2s} - \dots - \Phi_P L^{Ps}, \\ \theta(L) &= 1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q, \\ \Theta(L^s) &= 1 + \Theta_1 L^s + \Theta_2 L^{2s} + \dots + \Theta_Q L^{Qs}, \\ \Delta^d &= (1 - L)^d, \quad \Delta_s^D = (1 - L^s)^D, \\ L^k \mathcal{T}(n) &= \mathcal{T}_{n-k}.\end{aligned}$$

Алгоритм прогнозирования временного ряда следующий:

- 1) Применяем преобразование Бокса-Кокса [7] для уменьшения дисперсии.
- 2) Вычисляем сезонную разность (сдвиг на s дней) первого порядка.
- 3) Вычисляем вторую разность (сдвиг на 1 день) ряда, полученного в пункте 2.
- 4) Проверяем стационарность ряда из п. 3 критерием Дики-Фуллера.
- 5) Передаем соответствующие проделанным действиям параметры в модель $ARIMA(p, d, q)$ и подбираем остальные на основе минимизации информационного критерия Акаике. В качестве данных передается ряд из п. 1.
- 6) Полученная модель с настроенными гиперпараметрами используется для дальнейшего прогнозирования. Полученный результат подвергается обратному преобразованию Бокса-Кокса.

2.3. Модели на основе машинного обучения. Машинное обучение является мощным инструментом для поиска взаимосвязи между входными и выходными данными в случаях, когда аналитическое исследование затруднительно. Применение эвристических подходов для раннего обнаружения эпидемиологических рисков в некоторых случаях позволяет улучшить качество прогнозирования.

Методы случайный лес (Random Forest) и градиентный бустинг (Gradient Boosting) являются одними из самых эффективных и продвинутых алгоритмов машинного обучения для решения задач классификации и регрессии, особенно при работе с временными рядами [9, 10]. Данные методы являются глобальными моделями, что позволяет заимствовать статистическую силу у других, связанных временных рядов посредством обмена параметрами. Кроме того, они справляются с разреженностью целевых значений гораздо лучше, чем другие методы, требуют большого количества наблюдений (ближайшее сходство) в тренировочной выборке данных и работают корректно в случае, когда количество признаков меньше количества наблюдений в обучающих данных.

Одним из представителей моделей машинного обучения являются динамические байесовские сети – ориентированный граф, вершины которого соответствуют переменным модели, а ребра – вероятностным зависимостям между ними, которые заданы определенными законами распределения [11]. После обучения как на большом, так и на малом количестве исходных данных байесовские сети позволяют оценить вероятность наступления некоторого события при наблюдаемой последовательности явлений. Для прогнозирования заболеваемости используется простая форма скрытых марковских моделей, основной

идеей которых является сопоставление каждой случайной величины Y_t (например, количество выявленных, госпитализированных случаев с COVID-19) с ненаблюдаемой случайной величиной S_t (например, общее количество инфицированных индивидуумов), определяющей условное распределение Y_t [12]. Таким образом, величина Y_t зависит только от значения скрытой переменной S_t в момент времени t , а последовательность S_t обладает марковским свойством, то есть величина S_t зависит только от S_{t-1} (рис. 2а).

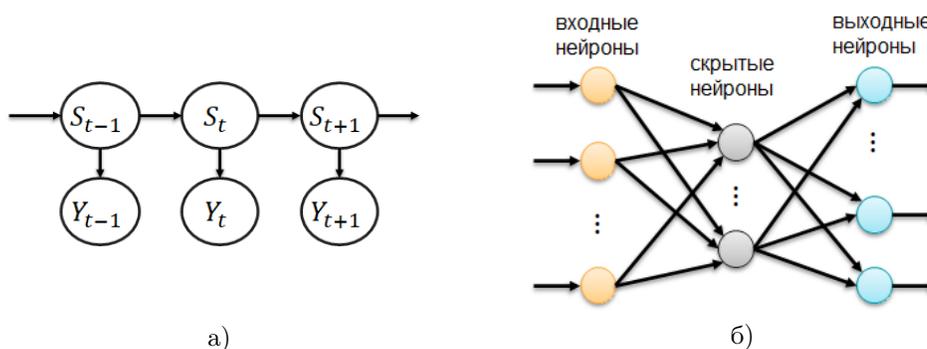


Рис. 2. Схема зависимостей в скрытой марковской модели (а) и нейронной сети с одним скрытым слоем (б).

Нейронные сети (НС) представляют собой направленный взвешенный граф, вершины которого моделируют функционирование биологических нейронов (рис. 2б). Обучение НС заключается в вычислении коэффициентов связей между вершинами, определяющих силу входящих сигналов, и выполняется на основе эмпирических данных: статистики заболеваемости и при наличии, значений факторов, ее предопределяющих. Для корректного обучения НС необходим большой объем исторических данных. В исследовании [13] рассматриваются возможности применения нейронных сетей для прогнозирования распространения COVID-19. Результаты работы нейронной сети прогнозирования количества ожидаемых выявленных случаев заболевания в следующий день для некоторых регионов достигали 87% (точность зависела от качества и полноты данных в начале пандемии).

2.4. Модели на основе фильтрации. Любые временные ряды заболеваемости можно рассматривать как случайный процесс, состоящий из сигнала, отражающего реальную эпидемическую обстановку, и шума. Фильтрация шума позволяет уточнить прогноз и может выполняться как в ходе предварительной обработки исходных данных, так и в составе самого алгоритма прогнозирования.

Одним из подходов является вейвлет-декомпозиция, в которой временной ряд представляется с помощью вейвлет-функций [14]. Однако, такой подход используется совместно с другими моделями.

Одной из таких моделей является экспоненциальное сглаживание, представляющее собой частный случай взвешенного скользящего среднего, а именно

значение заболеваемости y_t в момент времени t описывается взвешенной суммой последних наблюдений: $l_t = by_t + (1 - b)y_{t-1}$. Здесь $b \in (0, 1)$ – коэффициент сглаживания, который обеспечивает уменьшение веса по мере старения данных, которое может рассматриваться как отражение естественного процесса обучения. Такой метод построения модели не подходит для рядов, в поведении которых присутствуют отчетливый тренд или сезонность. Для этих целей используются обобщенные модели [15], например, сезонная модель Хольта-Винтерса (L – длина сезона) [8]

$$\begin{aligned}\ell_x &= \alpha(y_x - s_{x-L}) + (1 - \alpha)(\ell_{x-1} + b_{x-1}), \\ b_x &= \beta(\ell_x - \ell_{x-1}) + (1 - \beta)b_{x-1}, \\ s_x &= \gamma(y_x - \ell_x) + (1 - \gamma)s_{x-L}, \\ \hat{y}_{x+m} &= \ell_x + mb_x + s_{x-L+1+(m-1)\text{mod}L}.\end{aligned}$$

Параметры α , β , γ являются параметрами уровня, тренда и сезонности, соответственно. Результаты прогнозирования ряда ежедневных ПЦР-тестов в Новосибирской области в рамках данной модели представлены на рис. 11 в Разделе 7.1.

Любые эпидемические процессы можно описать следующей системой разностных уравнений:

$$\begin{aligned}\mathbf{x}_t &= \mathbf{A}\mathbf{x}_{t-1} + \mathbf{w}_t, \\ \mathbf{y}_t &= \mathbf{H}\mathbf{x}_t + \mathbf{D}\mathbf{f}_t + \mathbf{v}_t,\end{aligned}$$

где \mathbf{x}_t – вектор переменных состояний системы в момент времени t , \mathbf{y}_t – вектор наблюдений, \mathbf{f}_t – вектор значений внешних факторов, \mathbf{w}_t и \mathbf{v}_t – белый шум. Матрицы параметров \mathbf{A} , \mathbf{H} , \mathbf{D} определяют модель эпидемического процесса и выбираются исходя из решаемой задачи – краткосрочного или долгосрочного прогнозирования.

Такая форма записи позволяет предложить обобщенные модели распространения заболевания, в частности, модели на основе калмановской фильтрации [16].

2.5. Выводы. Методы машинного обучения и регрессионные модели являются бурноразвивающимся направлением для анализа и прогнозирования временных рядов, в том числе и в эпидемиологии. На Рис. 1 представлена лишь малая часть существующих методов прогнозирования временных рядов. Важным этапом в построении таких моделей является их комбинация друг с другом (например, ARIMA + модель Хольта-Винтерса + градиентный бустинг) с целью уменьшения ошибки прогнозирования изучаемых процессов. Отметим хорошую обзорную работу по моделям временных рядов [17], а также сборник работ в журнале *International Journal of Forecasting*, V. 38(2), 2022, 407–704, посвященных построению и анализу моделей временных рядов, описывающих распространение новой коронавирусной инфекции.

В данном обзоре мы не заостряли внимание на построении комбинированной модели анализа временных рядов, а сосредоточились на моделях дифференциальной и агентной структуры, о которых речь пойдет далее. Результаты, полученные на основе временных рядов, использованы нами в дифференциальных и агентных моделях. Прогнозирование ежедневных ПЦР-тестов и индекса самоизоляции позволяют строить сценарии развития COVID-19 в регионе в зависимости от введения ограничительных мер, а именно количество ожидаемых выявленных, умерших, госпитализированных, критических случаев и индекса

репродукции вируса (экстраполяция ряда количества ПЦР-тестов приведена в Разделе 7.1, а численные расчеты с использованием экстраполированного временного ряда приведены в Разделе 7.2).

3. МОДЕЛИ, ОСНОВАННЫЕ НА ДИФФЕРЕНЦИАЛЬНЫХ УРАВНЕНИЯХ

В данном разделе приведен обзор математических моделей, основанных на обыкновенных дифференциальных уравнениях (ОДУ), приведенных в Разделе 3.2, на уравнениях в частных производных (УРЧП), описанных в Разделе 3.3, и на стохастических дифференциальных уравнениях (СДУ), приведенных в Разделах 3.4 и 5, а также описание ключевой характеристики распространения заболевания – базовый индекс репродукции вируса (п. 3.5). В качестве примера будет представлена SEIR-HCD модель, описанная в п. 3.5.2.

3.1. Введение. Дифференциальные модели основаны на законе действующих масс и особенностях передачи инфекции. Закон действующих масс заключается в следующем: в замкнутой системе количество агентов постоянно во времени. Диаграмма 3 иллюстрирует развитие моделей эпидемиологии с 1760 по настоящее время.

Для описания распространения новой коронавирусной инфекции, вызванной вирусом штамма SARS-CoV-2 были использованы все основные достижения, отраженные в диаграмме 3. На рис. 4 приведена классификация существующих математических моделей динамики COVID-19 и их взаимосвязи. Одна из простейших моделей вспышки COVID-19 описывается логистическим уравнением (2). Добавление групп (бессимптомные, госпитализированные, критические, умершие случаи и т.д.) к логистическому уравнению приводит к камерным SIR-моделям (1) и (8). Учет пространственной неоднородности приводит к диффузионно-логистическому уравнению (3). Дальнейшие преобразования камерных моделей можно условно разделить на два направления: учет пространственной структуры в непрерывной (модель реакции-диффузии) и дискретной (модель конечных автоматов) постановках. Добавление управления системами состояний формирует новый класс моделей ИСП (15)-(16), а учет индивидуальных характеристик агентов – к АОМ. Отметим, что усреднение в АОМ по характеристикам агентов и их взаимодействию приводит к моделям среднего поля, а именно нелинейным цепям Маркова, в которых вероятности перехода зависят от распределения состояний агентов (подробнее см. Раздел 5).

3.2. SIR-модели. Математическое моделирование в эпидемиологии началось с работы D. Bernoulli в 1760 году, в которой была продемонстрирована эффективность вакцинации населения против ветряной оспы [18]. Впоследствии появилась серия математических моделей, основанные на законе действующих масс (см. обзорные статьи [19, 20] и приведенную в них литературу). Работы R. Ross в 1911 [21], А.А. Лотка в 1920 [22] и V. Volterra в 1926 [23] (модель «хищник-жертва»), привели к созданию камерной SIR-модели W.O. Kermack

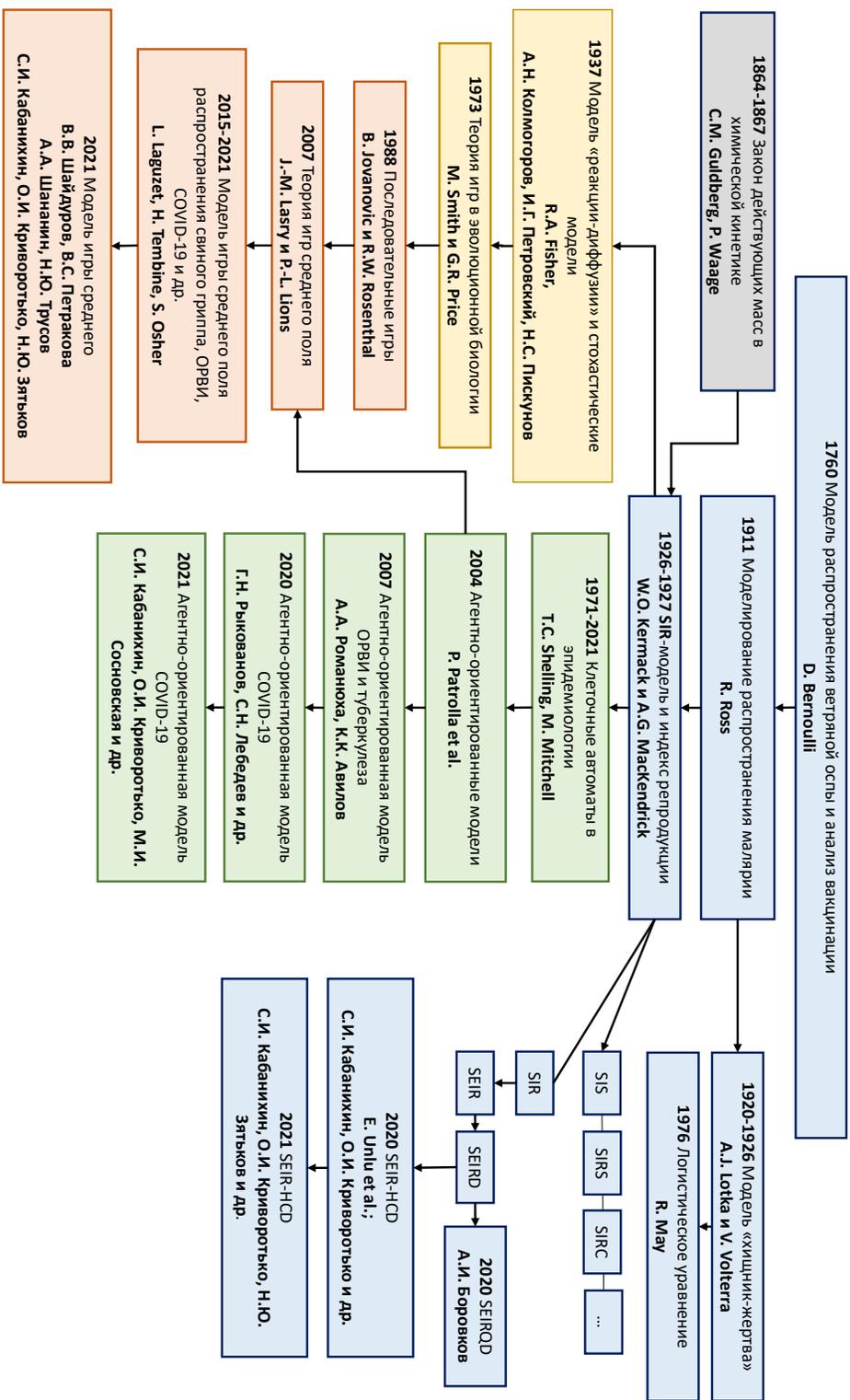


Рис. 3. Диаграмма развития математических моделей эпидемиологии с 1760 года на основе камерных, имитационных моделей и их комбинации. В синих рамках приведены результаты по применению ОДУ к описанию эпидемий, в желтых рамках – УРЧП, в оранжевых – модели ИСП и в зеленых рамках – АОМ.

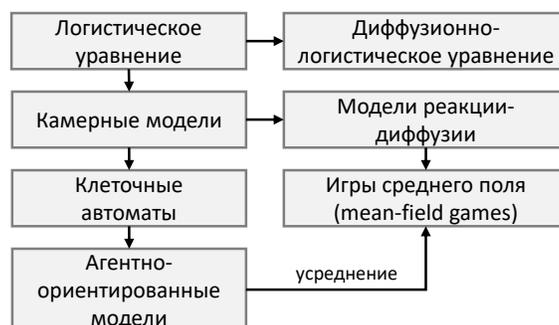


Рис. 4. Взаимосвязь математических моделей распространения COVID-19 на основе камерного подхода и имитационного моделирования.

и A.G. McKendrick [24, 25]

$$(1) \quad \begin{cases} \frac{dS}{dt} = -\alpha \frac{SI}{N}, & t > 0, \\ \frac{dI}{dt} = \alpha \frac{SI}{N} - \beta I, \\ \frac{dR}{dt} = \beta I, \end{cases}$$

в которой популяция из N особей разделена на три группы (камеры): S – восприимчивые, I – инфицированные ($I \ll N$) и R – вылеченные и умершие, связанные между собой вероятностными переходами $\alpha, \beta \in (0, 1)$ (схема модели приведена на Рис. 5а). Одним из важных результатов работы [25] было введение индекса репродукции (заразности)

$$\mathcal{R}_0 = \frac{\alpha}{\beta},$$

которая является важнейшей характеристикой заболевания и параметром распространения эпидемии (см. подробнее раздел 3.5).

Отметим, что частным случаем модели (1) при условии отсутствия иммунитета после перенесенного заболевания в рамках исследуемых временных промежутков (например, сезонный грипп) является SI-модель, решение которой сводится к логистическому уравнению:

$$(2) \quad \frac{dI}{dt} = (\alpha - \beta)I(t) - \frac{\alpha}{N}I^2(t).$$

Важным свойством модели (1) является выполнение закона действующих масс (закон сохранения), в рамках которого моделируемая популяция постоянна в течение всего времени $N = S(t) + I(t) + R(t)$.

В работах Е.Н. Пелиновского и его коллег [26, 27, 28] для моделирования распространения COVID-19 применяется обобщенное логистическое уравнение, описывающее рост численности заболевших. Предположение о единственности пика вспышки $I(t)$ эпидемии ограничивает применение логистической модели для описания длительного периода пандемии и учета ограничительных мер.

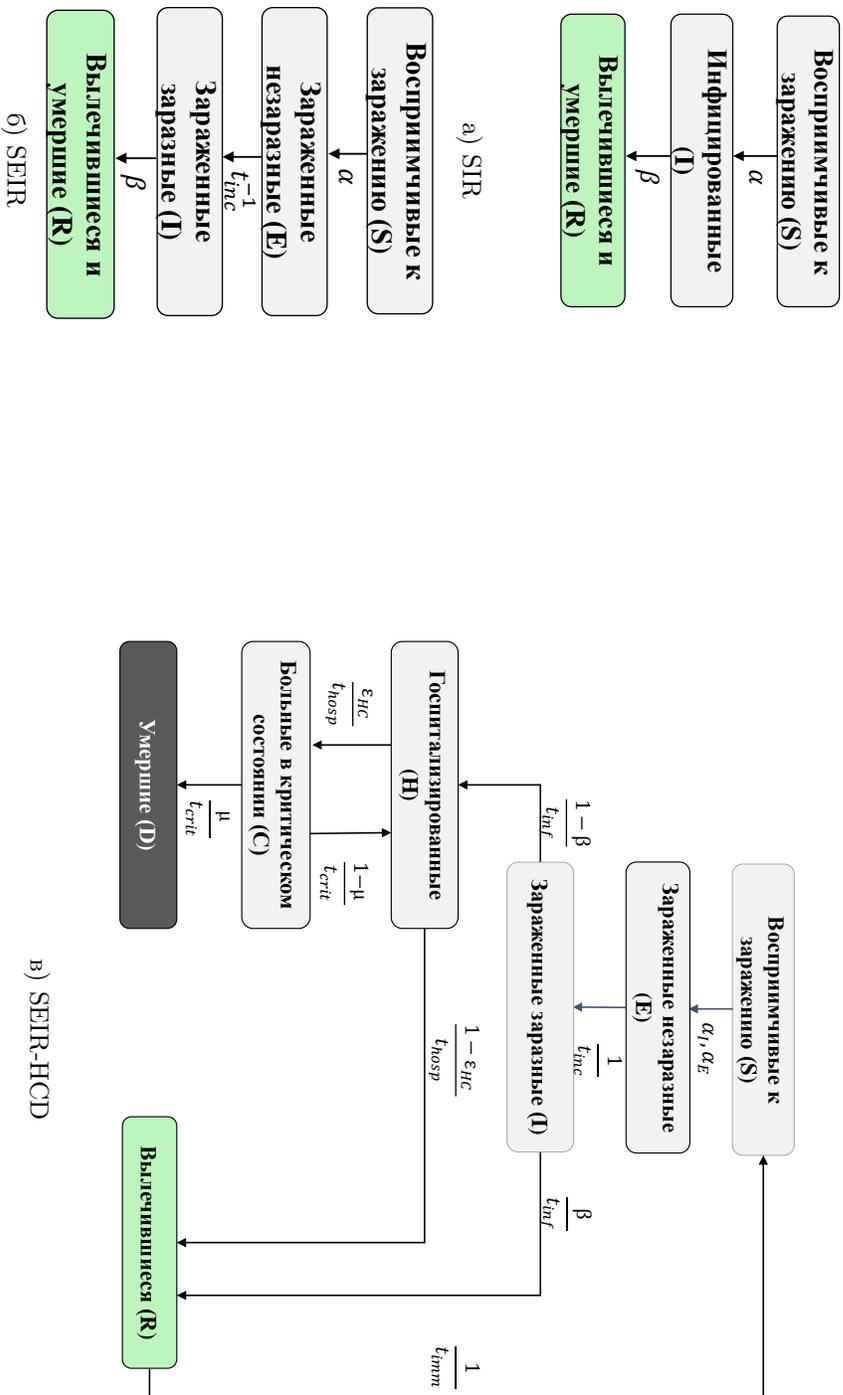


Рис. 5. Схемы камерных моделей (а) SIR, (б) SEIR и (в) SEIR-HCD.

