

МОДЕЛИРОВАНИЕ СЦЕНАРИЕВ РАСПРОСТРАНЕНИЯ
COVID-19 В РЕСПУБЛИКЕ КАЗАХСТАН
НА ОСНОВЕ РЕГУЛЯРИЗАЦИИ АГЕНТНОЙ МОДЕЛИ

О. И. Криворотько^{1,2,3, a}, С. И. Кабанихин^{2,3, b},
М. А. Бектемесов^{4, c}, М. И. Сосновская^{3, d}, А. В. Неверов^{1,3, e}

¹ Институт вычислительной математики и математической геофизики,
пр. Акад. Лаврентьева, 6, 630090 Новосибирск, Россия

² Институт математики им. С. Л. Соболева,
пр. Акад. Коптюга, 4, 630090 Новосибирск, Россия

³ Новосибирский гос. университет,
ул. Пирогова, 2, 630090 Новосибирск, Россия

⁴ Казахский национальный педагогический университет им. Абая,
пр. Достык, 13, 050010 Алматы, Казахстан

E-mail: ^akrivorotko.olya@mail.ru, ^bksi52@mail.ru, ^cmactagali@mail.ru,
^dm.sosnovskaya@alumni.nsu.ru, ^ea.neverov@g.nsu.ru

Аннотация. В работе построен алгоритм расчёта сценариев динамики выявленных случаев COVID-19 в Республике Казахстан, в основе которого лежат обработка неполных эпидемиологических данных и решение обратной задачи восстановления параметров агентной модели по совокупности доступных эпидемиологических данных. Основным инструментом построения модели является открытая библиотека Covasim. В случае резкого изменения ситуации (появление нового штамма, отмена или введение ограничительных мер и т. п.) параметры модели обновляются с учётом дополнительной информации за предыдущий месяц (оперативное усвоение данных). Обратная задача решалась методом стохастической глобальной оптимизации (древовидных оценок Парзена). В качестве примера приведены два сценария распространения COVID-19, рассчитанных 12 декабря 2021 г. на период до 20 января 2022 г. Сценарий, в котором учитывались новогодние праздники (опубликован 12 декабря 2021 г. на сайте covid19-modeling.ru), практически совпал с тем,

Исследование выполнено при финансовой поддержке Совета по грантам при Президенте Российской Федерации (проект № МК-4994.2021.1.1), Российского фонда фундаментальных исследований (проект № 21-51-10003) и Министерства образования и науки Республики Казахстан (проект № ИРН AP09260317).

© О. И. Криворотько, С. И. Кабанихин, М. А. Бектемесов, М. И. Сосновская, А. В. Неверов, 2023

что произошло в реальности (погрешность составила 0,2%). Табл. 3, ил. 6, библиогр. 33.

Ключевые слова: агентно-ориентированная модель, COVID-19, обратная задача, оптимизация, регуляризация, сценарий развития, индекс репродукции вируса.

Введение

С декабря 2019 г. по сентябрь 2022 г. в мире было зарегистрировано более 604 млн случаев заражения COVID-19, из которых около 6,5 млн оказались летальными. По состоянию на 5.09.2022 г. Республика Казахстан (РК) занимает 51 место среди 216 стран по количеству выявленных случаев COVID-19 и 119 место по доле вакцинированного от COVID-19 населения. Для мониторинга эпидемиологической ситуации в регионах и странах, а также анализа эффективности сдерживающих мер применяются агентные модели распространения инфекционных заболеваний.

Агентная модель (АМ) распространения COVID-19 позволяет

- учитывать демографическую информацию по конкретной стране (численность и возрастную структуру населения);
- строить реалистичные сети передачи инфекции в различных социальных слоях, включая домохозяйства, школы, организации, общественные места;
- учитывать возрастные особенности развития заболевания;
- определять вирусную нагрузку агента, включая скорость передачи инфекции;
- учитывать физическое дистанцирование и ношение масок, вакцинацию, тестирование (включая бессимптомных), изоляцию, мониторинг контактов, карантин в регионах с населением от 100 до 20 млн человек [1–3].

В агентной модели каждый действующий агент наделяется признаками (возраст, социальный статус, восприимчивость к заболеванию и т. д.).

Краткий обзор работ по АМ COVID-19. Приведём краткий обзор работ по агентному моделированию COVID-19, во многом опирающихся на более ранние работы по моделированию инфекционных заболеваний 2013 г. [4] и 2020 г. [5].

В [6] АМ применялась для описания распространения COVID-19 в Бостоне. Показано, что система реагирования, основанная на расширенном тестировании и отслеживании контактов, может играть важную роль в ослаблении ограничений по соблюдению социальной дистанции при отсутствии коллективного иммунитета против SARS-CoV-2.

В [7] авторы подсчитали на примере штата Джорджия в США, что инфицированные люди возраста до 60 лет могут быть в 2,78 раза более

заразными, чем пожилые люди, и они, как правило, являются основной движущей силой сверхраспространения.

В [8–10] на основе АМ распространения COVID-19 проанализированы противоэпидемические программы в различных регионах Франции и Великобритании. Показано, что для контроля новой волны вспышки COVID-19 достаточно использовать данные о контактах и изоляции в течение ближайших трёх месяцев.

В [11] построена АМ, в которой источниками заражения выступали суперраспространители, и показано, что сверхраспространение резко усиливает значимость ограничений личных контактов.

Группа американских учёных в 2021 г. [3] разработала программный комплекс Covasim [12], основу которого составляет агентный подход моделирования эпидемии с учётом особенностей заболевания, фармацевтических (вакцинация) и социальных (ограничения посещений, ношение масок) мер. Этот программный комплекс применялся для построения сценариев развития эпидемии COVID-19, изучения динамики пандемии и поддержке принятия административных решений более чем в десятке стран Африки, Азиатско-Тихоокеанского региона, Европы и Северной Америки.

Открытые программные комплексы. Приведём краткий обзор программных комплексов распространения COVID-19.

1) covid19-scenarios.org, Базельский университет, Швейцария. Реализуется структурированная по возрастам компартментная модель распространения коронавирусной инфекции, основанная на 9 дифференциальных уравнениях, с возможностью варьирования параметров модели [13].

2) covid19.biouml.org, ИВТ СО РАН, Новосибирск. Моделируется распространение COVID-19 в Москве, Новосибирской области, Германии, Франции и Италии. В комплексе применяются расширенная камерная и агентная модели, параметры которых идентифицируются на основе опубликованной статистики. Строятся прогнозы не только числа регистрируемых заболевших, вылечившихся и умерших, но и количества свободных койко-мест, аппаратов искусственной вентиляции лёгких (ИВЛ) и других характеристик, необходимых для эффективного управления ситуацией в условиях эпидемии.

3) anylogic.com/healthcare, Богота, Колумбия. Комплекс основан на агентной модели с учётом географических особенностей города (расположение школ, медицинских учреждений, общественных мест) и социальной дистанции между агентами с возможностью варьирования параметров.

4) github.com/kausaltech/reina-model, Хельсинки, Финляндия. На основе агентной модели учитывается различный возраст агентов и семь возможных состояний прогрессирования заболевания, а также рассматривается случайная структура контактов, т. е. агенты взаимодействуют между собой хаотично [14].

5) github.com/institutefordiseasemodeling/covasim, США. Комплекс основан на агентной модели со случайной структурой с возможностью идентификации параметров для конкретного региона [3]. Он послужил основой создания собственного программного комплекса (см. п. 6), а также был адаптирован для получения результатов для Республики Казахстан в рамках данной работы.

6) covid19-modeling.ru, ИВМиМГ СО РАН, Новосибирск. Комплекс на основе комбинации камерной и агентной моделей предоставляет возможность построения сценариев развития с помощью решения обратных задач [15, 16].

Обратные задачи для АМ. В разных регионах один и тот же вирус может распространяться и воздействовать на людей по-разному. В силу новизны и сложности заболевания COVID-19 (частые изменения заразности вируса и среднего возраста тяжёлого течения заболевания, длительность инкубационного периода и др.) параметры большинства математических моделей, как правило, неизвестны, что затрудняет адаптацию существующих программных решений для анализа ситуации в конкретном регионе с учётом введения ограничений в разные периоды и фармацевтических вмешательств. Основные проблемы моделирования распространения COVID-19 следующие.