Для учета инкубационного периода течения COVID-19 используется модификация модели Кермака-Маккендрика SEIR-типа (схема SEIR-модели приведена на Рис. 5б), которых на сегодняшний день разработано более 100 моделей (см. например, работы [29, 30, 31, 32, 33, 34, 38] и ссылки в них). В этих моделях популяция разделяется на группы (кроме S , I , R добавляются E – бессимптомные носители, H – госпитализированные, C – критические случаи, требующие подключения аппарата искусственной вентиляции легких (ИВЛ), D – умершие в результате COVID-19, Q – помещенные на карантин и другие). Это позволяет уточнить эпидемиологическую картину в регионе за счет варьирования более детального набора коэффициентов в уравнениях. Недостатком SIR-моделей является отсутствие гибкости – невозможность учета изменения параметров (новые мутации вируса и штамма, ограничительные меры, вакцинация). При попытке ввести в SIR-модели указанные изменения (например, сделать переменной скорость передачи инфекции $\alpha = \alpha(t)$) [35], мы сталкиваемся с неединственностью и неустойчивостью решения обратной задачи идентификации этого параметра $\alpha(t)$.

Отметим, что SIR-модели используют также для прогнозирования результатов управления развитием пандемии [36], т.е. в правую часть уравнений добавляют кусочно-постоянную функцию управления (ограничительные меры: ношение масок, социальная дистанция, карантин). Однако и в этих случаях проблема уточнения коэффициентов моделей SIR остается открытой и требует применения методов теории обратных задач.

Также в SIR-моделях можно учитывать влияние суперраспространителей на распространение COVID-19 (инфицированные индивидуумы, у которых повышенная вирусная концентрация) [37]. Однако теоретическое определение этого явления требует моделирования в масштабе отдельных людей (АОМ, см. Раздел 4.3).

Дальнейшее развитие математических моделей можно разделить на две составляющие: введение пространственной координаты в логистические уравнения и учет дискретной пространственной неоднородности. В первом случае мы получаем новый класс математических моделей «реакции-диффузии» (см. Раздел 3.3), а во втором – новый подход, в котором системы дифференциальных уравнений соединяются в пространстве графовой структурой (см. Раздел 4).

3.3. Модели «реакции-диффузии». В 1937 году британский ученый R.A. Fisher [39] и советские математики А.Н. Колмогоров, И.Г. Петровский и Н.С. Пискунов [40] предложили и обосновали математическую модель, основанную на уравнении в частных производных параболического типа, получившую впоследствии название модели «реакции-диффузии», которая была также применена при описании процессов в биологии, экологии и других приложениях:

$$(3) \quad \frac{\partial u}{\partial t}(x, t) = f(u(x, t)) + d \frac{\partial^2 u}{\partial x^2}(x, t).$$

Здесь $u(x, t)$ – плотность распределения групп популяции в точке x одномерного пространства и времени t , d – коэффициент диффузии, $f(u)$ – функция, характеризующая характер распространения заболевания в популяции, удовлетворяющая закону действующих масс и условиям

$$f(0) = f(1) = 0, \quad f(u) > 0, \quad \text{если } 0 < u < 1, \\ f'(0) > 0 \text{ и } f'(u) < f'(0), \quad \text{если } 0 < u \leq 1.$$

Авторы работы [40] строго доказали, что если начальное условие удовлетворяет следующим ограничениям

$$0 \leq u(x, 0) \leq 1, \quad u(x, 0) = 0 \quad \forall x < x_1 \quad \text{и} \quad u(x, 0) = 1 \quad \forall x > x_2 \geq x_1,$$

то динамика популяции по переменной t описывается скоростью $v^* = 2\sqrt{f'(0)d}$.

Отметим, что уравнение (3) можно записать в векторном виде, например, в случае SIR модели получим систему трех уравнений, для каждого из которых накладываются вышеуказанные ограничения. В этом случае переменная x может выступать как пространственная переменная, описывающая координату (x, y) , или возраст (одномерная переменная).

Учет пространственной неоднородности позволяет более точно моделировать эпидемию от очага распространения (крупного города в стране, столицы в регионе и пр.) при известных начальных условиях. Так, в работах [41, 42, 43, 44] получены оценки распространения COVID-19 в первые месяцы с начала эпидемии с учетом пассажиропотоков. Показано, что учет неоднородности влияет на характер распространения COVID-19 в крупных регионах и странах. В разделе 5 будут рассмотрены стохастические дифференциальные уравнения, учитывающие пространственную неоднородность и элементы управления (задачи ИСП). Однако использование модели для описания второй и последующих волн эпидемии требует добавления уравнений в (3), введения множественных известных источников распространения заболевания $u(x, 0)$ (обратная задача определения источника) и вычислительных ресурсов.

3.3.1. Пространственная SIR-модель распространения инфекционного заболевания. Приведем пример пространственной SIR-модели описания произвольного эпидемического процесса, основываясь на работах [45, 46]. Введем в пространстве Ω плотности основных подгрупп популяции в момент времени t

$$S(t) = \int_{\Omega} s(t, x) dx, \quad I(t) = \int_{\Omega} i(t, x) dx, \quad R(t) = \int_{\Omega} r(t, x) dx.$$

Запишем модель «реакции-диффузии» (3) SIR-типа, где описание динамики (реакция) использует соответствующие правые части SIR-моделей, а коэффициенты диффузии $d = (d_1, d_2, d_3)$ при вторых производных по пространственным переменным полагаются постоянными величинами:

$$\begin{cases} \frac{\partial s}{\partial t} = -\alpha \frac{si}{N} + d_1 s_{xx}, & t > 0, x \in \Omega, \\ \frac{\partial i}{\partial t} = \alpha \frac{si}{N} - \beta i + d_2 i_{xx}, \\ \frac{\partial r}{\partial t} = \beta i + d_3 r_{xx}, \end{cases}$$

3.4. Стохастические модели. Стохастические модели эпидемии изучались в различных сетях [47, 48, 49], а также применялись к моделированию пандемии COVID-19 [50].

В данном разделе пойдет речь о трех различных методах построения стохастических моделей эпидемий, которые непосредственно связаны с их детерминированными аналогами [51]: 1) дискретная временная марковская модель (Раздел 3.4.1), 2) модель Маркова с непрерывным временем (Раздел 3.4.2), и 3) модель, основанная на системе СДУ (Раздел 3.4.3). Эти стохастические процессы различаются по исходным предположениям относительно времени и

состояния переменных. В модели 1 время и переменные состояния дискретны. В модели 2 время является непрерывным, а переменная состояния – дискретной. Наконец, модель СДУ основана на диффузионном процессе, где и время, и переменные состояния являются непрерывными. Описание и свойства стохастических моделей эпидемий приведем для SIR-модели (1).

Одним из наиболее важных отличий между детерминированной и стохастическими моделями эпидемий является их асимптотическая динамика. В конечном итоге стохастические решения (выборочные пути) сходятся к состоянию без заболевания, даже если соответствующее детерминированное решение сходится к эндемическому равновесию. Однако стохастическим моделям описания эпидемий присущи следующие свойства: описание вероятности вспышки эпидемии, квазистационарное распределение вероятности, распределение конечного размера эпидемии и ожидаемая продолжительность эпидемии [51].

Всюду далее каллиграфические символы $\mathcal{S}(t)$, $\mathcal{I}(t)$, $\mathcal{R}(t)$ будут означать случайные величины (за исключением базового индекса репродукции \mathcal{R}_0).

3.4.1. Дискретная временная марковская модель. Пусть $\mathcal{S}(t)$, $\mathcal{I}(t)$ и $\mathcal{R}(t)$ – дискретные случайные величины, описывающие число восприимчивых, инфицированных и умерших/вылеченных индивидуумов в момент времени $t \in \{0, \Delta t, 2\Delta t, \dots\}$, соответственно, т.е. $\mathcal{S}(t), \mathcal{I}(t), \mathcal{R}(t) \in \{0, 1, 2, \dots, N\}$.

В данном примере мы имеем дело с двумя независимыми случайными величинами $\mathcal{S}(t)$ и $\mathcal{I}(t)$. Случайная величина $\mathcal{R}(t) = N - \mathcal{S}(t) - \mathcal{I}(t)$. Случайный двумерный процесс $\{\mathcal{S}(t), \mathcal{I}(t)\}_{t=0}^{\infty}$ характеризуется совместной функцией вероятности

$$p_{(s,i)}(t) = \text{Prob} \{ \mathcal{S}(t) = s, \mathcal{I}(t) = i \}.$$

Данный нелинейный процесс обладает свойством «марковости» и однородности по времени.

Определение 1. Свойство «марковости» стохастического процесса состоит в том, что процесс в момент времени $t + \Delta t$ зависит только от своего состояния в предыдущий момент времени t .

Определение 2. Если вероятность перехода $p_{ji}(t + \Delta t, t) = \text{Prob} \{ \mathcal{I}(t + \Delta t) = j \mid \mathcal{I}(t) = i \}$ из состояния i в состояние j не зависит от t , т.е. $p_{ji}(\Delta t)$, то случайный процесс $\{\mathcal{I}(t)\}_{t=0}^{\infty}$ является однородным по времени.

Вероятности перехода между состояниями основываются на предположениях SIR-модели (1). Предположим, что Δt достаточно мал, чтобы за это время произошло более одного события (заражение, выздоровление, смерть). Тогда вероятность перехода между состояниями имеет вид:

$$p_{(s+k,i+j),(s,i)}(\Delta t) = \text{Prob}\{(\Delta \mathcal{S}, \Delta \mathcal{I}) = (k, j) \mid (\mathcal{S}(t), \mathcal{I}(t)) = (s, i)\}.$$

Здесь $\Delta \mathcal{S} = \mathcal{S}(t + \Delta t) - \mathcal{S}(t)$. Пользуясь формулировкой SIR-модели (1), получим

$$p_{(s+k,i+j),(s,i)}(\Delta t) = \begin{cases} \frac{\alpha i s}{N} \Delta t, & (k, j) = (-1, 1), \\ \beta i \Delta t, & (k, j) = (0, -1), \\ 1 - \frac{\alpha i s}{N} \Delta t - \beta i \Delta t, & (k, j) = (0, 0), \\ 0, & \text{иначе.} \end{cases}$$

Таким образом, вероятность инфицирования (переход $(s, i) \rightarrow (s - 1, i + 1)$) равна $\frac{\alpha i s}{N} \Delta t$, а вероятность выздоровления или смерти (переход $(s, i) \rightarrow (s, i - 1)$) равна $\beta i \Delta t$. Наконец, вероятность отсутствия изменений (переход $(s, i) \rightarrow (s, i)$) имеет вид $1 - \frac{\alpha i s}{N} \Delta t - \beta i \Delta t$. Т.к. шаг по времени Δt достаточно мал, каждая вероятность перехода между состояниями лежит в интервале $[0, 1]$.

Применяя свойство «марковости» случайного процесса $\{\mathcal{S}(t), \mathcal{I}(t)\}_{t=0}^{\infty}$, получим дифференциальное уравнение для вероятности $p_{(s,i)}(t + \Delta t)$:

$$(4) \quad p_{(s,i)}(t + \Delta t) = p_{(s+1,i-1)}(t) \frac{\alpha}{N} (i-1)(s+1) \Delta t + p_{(s,i+1)}(t) \beta (i+1) \Delta t + p_{(s,i)}(t) \left(1 - \left[\frac{\alpha}{N} i s + \beta i \right] \Delta t \right).$$

Несложно показать, что состояние без заболевания $(N, 0)$ является поглощающим, т.е. все траектории процесса асимптотически стремятся к $(N, 0)$.

3.4.2. Модель Маркова с непрерывным временем. Пусть переменные состояния $\mathcal{S}(t), \mathcal{I}(t)$ и $\mathcal{R}(t)$ по-прежнему являются дискретными случайными величинами, т.е. $\mathcal{S}(t), \mathcal{I}(t), \mathcal{R}(t) \in \{0, 1, 2, \dots, N\}$, а время $t \in (0, T)$ непрерывно.

Вероятность перехода между состояниями процесса $\{\mathcal{S}(t), \mathcal{I}(t)\}_{t \in (0, T)}$ характеризуется следующей системой, как и в предыдущем случае, с добавлением $o(\Delta t)$:

$$p_{(s+k,i+j),(s,i)}(\Delta t) = \begin{cases} \frac{\alpha i s}{N} \Delta t + o(\Delta t), & (k, j) = (-1, 1), \\ \beta i \Delta t + o(\Delta t), & (k, j) = (0, -1), \\ 1 - \frac{\alpha i s}{N} \Delta t - \beta i \Delta t + o(\Delta t), & (k, j) = (0, 0), \\ 0, & \text{иначе.} \end{cases}$$

В результате мы получим схожее уравнение с (4) с добавлением $o(\Delta t)$ в правую часть. Полагая $\Delta t \rightarrow 0$ в итоговом уравнении для переходной вероятности, получим прямое уравнение Колмогорова:

$$\frac{dp_{(s,i)}}{dt} = p_{(s+1,i-1)} \frac{\alpha}{N} (i-1)(s+1) + p_{(s,i+1)} \beta (i+1) - p_{(s,i)} \left(\frac{\alpha}{N} i s + \beta i \right).$$

В работе 2011 года [52] показано, что время достижения вспышки эпидемии не сильно зависит от начальной траектории, описываемой в рамках SI марковской модели. В работе построена гибридная модель, состоящая из стохастического марковского процесса на начальном этапе моделирования с целью получения распределения исходных условий распределения S и I популяций, и детерминированной SIR модели для описания последующей эволюции системы. В работе тех же авторов 2016 года [53] построена модель динамики эпидемии в стохастических взаимодействующих центрах популяции в сочетании с случайной миграцией на основе марковских цепей SIR типа. Получено среднее приближение поля для общего процесса миграции и предложен приближенный метод, позволяющий вычислять статистические моменты для сетей с густонаселенными центрами. Рассматриваемые модели решаются численно на основе метода Монте Карло, что увеличивает их вычислительную сложность. В работе 2021 года [54] предложен гибридный метод численного решения задачи идентификации параметров кинетики жизненного цикла ВИЧ на клеточном

уровне, в основе которого лежит марковский процесс и анализ чувствительности стохастической модели. Авторы работы [55] с помощью детерминированной SIS модели, сведенной к стохастической на основе марковских случайных процессов с непрерывным временем и конечным числом состояний, описано распространение эпидемии внутри вагона железнодорожного состава и показано, что на скорость её развития влияет расположение изначально заражённых пассажиров. Показано также, что наибольший риск при распространении эпидемии представляют проводники и нерегламентированное использование общественных мест.

3.4.3. Стохастические дифференциальные уравнения. Пусть состояния SIR-системы и время непрерывны, т.е. $S(t), I(t), R(t) \in [0, N]$ и $t \in (0, T)$. Предположим, что независимые случайные величины $\{(S(t), I(t))\}$ имеют нормальное распределение (или асимптотически стремятся к такому).

Обозначим $\Delta X(t) = (\Delta S, \Delta I)^T$. Тогда математическое ожидание $\Delta X(t)$ порядка Δt записывается в виде

$$\mathbb{E}(\Delta X(t)) = \begin{pmatrix} -\frac{\alpha}{N}SI \\ \frac{\alpha}{N}SI - \beta I \end{pmatrix} \Delta t.$$

Матрица ковариации для $\Delta X(t)$ может быть записана в первом приближении:

$$V(\Delta X(t)) = \mathbb{E}(\Delta X(t)[\Delta X(t)]^T) - \mathbb{E}(\Delta X(t))\mathbb{E}(\Delta X(t))^T \approx \mathbb{E}(\Delta X(t)[\Delta X(t)]^T)$$

или

$$V(\Delta X(t)) = \begin{pmatrix} \frac{\alpha}{N}SI & -\frac{\alpha}{N}SI \\ -\frac{\alpha}{N}SI & \frac{\alpha}{N}SI + \beta I \end{pmatrix} \Delta t.$$

В силу того, что случайная величина $\Delta X(t)$ имеет нормальное распределение, то имеет место следующее разложение

$$(5) \quad X(t + \Delta t) = X(t) + \Delta X(t) \approx X(t) + \mathbb{E}(\Delta X(t)) + \sqrt{V(\Delta X(t))}.$$

В силу того, что $V > 0$ и симметричная, существует однозначный корень $\sqrt{V} = B\sqrt{\Delta t}$. Система уравнений (5) является аппроксимацией Эйлера системы СДУ в смысле Ито. При достаточно гладких коэффициентах решение $X(t)$ системы (5) сходится к решению системы СДУ:

$$(6) \quad \begin{cases} dS = -\frac{\alpha}{N}SI dt + B_{11}dW_1 + B_{12}dW_2, \\ dI = \left(\frac{\alpha}{N}SI - \beta I\right) dt + B_{21}dW_1 + B_{22}dW_2, \end{cases}$$

где W_1 и W_2 – два независимых винеровских процесса, $B = (B_{ij})$.

Определение 3. Случайный процесс W_t , $t \geq 0$, называется винеровским процессом, если

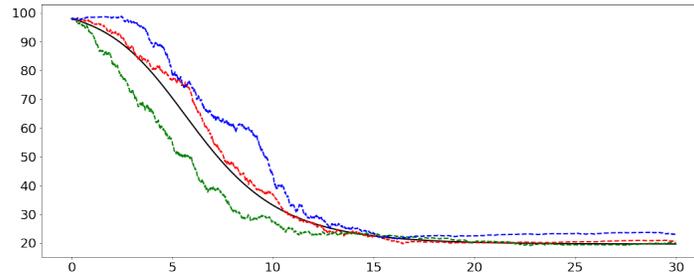
- (1) $W_0 = 0$ почти достоверно.
- (2) W_t – процесс с независимыми приращениями.
- (3) $W_{t+\Delta t} - W_t \sim \mathbf{N}(0, \Delta t)$. Здесь $\mathbf{N}(0, \Delta t)$ – нормальное распределение со средним 0 и дисперсией Δt .

На рисунке 6 приведены решения уравнений для SIR модели в детерминированной постановке (1) (отмечены черной линией) и в стохастической (6) для восприимчивых $S(t)$ (рис. 6а) и инфицированных $I(t)$ (рис. 6б) индивидуумов в случае размера моделируемой популяции $N = 100$ человек, $\alpha = 1$, $\beta = 0.5$, $S_0 = 98$, $I_0 = 2$, шаг по времени брался равномерным $\Delta t = 0.01$. Для решения системы (1) применялась схема Эйлера 1-го порядка аппроксимации, решение системы (6) строилось на основе схемы Эйлера-Маруямы:

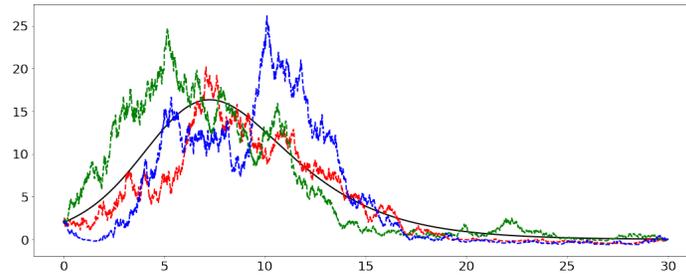
$$S_{i+1} = S_i - \frac{\alpha}{N} S_i I_i \Delta t + B_{11} \sqrt{\Delta t} \text{rand}_1 + B_{12} \sqrt{\Delta t} \text{rand}_2,$$

$$I_{i+1} = I_i + \left(\frac{\alpha}{N} S_i I_i - \beta I_i \right) \Delta t + B_{21} \sqrt{\Delta t} \text{rand}_1 + B_{22} \sqrt{\Delta t} \text{rand}_2,$$

где $\text{rand}_1 \sim \mathbf{N}(0, 1)$ и $\text{rand}_2 \sim \mathbf{N}(0, 1)$ – две независимые случайные величины.



а) Количество восприимчивых индивидуумов $S(t)$



б) Количество инфицированных индивидуумов $I(t)$

Рис. 6. Моделирование изменения $S(t)$ и $I(t)$ в детерминированной (1) (черная линия) и стохастической (6) (цветные линии) SIR модели при условиях: $N = 100$, $\alpha = 1$, $\beta = 0.5$, $S_0 = 98$, $I_0 = 2$, $\Delta t = 0.01$.

Траектории решения системы (6) достаточно близки к решению детерминированной SIR модели (1).

В работе [51] показано, что в случае достаточно большой популяции $N \geq 100$, индекса репродукции вируса \mathcal{R}_0 и если вспышка начинается с достаточным количеством инфицированных особей, то результаты для стохастической SIR

эпидемии находятся в близком согласии с предсказаниями детерминированной эпидемиологической SIR модели; болезнь становится эндемичной.

3.4.4. Случайные процессы в эпидемиологии. Руководствуясь случайностью изучаемых процессов, анализ эпидемии возможен и без явного построения стохастических моделей. Известно, что плотность потока частиц в размножающей среде при достаточно широких условиях асимптотически экспоненциальна по времени t с некоторым параметром λ , т.е. с показателем λt . В работах [56, 57] показано, что если среда случайна, то параметр λ – случайная величина, и для оценки временной асимптотики среднего (по реализациям среды) числа частиц можно в некотором приближении усреднять экспоненту по распределению λ . В предположении гауссовости этого распределения таким образом получается асимптотическая “сверхэкспоненциальная” оценка среднего потока, выражаемая экспонентой с показателем $t\mathbb{E}\lambda + t^2 D\lambda/2$. Для численной экспериментальной проверки такой оценки разработано вычисление вероятностных моментов случайного параметра λ на основе рандомизации фурье-приближений специальных нелинейных функционалов.

Отмечено, что результаты в статье имеют широкое приложение. В частности, согласно статистике Всемирной организации здравоохранения (ВОЗ), сверхэкспоненциальное поведение проявляла пандемия COVID-19, развивающаяся во всем мире. А именно, количество наблюдений (по дням) аппроксимируется с точностью до 2% с 9 марта 2020 года по 21 марта 2020 года.

3.5. Базовый индекс репродукции вируса \mathcal{R}_0 для дифференциальных моделей. Общее определение базового индекса репродукции (basic reproduction number) [58]: базовый индекс репродукции вируса \mathcal{R}_0 определяется как среднее количество людей, которых заражает активный инфицированный, попавший в полностью неиммунизированное окружение при отсутствии специальных эпидемиологических мер, направленных на предотвращение распространения заболевания.

В разделе 3.5.1 мы покажем, что базовый индекс репродукции \mathcal{R}_0 является границей устойчивости состояния равновесия SIR-системы при отсутствии инфицированных. Если $\mathcal{R}_0 > 1$, то на начальном этапе число заболевших будет расти экспоненциально. Если $\mathcal{R}_0 \in (0, 1)$, то небольшое количество инфицированных людей, попавших в полностью восприимчивую популяцию, в среднем не смогут сохранить свою группу, и эпидемии не будет.

3.5.1. Алгоритм вычисления индекса репродукции. Опишем метод вычисления индекса репродукции, предложенный van den Driessche и Watmough [59] для SIR-моделей. Разделим всю популяцию на две категории: n групп инфицированных и m групп не инфицированных. Векторы $x \in \mathbb{R}^n$ и $y \in \mathbb{R}^m$ определяют количество индивидуумов в каждой из двух категорий, например, в модели SEIR-HCD $x = (E, I, H, C) \in \mathbb{R}^4$, $y = (S, R, D) \in \mathbb{R}^3$ (см. Раздел 3.5.2). Тогда SIR-модель распространения инфекционного заболевания принимает вид

$$(7) \quad \begin{aligned} \dot{x}_i &= \mathcal{F}_i(x, y) - \mathcal{V}_i(x, y), & i &= 1, \dots, n, \\ \dot{y}_j &= g_j(x, y), & j &= 1, \dots, m. \end{aligned}$$

Здесь \mathcal{F}_i – скорость изменения количества инфицированных и \mathcal{V}_i – скорость изменения умерших, выздоровевших и тех, кто заболевает в i -й группе, $g_j(x, y)$ –

оставшиеся функции модели, согласованные с предыдущими в смысле закона действующих масс.

Вывод индекса репродукции заключается в линеаризации системы (7) в окрестности состояния равновесия в случае отсутствия инфекции $(0, y_0)$. В работе [59] были введены пять условий существования этого состояния равновесия, согласованных с законом действующих масс для модели модели (7). Вычислим собственные значения матрицы FV^{-1} , где

$$F_{ij} = \frac{\partial \mathcal{F}_i}{\partial x_j}(0, y_0), \quad V_{ij} = \frac{\partial \mathcal{V}_i}{\partial x_j}(0, y_0).$$

Тогда базовый индекс репродукции $\mathcal{R}_0 = \max \lambda_i(FV^{-1})$ является неотрицательным и соответствующий ему собственный вектор ω состоит из неотрицательных компонент [60]. Компоненты вектора ω можно интерпретировать как распределение инфицированных индивидумов, вызывающих наибольшее количество \mathcal{R}_0 вторичных инфекций в поколении.

В [59] доказана теорема о локальной устойчивости неинфицированного состояния равновесия системы (7), а именно состояние равновесия $(0, y_0)$ локально асимптотически устойчиво, если $\mathcal{R}_0 < 1$, и неустойчиво, если $\mathcal{R}_0 > 1$. В следующем разделе приведен вывод индекса репродукции для SEIR-HCD модели распространения COVID-19.

3.5.2. *Базовый индекс репродукции вируса для модели SEIR-HCD.* Используя описанный алгоритм, выведем базовый индекс репродукции для модели SEIR-HCD распространения COVID-19 [31, 32]

$$(8) \quad \begin{cases} \frac{dS}{dt} = -\tilde{a}(t) \left(\frac{\alpha_I(t)S(t)I(t)}{N} + \frac{\alpha_E(t)S(t)E(t)}{N} \right) + \frac{1}{t_{imm}} R(t), \\ \frac{dE}{dt} = \tilde{a}(t) \left(\frac{\alpha_I(t)S(t)I(t)}{N} + \frac{\alpha_E(t)S(t)E(t)}{N} \right) - \frac{1}{t_{inc}} E(t), \\ \frac{dI}{dt} = \frac{1}{t_{inc}} E(t) - \frac{1}{t_{inf}} I(t), \\ \frac{dR}{dt} = \frac{\beta(t)}{t_{inf}} I(t) + \frac{1 - \varepsilon_{HC}(t)}{t_{hosp}} H(t) - \frac{1}{t_{imm}} R(t), \\ \frac{dH}{dt} = \frac{1 - \beta(t)}{t_{inf}} I(t) + \frac{1 - \mu(t)}{t_{crit}} C(t) - \frac{1}{t_{hosp}} H(t), \\ \frac{dC}{dt} = \frac{\varepsilon_{HC}(t)}{t_{hosp}} H(t) - \frac{1}{t_{crit}} C(t), \\ \frac{dD}{dt} = \frac{\mu(t)}{t_{crit}} C(t) \end{cases}$$

с начальными условиями

$$(9) \quad \begin{aligned} S(t_0) &= N - E_0 - I_0 - R_0 - H_0 - C_0 - D_0, \quad E(t_0) = E_0, \\ I(t_0) &= I_0, \quad R(t_0) = R_0, \quad H(t_0) = H_0, \quad C(t_0) = C_0, \quad D(t_0) = D_0. \end{aligned}$$

Здесь $\tilde{a}(t) = \frac{5 - a(t - \tau)}{5}$. Схема модели (8) приведена на Рис. 5в, а описание и значения параметров и начальных условий для Новосибирской области приведены в таблице 1 (начальный момент времени полагается 15.04.2020). В SEIR-HCD модели происходит перемещение бессимптомной популяции $E(t)$ после t_{inc} дней в симптоматическую $I(t)$. Инфицированные индивидумы после t_{inf} дней выздоравливают с вероятностью $\beta(t)$ и госпитализируются $H(t)$

с вероятностью $1 - \beta(t)$. Затем госпитализированные могут выздоравливать или нуждаться в подключении аппарата ИВЛ $C(t)$. В модели только критические случаи могут умереть $D(t)$ с вероятностью $\mu(t)$. Отметим, что параметры предполагаются кусочно-постоянными ввиду изменчивости ситуации (введение ограничений, возникновение новых штаммов заболевания и др.), что будет учтено в численных расчетах (Раздел 7.2).

Вычислим базовый индекс репродукции в окрестности начального момента времени, когда все параметры можно считать постоянными.

Разделяя всю популяцию N в модели (8) на инфицированных $x = (E, I, H, C) \in \mathbb{R}^4$ и неинфицированных $y = (S, R, D) \in \mathbb{R}^3$, получим следующие вектор-функции, согласно (7):

$$F = \begin{pmatrix} \tilde{a}(t) \left(\frac{\alpha_I(t)S(t)I(t)}{N} + \frac{\alpha_E(t)S(t)E(t)}{N} \right) \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

$$V = \begin{pmatrix} \frac{1}{t_{inc}} E(t) \\ -\frac{1}{t_{inc}} E(t) + \frac{1}{t_{inf}} I(t) \\ -\frac{1-\beta}{t_{inf}} I(t) - \frac{1-\mu}{t_{crit}} C(t) + \frac{1}{t_{hosp}} H(t) \\ -\frac{\varepsilon_{HC}}{t_{hosp}} H(t) + \frac{1}{t_{crit}} C(t) \end{pmatrix}.$$

Состояние равновесия системы (8)-(9) в случае отсутствия инфицированных индивидуумов есть $\mathcal{E}_0 = (N, 0, 0, 0, 0, 0, 0)$. Тогда матрицы F и V имеют вид:

$$F = \begin{pmatrix} \tilde{a}\alpha_E & \tilde{a}\alpha_I & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad V = \begin{pmatrix} \frac{1}{t_{inc}} & 0 & 0 & 0 \\ -\frac{1}{t_{inc}} & \frac{1}{t_{inf}} & 0 & 0 \\ 0 & -\frac{1-\beta}{t_{inf}} & \frac{1}{t_{hosp}} & -\frac{1-\mu}{t_{crit}} \\ 0 & 0 & -\frac{\varepsilon_{HC}}{t_{hosp}} & \frac{1}{t_{crit}} \end{pmatrix}.$$

Вычисляя максимальное собственное значение матрицы FV^{-1} , получим базовый индекс репродукции для SEIR-HCD модели

$$(10) \quad \mathcal{R}_0(t) = \tilde{a}(t) (\alpha_E(t)t_{inc} + \alpha_I(t)t_{inf}),$$

первое слагаемое в котором характеризует заразность от бессимптомной части популяции, а второе слагаемое – от симптомной.

Лемма 1. В случае постоянных параметров $\alpha_E, \alpha_I, \beta, \varepsilon_{HC}, \mu$ модели (8) состояние равновесия \mathcal{E}_0 системы неустойчиво при $\mathcal{R}_0 > 1$.

При отсутствии слагаемого $\frac{R(t)}{t_{imm}}$, характеризующего потерю иммунитета после выздоровления, состояние равновесия \mathcal{E}_0 устойчиво по Ляпунову при $\mathcal{R}_0 < 1$, определяемым формулой (10).

3.6. Выводы. Преимущество использования дифференциальных моделей для описания распространения эпидемий (в том числе COVID-19) состоит в учете особенностей инфекции (наличие инкубационного периода, взаимосвязь инфицированных и критических случаев и т.д.) в параметрах моделей и закона действующих масс (размера популяции), а в случае добавления переменной x (модели «реакции-диффузии») – учет пространственной неоднородности. Такие

ТАБЛИЦА 1. Описание и значения параметров SEIR-HCD модели для Новосибирской области

Параметр	Описание	Значение
$a(t)$	Индекс самоизоляции от Яндекса	(0, 5)
$\alpha_E(t)$	Параметр заражения между бессимптомной и восприимчивой группами населения ($\alpha_E \gg \alpha_I$)	(0, 1)
$\alpha_I(t)$	Параметр заражения между инфицированным и восприимчивым населением	(0, 1)
$\beta(t)$	Доля инфицированных, которая переносит заболевание без осложнений	(0, 1)
$\varepsilon_{HC}(t)$	Доля госпитализированных случаев, которым требуется подключение ИВЛ	(0, 1)
$\mu(t)$	Доля смертельных случаев в результате COVID-19	(0, 0.5)
τ	Латентный период (характеризует запаздывание наступления заразности)	2 дня
t_{inc}	Длительность инкубационного периода	2-14 дня
t_{inf}	Длительность периода инфицирования	2.5-14 дня
t_{hosp}	Длительность периода госпитализации	4-5 дня
t_{crit}	Длительность использования аппарата ИВЛ	10-20 дня
t_{imm}	Длительность гуморального иммунитета	120-180 дней
N	Население в Новосибирской области (человек)	2798170
E_0	Начальное количество бессимптомных носителей	(1, 5000)
I_0	Начальное количество инфицированных случаев	(1, 5000)
R_0	Начальное количество вылеченных случаев	(1, 100)
H_0	Начальное количество госпитализированных	133
C_0	Начальное количество критических случаев	1
D_0	Начальное количество смертей	1

модели качественно описывают вспышку эпидемии, но не обладают достаточной гибкостью. Учет изменения параметров для описания ограничительных мер, новых мутаций вируса приводит к неединственности и неустойчивости решения задачи их идентификации.

В стохастических SIR моделях эндемическое равновесие не достигается, поскольку существует конечная вероятность того, что число инфицированных особей в системе упадет ниже единицы. Таким образом, в данной системе число инфицированных индивидумов может не увеличиваться при ненулевом начальном условии. В случае достаточно большой популяции, индекса репродукции вируса $\mathcal{R}_0 > 1$ и если вспышка начинается с достаточным количеством инфицированных особей, то результаты для стохастической SIR эпидемии находятся в близком согласии с предсказаниями детерминированной эпидемической SIR модели; болезнь становится эндемичной.

4. АГЕНТНО-ОРИЕНТИРОВАННЫЕ МОДЕЛИ

В данном разделе будет приведен краткий обзор имитационных моделей, основанные на клеточных автоматах, и агентно-ориентированные модели (АОМ).

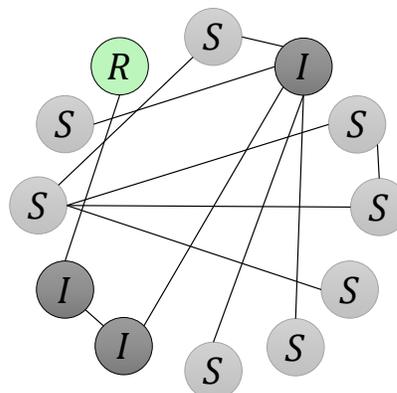
Подробное построение АОМ распространения COVID-19 в конкретной регионе приведено в конце раздела 4.3.1.

4.1. Введение. Значимым преимуществом моделей, базирующихся на аппарате дифференциальных уравнений, является возможность их аналитического исследования. Тем не менее для всех таких моделей характерно допущение – характеристики и поведение всех индивидов, отнесенных к одной подгруппе, считаются одинаковыми. Имитационные модели (клеточные автоматы, сетевые модели и АОМ) позволяют ослабить указанные ограничения.

4.2. Клеточные автоматы. Т.С. Shelling в 1971 [61] и М. Mitchel в 1993 [62] предложили теорию клеточных автоматов для моделирования локальных характеристик восприимчивых популяций вместе со стохастическими параметрами, которые отражают вероятностный характер передачи болезни. Клеточные автоматы представляют собой совокупность квадратных ячеек, объединенных в прямоугольную решетку, каждая из которых принимает состояние из конечного множества. Узлы решетки моделируют индивидов, каждый из которых имеет фиксированное положение в пространстве (схема клеточного автомата представлена на Рис. 7а). Так, анализ социальной дистанции, ношение масок при распространении COVID-19 в локальной популяции может описываться с помощью клеточных автоматов [63]. В работе [64] SEIR модель описывается в терминах вероятностных клеточных автоматов и обыкновенных дифференциальных уравнений передачи COVID-19, достаточно гибких для моделирования различных сценариев социальной изоляции.

S	S	S	S	S	S	S	S	S
S	S	S	S	S	S	S	S	S
S	S	I	I	I	S	S	S	S
S	S	I	R	R	I	S	S	S
S	I	I	R	R	I	S	S	S
S	I	I	I	I	I	S	S	S
S	S	I	I	S	S	S	S	S
S	S	S	I	S	S	S	S	S

а)



б)

Рис. 7. Представление распространения инфекции клеточным автоматом (а) и сетевой моделью (б).

4.3. Агентно-ориентированные модели. После публикации статьи Р. Patralla в 2004 году [65] была предложена агентно-ориентированная модель (АОМ), в которой расширены возможности клеточных автоматов отслеживания распространения болезни и контактов между каждым человеком в социальной группе, расположенной в географической области. АОМ позволяют взаимодействовать между людьми и способны преодолевать ограничения различных подходов,

обращаясь к естественной стохастической природе эпидемического процесса. АОМ представляют схему возможных контактов в виде динамического или статического графа, в котором вершины – объекты с набором индивидуальных свойств, сколь угодно детализировано описывающие состояние отдельных индивидов (упрощенная схема графа в АОМ приведена на Рис. 76).

Группа под руководством академика Г.Н. Рыкованова [66, 67] разработали и проанализировали камерную SEIR-D и АОМ для описания распространения COVID-19. Разработанная статистическая АОМ, несмотря на некоторую упрощенность модели поведения людей, позволяет расчетно анализировать такие факторы, как, например, введение карантина в отношении отдельных социальных групп или в отдельных сферах деятельности (работа, транспорт, магазины). Авторы отметили (стр. 27-28), что достоинства статистической модели влекут за собой и некоторые её недостатки. Так, чтобы достоверно моделировать те или иные факторы, в основу расчета должны закладываться адекватные исходные данные – от численности населения и его распределения по социальным группам до загруженности различных видов транспорта или магазинов.

Группа американских ученых [68] разработали открытый программный комплекс Covasim [69], основу которого составляет агентный подход моделирования эпидемии с учетом особенностей заболевания, фармацевтических (вакцинация) и политических (физические ограничения, ношение масок) вмешательств. Этот программный комплекс применялся для построения сценариев развития эпидемии COVID-19, изучения динамики пандемии и поддержке принятия политических решений более чем в десятке стран Африки, Азиатско-Тихоокеанского региона, Европы и Северной Америки. В статье [70] показано, что система реагирования, основанная на расширенном тестировании и отслеживании контактов, может играть важную роль в ослаблении интервенций социального дистанцирования при отсутствии коллективного иммунитета против SARS-CoV-2. Авторы [71] подсчитали, что инфицированные люди не пожилого возраста (<60 лет) могут быть в 2,78 раза более заразными, чем пожилые люди, и первые, как правило, являются основной движущей силой сверхраспространения. В работах [72, 73, 74] в рамках АОМ распространения COVID-19 проанализированы противоэпидемические программы в различных регионах, в результате чего получено понимание эффективных мер для разных географических и демографических условий, а также текущих штаммов SARS-CoV-2. В работе [75] авторы построили АОМ, в котором источники заражения выступали суперраспространителями. Они показали, что сверхраспространение резко усиливает значимость ограничений личных контактов.

В работе [76] проведено численное исследование динамики когорт индивидов некоторого региона для наборов параметров, отражающих различные варианты передачи инфекции между индивидами на основе статистического моделирования и метода Монте-Карло. Представленный в работе подход позволяет перейти к марковскому случайному процессу в расширенном пространстве состояний, включающем описание когорт индивидов с помощью целочисленных и вещественных переменных, и расширять модель с точки зрения введения новых факторов.

В следующем разделе подробно описан процесс построения популяции на основе пакета Covasim, инфицирования вирусом штамма SARS-CoV-2 на основе графов, тестирования и вакцинации в Российской Федерации на основе статистических данных.

4.3.1. *АОМ распространения COVID-19.* Опишем структуру АОМ, лежащую в основе пакета Covasim [69]:

(1) **Инициация популяции.** Формируются четыре структуры контактов: домохозяйства, образовательные учреждения, рабочие и общественные места.

(а) Постоянные характеристики агента:

- возраст (все агенты делятся на возрастные группы по 10 лет: 0–9 лет, 10–19, ..., 80+),
- пол,
- социальный статус,
- вероятности прогрессирования заболевания зависят от возраста агента (возникновения симптомов, тяжелых и критических случаев, смертность).

(б) Переменные характеристики агента (пересчитываются к концу дня – шаг по времени):

- эпидемиологический статус. Каждый агент в определенный момент времени может находиться в одной из стадий заболевания: к описанным в Разделе 3.5.2 состояниям S , E , I , R , H , C , D добавлены бессимптомные больные A и больные в легкой форме M ,
- шанс быть протестированным.

Домохозяйства заполняются агентами согласно статистическим данным ООН [77] о среднем размере семьи в регионе (2.6 человек). В зависимости от возраста агенты контактируют друг с другом в контактных сетях, представляющие собой полносвязные графы, количество вершин которых является пуассоновской случайной величиной с средним:

- для домохозяйства – размер семьи,
- для общественных мест и образовательных учреждений – 20.
- для работы – 8.

Все агенты имеют контакты в домохозяйствах и в общественных местах, агенты в возрасте 6-21 лет также могут контактировать в образовательных учреждениях с агентами своего возраста, агенты в возрасте 22-65 лет – на работе. Пример графов в разных структурах контактов в АОМ приведен на Рис. 8.

(2) **Заражение.** В рамках модели предполагается, что вирус передается между агентами, соединенными ребром графа. Заражение при близком контакте описывается кусочно-постоянным параметром $\alpha(t)$. В зависимости от структуры контакта, параметр α умножается на соответствующую константу w_α (для домохозяйств $w_\alpha = 3$, для образовательных учреждений и работы $w_\alpha = 0.6$, для общественных мест $w_\alpha = 0.3$). Таким образом, вероятность передачи вируса для каждой контактной сети различная. Симптомные и бессимптомные агенты передают вирус одинаково.

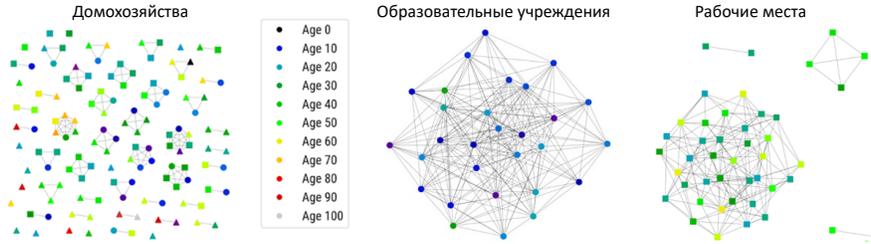


Рис. 8. Пример схем связей для подвыборки 127 человек из 10 000. Все люди присутствуют в домохозяйствах (слева), в том числе некоторые не имеют связи с домом. Как правило, эти люди, включая учителей, присутствуют в школьной сети (кружки); другое подмножество присутствует в сетях рабочих мест (квадраты); некоторые люди не входят ни в школьную, ни в рабочую сеть (треугольники). Цветом отмечена возрастная категория агента.