1. Данные для решения обратной задачи неполные и зашумлённые, а также представляют собой большие данные (ежедневные сводки о заболевших, заразившихся, вакцинированных и т. д.).

2. Параметры меняются со временем: contagiозность вируса $\beta(t)$, вероятность появления тяжёлых случаев $p_{sev}(t)$, смертность $p_{death}(t)$ и т. д.

3. Процесс распространения COVID-19 существенно изменяется при введении или отмене ограничительных мер (маски, социальная дистанция, перевод на удалённый рабочий режим, закрытие школ, предприятий, районов и городов).

Указанные проблемы приводят к необходимости рассматривать и решать обратные задачи идентификации неизвестных эпидемиологических параметров в конкретном регионе с учётом мутации вируса и различных вмешательств (административные, фармацевтические) по дополнительной информации о количестве проведённых ПЦР-тестов, выявленных случаях, госпитализированных, критических больных и умерших от COVID-19. Ввиду некорректности обратных задач (решение может

быть неединственным и/или неустойчивым) используется регуляризация с учётом ограничений на искомые параметры, которые получены по данным Всемирной организации здравоохранения (ВОЗ) и Министерства здравоохранения исследуемого региона (в нашем случае РК).

Особенность АМ в Республике Казахстан. Особенность представленной нами работы заключается в адаптации агентной модели к РК с учётом распределения населения (по возрастам, регионам), ограничительных мер и типа и качества статистических данных (неполные и неточные данные). По состоянию на 2021 г. 40,8% населения Республики Казахстан сельское, а его плотность сравнима с плотностью городского населения.

В работе предполагалось, что городское и сельское население РК находится в равных условиях по тестированию методом полимеразной цепной реакции (ПЦР) и скорости распространения COVID-19. В работе предложен алгоритм поэтапного уточнения эпидемиологических параметров АМ распространения COVID-19 в РК, а также алгоритм построения сценариев развития эпидемии в регионе на основе комбинации методов машинного обучения и решения обратных задач.

Структура статьи. Для достоверного моделирования тех или иных факторов распространения COVID-19 в основу расчётов должны закладываться адекватные исходные данные — от численности населения и его распределения по социальным группам до загруженности различных видов транспорта или магазинов [1]. В силу неполных и зашумлённых статистических данных в работе применяются методы машинного обучения и регрессионного анализа обработки и анализа временных рядов. Временной ряд (количество проведённых ПЦР-тестов) экстраполирован на основе статистических моделей с целью построения сценариев распространения COVID-19, используемых в агентном подходе (разд. 1). Исходя из особенностей данных для Республики Казахстан, в разд. 2 строится агентная модель, основная цель которой — составление сценариев развития и оценка влияния различных вмешательств на эпидемию, формулируются прямая и обратная задачи для АМ. В разд. 3 описан алгоритм решения обратной задачи на основе метода древовидных оценок Парзена минимизации абсолютного целевого функционала и усвоения данных. В разд. 4 построены и проанализированы сценарии распространения COVID-19 в Республике Казахстан с учётом ограничительных мер с 13.12.2021 г. по 20.01.2022 г. Показано, что увеличение концентрации агентов в новогодние праздники в общественных местах (магазины, театры, парки) увеличивает количество выявленных случаев COVID-19 (к 15.01.2022 г. увеличилось в 3,5 раза по сравнению с 1.12.2021 г.). В заключении приведены основные выводы.

1. Анализ данных по Республике Казахстан

Представим основные демографические и эпидемиологические данные по Республике Казахстан, использованные в последующем построении агентной модели распространения COVID-19, описанной в разд. 2.

1.1. Демографические данные. Статистические данные о численности населения по возрастным категориям (возрастные группы разбиты на периоды по 10 лет) были предоставлены М. А. Бектемесовым (табл. 1). Такого рода данные используются при инициации искусственной популяции АМ (см. разд. 2.2). Общая численность населения в РК составляет 18 879 552 человек.

Таблица 1

**Распределение населения по возрастам
в Республике Казахстан на 1.10.2021 г.**

Возрастная категория (лет)	Количество (чел.)
0–9	3 890 241
10–19	2 855 522
20–29	2 537 252
30–39	2 987 296
40–49	2 292 980
50–59	1 994 801
60–69	1 439 848
70–79	588 375
80+	293 237

1.2. Эпидемиологические данные. Источники данных:

- ourworldindata.org (OWD),
- worldhealthorg.shinyapps.io/covid (WHO),
- coronavirus2020.kz/ru (CV2020),
- kt.kz (КТ).

Для сбора и обработки эпидемиологических данных была создана программа, написанная на языке программирования Python. Схема работы программы следующая.

1. Задаётся URL ресурса, с которого будет происходить сбор данных.
2. Для новостных сайтов проводится поиск по статьям инструментами сайта и выделяются необходимые статистические данные.
3. В разметке html страницы по заранее подобранным ключам выделяются элементы, содержащие следующие данные.
 - Для новостных сайтов данные находятся по заданным ключевым словам в предложениях.

- Для сайтов со статистикой данные не требуют дополнительного поиска и выделяются в заранее заданном порядке.

Были собраны и обработаны следующие эпидемиологические данные о распространённости COVID-19 в день t :

- количество проведённых ПЦР тестов $\mathcal{T}(t)$ (рис. 1) — OWD,
- количество диагностированных случаев COVID-19 (пациентов с положительным ПЦР тестом) $f(t)$ (рис. 2) — OWD, WHO, CV2020, КТ,
- количество вакцинированных — OWD, WHO, КТ,
- количество госпитализированных с COVID-19 $H(t)$ — OWD (только количество коек-мест), КТ, CV2020,
- количество пациентов, подключённых к аппарату ИВЛ $C(t)$ — КТ, CV2020,
- число умерших от COVID-19 $D(t)$ — OWD, WHO, CV2020, КТ.

Некоторые данные, размещённые на указанных сайтах, неполны (пропущены некоторые дни, недели, месяцы). Пропущенные данные промежуточных значений во временных рядах были интерполированы с помощью кубических сплайнов методом *interpolate* библиотеки *pandas* [17].

1.3. Регрессионная модель обработки и экстраполяции сезонных временных рядов. Статистические данные $\mathcal{T}(t)$ о количестве проведённых ПЦР тестов в РК имеют пропуски с 6.08.2021 г., а также свойственную недельную сезонность (периодичность), поэтому для построения сценариев распространения COVID-19 на базе АМ был использован экстраполированный временной ряд $\mathcal{T}(t)$ сезонной авторегрессионной моделью SARIMA, являющейся модификацией модели ARIMA (AutoRegressive Integrated Moving Average) [18], которая описывает одномерные временные ряды с сезонной компонентой.

Модель SARIMA(p, d, q)(P, D, Q) $_s$ для нестационарного временного ряда $\mathcal{T}(n)$ имеет вид [19]

$$\Phi(L^s)\varphi(L)\Delta^d\Delta_s^D\mathcal{T}(n) = \theta_0 + \Theta(L^s)\theta(L)\varepsilon(n).$$

Здесь параметры p, d, q отвечают за несезонную часть временного ряда, а P, D, Q соответствуют сезонным компонентам ряда, $s = 7$ — длина сезона, $\varepsilon(n)$ — стационарный временной ряд белого шума, 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.), 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, n — временной параметр (день).

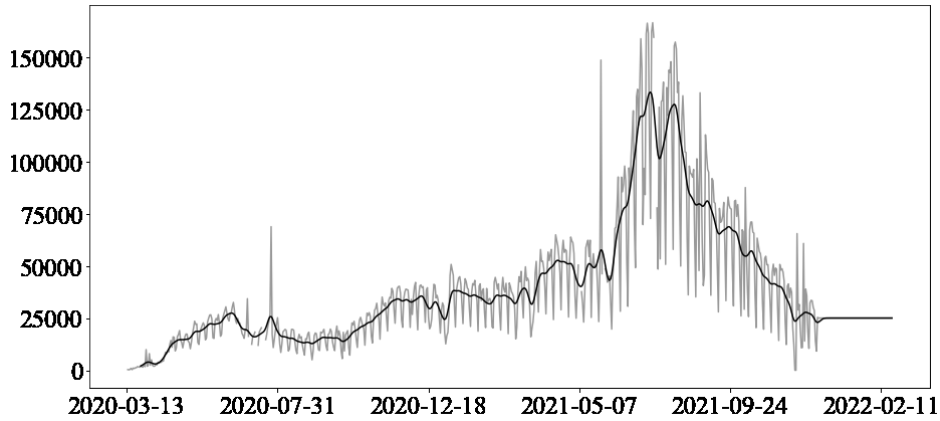


Рис. 1. Статистические данные о количестве проведённых ПЦР тестов $\mathcal{T}(t)$ в Республике Казахстан, используемые в построении модели

Алгоритм экстраполяции временного ряда следующий.