- (3) **Прогрессирование заболевания.** Переход из одной стадии заболевания $S, E, A, I, M, H, C, R, D$ в другую в момент времени t контролируется параметрами, зависящими от возраста, т.е. чем старше агент, тем он более уязвим (например, вероятность проявлять симптомы после заражения $p_{sym} = 0.5 + 0.05i \in [0.5, 0.9]$, где i – номер возрастной группы). Взаимосвязь эпидемиологических состояний обозначена на Рис. 9. Продолжительность каждой стадии заболевания представляет собой случайную логнормальную величину с различными средними и параметрами дисперсии (см. Таблицу 2).

Параметр	Описание	Распределение
t_{inc}	Количество дней с момента контакта до того, как агент станет заразным.	$\text{LogN}(4.6, 4.8)$ [78]
t_{sym}	Количество дней с момента, когда агент стал заражен, до проявления симптомов.	$\text{LogN}(1, 0.9)$ [78]
t_{rec1}	Продолжительность болезни для бессимптомных и легких случаев.	$\text{LogN}(8, 2)$ [79]
t_{rec2}	Продолжительность болезни для тяжелых и критических случаев.	$\text{LogN}(14, 2.4)$ [80]
t_{inf}	Количество дней, за которое агент переходит из легкого состояния в тяжелое.	$\text{LogN}(6.6, 4.9)$ [78]
t_{hosp}	Количество дней, за которое агент переходит из тяжелого состояния в критическое.	$\text{LogN}(3, 7.4)$ [81]
t_{crit}	Длительность пребывания агента в критическом состоянии.	$\text{LogN}(6.2, 1.7)$ [80]

ТАБЛИЦА 2. Параметры продолжительность болезни в днях в АОМ.

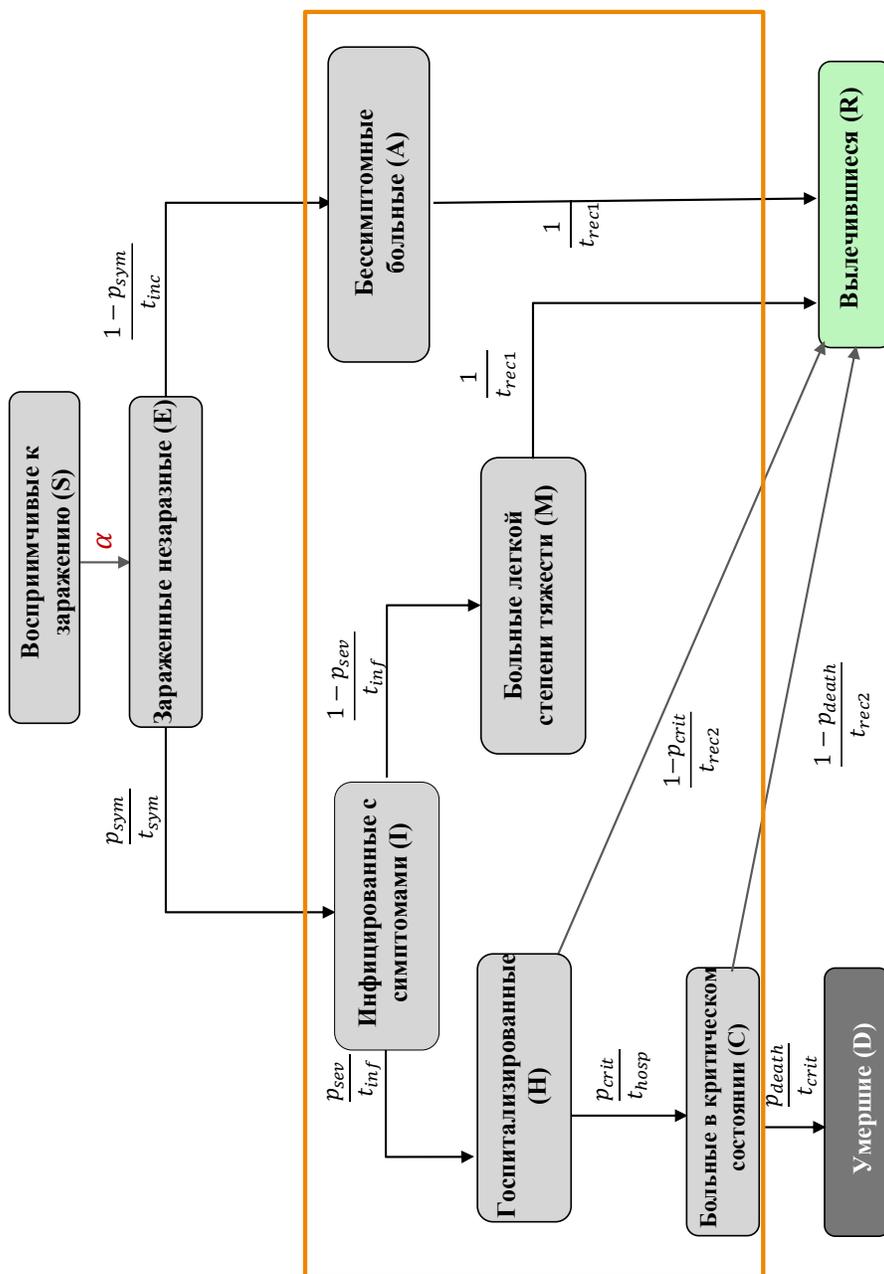


Рис. 9. Диаграмма состояний агентов в АОМ. Оранжевой рамкой обозначены те состояния, находясь в которых агент имеет возможность получить положительный тест на COVID-19.

- (4) **Тестирование агентов** проводится согласно ежедневным статистическим данным о количестве проведенных тестов в регионе. Шанс быть протестированным на COVID-19 $p(t, i)$ зависит от эпидемиологического статуса агента и определяется в ходе решения обратной задачи (см. Раздел 6.2.2). Положительный результат могут получить агенты, находящиеся в симптомном, бессимптомном, в легкой форме, госпитализированном, критическом состояниях (на рис. 9 эти состояния обведены в оранжевую рамку). В модели предполагается, что вероятность тестирования агентов с симптомами выше, чем у бессимптомных больных.
- (5) **Введение сдерживающих эпидемию мер.** В модели возможно введение карантинных мер как для всех контактных слоев, так и для каждого в отдельности. Это может быть сделано двумя способами: либо изменением значения параметра $\alpha(t)$ (в случае введения обязательной меры ношения масок или социального дистанцирования), либо удалением ребер в графах (в случае введения самоизоляции и дистанционной работы).

Обзор [82] и анализ литературы в нем в пользу эффективности ношения масок показал, что места и периоды времени где использование масок является обязательным или широко распространенным, показали значительно более низкий уровень передачи вируса в сообществе. Имеющиеся данные свидетельствуют о том, что практически повсеместное использование немедицинских масок при нахождении в общественных местах, в сочетании с дополнительными мерами общественного здравоохранения, может успешно снизить базовый индекс репродукции вируса до уровня ниже 1, тем самым снижая распространение в обществе. Экономический анализ показывает, что обязательное ношение масок может добавить 1 триллион долларов к ВВП США [83].

Загружаются все необходимые параметры и статистические данные, создается искусственная популяция с учетом распределения по возрастам в регионе. Далее агенты соединяются в контактные сети. Затем начинается цикл по времени: на каждом шаге (временной интервал равен одному дню) обновляется эпидемиологический статус агента с учетом его структуры контактов и введенных ограничительных мер (самоизоляция, закрытие общественных мест, ношение масок и т.д.).

Модель основана на нескольких предположениях:

- (1) Изначально иммунитета нет ни у одного агента.
- (2) Часть агентов находятся в инкубационном периоде $E(0)$.
- (3) Умереть могут только пациенты, находящиеся на ИВЛ (критическое состояние $C(t)$).
- (4) Состояние агента обновляется в конце дня с учетом усреднения возможного пребывания в разных структурах контактов (например, 8 часов на работе, 4 часа в общественном месте и 12 часов дома за весь моделируемый день).

4.4. Базовый индекс репродукции вируса \mathcal{R}_0 для АОМ. В АОМ базовый индекс репродукции вируса \mathcal{R}_0 определяется следующим образом:

$$\mathcal{R}_0(t) = \frac{d \cdot I_{new}(t)}{I_{active}(t)}.$$

Здесь $I_{new}(t)$ – количество новых случаев инфицирования в день t , $I_{active}(t)$ – количество людей с активным инфекционным заболеванием в день t , d – среднее время инфицирования.

4.5. Выводы. АОМ позволяют преодолевать ограничения дифференциальных моделей, используя стохастическую природу эпидемического процесса, возможности введения ограничительных мер в модель, особенности моделируемого региона и учета индивидуальности индивидуума (агента). На основе результатов моделирования в рамках АОМ получено понимание эффективности мер для разных географических и демографических условий для каждого выявленного штамма SARS-CoV-2. Для повышения достоверности результатов необходимо использовать адекватные статистические данные (эпидемиологические, демографические, транспортные), уточнять границы чувствительных параметров моделей, иметь большие вычислительные мощности для реализации моделирования и прогнозирования.

5. МОДЕЛИ «ИГРЫ СРЕДНЕГО ПОЛЯ»

В данном разделе приведено описание комбинации SIR-модели (с учетом пространственной неоднородности) и агентной модели. Распределение агентов в пространственной SIR-модели описывается уравнением Колмогорова-Фоккера-Планка (КФП) (20) (Раздел 5.2), а управление агентов описывается уравнением Гамильтона-Якоби-Беллмана (ГЯБ) (25) (Раздел 5.3). Совместное рассмотрение КФП и ГЯБ образуют систему уравнений игр среднего поля (ИСП) (Раздел 5.4).

5.1. Введение. J.M. Smith и G.R. Price в 1973 году [84] разработали концепцию эволюционной устойчивой стратегии, которая является центральной концепцией теории игр. Затем V. Jovanovic и R.W. Rosenthal в 1988 году [85] доказали существование равновесия в анонимных последовательных играх. Только в 2006-2007 годах J.-M. Lasry и P.-L. Lions [86] предложили и обосновали концепцию ИСП, которая используется в физике, экономике, метеорологии, социальных и биологических процессах. Основные задачи, которые рассматриваются в теории ИСП:

- исследование существования и единственности решения ИСП, при этом ИСП формализуется в виде системы двух уравнений в частных производных первого порядка и задачи управления для некоторой динамической системы, используемая для вывода уравнения ГЯБ, соответствующего модели поведения агентов (подробнее см. Раздел 5.4);
- построение приближенного равновесия в игре конечного числа агентов;
- исследование сходимости приближенных равновесий в играх конечного числа лиц к решению ИСП при стремлении числа игроков к бесконечности;
- анализ конкретных эпидемиологических задач при помощи методологии ИСП.

В работе [87, 88] исследуется сходимость решений для игр с нелинейными процессами устойчивого типа с переменными коэффициентами. Концепция ИСП находится на пересечении теории среднего поля и нелинейного марковского управления [89, 90]

С 2015 года опубликовано множество работ, использующих ИСП для описания эпидемических процессов распространения гриппа, острая респираторная вирусная инфекция, COVID-19 [91, 92, 93, 94].

5.2. Уравнение Колмогорова-Фоккера-Планка. Рассмотрим однородные марковские случайные процессы в \mathbb{R}^n , для которых оператор переходных вероятностей задаётся переходной плотностью $p(t, x, y)$. Зададим оператор \mathbf{Q} , действующий на функцию $f(x)$ в \mathbb{R}^n :

$$(\mathbf{Q}f)(x) = \int_{\mathbb{R}^n} q(x, y) f(y) dy.$$

При $t > 0$, $t \rightarrow 0$ переходная плотность $p(t, x, y)$ стремится к δ -функции (в смысле слабого предела обобщенных функций): $\lim_{t \rightarrow 0} p(t, x, y) = \delta(x - y)$. Если существует предел (также обобщенная функция)

$$q(x, y) = \lim_{h \rightarrow 0} \frac{p(h, x, y) - \delta(x - y)}{h},$$

то уравнение Колмогорова принимает вид:

$$\frac{\partial p(t, x, y)}{\partial t} = \int_{\mathbb{R}^n} p(t, x, z) q(z, y) dz.$$

В случае, когда \mathbf{Q} – дифференциальный оператор второго порядка с непрерывными коэффициентами (это означает, что $q(x, y)$ есть линейная комбинация первых и вторых производных $\delta(x - y)$ с непрерывными коэффициентами), а матрица коэффициентов σ^{ij} перед вторыми производными является симметричной и положительно определенной в каждой точке, то уравнение Колмогорова будет совпадать с уравнением Фоккера-Планка [95]:

$$(11) \quad \frac{\partial p(t, x, y)}{\partial t} = \frac{1}{2} \sum_{i,j} \frac{\partial^2}{\partial y^i \partial y^j} (\sigma^{ij}(y) p(t, x, y)) - \sum_j \frac{\partial}{\partial y^j} (b^j(y) p(t, x, y)).$$

Вектор b^j в физической литературе называется вектором сноса, а матрица σ^{ij} – тензором диффузии.

Уравнение КФП (11) используется для расчёта плотности вероятности в СДУ:

$$(12) \quad d\mathbf{X}_t = \mathbf{b}(\mathbf{X}_t, t) dt + \boldsymbol{\sigma}(\mathbf{X}_t, t) d\mathbf{B}_t.$$

Здесь $\mathbf{X}_t \in \mathbb{R}^M$ – функция состояния системы, а $\mathbf{B}_t \in \mathbb{R}^M$ – стандартное M -мерное броуновское движение. В нашем случае вектор \mathbf{X}_t описывает состояния агентов в каждой из камер и является решением системы (12). Более подробно см. систему (15). Если начальное распределение задано как $\mathbf{X}_0 \sim p(0, \mathbf{x})$, то плотность вероятности $p(t, \mathbf{x})$ состояния системы \mathbf{X}_t является решением уравнения КФП (11).

5.3. Уравнение Гамильтона-Якоби-Беллмана. Уравнение ГЯБ – дифференциальное уравнение в частных производных, играющее центральную роль в теории оптимального управления. Рассмотрим задачу оптимального управления на промежутке времени $[0, T]$:

$$\psi = \min_u \left\{ \int_0^T C[x(t), u(t)] dt + G[x(T)] \right\}.$$

Здесь C и G – липшиц-непрерывные функции стоимости, определяющие соответственно интегральную и терминальную часть функционала, $x(t)$ – вектор,

определяющий состояние системы в каждый момент времени с заданным начальным значением $x(0)$, $u(t)$ – вектор управления.

Эволюция системы под действием управления $u(t)$ описывается следующим образом:

$$\dot{x}(t) = F[x(t), u(t)],$$

и уравнения ГЯБ принимают следующий вид:

$$(13) \quad \frac{\partial \psi(x, t)}{\partial t} + \min_u \{ \nabla \psi(x, t) \cdot F(x, u) + C(x, u) \} = 0,$$

с начальным значением в конечный момент времени T

$$(14) \quad \psi(x, T) = G(x).$$

Неизвестная в этом уравнении беллмановская функция значения $\psi(x, t)$, отвечающая максимальному значению, которое можно получить, ведя систему из состояния (x, t) оптимальным образом до момента времени T . Тогда оптимальное решение в начальный момент времени – значение $\psi = \psi(x(0), 0)$.

Уравнение (13) называется уравнением Беллмана [96], также известное как уравнение динамического программирования. Также Беллманом было введено понятие *принципа оптимальности*: оптимальная стратегия имеет свойство, что какими бы ни были начальное состояние и начальное решение, последующие решения должны составлять оптимальный курс действий по отношению к состоянию, полученному в результате первого решения. Иными словами, оптимальная стратегия зависит только от текущего состояния и цели, и не зависит от предыстории.

При рассмотрении задачи с непрерывным временем полученные уравнения могут рассматриваться как продолжение более ранних работ в области теоретической механики, связанных с уравнением Гамильтона-Якоби.

5.4. Модель игры среднего поля. Предположение, что агенты в популяции рациональны (то есть обладают способностью находиться в состоянии $X(t)$ и иметь возможность его изменить), позволяет рассматривать задачу управления с большим количеством участников. Динамика отдельного индивидуума удовлетворяет дифференциальному уравнению Ито

$$dX_i^N(t) = b(t, X_i^N(t), \theta^N(t, X_i^N(t)), u_i)dt + \sigma(t, X_i^N(t), \theta^N(t, X_i^N(t)))dW_i^N(t).$$

Здесь $i \in 1, \dots, N$, W_i^N – независимые стандартные винеровские процессы, $u_i(t, X_i^N(t))$ – стратегия i -го агента и $\theta^N(t, X_i^N(t))$ – эмпирическая мера распределения агентов в системе в момент времени t [97]. В предположении, что функции b и σ непрерывны во времени и одинаковы для всех агентов, в работе [97] показано, что когда количество агентов в системе чрезвычайно велико $N \rightarrow \infty$, мы можем заменить массу отдельных индивидов репрезентативным агентом, состояние которого определяется следующим уравнением

$$(15) \quad dX(t) = b(t, X(t), p(t, X(t)), u(t, X(t)))dt + \sigma(t, X(t), p(t, X(t)))dW(t).$$

Здесь $X(t) : [0, T] \rightarrow \Omega$; $p(t, X(t)) : \theta^N(t, X_i^N(t)) \xrightarrow{N \rightarrow \infty} p(t, X(t))$ – это распределение агента по пространству состояний Ω в момент времени t и $u(t, X(t))$ – стратегия репрезентативного агента, обеспечивающая равновесие по Нэшу системы взаимодействующих агентов (ни один агент не может увеличить выигрыш, изменив свою стратегию, если другие агенты своих стратегий не меняют)

и минимизирующая функционал

$$(16) \quad J(p, u) = \mathbf{E} \left[\int_0^T C(t, X(t), p(t, X(t)), u(t, X(t))) dt + G(X(T), p(T, X(T))) \right].$$

Данный подход к контролю над популяцией с большим количеством взаимодействующих агентов получил название ИСП [86, 98, 97]. Модель ИСП предполагает, что каждый агент выбирает свою рациональную стратегию $u(t, x)$ с учетом своего положения и распределения $p(t, x) : [0, T] \times \Omega \rightarrow \mathbb{R}$ других агентов. В работе [98] было показано, что при постоянном σ в уравнении (15), характеризующем стохастический характер равновесия процесса взаимодействия агентов, распределение агентов $p(t, x)$ подчиняется уравнению КФП

$$(17) \quad \frac{\partial p}{\partial t} - \frac{1}{2} \sigma^2 \Delta p + \nabla(pu) = 0 \quad \text{в } [0, T] \times \Omega$$

с начальными условиями

$$(18) \quad p(0, x) = p_0(x) \quad \text{на } \Omega$$

и граничными условиями типа Неймана

$$(19) \quad \frac{\partial p}{\partial x} = 0 \quad \forall t \text{ и } x \in \Gamma_\Omega.$$

Здесь Γ_Ω – гладкая граница области Ω .

Исследованы модели с динамикой каждого игрока, задаваемой марковской цепью с непрерывным временем и конечным числом состояний [99], а также марковским процессом общего вида [100, 101]. В этом случае уравнение (17) и ГЯБ заменяются на специально построенные задачи оптимизации. Альтернативный подход к ИСП называется вероятностным и связан с исследованием задачи управления для динамической системы, описываемой нелинейным марковским процессом. В этом случае динамика и интегральная часть выигрыша зависят от распределения игроков в текущий момент времени. Распределение игроков определяется по решению задачи оптимизации [102].

Недостатком вероятностного подхода является зависимость определения решения от выбора сопутствующего вероятностного пространства. Этот недостаток пытается преодолеть минимаксный подход, в рамках которого задача поиска решения ИСП сводится к решению игры бесконечного числа лиц, при этом динамика определяется распределением оптимальных траекторий.

ИСП достаточно гибки, чтобы улавливать межклассовое взаимодействие при распространении эпидемии, при котором несколько органов власти осведомлены о рисках лиц, принимающих локальные решения, а отдельные лица осведомлены о рисках агентов (государство, региональное правительство), принимающих глобальные решения [93, 94]. В следующих разделах приведены уравнения КФП и ГЯБ для описания распределения и управления SIR-модели распространения эпидемии.

5.4.1. *Уравнение КФП для SIR-модели.* Для достаточно больших ($N \rightarrow \infty$) популяций SIR-модель можно интерпретировать как приближение среднего поля вероятностной модели клеточного автомата [103, 104]. Запишем аналог начально-краевой задачи для уравнения КФП (17)-(19) в случае SIR-модели (1).

Введем плотность $p_i(t, x) : [0, T] \times [0, 1] \rightarrow \mathbf{R}$ распределения агентов в группах S, I и R , где $i \in \{S, I, R\}$. Переменная $x \in \Omega = [0, 1]$ характеризует физическое дистанцирование следующим образом: $x = 0$ означает, что агент склонен соблюдать физическую дистанцию, а $x = 1$ означает противоположное. Определим функции $u_i(t, x) : [0, T] \times [0, 1] \rightarrow \mathbf{R}$, $i \in \{S, I, R\}$, которые описывают скорость перемещения в пространственной области репрезентативного агента в каждой группе населения. Принимая во внимание стохастический характер взаимодействия агентов (15), уравнения КФП (17)-(19) для функций $p_i(t, x)$ запишутся в виде системы уравнений в частных производных:

$$(20) \quad \begin{cases} \frac{\partial p_S}{\partial t} + \nabla(p_S u_S) + \alpha p_S p_I - \sigma_S^2 \Delta p_S / 2 = 0, \\ \frac{\partial p_I}{\partial t} + \nabla(p_I u_I) - \alpha p_S p_I + \beta p_I - \sigma_I^2 \Delta p_I / 2 = 0, \\ \frac{\partial p_R}{\partial t} + \nabla(p_R u_R) - \beta p_I - \sigma_R^2 \Delta p_R / 2 = 0. \end{cases}$$

Здесь $\sigma_i \geq 0$, $i \in \{S, I, R, C\}$. Начальные (18) и граничные (19) условия остаются без изменений

$$(21) \quad p_i(0, x) = p_{i0}(x) \quad \forall x \in [0, 1].$$

$$(22) \quad \frac{\partial p_i}{\partial x} = 0 \quad \forall t \in [0, T] \text{ и } x = 0, 1.$$

Начальные условия p_{i0} совпадают с плотностью нормального распределения со средним, демонстрирующим склонность соблюдать физическую дистанцию, и дисперсией, описывающей строгость соблюдения ограничений. Например, восприимчивая популяция m_S не склонна соблюдать ограничения, поэтому среднее в распределении сместится к 1, а инфицированная m_I , наоборот, стремится соблюдать ограничения и карантин.

Отметим, что система (20) удовлетворяет закону действующих масс, а именно

$$\frac{\partial}{\partial t} \int_0^1 (p_S(t, x) + p_I(t, x) + p_R(t, x)) dx = 0.$$

5.4.2. Уравнение ГЯБ для SIR-модели. Определим целевой функционал (16) следующим образом [92]:

$$(23) \quad J(p, u) = \int_0^T \int_0^1 \left(\sum_{i \in \{S, I, R\}} \frac{r_i}{2} p_i |u_i|^2 + \frac{c}{2} (p_S + p_I + p_R)^2 \right) dx dt + \frac{1}{2} \int_0^1 p_I^2(T, x) dx.$$

Здесь r_i и c неотрицательные константы. Первое слагаемое, характеризующее кинетическую энергию, описывает цену перемещения агентов со скоростью u_i за весь промежуток времени $[0, T]$. Чем выше значение r_i , тем более затратно агентам перемещаться между группами (например, $r_S = r_R = 1$, $r_I = 10$ означает, что агентам из инфицированной группы труднее передвигаться). Второе слагаемое в (23) контролирует скопление всего населения в одном месте. Это может повысить риск вспышек заболевания и их более быстрого и широкого распространения.

6. АЛГОРИТМЫ ЧИСЛЕННОГО РЕШЕНИЯ

Математические модели в эпидемиологии характеризуются своими коэффициентами и начальными условиями, которые индивидуальны для каждой моделируемой популяции. В следующих разделах будут приведены алгоритмы, входящие в комплекс программ COVID-19 моделирования распространения коронавирусной инфекции в Новосибирской области. В Разделе 6.1 приведены алгоритмы численного решения прямых задач, в которых при заданных эпидемиологических параметрах и начальных условиях требуется определить распределение различных категорий населения (восприимчивые, инфицированные, госпитализированные, вылеченные и т.п.) в течение всего времени моделирования. В Разделе 6.2 описываются алгоритмы численного решения обратных задач, в которых требуется определить неизвестные параметры моделей по дополнительной информации о количестве выявленных инфицированных, протестированных, госпитализированных и умерших в фиксированные моменты времени.

6.1. Прямые задачи. Прямая задача для модели эпидемиологии состоит в определении количества или плотности восприимчивой, протестированной, инфицированной, госпитализированной и т.п. групп населения по заданным эпидемиологическим коэффициентам и начальным условиям. Численные методы решения прямых задач для математических моделей, основанных на дифференциальных уравнениях (Раздел 3), характеризуются конечно-разностной структурой (методы Эйлера, конечных разностей, элементов, объемов). Полученное численное решение чувствительно к разбиению области моделирования (по времени, по пространству и времени). Обзор численных методов решения задач игр среднего поля приведен в работах [105, 106, 107].

В комплексе программ *SBRAS-COVID-19* для численного решения прямых задач использованы:

- для моделей $SIRC_T$ и $SEIR-HCD$: метод Рунге-Кутты 4-го порядка.
- для агентных моделей: методы графов, Монте-Карло и программные комплексы Covasim [68].
- для моделей ИСП: метод конечных разностей дробных шагов, конечных объемов.

6.2. Обратные задачи. Параметры моделей в эпидемиологии (базовый индекс репродукции вируса $\mathcal{R}_0(t)$, вероятность госпитализации, тестирования, выздоровления, количество бессимптомных носителей и т.п.) заданы приближенно и нуждаются в уточнении в каждом конкретном регионе для увеличения точности моделирования и прогнозирования распространения эпидемии. С этой целью мы решаем обратную задачу, которая состоит в определении вектора неизвестных параметров \mathbf{q} математической модели эпидемиологии по дополнительной информации f_k^i о количестве выявленных ($i = 1$), госпитализированных ($i = 2$), умерших ($i = 3$) в фиксированные моменты времени t_k , $k = 1, \dots, K$ (подробнее о постановке обратной задач для моделей $SEIR-HCD$ и АОМ см. Разделы 6.2.1 и 6.2.2 соответственно). Обратная задача может быть сформулирована в виде задачи минимизации целевого функционала [108, 109]:

$$(27) \quad J(\mathbf{q}) = \langle A(\mathbf{q}) - \mathbf{f}, A(\mathbf{q}) - \mathbf{f} \rangle.$$

Здесь A – нелинейный оператор обратной задачи, $\mathbf{f} = \left\{ f_k^i \right\}_{\substack{i \in \mathcal{I}, \\ k=1, K}}$, \mathcal{I} – множество измеряемых состояний системы.

Обратная задача является некорректной, а именно ее решение может быть неединственным и/или неустойчивым [110, 111, 109]. Для разработки алгоритма регуляризации решения обратной задачи проводится анализ чувствительности параметров, который позволяет упорядочить параметры q_i по степени чувствительности по отношению к вариации данных обратной задачи [112, 113, 114, 115].

В комплексе программ для численного решения обратных задач использованы:

- Поиск глобального минимума функционала: генетический алгоритм, метод дифференциальной эволюции, метод имитации отжига, метод роя частиц, метод древовидных оценок Парзена, тензорная оптимизация, стохастический градиентный спуск, глубокие нейронные сети, обучение с подкреплением.
- Для уточнения минимума: методы минимальных ошибок, наискорейшего спуска, Левенберга-Марквардта, Бройдена-Флетчера-Гофбардто-Шанно, Нелдера-Мида.

Для исследования прямых и обратных задач программного комплекса *SBRAS-COVID-19* использованы теоретические результаты:

(1) SIR-модели:

- Теория нелинейных операторных уравнений Вольтерра в банаховых и гильбертовых пространствах – локальная корректность, корректность в окрестности точного решения, единственность и условная устойчивость, сходимость дискретных аналогов к точному решению.
- Устойчивость обратных задач – А.Н. Тихонов [116], М.М. Лаврентьев [117], С.К. Годунов [118], С.И. Кабанихин [119].
- Методы идентифицируемости – структурная (дифференциальной алгебры, передаточной функции, разложения в ряд Тейлора) и практическая (Монте-Карло, корреляционной матрицы) [112].
- Методы анализа чувствительности: ортогональный, сингулярного разложения, собственных значений [112, 113, 115].
- Теория регуляризации – методы М.М. Лаврентьева [120], А.Н. Тихонова [121] и В.К. Иванова [122], итерационная регуляризация [123, 124].
- Методы учета априорной информации – И.И. Еремин, В.В. Васин, А.Г. Ягола.
- Исследование сходимости природоподобных алгоритмов – теорема стохастической сходимости [125].
- Теория больших данных – тензорное разложение [126, 127, 128, 129].

(2) Агентные модели:

- Методы анализа чувствительности – баесовский подход, регрессионный анализ [114, 130].
- Методы высокопроизводительных вычислений – распределенные вычисления, MPI.

6.2.1. *Алгоритм решения обратной задачи для SEIR-HCD.* Предположим, что известна дополнительная информация о количестве симптомных выявленных случаях $(1 - b_k)h_k$, критических C_k и умерших g_k в SEIR-HCD модели в фиксированные моменты времени t_k , $k = 1, \dots, K$:

$$(28) \quad E(t_k; \mathbf{q}) = (1 - b_k)h_k, \quad C(t_k; \mathbf{q}) = C_k, \quad D(t_k; \mathbf{q}) = g_k, \quad k = 1, \dots, K.$$

Здесь $\mathbf{q} = (\alpha_E(t), \alpha_I(t), \varepsilon_{HC}, \mu, E_0, I_0)$ – вектор неизвестных параметров модели, h_k – количество выявленных случаев в Новосибирской области (см. Рис. А.1 в Приложении А), b_k – процент бессимптомных случаев по результатам ПЦР.

Обратная задача для SEIR-HCD модели (8)-(9), (28) состоит в определении вектора параметров \mathbf{q} по дополнительной информации (28). Обратная задача для SEIR-HCD модели была сведена к задаче минимизации целевого функционала

$$(29) \quad J(\mathbf{q}) = \sum_{k=1}^K \left[\frac{1}{(1 - b_k)^2 h_k^2} \left(\frac{1}{t_{inc}} E(t_{k-1}; \mathbf{q}) - (1 - b_k)h_k \right)^2 + \frac{(C(t_k; \mathbf{q}) - C_k)^2}{C_k^2} + \frac{(D(t_k; \mathbf{q}) - g_k)^2}{g_k^2} \right].$$

Здесь первое слагаемое $t_{inc}^{-1} E(t_{k-1}; \mathbf{q})$ описывает количество бессимптомных носителей вируса COVID-19, которые в день t_k попадают в состояние выявленных симптомных инфицированных (см. Рис. 5в). В функционал $J(\mathbf{q})$ также включена информация о критических $C(t_k; \mathbf{q})$ и умерших $D(t_k; \mathbf{q})$ в результате COVID-19 случаев, информация о которых доступна в открытых источниках (см. Приложение А).

Анализ идентифицируемости модели (8)-(9), (28) показал, что параметр β , описывающий долю инфицированных индивидуумов, которые переносят заболевание без осложнений, является наименее идентифицируемым [131], поэтому в качестве дополнительной информации использованы данные медицинского центра «Инвитро» (см. Приложение А), которые в полтора раза уменьшили доверительный интервал по количеству выявленных случаев для полученных параметров после решения обратной задачи.

Для численного решения задачи минимизации $J(\mathbf{q})$ (29) использовалась следующая последовательность шагов:

- (1) Подготовка и обработка реальных данных для вычисления обратной задачи:
 - 1.1. Данные по новым выявленным случаям заражения COVID-19 h_k , критическим (требующих подключения аппарата ИВЛ) C_k и умершим g_k .
 - 1.2. Данные по проценту бессимптомных выявленных случаев b_k от общего числа выявленных случаев заражения COVID-19, а также его прогноз на период моделирования, построенного с помощью полносвязной нейронной сети с 30 входных нейронов (входной слой), которые соответствуют 30 временным точкам в обучающей выборке, 64 нейронов в скрытом слое с функцией активации $ReLU(x) = x$, $x \geq 0$, и 1 нейроном в выходном слое (предсказание на 1 день вперед).

- 1.3. Индекс самоизоляции от Яндекса $a(t)$, а также его прогноз на период моделирования, построенного с помощью полносвязной нейронной сети (см. выше ее описание).
 - 1.4. Данные по проценту индивидуумов с антителами к COVID-19 от медицинского центра «Инвитро» $\beta(t)$, а также его прогноз на период моделирования, построенного с помощью полносвязной нейронной сети, организация которой указана выше.
- (2) Определение границ параметров для искомого вектора \mathbf{q} (см. Таблицу 1, колонку 3).
 - (3) Уточнение вектора неизвестных параметров \mathbf{q} путём решения обратной задачи алгоритмом усвоения данных. Обратная задача решается для каждого 30-дневного отрезка реальных данных (тренировочные данные). Далее, по восстановленным параметрам \mathbf{q} для текущего 30-дневного отрезка осуществляется прогноз эпидемиологических данных на следующие 7 дней (валидационные данные). Прогноз осуществляется путём решения прямой задачи (3.5.2) по схеме, указанной на рис. 10 для постоянного набора восстановленных на последнем временном промежутке параметров \mathbf{q} . Далее новые тренировочный и валидационный периоды смещаются на 7 дней и для этих периодов решается новая обратная задача. И так далее до тех пор пока не закончатся реальные данные.

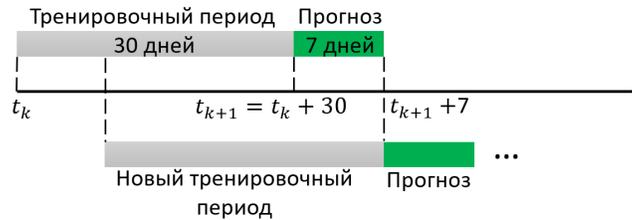


Рис. 10. Алгоритм усвоения данных при решении обратной задачи для SEIR-HCD модели (8)-(9), (28). Тренировочный период (30 дней) – уточнение параметров \mathbf{q} . Прогноз (7 дней) рассчитывается при найденных \mathbf{q} . Новый период сдвигается на 7 дней и снова решается обратная задача.

- 3.1. Для решения задачи минимизации целевого функционала (29) использовался пакет глобальной оптимизации OPTUNA, в реализации которого лежат природоподобные алгоритмы, метод древовидных оценок Парзена и тензорная оптимизация.
- 3.2. Полученный на предыдущем шаге глобальный оптимум уточнялся с помощью локальных методов градиентного типа [123].

Для применения градиентных алгоритмов важно использовать выпуклую функцию оптимизации, которая корректно работает со сглаженными данными. Так как для модели SEIR-HCD не требуется большого количества данных, а сама модель ориентирована на динамику крупных групп в популяции, то сглаженные данные позволяют определить решение обратной задачи в допустимой погрешности.

6.2.2. *Алгоритм решения обратной задачи для АОМ.* Предположим, что для АОМ, описанной в разделе 4.3.1, известна дополнительная информация о количестве ежедневно выявленных случаев h_k , проведенных ПЦР-тестов в регионе $T(t_k)$, критических C_k и умерших g_k случаев в фиксированные моменты времени $t_k, k = 1, \dots, K$. Обратная задача для модели АОМ состоит в определении вектора параметров $\mathbf{q} = (\alpha, \alpha_d(t), \alpha_c(t), p(t), E_0)$ по дополнительной информации $h_k, T(t_k), C_k, g_k, k = 1, \dots, K$. Здесь α – параметр контагиозности вируса, $\alpha_d(t)$ – дни изменения параметра α , $\alpha_c(t)$ – значения, на которые изменяется параметр α в дни $\alpha_d(t)$, $p(t)$ – шанс быть протестированным (зависит от возрастной группы), E_0 – начальное количество бессимптомных инфицированных.

АОМ для каждой возрастной группы и состояния показывает динамику изменения, что требует большого количества данных. Сглаженные данные приводят к огрублению модели с сведением ее к дифференциальному аналогу. Важно использовать несглаженные данные. В связи с этим обратная задача сводится к задаче минимизации целевого функционала:

$$(30) \quad J(\mathbf{q}) = \sum_{k=1}^K \left[\frac{|Y(t_k; \mathbf{q}) - h_k|}{h_k} + \frac{|C(t_k; \mathbf{q}) - C_k|}{C_k} + \frac{|D(t_k; \mathbf{q}) - g_k|}{g_k} \right].$$

Здесь $Y(t; \mathbf{q})$ – количество моделируемых выявленных случаев COVID-19 в результате ПЦР тестирования.

Была исследована идентифицируемость агентной модели на основе байесовского подхода [114] с тремя неизвестными параметрами: параметр контагиозности α , начальное количество инфицированных E_0 и параметр тестирования p по статистическим данным о количестве выявленных h_k , смертей g_k в результате COVID-19 и пациентов, находящихся в отделении интенсивной терапии C_k . В результате удалось уменьшить границы поиска параметра α более чем в 2 раза, в то время как границы параметров p и E_0 остались неизменными.

Алгоритм численного решения задачи минимизации функционала $J(\mathbf{q})$ (30):

- (1) Подготовка и обработка реальных данных для вычисления обратной задачи (файлы с данными доступны на сайте <http://covid19-modeling.ru/> в разделе "Данные"):
 - 1.1. Заполнение пропусков в данных о ежедневно выявленных случаях, проведенных ПЦР-тестов, критических и умерших методами обратной и прямой экстраполяции.
 - 1.2. Сглаживание данных с помощью Гауссовского фильтра для уменьшения флуктуации данных перед использованием их в решении обратной задачи.
- (2) Уточнение границ изменения неизвестных параметров \mathbf{q} методами анализа чувствительности.
- (3) Поэтапное восстановление вектора параметров \mathbf{q} с шагом 30 дней, в ходе которого значения кусочно-постоянного параметра α ($\alpha_d(t), \alpha_c(t)$) находились последовательно один за другим. Таким образом, значения параметров, восстановленные на предыдущем шаге, использовались в последующем запуске алгоритма минимизации функционала, который представлял собой комбинацию методов глобального (библиотека <https://optuna.org/OPTUNA>) и локального (градиентные) метода оптимизации.

- (4) Построение сценариев распространения количества выявленных случаев COVID-19:
 - 4.1. Экстраполяция количества ожидаемых ПЦР-тестов в регионе $T(t)$ с помощью регрессионных моделей SARIMA (Раздел 2.2) в комбинации с методами машинного обучения (Раздел 2.3) на 45 дней.
 - 4.2. Для уточненных параметров и ограничительных мер (введение ограничительных мер описано в Разделе 4.3.1) решаем прямую задачу.
- (5) Построение доверительных интервалов:
 - 5.1. Решаем прямую задачу с идентифицированными параметрами 10000 раз с помощью метода Монте-Карло для данных с добавлением аддитивного гауссовского шума в 10% (вычислительное время решения прямой задачи для АОМ на кластере составляет 10 сек).
 - 5.2. Для каждого дня считаем квантили уровня 0.1, 0.5, 0.9 (строим функцию распределения случайной величины выявленных случаев и выбираем ее значения в точках 0.1, 0.5 и 0.9).
 - 5.3. Получаем 3 массива точек с шагом один день, по которым строим медианное значение (квантиль уровня 0.5, сплошная линия на графиках в Разделе 7.2).

7. РЕЗУЛЬТАТЫ ЧИСЛЕННЫХ РАСЧЕТОВ И ЗАКЛЮЧЕНИЕ

В данном разделе будет проведен пример экстраполяции временного ряда, описывающего количество ожидаемых ПЦР-тестов в Новосибирской области (Раздел 7.1), анализ численных расчетов для АОМ и SEIR-HCD модели в Новосибирской области (Раздел 7.2), влияние социальной дистанции на распространение эпидемии в Новосибирской области на основе ИСП (Раздел 7.3), выводы (Раздел 7.4) и направления дальнейшей работы (Раздел 7.5).

7.1. Прогнозирование количества ожидаемых ПЦР-тестов. Был построен прогноз временного ряда количества ежедневных ПЦР-тестов $\mathcal{T}(n)$ в Новосибирской области с 18.01.2021 по 17.02.2021 на основе исторических наблюдений с 05.05.2020 по 17.01.2021 с помощью моделей SARIMA(1, 1, 2)(0, 1, 1)₇, Хольта-Винтерса [8] (подробнее см. Раздел 2.4) и машинного обучения линейной регрессии (см. Рис. 11).

После проведения кросс-валидации были использованы следующие параметры моделей (кроме модели SARIMA, так как она имеет большое число вспомогательных параметров) и характеристики:

- В модели линейной регрессии использовались признаки: значения тестов за последние 7 дней, средние значения за последние 7, 14 и 21 день, средние значения по дням недели, среднее значение за все предыдущие дни.
- Модель Хольта-Винтерса: $\alpha = 0.0216$, $\beta = 0.099$, $\gamma = 0.3987$.

SARIMA показала на предсказаниях разных промежутков в среднем не лучшие, но наиболее устойчивые результаты (усредненная абсолютная ошибка для модели SARIMA равна 1454.73, для модели Хольта-Винтерса – 1646.49, для модели линейной регрессии – 1586.34).

В дальнейших расчетах была использована модель SARIMA для экстраполяции временных рядов статистических данных в эпидемиологии, которые

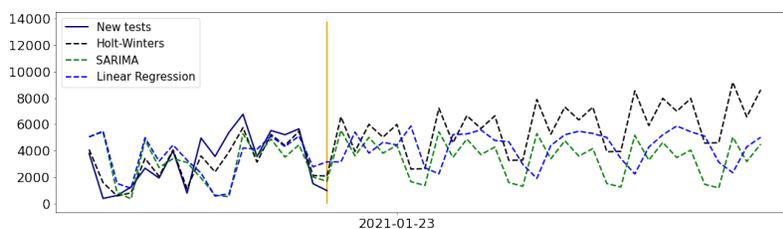


Рис. 11. Результаты предсказания временного ряда ежедневно проводимых ПЦР-тестов $\mathcal{T}(t)$ в Новосибирской области (сплошная синяя линия) на месяц вперед с 18.01.2021 (вертикальная оранжевая линия) моделями Хольта-Винтерса (черная пунктирная линия), линейной регрессии (синяя пунктирная линия) и SARIMA (зеленая пунктирная линия).

являются входными параметрами для дифференциальных и агентных моделей.

7.2. Анализ расчета сценариев распространения COVID-19 в Новосибирской области. Численные расчеты данного раздела приведены для Новосибирской области. Данные, использованные при анализе и построении моделей, описаны в Приложении А.

В Табл. 3 приводятся восстановленные параметры α (АОМ) и $\alpha_I, \alpha_E, \varepsilon_{НС}, \mu$ (SEIR-HCD), характеризующие распространение COVID-19 в Новосибирской области. Дата изменения $\alpha_d(t)$ параметров (результат решения обратной задачи для АОМ, см. Раздел 6.2.2) коррелируют с ограничительными мерами, принятыми в регионе (с интервалом 5-14 дней). Например, возобновление удаленного режима с 1 ноября 2020 года в учебных заведениях повлекло уменьшение значений параметра контагиозности α в структурах с 14 ноября 2020 года.

В последней колонке Табл. 3 приведены 95% доверительные интервалы (95% CI) полученных значений, вычисленные по 30 испытаниям оптимизационного алгоритма OPTUNA. В АОМ параметр контагиозности вируса $\alpha(t)$ имеет различные значения для каждой структуры контактов (домохозяйства, образовательные учреждения, работа и общественные места). Доверительные интервалы для каждой структуры контактов затруднительно оценить ввиду отсутствия детализированных данных в исследуемом регионе. В SEIR-HCD модели восстановленные значения параметров являются усредненными в регионе, так как разделения на структуры контактов не подразумевалось.