ШАГ 1. Применяем преобразование Бокса — Кокса [18] для уменьшения дисперсии.

ШАГ 2. Вычисляем сезонную разность (сдвиг на $s = 7$ дней) первого порядка.

ШАГ 3. Вычисляем вторую разность (сдвиг на 1 день) ряда, полученного на ШАГЕ 2.

ШАГ 4. Проверяем стационарность ряда, полученного на ШАГЕ 3, критерием Дики — Фуллера [20].

ШАГ 5. Параметры, соответствующие ШАГАМ 1–4, передаём в модель SARIMA(1, 1, 2)(0, 1, 1)₇ и подбираем остальные на основе минимизации информационного критерия Акаике. В качестве данных передаётся ряд с ШАГА 1.

ШАГ 6. Применяем обратное преобразование Бокса — Кокса к полученной модели с настроенными гиперпараметрами.

1.4. Результат экстраполяции временного ряда. Прогноз временного ряда $\mathcal{T}(t)$ количества ежедневных ПЦР-тестов в Республике Казахстан с 6.08.2021 г. по 20.01.2022 г. представлен на рис. 1.

Ввиду высокой флуктуации из-за сезонности, а также скачков во временных рядах в результате некорректного сбора данных, было решено провести предварительное сглаживание рядов перед передачей в модель с помощью гауссовского фильтра [21]. Таким образом мы избавляемся

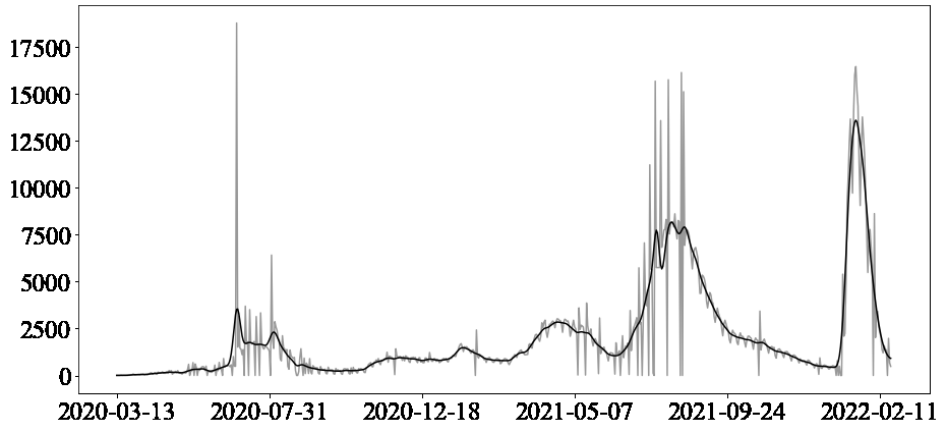


Рис. 2. Статистические данные по выявленным случаям COVID-19 $f(t)$ в Республике Казахстан, используемые в построении модели

от возможной неустойчивости и неинтерпретируемости результатов модели ввиду резких скачков в данных, полученных из открытого источника OWD (серая линия на рис. 1 и 2). Сглаженные данные, которые были переданы в модель, представлены чёрной линией на рис. 1 и 2.

2. Агентная модель

Агентное моделирование основано на исследовании динамики развития заболевания путём изучения взаимодействия между отдельно взятыми индивидуумами, а глобальные изменения в системе возникают как результат деятельности множества агентов (моделирование «снизу вверх»). Общее описание АМ распространения COVID-19 в регионе приведено в п. 2.1. Оно включает в себя инициацию популяции (п. 2.2), правила распространения заболевания (п. 2.3) и тестирования агентов (п. 2.4). В п. 2.5 приведена постановка прямой задачи для АМ, а в п. 2.6 — постановка обратной задачи для АМ.