Дата изменения	Символ	Восстановленное значение			Доверительный интервал
		Домохоз-ва	Образование и работа	Обществ. места	
23-05-2020 (начало	α	0.02524	0.00505	0.00252	(0.02879, 0.09887) (0.03673, 0.15023)
	α_I		0.06383		
	α_E		0.09348		

периода)	ε_{HC}		0.00824		(0.00452, 0.01196)
	μ		0.22189		(0.06458, 0.3792)
17-06-2020	α	0.0223	0.00446	0.00223	
	α_I		0.06701		(0.0409, 0.09313)
	α_E		0.08834		(0.02224, 0.15445)
	ε_{HC}		0.01607		(0.01461, 0.01753)
	μ		0.24682		(0.09869, 0.39495)
12-07-2020	α	0.02788	0.00558	0.00279	
	α_I		0.0643		(0.04158, 0.08702)
	α_E		0.07852		(0.02035, 0.13669)
	ε_{HC}		0.00896		(0.00812, 0.0098)
	μ		0.2681		(0.10798, 0.42823)
06-08-2020	α	0.02181	0.00436	0.00218	
	α_I		0.04452		(0.03538, 0.05367)
	α_E		0.03547		(0.00641, 0.06453)
	ε_{HC}		0.00208		(0.00133, 0.00283)
	μ		0.21982		(0.07017, 0.36946)
31-08-2020	α	0.02466	0.00493	0.00247	
	α_I		0.06572		(0.04372, 0.08773)
	α_E		0.08158		(0.02038, 0.14278)
	ε_{HC}		0.00115		(0.00103, 0.00128)
	μ		0.24302		(0.08012, 0.40592)
25-09-2020	α	0.02948	0.0059	0.00295	
	α_I		0.07915		(0.03855, 0.11974)
	α_E		0.13327		(0.05215, 0.2144)
	ε_{HC}		0.00854		(0.00784, 0.00925)
	μ		0.22748		(0.07004, 0.38493)
20-10-2020	α	0.02948	0.0059	0.00295	
	α_I		0.09853		(0.05844, 0.13862)
	α_E		0.13321		(0.05673, 0.20969)
	ε_{HC}		0.00979		(0.00929, 0.01029)
	μ		0.25076		(0.07987, 0.42165)
14-11-2020	α	0.02754	0.00551	0.00275	
	α_I		0.07429		(0.04941, 0.09917)
	α_E		0.09587		(0.0368, 0.15495)
	ε_{HC}		0.01365		(0.01239, 0.01491)
	μ		0.22794		(0.08333, 0.37254)
09-12-2020	α	0.02447	0.00489	0.00245	
	α_I		0.066		(0.03925, 0.09274)
	α_E		0.09382		(0.02578, 0.16187)
	ε_{HC}		0.01057		(0.00983, 0.01131)
	μ		0.2593		(0.09546, 0.42313)
03-01-2021	α	0.02447	0.00489	0.00245	
	α_I		0.11286		(0.09334, 0.13239)
	α_E		0.05064		(0.00117, 0.1001)
	ε_{HC}		0.00995		(0.00948, 0.01043)
	μ		0.20234		(0.06234, 0.34234)
	α	0.02332	0.00466	0.00233	

28-01-2021

	α_I		0.06664		(0.04204, 0.09123)
	α_E		0.07783		(0.01485, 0.1408)
	ε_{HC}		0.01158		(0.01093, 0.01223)
	μ		0.2535		(0.11234, 0.39465)
22-02-2021	α	0.02332	0.00466	0.00233	
	α_I		0.05395		(0.02886, 0.07904)
	α_E		0.1069		(0.03337, 0.18044)
	ε_{HC}		0.00858		(0.00796, 0.0092)
	μ		0.24984		(0.1092, 0.39048)
19-03-2021	α	0.0398	0.00796	0.00398	
	α_I		0.05269		(0.03706, 0.06832)
	α_E		0.07891		(0.03338, 0.12445)
	ε_{HC}		0.0084		(0.00773, 0.00906)
	μ		0.25176		(0.11776, 0.38576)
13-04-2021	α	0.02788	0.00558	0.00279	
	α_I		0.04975		(0.02219, 0.07731)
	α_E		0.10684		(0.03534, 0.17834)
	ε_{HC}		0.00381		(0.00222, 0.00539)
	μ		0.27165		(0.11115, 0.43214)
08-05-2021	α	0.02175	0.00435	0.00217	
	α_I		0.05668		(0.03625, 0.0771)
	α_E		0.09091		(0.03979, 0.14203)
	ε_{HC}		0.00271		(0.00195, 0.00346)
	μ		0.28275		(0.14527, 0.42023)
02-06-2021	α	0.0406	0.00812	0.00406	
	α_I		0.08184		(0.04872, 0.11495)
	α_E		0.09535		(0.02064, 0.17007)
	ε_{HC}		0.00527		(0.00333, 0.0072)
	μ		0.20956		(0.07153, 0.34758)
27-06-2021	α	0.03063	0.00613	0.00306	
	α_I		0.09507		(0.06195, 0.12818)
	α_E		0.08282		(0.02295, 0.14268)
	ε_{HC}		0.02107		(0.02011, 0.02203)
	μ		0.22876		(0.07595, 0.38158)
22-07-2021	α	0.02843	0.00569	0.00284	
	α_I		0.06641		(0.04038, 0.09245)
	α_E		0.08157		(0.02461, 0.13852)
	ε_{HC}		0.00774		(0.00641, 0.00907)
	μ		0.25786		(0.0942, 0.42151)
16-08-2021	α	0.02449	0.0049	0.00245	
	α_I		0.05153		(0.02638, 0.07667)
	α_E		0.08968		(0.0224, 0.15697)
	ε_{HC}		0.00366		(0.00232, 0.005)
	μ		0.21537		(0.09108, 0.33965)
10-09-2021	α	0.03273	0.00655	0.00327	
	α_I		0.06993		(0.04723, 0.09263)
	α_E		0.0706		(0.01428, 0.12692)
	ε_{HC}		0.01063		(0.00999, 0.01128)

	μ	0.24317			(0.07392, 0.41242)
08-10-2021	α	0.04294	0.00859	0.00429	
	α_I	0.08198			(0.05147, 0.11248)
	α_E	0.13857			(0.0659, 0.21124)
	ε_{HC}	0.01178			(0.01126, 0.01231)
	μ	0.24395			(0.09619, 0.3917)

Таблица 3: Восстановленные параметры распространения COVID-19 в Новосибирской области с 13.03.2020 по 30.09.2021. Здесь α – параметр контагиозности, α_E – доля заражения между бессимптомной и восприимчивой группами населения, α_I – доля заражения между инфицированным и восприимчивым населением, ε_{HC} – доля госпитализированных случаев, которым требуется подключение ИВЛ, μ – доля смертельных случаев в результате COVID-19.

На рисунке 12 представлены численные результаты моделирования распространения ежедневно выявленных случаев. Зелеными точками отмечены реальные данные с 12.03.2020 по 16.10.2021, которые участвовали в решении обратной задачи. Красной линией представлены результаты для АОМ, синей – для камерной SEIR-HCD модели. При моделировании был сделан прогноз на 40 дней (с 17.10.2021 по 26.11.2021) с учетом сохранения карантинных мер. Результат прогнозирования был сравнен с реальными данными 08.11.2021: количество выявленных случаев согласно реальным данным – 400 человек, результатам АОМ – 362, камерной SEIR-HCD модели – 387.



Рис. 12. Моделирование распространения ежедневно выявленных случаев COVID-19 в Новосибирской области с 12.03.2020 по 30.11.2021 (расчеты проведены 16.10.2021). Синяя линия – SEIR-HCD модель, красная линия – АОМ, зеленая линия – реальные данные с 12.03.2020 по 16.10.2021.

Для анализа эффективности ограничительных мер были изучены следующие сценарии распространения ежедневно выявленных случаев COVID-19 (рис. 13):

- Нерабочие дни с 30.10 по 07.11, в течение которых уменьшается на 40% людей на работе и учебе, потом увеличивается заболеваемость из-за привезенных случаев в 2 раза — синяя линия;
- Локдаун с 30.10.2021 по 14.11.2021, в течение которого полностью закрыты образовательные учреждения, 50% общественных мест и 50% рабочих переведены на удаленный режим работы — красная линия;
- Ничего не предпринимать (базовый сценарий) — зеленая линия.

Результаты численных расчетов для сценариев развития представлены на рисунке 13. Так, введенная мера о нерабочих днях (синяя линия) может привести к самому неблагоприятному сценарию развития ситуации, при которой к 8 ноября ожидается 416 выявленных случаев. Базовый сценарий развития (зеленая линия) предполагает 356 выявленных случаев к 8 ноября, в рамках которого количество людей в общественных местах уменьшится из-за введения QR-кодов. Наиболее оптимистичный сценарий стоит ждать в случае более серьезного локдауна (красная линия), при котором к 8 ноября ожидается 349 выявленных случаев.



Рис. 13. Сценарии распространения ежедневно выявленных случаев COVID-19 в Новосибирской области с 17.10.2021 по 30.11.2021.

Синяя линия – нерабочие дни с 30.10 по 07.11 с учетом оттока 40% населения в эндемичные районы, красная линия – локдаун с 30.10.2021 по 14.11.2021, зеленая линия – базовый сценарий на 16.10.2021.

Синие точки – реальные данные выявленных случаев COVID-19 по 16.10.2021, используемые в моделировании, черные точки – реальные данные выявленных случаев COVID-19 с 17.10.2021 по 08.11.2021, используемые для валидации.

Для каждого из сценариев были проведены расчеты индекса репродукции вируса R_0 для АОМ, описанного в разделе 3.5. Результаты численных расчетов представлены на рис. 14. В случае сценария «объявления нерабочей недели» базовый индекс репродукции ожидается наибольшим по значению, что впоследствии может привести к повышенному числу выявленных и умерших случаев, а также нагрузку на систему здравоохранения (синяя линия). Более серьезные ограничения до 14.11 (красная линия) значительно уменьшат значение индекса репродукции, однако через 2 недели он снова может увеличиться и переступить порог в 1. Базовый сценарий развития (черная линия) к 30 ноября предполагает стабильную ситуацию в регионе.

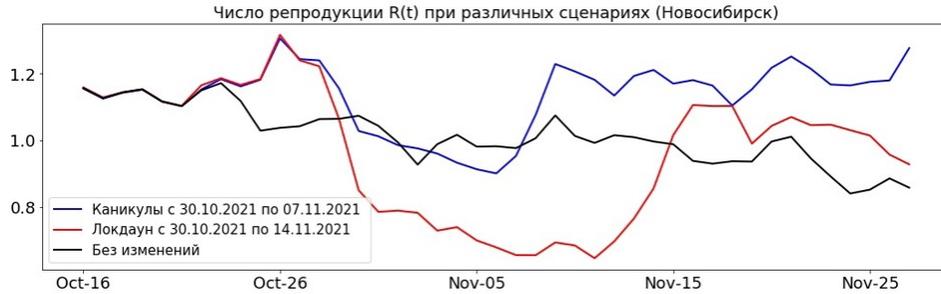


Рис. 14. Сценарии изменения индекса репродукции вируса $\mathcal{R}_0(t)$ для Новосибирской области с 17.10.2021 по 30.11.2021. Синяя линия – объявление нерабочих дней с 30.10 по 07.11 и отток 40% населения из региона, красная линия – закрытие учебных заведений и 50% общественных и рабочих мест с 30.10.2021 по 14.11.2021, черная линия – базовый сценарий на 16.10.2021.

7.3. Анализ влияния социальной дистанции на распространение эпидемии. На основе SIR-модели была введена модель $SIRC_r$ [132], в которой $C_r(t)$ – группа населения, имеющая перекрестный иммунитет (т.е. иммунную память к коронавирусу):

$$(31) \quad \begin{cases} \frac{dS}{dt} = -\alpha S(t)I(t) + \gamma C_r(t), & t > 0, \\ \frac{dI}{dt} = \alpha S(t)I(t) + \varepsilon \alpha C_r(t)I(t) - \beta I(t), \\ \frac{dR}{dt} = (1 - \varepsilon)\alpha C_r(t)I(t) + \beta I(t) - \delta R(t), \\ \frac{dC_r}{dt} = \delta R(t) - \alpha C_r(t)I(t) - \gamma C_r(t). \end{cases}$$

Значения параметров для Новосибирской области, определенные с помощью решения обратной задачи для модели (31) по дополнительной информации о количестве ежедневно выявленных в результате ПЦР-теста h_k и вылеченных от COVID-19 R_k , $k = 1, \dots, 100$, с 1 мая 2020 года по 8 августа 2020 года приведены в Табл. 4. Ввиду того, что моделируемый период характеризуется появлением нового штамма, параметры восстанавливались для каждого из двух интервалов: 1 мая - 30 июня и 1 июля - 8 августа 2020 года. Доверительный интервал вычислялся на основе средних значений при 30 испытаниях алгоритма древовидных оценок Парзена, реализованного в пакете программ OPTUNA [133].

Численные расчеты данного раздела основываются на работе [134]. Для сведения модели (31) к модели ИСП (см. Раздел 5.4) вместо численности населения в каждой группе $S(t)$, $I(t)$, $R(t)$, $C_r(t)$ введем плотность распределения людей внутри этих групп $p_S(t, x)$, $p_I(t, x)$, $p_R(t, x)$, $p_{C_r}(t, x)$. Переменная x изменяется в пределах $[0, 1]$ и означает соблюдение карантинных мер: $x = 1$ означает, что человек полностью придерживается карантинных ограничений, включая вакцинацию, а $x = 0$ означает, что он не придерживается никаких ограничений.

Описание параметров	Символ	01.05–30.06		01.07–08.08	
		Ср. знач.	95% ДИ	Ср. знач.	95% ДИ
Скорость передачи инфекции	α	0.56097	[0.46295, 0.65899]	0.58534	[0.4957, 0.67498]
Скорость выздоровления инфицированного населения	β	0.52564	[0.42824, 0.62304]	0.61078	[0.5216, 0.69996]
Вероятность приобретения перекрестного иммунитета	δ	0.40716	[0.30428, 0.51004]	0.53123	[0.43801, 0.62445]
Скорость, с которой перекрестно-иммунная популяция снова становится восприимчивой	γ	0.4562	[0.35009, 0.56231]	0.49515	[0.40041, 0.58989]
Средняя вероятность повторного заражения человека с перекрестным иммунитетом	ε	0.5184	[0.42326, 0.61354]	0.47598	[0.38828, 0.56368]
Начальное количество инфицированных	I_0	451	[427, 475]	–	–
Начальное количество вылеченных от COVID-1	R_0	549	[460, 637]	–	–
Начальное количество индивидумов с перекрестным иммунитетом	C_{r0}	6	[5, 7]	–	–

ТАБЛИЦА 4. Эпидемиологические параметры для $SIRC_r$ модели (31) для Новосибирской области с 01.05.2020 по 08.08.2020, полученные с помощью решения обратной задачи. ДИ – доверительный интервал.

Введем функции $u_S(t, x)$, $u_I(t, x)$, $u_R(t, x)$, $u_{C_r}(t, x)$, обозначающие изменение состояния человека в шкале соблюдения карантинных мер, вызываемое окружающей обстановкой, воздействием СМИ, вакцинацией и организационными мероприятиями. Тогда уравнения КФП (17)-(19) примут вид:

$$(32) \quad \begin{cases} \frac{\partial p_S}{\partial t} + \nabla(p_S u_S) + \alpha p_S p_I - \gamma p_{C_r} - \sigma_S^2 \Delta p_S / 2 = 0, \\ \frac{\partial p_I}{\partial t} + \nabla(p_I u_I) - \alpha p_S p_I - \varepsilon \alpha p_{C_r} p_I + \beta p_I - \sigma_I^2 \Delta p_I / 2 = 0, \\ \frac{\partial p_R}{\partial t} + \nabla(p_R u_R) - (1 - \varepsilon) \alpha p_{C_r} p_I - \beta p_I + \delta p_R - \sigma_R^2 \Delta p_R / 2 = 0, \\ \frac{\partial p_{C_r}}{\partial t} + \nabla(p_{C_r} u_{C_r}) - \delta p_R + \alpha p_{C_r} p_I + \gamma p_{C_r} - \sigma_{C_r}^2 \Delta p_{C_r} / 2 = 0 \end{cases}$$

с начальными

(33)

$$p_i(0, x) = \frac{A_i}{B_i} \left(\exp \left(-\frac{(x - x_i^c)^2}{2(\sigma_i^c)^2} \right) / \sigma_i^c \sqrt{2\pi} + a_i x^2 + b_i (1 - x)^2 \right), \quad x \in [0, 1]$$

и краевыми условиями

$$(34) \quad \frac{\partial p_i}{\partial x}(t, 0) = \frac{\partial p_i}{\partial x}(t, 1) = 0, \quad t \in [0, T], \quad i \in \{S, I, R, C_r\}.$$

Здесь A_i – доля текущей фракции по отношению к общей численности населения в начальный момент времени; B_i – коэффициент нормировки, равный интегралу по $[0, 1]$ от выражения в скобках; $a_i = \exp\left(-\frac{(1-x_i^c)^2}{2(\sigma_i^c)^2}\right) (1 - x_i^c) / (2(\sigma_i^c)^3 \sqrt{2\pi})$ and $b_i = \exp\left(-\frac{(x_i^c)^2}{2(\sigma_i^c)^2}\right) (x_i^c) / (2(\sigma_i^c)^3 \sqrt{2\pi})$ для обеспечения выполнения граничных условий (19) для p_{0i} . В численных расчетах мы использовали следующие значения A_i , x_i^c и σ_i^c

$$\begin{aligned} A_S &= 0.999757; & x_S^c &= 0.8; & \sigma_S^c &= 0.1; \\ A_I &= 0.000204; & x_I^c &= 0.2; & \sigma_I^c &= 0.1; \\ A_R &= 0.000039; & x_R^c &= 0.7; & \sigma_R^c &= 0.2; \\ A_{C_r} &= 0.000000; & x_{C_r}^c &= 0.3; & \sigma_{C_r}^c &= 0.2. \end{aligned}$$

Функционал стоимости организационных мероприятий и социально-экономических потерь

$$(35) \quad J(p_{sirc}, u_{sirc}) = \int_0^T \int_0^1 \left(\tilde{C}(t, x, u_{sirc}(t, x)) + G(t, x, p_{sirc}(t, x)) \right) dx dt.$$

Здесь $\tilde{C}(t, x, u_{sirc}(t, x))$ – стоимость проводимых организационных мероприятий: вакцинация, карантинные мероприятия, введение QR-кодов и другие, $G(t, x, p_{sirc}(t, x))$ – стоимость социальных и экономических потерь для группы населения в позиции x на момент времени t , $p_{sirc} = (p_S, p_I, p_R, p_{C_r})$, $u_{sirc} = (u_S, u_I, u_R, u_{C_r})$.

Минимум функционала $J(p_{sirc}, u_{sirc})$ (35) при выполнении уравнений КФП определяется методом множителей Лагранжа $\psi_S(t, x)$, $\psi_I(t, x)$, $\psi_R(t, x)$, $\psi_{C_r}(t, x)$, которые удовлетворяют системе ГЯБ типа (13):

$$(36) \quad \begin{cases} \frac{\partial \psi_S}{\partial t} + \frac{\sigma_S^2}{2} \Delta \psi_S + u_S \cdot \nabla \psi_S = \tilde{C}_S(t, x, p_{sirc}, \psi_{sirc}), \\ \frac{\partial \psi_I}{\partial t} + \frac{\sigma_I^2}{2} \Delta \psi_I + u_I \cdot \nabla \psi_I = \tilde{C}_I(t, x, p_{sirc}, \psi_{sirc}), \\ \frac{\partial \psi_R}{\partial t} + \frac{\sigma_R^2}{2} \Delta \psi_R + u_R \cdot \nabla \psi_R = \tilde{C}_R(t, x, p_{sirc}, \psi_{sirc}), \\ \frac{\partial \psi_{C_r}}{\partial t} + \frac{\sigma_{C_r}^2}{2} \Delta \psi_{C_r} + u_{C_r} \cdot \nabla \psi_{C_r} = \tilde{C}_{C_r}(t, x, p_{sirc}, \psi_{sirc}) \end{cases}$$

с начальными условиями на конце периода времени оптимизации T

$$\psi_i(T, x) = 0, \quad x \in [0, 1], \quad i \in \{S, I, R, C_r\}$$

и краевыми условиями

$$\frac{\partial \psi_i}{\partial x}(t, 0) = \frac{\partial \psi_i}{\partial x}(t, 1) = 0, \quad t \in [0, T], \quad i \in \{S, I, R, C_r\}.$$

Дифференциальные части задач КФП и ГЯБ сопряжены между собой и их коэффициенты связаны алгебраическими уравнениями

$$(37) \quad \frac{\partial \tilde{C}}{\partial u_i}(t, x, u_{sirc}) = -\frac{\partial \psi_i}{\partial x}(t, x), \quad i \in \{S, I, R, C_r\}$$

из которых определяются параметры u_S, u_I, u_R, u_{C_r} , отражающие текущее изменение позиций групп населения по отношению к соблюдению карантинных мер.

Решение задач для уравнений КФП (32)-(34) и ГЯБ (36) совместно с уравнением связи 37 дает решение задачи оптимального поведения агентов при эпидемии с указанием строгости мер или возможности ослабления карантинных ограничений. На рис. 15 приведено сравнение результатов моделирования числа инфицированного населения для моделей $SIRC_r$ (31) и ИСП (уравнения КФП и ГЯБ) со статистическими данными в течение 100 дней с 1 мая по 8 августа 2020 года в Новосибирской области.