2.1. Общее описание модели. В рамках данного исследования была реализована стохастическая АМ для Республики Казахстан. Основным инструментом для создания модели была библиотека Covasim [3], реализованная на языке Python (открытый код доступен по ссылке [12]) и созданная для исследования агентных моделей COVID-19 с нетривиальными структурами. Общий алгоритм выглядит следующим образом: загружаются все необходимые параметры и статистические данные, создаётся искусственная популяция с учётом распределения по возрастам в регионе. Далее агенты соединяются в контактные сети, представляющие собой полные графы. Затем начинается цикл по времени: на каждом

шаге (временной интервал равен одному дню) обновляется эпидемиологический статус агента как суперпозиция вероятностей с учётом его структуры контактов и введённых ограничительных мер (самоизоляция, закрытие общественных мест, ношение масок и т. д.).

2.2. Инициация популяции. Инициация искусственной популяции производится на основании статистических данных в регионе и зависит от следующих параметров агентов.

- Возраст (t^*). Все агенты делятся на возрастные группы по 10 лет (0–9, 10–19, ..., 80+ лет) согласно статистическим данным Республики Казахстан (табл. 1).

- Социальный статус (рабочий, студент, ребёнок, пенсионер), который зависит от возраста агента t^* .

Домохозяйства заполняются агентами согласно статистическим данным о среднем размере семьи в регионе. В зависимости от возраста агенты контактируют друг с другом в контактных сетях, представляющих собой полные графы, степень которых определяется пуассоновской случайной величиной с параметром λ :

- для домохозяйств $\lambda = 3,496$ — средний размер семьи (чел.) [22];
- для организаций $\lambda = 8$;
- для общественных мест и образовательных учреждений $\lambda = 20$.

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. Сети контактов строятся на основе открытого алгоритма SynthPops [23], способного генерировать реалистичные контактные сети для популяций. Этот метод основывается на ранее опубликованных моделях и эмпирических исследованиях, позволяющих определить характерное для конкретных возрастных групп модели количество контактов.

В случае образовательных учреждений алгоритм выбирает ученика или студента и в зависимости от возраста формирует матрицу смешения возрастов в образовательном учреждении для определения вероятного возраста в контактной сети. Ученики выбираются из упорядоченного списка домохозяйств, так что они воспроизводят приближение динамики соседства детей, посещающих районные образовательные организации вместе. Преподаватели и другой вспомогательный персонал отбираются из числа взрослого населения, входящего в состав рабочей силы, и распределяются по мере необходимости по школам, что отражает средние данные о соотношении учащихся и учителей и учащихся и персонала. В крупных образовательных учреждениях близкие контакты моделируются случайным набором n контактов из числа возможных

в учреждении, где n определяется как случайная величина с распределением Пуассона с параметром $\lambda = 20$, соответствующим среднему размеру класса или группы.

Численность рабочей силы рассчитывается с использованием коэффициентов занятости с разбивкой по возрасту, а лица, не относящиеся к числу занятых, распределяются по рабочим местам с использованием данных о размерах организаций. Первичный эталонный работник отбирается из числа рабочей силы, а его коллеги выводятся на основе возрастных моделей смещения рабочей силы. Все работники (включая учителей) отбираются произвольно из числа населения, чтобы отразить общее смещение взрослых из разных районов на работе. Близкие контакты моделируются случайным набором n контактов в организации, где n — случайная величина с распределением Пуассона с параметром $\lambda = 8$, равным предполагаемому максимальному числу тесных контактов на рабочем месте.

Для моделирования контактов в общественных местах для каждого человека используются n случайных контактов в популяции, где n имеет распределение Пуассона с параметром $\lambda = 20$. Соединения в этом уровне отражают природу контактов в парках и зонах отдыха, торговых центрах, общественном транспорте и т. п. Все связи между индивидами считаются неориентированными, чтобы отразить способность любого индивида в паре заражать другого.