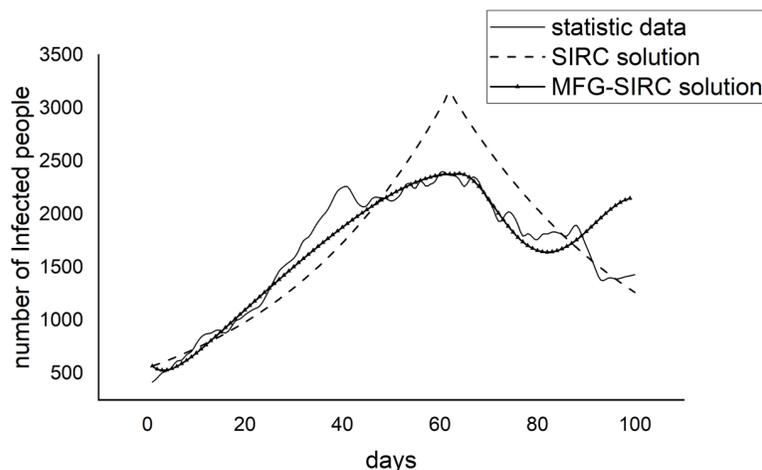


Рис. 15. Сравнение результатов моделирования числа выявленных случаев COVID-19 в Новосибирской области со статистическими данными с 01.05.2020 по 08.08.2020.

Сплошная линия – статистические данные, пунктирная линия – $SIRC_r$ модель, линия с треугольниками – ИСП модель.

7.4. Выводы. В работе проанализированы различные классы моделей распространения инфекционных заболеваний, имеющие различный горизонт предсказуемости эпидемиологических процессов:

- Модели машинного обучения (ИНС, регрессионные модели) прогнозируются на основе имеющегося набора данных эпидемиологическую ситуацию на 1 день вперед. Последующие предсказания (более 1 дня) опираются на полученные ранее прогнозы, что увеличивает доверительный интервал полученных предсказаний. Надежность прогноза на следующий день зависит от качества подобранных обучающих выборок и характеристик (наиболее показательные комбинации данных).

- Дифференциальные SIR модели качественно описывают вспышку заболеваемости с учетом диффузионных процессов в случае использования уравнений в частных производных (и без учета в случае систем ОДУ). При отсутствии информации о возможном появлении нового штамма заболевания, который можно учесть в параметрах модели, такие модели не способны описать новую вспышку заболеваемости.
- Агентно-ориентированные модели при условии достаточно подробных данных способны получать разнообразные сценарии развития эпидемической ситуации на несколько дней вперед (в работе максимальная «дальность» предсказаний равна 45 дням с учетом изменчивости вируса SARS-CoV-2). Такие модели способны уловить новую вспышку заболевания при наличии дополнительной информации о новом штамме (например, если данный штамм появился в некотором регионе/стране ранее).
- Модели игры среднего поля для определенного набора параметров распространения заболевания (коэффициенты и начальные условия уравнений КФП и ГЯБ) оценивают качество влияния ограничительных мер на сдерживание распространения эпидемии. Для получения предсказаний на основе моделей ИСП необходимо решать задачу уточнения параметров с учетом оптимального управления, что является сложной задачей.

Сочетание моделей SIR и АОМ позволяет строить более разнообразные сценарии распространения пандемии COVID-19. А именно, используя результаты решения обратной задачи SEIR-HCD модели как дополнительную информацию для АОМ, мы уточняем сценарии с учетом ограничительных мер с помощью АОМ. Рассчитанные сценарии используются как дополнительная информация для дальнейшего уточнения параметров SEIR-HCD модели. Также можно перевести на материалистический язык разные с содержательной точки зрения сценарии и сравнить их.

При усложнении SIR-модели, вводя возрастные разграничения в популяции и пространственные перемещения, мы получим первое приближение АОМ.

Для более точных рекомендаций поддержки принятия решений и более комплексного моделирования мы применяем подход ИСП и ИСПУ (управление играми среднего поля, mean-field-type control). А именно, влияние вакцинации, характеризующее социальные настроения в регионе, учитывается при построении распределений населения в эпидемиологических группах.

История заболевания (временные ряды эпидемиологических данных) составляет обучающие множества для методов машинного обучения.

7.5. Направления дальнейшей работы. Необходимо совершенствовать методы искусственного интеллекта в приложении к моделированию COVID-19 и расчетов сценариев выхода из пандемии с учетом социальных, экономических и экологических процессов, включая анализ ситуации в различных регионах и группах, обработку данных, а также создавать комплексную модель на основе SIR, АОМ и ИСП (ИСПУ).

ФИНАНСОВАЯ ПОДДЕРЖКА

Работа выполнена при поддержке Математического Центра в Академгородке, соглашение с Министерством науки и высшего образования Российской Федерации № 075-15-2019-1675, Российского научного фонда (проект № 18-71-10044) и Российского фонда фундаментальных исследований (проект № 20-51-54004).

БЛАГОДАРНОСТИ

В различные годы исследования по моделированию в эпидемиологии (туберкулеза, ВИЧ, COVID-19) выполнялись по грантам:

- Российского научного фонда (проект № 18-71-10044);
- Российского фонда фундаментальных исследований (проекты № 21-51-10003, № 20-51-54004, № 18-31-20019, № 17-51-540004);
- Президента РФ (№ МК-814.2019.1, № МК-4994.2021.1.1);
- Мэрии города Новосибирска на предоставление грантов в форме субсидий в сфере научной и инновационной деятельности 2021;
- Министерства образования Республики Казахстан (№ ИРН АР09260317).

REFERENCES

- [1] М.А. Кондратьев, *Методы прогнозирования и модели распространения заболеваний*, Компьютерные исследования и моделирование, **5:5** (2013), 863-882.
- [2] В.В. Захаров, Ю.Е. Балыкина, *Балансовая модель эпидемии COVID-19 на основе процентного прироста*, Информатика и автоматизация, **20:5** (2021), 1034-1064.
- [3] H.S. Burkom, S.P. Murphy, G. Shmueli, *Automated time series forecasting for biosurveillance*, Statistics in Medicine, **26:22** (2007), 4202-4218.
- [4] R.E. Serfling, *Methods for current statistical analysis of excess pneumonia-influenza deaths*, Public Health Reports, **78:6** (1963), 494-506.
- [5] Д. Бокс, Г. Дженкинс, *Анализ временных рядов: прогноз и управление*, Выпуск **1**, Мир, Москва (1974).
- [6] P.P. Dabral, M.Z. Murry, *Modelling and forecasting of rainfall time series using SARIMA*, Environmental Processes, **4** (2017), 399-419 .
- [7] G.E.P. Box, D.R. Cox., J. Roy, *An analysis of transformations. (With discussion)*, Journal of the Royal Statistical Society. Series B (Methodological), **26** (1964), 211-252.
- [8] T. Williams, *Adaptive Holt-Winters forecasting*, The Journal of the Operational Research Society, **38** (1987), 553-560.
- [9] Y. In, J. Jung, *Simple averaging of direct and recursive forecasts via partial pooling using machine learning*, International Journal of Forecasting, **38(4)** (2022), 1386-1399.
- [10] T. Januschowski, Y. Wang, K. Torkkola, T. Erkkilä, H. Hasson, J. Gasthaus, *Forecasting with trees*, International Journal of Forecasting, **38(4)** (2022), 1473-1481.
- [11] P. Sebastiani, K.D. Mandl, P. Szolovits, I.S. Kohane, M.F. Ramoni, *A Bayesian dynamic model for influenza surveillance*, Statistics in Medicine, **25:11** (2006), 1803-1816.
- [12] Y.Le Strat, F. Carrat, *Monitoring epidemiologic surveillance data using hidden Markov models*, Statistics in Medicine, **18:24** (1999), 3463-3478.
- [13] M. Wiecek, J. Silka, M. Woźniak, *Neural network powered COVID-19 spread forecasting model*, Chaos, Solitons & Fractals, **140** (2020), 110203.
- [14] G. Shmueli, S.E. Fienberg, *Current and potential statistical methods for monitoring multiple data streams for biosurveillance*, Statistical Methods in Counterterrorism: Game Theory, Modeling, Syndromic Surveillance, and Biometric Authentication, Springer Science + Business Media, New York, (2006), 109-140.
- [15] E.S. Gardner, *Exponential smoothing: the state of the art*, Journal of Forecasting, **4:1** (1985), 1-28.
- [16] J.D. Hamilton, *Time Series Analysis*, Princeton University Press, Princeton, NJ, 1994, 820.

- [17] F. Petropoulos, D. Apiletti, V. Assimakopoulos et al., *Forecasting: theory and practice*, International Journal of Forecasting, **38**(3) (2022), 705-871.
- [18] D. Bernoulli, *Essai d'une nouvelle analyse de la mortalité causée par la petite vérole et des avantages de l'inoculation pour la prévenir*, Mem. Math. and Phys. de l'Acad. Roy. Sci., in: Hist. de l'Acad. Roy. Sci., Ann., Paris, 1760.
- [19] N. Bacaër, *A Short History of Mathematical Population Dynamics*, Springer, London, 2011.
- [20] F. Brauer, *Mathematical epidemiology: Past, present, and future*, Infectious Disease Modelling, **2** (2017), 113–127.
- [21] R. Ross, *The Prevention of Malaria* - 2nd edn, John Murray, London, 1911.
- [22] A.J. Lotka, *Undamped oscillations derived from the law of mass action*, J. Amer. Chem. Soc., **42** (1920), 1595–1599.
- [23] V. Volterra, *Fluctuations in the abundance of a species considered mathematically*, Nature, **118** (1926), 558–560.
- [24] A.G. McKendrick, *Applications of mathematics to medical problems*, Proc. Edinb. Math. Soc, **13** (1926), 98–130.
- [25] W.O. Kermack, A.G. McKendrick, *A contribution to the mathematical theory of epidemics*, Proc. R. Soc. Lond. A, **115** (1927), 700–721.
- [26] E. Pelinovsky, A. Kurkin, O. Kurkina, M. Kokoulina, A. Epifanova, *Logistic equation and COVID-19*, Chaos, Solitons & Fractals, **140** (2020), 110241.
- [27] P. Wang, X. Zheng, J. Li, B. Zhu, *Prediction of epidemic trends in COVID-19 with logistic model and machine learning technics*, Chaos, Solitons & Fractals, **139** (2020), 110058.
- [28] E.M. Koltsova, E.S. Kurkina, A.M. Vasetsky, *Mathematical modeling of the spread of COVID-19 in Moscow*, Computational nanotechnology, **7**:1 (2020), 99-105. In Russian.
- [29] Y. Chen, J. Cheng, Y. Jiang, K. Liu, *A time delay dynamical model for outbreak of 2019-nCoV and the parameter identification*, Journal of Inverse and Ill-posed Problems, **28**:2 (2020), 243-250.
- [30] M.V. Tamm, *COVID-19 in Moscow: prognoses and scenarios*, FARMAKOEKONOMIKA. Modern Pharmacoeconomic and Pharmacoepidemiology, **13**:1 (2020), 43-51.
- [31] E. Unlu, H. Leger, O. Motornyi, A. Rukubayihunga, T. Ishacian, M. Chouiten, *Epidemic analysis of COVID-19 Outbreak and Counter-Measures in France*, MedRxiv, 2020.
- [32] O.I. Krivorotko, S.I. Kabanikhin, N.Y. Zyatkov et al, *Mathematical modeling and forecasting of COVID-19 in Moscow and Novosibirsk region*, Numer. Analys. Appl., **13** (2020), 332–348.
- [33] A.I. Borovkov, M.V. Bolsunovskaya, A.M. Gintciak, T.Yu. Kudryavtseva, *Simulation modelling application for balancing epidemic and economic crisis in the region*, International Journal of Technology, **11**:8 (2020), 1579-1588.
- [34] H.M. Yang, L.P. Junior Lombardi, F.F.M. Castro, A.C. Yang, *Mathematical modeling of the transmission of SARS-CoV-2—Evaluating the impact of isolation in São Paulo State (Brazil) and lockdown in Spain associated with protective measures on the epidemic of CoViD-19*, PLoS ONE, **16**:6 (2021), e0252271.
- [35] S. Margenov, N. Popivanov, I. Ugrinova, S. Harizanov, T. Hristov, *Mathematical and computer modeling of COVID-19 transmission dynamics in Bulgaria by time-dependent inverse SEIR model*, AIP Conference Proceedings, **2333**:090024, (2021).
- [36] C.J. Silva, C. Cruz, D.F.M. Torres, et al, *Optimal control of the COVID-19 pandemic: controlled sanitary deconfinement in Portugal*, Scientific Reports, **11**:3451 (2021).
- [37] F. Ndairou, I. Area, J.J. Nieto, D.F.M. Torres, *Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan*, Chaos Solitons & Fractals, **135**: 109846 (2020).
- [38] I.N. Kiselev, I.R. Akberdin, F.A. Kolpakov, *A delay differential equation approach to model the COVID-19 pandemic*, MedRxiv (2021).
- [39] R.A. Fisher, *The wave of advance of advantageous genes*, Ann. Eugen. **7** (1937) 355–369.
- [40] A.N. Kolmogorov, I.G. Petrovskii, N.S. Piskunov, *A study of the diffusion equation with increase in the amount of substance, and its application to a biological problem*, Bull. Moscow Univ. Math. Mech., **1**:6 (1937), 1–26.
- [41] A. Viguier, A. Veneziani, G. Lorenzo et al, *Diffusion–reaction compartmental models formulated in a continuum mechanics framework: application to COVID-19, mathematical analysis, and numerical study* Comput. Mech., **66** (2020) 1131–1152.
- [42] V.V. Aristov, A.V. Stroganov, A.D. Yastrebov, *Simulation of spatial spread of the COVID-19 pandemic on the basis of the kinetic-advection model*, Physics, **3**:1 (2021), 85-102.

- [43] G. Bärwolff, *A local and time resolution of the COVID-19 propagation – a two-dimensional approach for Germany including diffusion phenomena to describe the spatial spread of the COVID-19 pandemic*, *Physics*, **3** (2021), 536-548.
- [44] Z. Lau, I.M. Griffiths, A. English, K. Kaouri, *Predicting the spatially varying infection risk in indoor spaces using an efficient airborne transmission model*, ArXiv:2012.12267v3 (2021).
- [45] V. Capasso, R.E. Wilson, *Analysis of a reaction-diffusion system modeling man-environment-man epidemics*, *SIAM Journal on Applied Mathematics*, **57**:2, (1997) 327–346.
- [46] A.S. Bratus', V.P. Posvyanskii, A.S. Novozhilov, *Existence and stability of stationary solutions to spatially extended autocatalytic and hypercyclic systems under global regulation and with nonlinear growth rates*, *Nonlinear Analysis: Real World Applications*, **11**:3 (2010), 1897-1917.
- [47] R.M. May, A.L. Lloyd, *Infection dynamics on scale-free networks*, *Phys. Rev. E Stat. Nonlin. Soft. Matter Phys.*, **64**(6 Pt 2):066112 (2001).
- [48] R. Pastor-Satorras, A. Vespignani, *Epidemic spreading in scale-free networks*, *Physical Review Letters*, **86**:14 (2001), 3200–3203.
- [49] M.E. Newman, *Spread of epidemic disease on networks*, *Phys. Rev. E Stat. Nonlin. Soft. Matter Phys.*, **66**:(1 Pt 2) (2002), 016128.
- [50] F. Wong, J.J. Collins, *Evidence that coronavirus superspreading is fat-tailed*, In: Proceedings of the National Academy of Sciences of the United States of America, **117**:47 (2020), 29416–29418.
- [51] L.J.S. Allen, *An Introduction to Stochastic Epidemic Models*, In: Brauer, F., van den Driessche, P., Wu, J. (eds) *Mathematical Epidemiology*. Lecture Notes in Mathematics, Springer, Berlin, Heidelberg, **1945** (2008).
- [52] I. Sazonov, M. Kelbert, M.B. Gravenor, *A two-stage model for the SIR outbreak: Accounting for the discrete and stochastic nature of the epidemic at the initial contamination stage*, *Mathematical Biosciences*, **234**:2 (2011), 108-117.
- [53] I. Sazonov, M. Kelbert, M.B. Gravenor, *Random migration processes between two stochastic epidemic centers*, *Mathematical Biosciences*, **274** (2016), 45-57.
- [54] I. Sazonov, D. Grebennikov, A. Meyerhans, G. Vocharov, *Markov Chain-Based Stochastic Modelling of HIV-1 Life Cycle in a CD4 T Cell*, *Mathematics*, **9** (2021), 2025.
- [55] Братусь А.С., Очеретяная А.С., *Математическая модель распространения эпидемии внутри железнодорожного купейного вагона*, *Мир транспорта*, **18**:6 (2020), 6-29.
- [56] G.Z. Lotova, G.A. Mikhailov, *Numerically statistical investigation of the partly super-exponential growth rate in the COVID-19 pandemic (throughout the world)*, *J. Inverse Ill-Posed Probl.*, **28**:6 (2020), 877-879.
- [57] G.Z. Lotova, G.A. Mikhailov, *Numerical-statistical and analytical study of asymptotics for the average multiplication particle flow in a random medium*, *Comput. Math. and Math. Phys.*, **61** (2021), 1330-1338.
- [58] https://en.wikipedia.org/wiki/Basic_reproduction_number
- [59] P. van den Driessche, J. Watmough, *Further notes on the basic reproduction number*, In: Brauer F., van den Driessche P., Wu J. (eds) *Mathematical epidemiology*, Springer, (2008), 159–178.
- [60] A. Berman, R.J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, Academic, New York (1970).
- [61] T.C. Schelling, *Dynamic models of segregation*, *J. Math. Sociol.*, **1**:2 (1971), 143–186.
- [62] M. Mitchell, P.T. Hraber, J.P. Crutchfield, *Revisiting the edge of chaos: Evolving cellular automata to perform computations*, *Complex Systems*, **7** (1993), 89-130.
- [63] P.H.T. Schimit, *A model based on cellular automata to estimate the social isolation impact on COVID-19 spreading in Brazil*, *Computer Methods and Programs in Biomedicine*, **200** (2021), 105832.
- [64] J. Dai, C. Zhai, J. Ai, J. Ma, J. Wang, W. Sun, *Modeling the Spread of Epidemics Based on Cellular Automata*, *Processes*, **9** (2021), 55.
- [65] P. Patlolla, V. Gunupudi, A.R. Mikler, R.T. Jacob, *Agent-Based Simulation Tools in Computational Epidemiology*, In: 4th International Workshop, International Conference on Innovative Internet Community Systems (I2CS '04), Springer Berlin/Heidelberg (2004), 212–223.

- [66] V.A. Adarchenko, S.A. Baban, A.A. Bragin, K.F. Grebenkin et al, *Modeling the development of the coronavirus epidemic using differential and statistical models*, Preprint, RFNC-VNIITF, **264** (2020).
- [67] A.I. Vlad, T.E. Sannikova, A.A. Romanyukha, *Transmission of acute respiratory infections in a city: agent-based approach*, *Mathematical Biology and Bioinformatics*, **15:2** (2020), 338-356.
- [68] C.C. Kerr, R.M. Stuart, D. Mistry, R.G. Abeyesuriya, K. Rosenfeld, G.R. Hart, et al, *Covasim: An agent-based model of COVID-19 dynamics and interventions*, *PLoS Comput. Biol.*, **17:7** (2021), e1009149.
- [69] Covasim documentation: <https://docs.idmod.org/projects/covasim/en/latest/index.html>
- [70] A. Aleta, D. Martin-Corral, Y. Pastore, A. Piontti, M. Ajelli, M. Litvinova, et al, *Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19*, *Nat. Hum. Behav.*, **4:9** (2020), 964-971.
- [71] M.S.Y. Lau, B. Grenfell, M. Thomas, M. Bryan, K. Nelson, B. Lopman, *Characterizing superspreading events and age-specific infectiousness of SARS-CoV-2 transmission in Georgia, USA*, *Proc. Natl. Acad. Sci. U S A*, **117:3** (2020), 22430-22435.
- [72] A.J. Kucharski, P. Klepac, A.J.K. Conlan, S.M. Kissler, M.L. Tang, H. Fry, et al, *Effectiveness of isolation, testing, contact tracing, and physical distancing on reducing transmission of SARS-CoV-2 in different settings: a mathematical modelling study*, *Lancet Infect. Dis.*, **20:1** (2020), 1151-1160.
- [73] N. Hoertel, M. Blachier, C. Blanco, M. Olsson, M. Massetti, M.S. Rico, et al, *A stochastic agent-based model of the SARS-CoV-2 epidemic in France*, *Nat. Med.*, **26:9** (2020), 1417-1421.
- [74] J. Hellewell, S. Abbott, A. Gimma, N.I. Bosse, C.I. Jarvis, T.W. Russell, et al, *Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts*, *Lancet. Glob. Health*, **8:4** (2020), e488-e496.
- [75] B.F. Nielsen, K. Sneppen, *COVID-19 superspreading suggests mitigation by social network modulation*, *medRxiv*, **2020.09.15.20195008** (2020).
- [76] K.K. Loginov, N.V. Pertsev, *Direct statistical modeling of spread of epidemic based on a stage-dependent stochastic model*, *Mat. Biolog. Bioinform.*, **16:2** (2021), 169-200.
- [77] Household Size, 2019, UN. <https://population.un.org/Household/#/countries/840>
- [78] S.A. Lauer, K.H. Grantz, Q. Bi et al, *The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application*, *Ann. Intern. Med.*, **172:9** (2020), 577-582.
- [79] R. Wölfel, V.M. Corman, W. Guggemos et al, *Virological assessment of hospitalized patients with COVID-2019*, *Nature*, **581** (2020), 465-469.
- [80] R. Verity, L.C. Okell, I. Dorigatti et al, *Estimates of the severity of coronavirus disease 2019: a model-based analysis*, *The Lancet Infectious Diseases*, **20:6** (2020), 669-677.
- [81] D. Wang, B. Hu, C. Hu et al, *Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China*, *JAMA*, **323:11** (2020), 1061-1069.
- [82] J. Howard, A. Huang, Z. Li et al, *An evidence review of face masks against COVID-19*, *Proceedings of the National Academy of Sciences*, **118:4** (2021), e2014564118.
- [83] J. Hatzius, D. Struyven, I. Rosenbery, *Face masks and GDP*, <https://www.goldmansachs.com/insights/pages/face-masks-and-gdp.html>.
- [84] M.J. Smith, G.R. Price, *The logic of animal conflict*, *Nature*, **246** (1973), 15-18.
- [85] B. Jovanovic, R.W. Rosenthal, *Anonymous sequential games*, *Journal of Mathematical Economics*, **17:1** (1988), 77-87.
- [86] J.-M. Lasry, P.-L. Lions, *Mean field games*, *Jpn. J. Math.*, **2:1** (2007), 229-260.
- [87] В.Н. Колокольцов, М.С. Троева, В. Янг, *Игры среднего поля, связанные с процессами устойчивого типа*, *МТИП*, **5:4** (2013), 33-65.
- [88] V.N. Kolokoltsov, J. Li., W. Yang, *Mean field games and nonlinear Markov processes*, *ArXiv:1112.3744* (2012).
- [89] D. Andersson, B. Djehiche, *A maximum principle for SDEs of meanfield type*, *Appl. Math. Optim.*, **63** (2011), 341-356.
- [90] R. Buckdahn, B. Djehiche, J. Li, S. Peng, *Mean-field backward stochastic differential equations: a limit approach*, *Ann. Prob.*, **37:4** (2009), 1524-1565.

- [91] L. Laguzet, G. Turinici., *Global optimal vaccination in the SIR model: Properties of the value function and application to cost-effectiveness analysis*, *Mathematical Biosciences*, **263** (2015), 180-197.
- [92] W. Lee, S. Liu, H. Tembine, W. Li, S. Osher, *Controlling propagation of epidemics via mean-field control*, ArXiv:2006.01249 (2020).
- [93] H. Tembine, *COVID-19: Data-Driven Mean-Field-Type Game Perspective*, *Games*, **11**:4 (2020), 51.
- [94] W. Lee, S. Liu, W. Li, S. Osher, *Mean field control problems for vaccine distribution*, *Research in the Mathematical Sciences*, **9**:51, (2022).
- [95] Н.Н. Боголюбов, Н.М. Крылов, *Об уравнениях Фоккера-Планка, которые выводятся в теории возмущений методом, основанным на спектральных свойствах возмущённого гамильтониана* Записки кафедры математической физики Института нелинейной механики АН УССР, **4** (1939), 5-80.
- [96] R.E. Bellman, *Dynamic Programming*, Princeton, NJ, 1957.
- [97] M. Fischer, *On the connection between symmetric N-player games and mean field games*, *The Annals of Applied Probability*, **27**:2 (2017), 757-810.
- [98] A. Bensoussan, J. Frehse, Ph. Yam, *Mean Field Games and Mean Field Type Control Theory*, Springer Briefs in Mathematics, Springer, New York, 2013.
- [99] D. Gomes, J. Mohr, R.R. Souza, *Continuous time finite state mean-field games*, *Appl. Math. and Opt.*, **68**:1 (2013), 99-143.
- [100] V.V. Kolokoltsov, J.J. Li, W. Yang, *Mean field games and nonlinear Markov processes*, ArXiv:1112.3744v2 (2011).
- [101] V.V. Kolokoltsov, W. Yang, *Sensitivity analysis for HJB equations with an application to a coupled backward-forward system*, ArXiv:1303.6234 (2013).
- [102] R. Carmona, F. Delarue, *Probabilistic analysis of mean-field games*, *SIAM J. Control Optim.*, **51**:4 (2013), 2705-2734.
- [103] L. Berec, *Techniques of spatially explicit individual-based models: Construction, simulation, and mean-field analysis*, *Ecological Modelling*, **150**:1-2 (2002), 55-81.
- [104] P.H.T. Schimit, L.H.A. Monteiro. *On the basic reproduction number and the topological properties of the contact network: An epidemiological study in mainly locally connected cellular automata*, *Ecological Modelling*, **220**:7 (2009), 1034-1042.
- [105] Y. Achdou, I. Capuzzo-Dolcetta, *Mean field games: numerical methods*, *SIAM J. Numer. Anal.*, **48**:3 (2010), 1136-1162.
- [106] Y. Achdou, *Finite difference methods for mean field games*, *Hamilton-Jacobi Equations: Approximations, Numerical Analysis and Applications*, Springer, 1-47 (2013).
- [107] V.V. Shaydurov, S. Zhang, V.S. Kornienko, *A finite-difference solution of mean field problem with the fractional derivative for subdiffusion*, *AIP Conference Proceeding*, **2302**:110002 (2020).
- [108] B. Kaltenbacher, A. Neubauer, O. Scherzer, *Iterative regularization methods for nonlinear ill-posed problems*, De Gruyter, NY, 2008.
- [109] S. Kabanikhin, *Definitions and examples of inverse and ill-posed problems*, *J.Inverse Ill-Posed Probl*, **16**:4 (2009), 317-357.
- [110] А.Н. Тихонов, А.В. Гончарский, В.В. Степанов, А.Г. Ягола, *Регуляризирующие алгоритмы и априорная информация*, Наука, Москва, 1983.
- [111] А.Н. Тихонов, А.С. Леонов, А.Г. Ягола, *Нелинейные некорректные задачи*, Наука, Москва, 1995.
- [112] M. Hongyu, X. Xiaohua, A.S. Perelson, W. Hulin, *On identifiability of nonlinear ODE models and applications in viral dynamics*, *SIAM Rev. Soc. Ind. Appl. Math.*, **53**:1 (2011), 3-39.
- [113] B.M. Adams, H.T. Banks et. al, *On HIV dynamics: modeling, data analysis, and optimal treatment protocols*, *Journal of Computational and Applied Mathematics*, **184** (2015), 10-49.
- [114] I. Andrianakis, I.R. Vernon, N. McCreech, T.J. McKinley, J.E. Oakley, R.N. Nsubuga, M. Goldstein, R.G. White, *Bayesian history matching of complex infectiousdisease models using emulation: a tutorial and a case study on HIV in Uganda*, *PLOS Computational Biology*, **11**:1 (2015), e1003968.
- [115] O.I. Krivorotko, D.V. Andornaya, S.I. Kabanikhin, *Sensitivity analysis and practical identifiability of some mathematical models in biology*, *J. Appl. Ind. Math.*, **14** (2020), 115-130.

- [116] А.Н. Тихонов, *Об устойчивости обратных задач*, Докл. АН СССР, **39**:5 (1943), 195-198.
- [117] М.М. Лаврентьев, *К вопросу об улучшении точности решения системы линейных уравнений*, Доклады Академии наук СССР, **92**:5 (1953), 885-886.
- [118] С.К. Годунов, А.Г. Антонов, О.П. Кирилюк, В.И. Костин, *Гарантированная точность решения систем линейных уравнений в евклидовых пространствах*, Наука, Новосибирск, 1992.
- [119] S.I. Kabanikhin, O.I. Krivorotko, *Singular value decomposition in an inverse source problem*, Numerical Analysis and Applications, **5**:2 (2012), 168-174.
- [120] М.М. Лаврентьев, *О некоторых некорректных задачах математической физики*, Наука, Новосибирск, 1962.
- [121] А.Н. Тихонов, *О решении некорректно поставленных задач*, Докл. АН СССР, **151**:3 (1963), 501-504.
- [122] В.К. Иванов, *О некорректно поставленных задачах*, Матем. сб., **61**:2 (1963), 211-223.
- [123] S.I. Kabanikhin, O.I. Krivorotko, *Identification of biological models described by systems of nonlinear differential equations*, J. Inverse Ill-Posed Probl, **23**:5 (2015), 519-527.
- [124] O. Krivorotko, S. Kabanikhin, Sh. Zhang, V. Kashtanova, *Global and local optimization in identification of parabolic systems*, J. Inverse Ill-Posed Probl, **28**:6 (2020), 899-913.
- [125] А.А. Жигляевский, А.Г. Жилинскас, *Методы поиска глобального экстремума*, Наука, Москва, 1991.
- [126] Е.Е. Тыртышников, *Новые теоремы о распределении собственных и сингулярных чисел многоуровневых теплицевых матриц*, Доклады Академии наук, **333**:3 (1993), 300-303.
- [127] I.V. Oseledets, *Tensor-train decomposition*, SIAM J. Sci. Comput., **33** (2011), 2295-2317.
- [128] Д.А. Желтков, И.В. Оферкин, Е.В. Каткова, А.В. Сулимов, В.Б. Сулимов, Е.Е. Тыртышников, *TTDock: метод дожинга на основе тензорных поездов*, Вычислительные методы и программирование, **14**:3 (2013), 279-291.
- [129] V.V. Zheltkova, D.A. Zheltkov, Z. Grossman, G.A. Bocharov, E.E. Tyrtshnikov, *Tensor based approach to the numerical treatment of the parameter estimation problems in mathematical immunology*, J. Inverse Ill-Posed Probl, **26**:1 (2018), 51-66.
- [130] O.I. Krivorotko, M.I. Sosnovskaia, I. Vashchenko, *Agent-based mathematical model of COVID-19 spread in Novosibirsk region: identifiability, optimization and forecasting*, J. Inverse Ill-Posed Probl, In print (2021).
- [131] O.I. Krivorotko, S.I. Kabanikhin, M.I. Sosnovskaya, D.V. Andornaya, *Sensitivity and identifiability analysis of COVID-19 pandemic models*, Vavilov journal of genetics and breeding, **25**:1 (2021), 82-91.
- [132] R. Casagrandi, L. Bolzoni, S. Levin, V. Andreasen, *The SIRC model and influenza A*, Math. Biosci., **200** (2006), 152-169.
- [133] OPTUNA: hyperparameter optimization framework: <https://optuna.org/>
- [134] V. Petrakova, O. Krivorotko, *Mean field game for modeling of COVID-19 spread*, Journal of Mathematical Analysis and Applications, **514**:1 (2022), 126271.
- [135] Федеральная служба государственной статистики, Новосибирская область, <https://novosibstat.gks.ru/folder/31729>.

ПРИЛОЖЕНИЯ

А. Обработка эпидемиологических данных. Для анализа данных, описанного в разделах 2.2-2.3 и дальнейшего построения SEIR-HCD (Раздел 3.5.2) модели и АОМ (Раздел 4.3.1), были использованы следующие показатели для Новосибирской области:

- (1) Официальные статистические данные о выявленных случаях заражения COVID-19, проведенных ПЦР тестов, критических пациентах (подключение аппарата ИВЛ) и умерших были получены более чем из 1000 различных открытых источников с помощью разработанной нами программы автоматического сбора данных. Эта информация доступна на сайте группы COVID-19 СО РАН по ссылке: <http://covid19-modeling.ru/data/novosibirsk-region-data.csv>

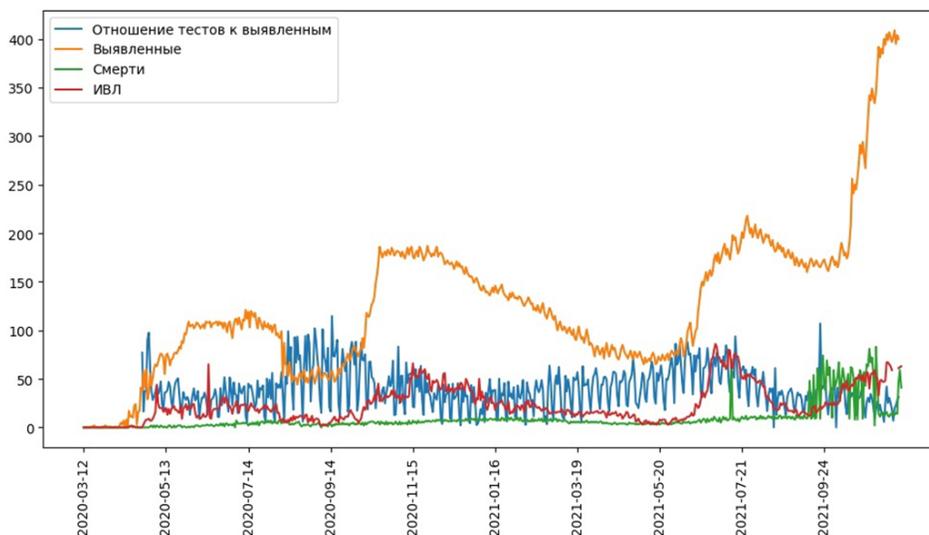


Рис. А.1. Статистические данные COVID-19 по Новосибирской области с 12.03.2020 по 30.11.2021 (стопкоронавирус.рф): Синяя линия – отношение ежедневных тестов ПЦР к выявленным случаям, оранжевая линия – выявленные случаи COVID-19, зеленая линия – умершие от COVID-19, красная линия – критические случаи, находящиеся на ИВЛ.

- (2) Информация о распределении населения по возрастным группам доступна на сайте Федеральной службы государственной статистики [135]);
- (3) Сведения о среднем размере семьи: 2,5 человека согласно отчету Росстата.

Для построения SEIR-HCD модели, описанной в разделе 3, были использованы следующие данные по Новосибирской области:

- (1) Данные по новым выявленным случаям заражения COVID-19, госпитализированным, критическим (требующих подключения аппарата ИВЛ) и умершим. Информация была агрегирована с открытых источников

Возрастная группа, лет	Количество человек
0-9	357814
10-19	279706
20-29	316949
30-39	493491
40-49	391877
50-59	343950
60-69	353261
70-79	157762
80+	103360

ТАБЛИЦА А.1. Распределение численности населения Новосибирской области по возрастным группам на 1 января 2020 года.

сети Интернет с помощью разработанной программы для автоматического сбора данных и доступна для скачивания по ссылке: <http://covid19-modeling.ru/data/novosibirsk-region-data.csv>.

- (2) Данные по проценту бессимптомных выявленных случаев от общего числа выявленных случаев заражения COVID-19, а также его прогноз на период моделирования. Данные получены из ежедневных сводок оперативного штаба Москвы, публикуемые в их официальном Телеграм-канале и доступные по ссылке: https://t.me/COVID2019_official.
- (3) Индекс самоизоляции от Яндекса, а также его прогноз на период моделирования. Индекс публикуется компанией Яндекс и доступен по ссылке: <https://yandex.ru/company/researches/2020/podomam>.
- (4) Данные по проценту индивидуумов с антителами к COVID-19 от медицинского центра «Инвитро», а также его прогноз на период моделирования. Ежедневная динамика процента индивидуумов с антителами к COVID-19 по лабораторным данным медицинского центра «Инвитро» в Новосибирске доступна по следующей ссылке: https://www.invitro.ru/1/invitro_monitor.

В силу того, что данные по смертности с официальных источников не коррелируют с остальными статистиками, в численных расчетах использовались данные по захоронениям в Новосибирске Муниципальной информационной системы «Ритуал» (Мэрия города Новосибирска), предоставленным директором Центра по взаимодействию с органами власти и промышленными партнерами Новосибирского государственного университета, к.ф.-м.н. А.Н. Люлько (см. Рис. А.2, А.3 и А.4). На Рис. А.2 изображены статистические данные с сайта стопкоронавирус.рф по количеству госпитализированных (красная линия) и критических (синяя линия) случаев COVID-19 и данные по захоронениям в городе Новосибирске по причине COVID-19 (черная линия), сглаженные 14-дневной экспоненциальной скользящей средней. Наблюдается хорошая корреляция данных. Однако данные по смертности от COVID-19 (красная линия на Рис. А.3) в Новосибирской области и количество захоронений в городе Новосибирске (черная линия) слабо коррелируют друг с другом (см. Табл. А.2).



Рис. А.2. Количество госпитализированных (красная линия) и критических (синяя линия) случаев COVID-19 по Новосибирской области с 24.04.2020 по 06.12.2021 и количество захоронений в городе Новосибирске (черная линия).



Рис. А.3. Количество умерших в результате COVID-19 по Новосибирской области (данные Оперштаба стопкоронавирус.рф, красная линия) и захоронений в городе Новосибирске (черная линия) в результате COVID-19 с 24.04.2020 по 06.12.2021.

Сравнение графиков данных по захоронениям в городе Новосибирске по причине COVID-19 и новым выявленным случаям заражения COVID-19 (красная линия на Рис. А.4) приведено на Рис. А.4. Наблюдается частичная корреляция данных в смысле градиентного роста/падения. Однако в периоды 20.10.2020–06.12.2020 и 25.06.2021–30.07.2021 количество выявленных случаев в

Новосибирской области резко перестало расти, а количество захоронений продолжало увеличиваться. Данный эффект возможен из-за небольшого снижения числа проведенных ПЦР-тестов в указанные периоды (см. Рис. А.1, синяя линия).



Рис. А.4. Количество захоронений в городе Новосибирске (черная линия) в результате COVID-19 и новые выявленные случаи заражения COVID-19 (данные Оперштаба стопкоронавирус.рф, красная линия) с 24.04.2020 по 06.12.2021.

	Госпитализированные	Критические	Захоронения в Нск	Смерти (Оперштаб)	Новые случаи
Госпитализированные	1.00	0.89	0.95	0.54	0.73
Критические	0.89	1.00	0.88	0.58	0.79
Захоронения в Нск	0.95	0.88	1.00	0.37	0.58
Смерти (Оперштаб)	0.54	0.58	0.37	1.00	0.87
Новые случаи	0.73	0.79	0.58	0.87	1.00

ТАБЛИЦА А.2. Коэффициенты корреляция скользящих средних по статистическим данным о количестве госпитализированных, критических, выявленных и смертей от COVID-19 в Новосибирской области и количества захоронений от COVID-19 в городе Новосибирске (графики изображены на Рис. А.2, А.3 и А.4).

Для более полного анализа и последующего построения более точных сценариев развития эпидемиологической ситуации в регионе необходим анализ следующих данных, которые не публикуются в открытых источниках:

- Информация о количестве завезенных случаев COVID-19 (из других регионов РФ или из-за границы);
- Информация о доле больничных коек, занятых больными COVID-19, от всех доступных в учреждениях здравоохранения;
- Определение штамма вируса при выявлении заражения (см. Табл. А.3);

Название штамма	Линия Панго	Место обнаружения	Дата
Альфа	V.1.1.7	Великобритания	20.09.2020
Бета	V.1.351	ЮАР	05.2020
Гамма	P.1	Бразилия	11.2020
Дельта	V.1.617.2	Индия	10.2020
Омикрон	V.1.1.529	ЮАР, Ботсвана	09.11.2021

ТАБЛИЦА А.3. Наиболее значимые штаммы вируса SARS-CoV-2 по версии ВОЗ на 31.12.2021.

- Идентификация суперраспространителей COVID-19 при проведении тестирования.

В. Комплекс программ. Для построения агентно-ориентированной модели, описанной в разделе 4.3.1, был использован программный комплекс Covasim (<https://github.com/InstituteforDiseaseModeling/covasim>), разработанный институтом Institute for Disease Modeling (<https://www.idmod.org/>). Данная библиотека написана на языке Python и создана для исследования агентных моделей COVID-19 с нетривиальными структурами. Covasim использовался для анализа эпидемиологической ситуации в более чем десяти странах, а также являлся одним из инструментов для принятия решений о введении ограничительных мер в США, Великобритании и Австралии.

Для решения обратных задач для агентно-ориентированной модели (раздел 6.2, был подготовлен программный код на языке Python для интеграции разработанного поэтапного метода восстановления неизвестных параметров с библиотекой Covasim. Исходные файлы и примеры использования доступны по ссылке на Github (<https://github.com/msosnovskaya/autocalibration-covasim-pub>). Также данный программный код был зарегистрирован в Роспатент (Номер свидетельства: 2021614740).

ОБОЗНАЧЕНИЯ

- ARIMA AutoRegressive Integrated Moving Average
- COVID-19 коронавирусная инфекция, вызванная штаммом вируса SARS-CoV-2
- SARIMA учет сезонной составляющей в модели ARIMA
- SEIR камерная модель, основанная на системе из 4 обыкновенных дифференциальных уравнений, где к классическим группам SIR модели добавлена группа E – бессимптомные носители инфекции (Exposed)
- SEIR-HCD камерная модель, основанная на системе из 7 обыкновенных дифференциальных уравнений, где к классическим группам SIR модели добавлены E – бессимптомные носители инфекции (Exposed), H – госпитализированные (Hospitalized), C – критические случаи (Critical), нуждающиеся в подключении аппарата ИВЛ, D – умершие (Death)
- SIR модель, основанная на законе действующих масс, в которой вся популяция разделена на 3 группы: S – восприимчивые (Susceptible), I – инфицированные (Infected) и R – вылеченные и умершие (Removed)
- АОМ агентно-ориентированная модель
- ВВП валовой внутренний продукт
- ВОЗ Всемирная организация здравоохранения
- ГЯБ уравнение Гамильтона-Якоби-Беллмана
- ИВЛ искусственная вентиляция легких
- ИНС искусственные нейронные сети
- ИСП игра среднего поля
- КФП уравнение Колмогорова-Фоккера-Планка
- ОДУ обыкновенные дифференциальные уравнения
- ПЦР полимеразная цепная реакция – экспериментальный метод молекулярной биологии, способ значительного увеличения малых концентраций определённых фрагментов нуклеиновой кислоты (ДНК) в биологическом материале
- СДУ стохастические дифференциальные уравнения
- УРЧП уравнения в частных производных

OLGA IGOREVNA KRIVOROTKO
 SOBOLEV INSTITUTE OF MATHEMATICS,
 PR. КОРТУГА, 4,
 630090, NOVOSIBIRSK, RUSSIA
 Email address: krivorotko.olya@mail.ru