2.3. Распространение заболевания в АМ. В рамках модели предполагается, что вирус передаётся между агентами, соединёнными ребром графа. Заражение при близком контакте описывается кусочно-постоянным параметром $\beta(t)$, который в зависимости от структуры контакта умножается на соответствующую константу w_β (для домохозяйств

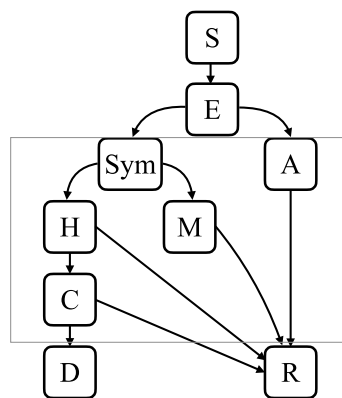


Рис. 3. Диаграмма состояний агентов в Covasim

$w_\beta = 3$, для образовательных учреждений и работы $w_\beta = 0,6$, для общественных мест $w_\beta = 0,3$). Таким образом, вероятность передачи вируса для каждой контактной сети различная.

Каждый агент может находиться в 9 стадиях заболевания: S — восприимчивые к заражению, E — заражённые незаразные, A — бессимптомные больные, Sym — больные с симптомами, M — больные лёгкой степени тяжести, H — госпитализированные, C — больные в критическом состоянии (нуждаются в реанимации), R — вылечившиеся, D — умершие (рис. 3). В рамку обведены те состояния, находясь в которых агент имеет возможность получить положительный тест на COVID-19. Переход из одной стадии заболевания в другую контролируется параметрами, зависящими от возраста (т. е. чем старше агент, тем он более уязвим): p_{sym} — вероятность проявлять симптомы после заражения, p_{sev} — вероятность перехода больного с симптомами в тяжёлое состояние (нуждается в госпитализации), p_{crit} — вероятность перехода больного из тяжёлого состояния в критическое (нуждается в реанимации), p_{death} — вероятность смерти для больного, находящегося в реанимации. Численные значения параметров для каждой возрастной группы приведены в табл. 2 [3].

Таблица 2

**Вероятности перехода между состояниями заболевания
в зависимости от возрастной группы**

Пар-р \ Возраст	0–9	10–19	20–29	30–39	40–49
p_{sym}	0,5	0,55	0,6	0,65	0,7
p_{sev}	$5 \cdot 10^{-4}$	0,0016	0,0072	0,0208	0,0343
p_{crit}	$3 \cdot 10^{-5}$	$8 \cdot 10^{-5}$	$4 \cdot 10^{-4}$	0,001	0,0022
p_{death}	$2 \cdot 10^{-5}$	$2 \cdot 10^{-5}$	10^{-4}	$3 \cdot 10^{-4}$	0,0098
Пар-р \ Возраст	50–59	60–69	70–79	80+	
p_{sym}	0,75	0,8	0,85	0,9	
p_{sev}	0,0765	0,133	0,207	0,246	
p_{crit}	0,0093	0,0364	0,089	0,174	
p_{death}	0,0026	0,008	0,024	0,082	

Продолжительность каждой стадии заболевания представляет собой случайную логнормальную величину с различными средними и параметрами дисперсии, согласованными со статистическими оценками ВОЗ (средние и дисперсии распределений представлены в табл. 3).

Таким образом, восприимчивый к заражению агент (S) при контакте с заражёнными агентами, соединёнными ребром графа, переходит в стадию заражённого незаразного (E) с вероятностью β в момент времени t .

Таблица 3

**Продолжительность стадий заболевания
в каждом эпидемиологическом состоянии**

Параметр	Описание	Распределение
t_{inc}	количество дней с момента контакта до того, как агент станет заразным	$\text{LogN}(4,6,4,8)$ [24]
t_{sym}	количество дней с момента, когда агент стал заражен, до проявления симптомов	$\text{LogN}(1,0,9)$ [24]
t_{rec1}	продолжительность болезни для бессимптомных и лёгких случаев	$\text{LogN}(8,2)$ [25]
t_{rec2}	продолжительность болезни для тяжёлых и критических случаев	$\text{LogN}(14,2,4)$ [26]
t_{inf}	количество дней, за которое агент переходит из лёгкого состояния в тяжёлое	$\text{LogN}(6,6,4,9)$ [24]
t_{hosp}	количество дней, за которое агент переходит из тяжёлого состояния в критическое	$\text{LogN}(3,7,4)$ [27]
t_{crit}	длительность пребывания агента в критическом состоянии	$\text{LogN}(6,2,1,7)$ [26]

Затем агент в состоянии заражённого незаразного (E) может перейти в состояние инфицированного с симптомами (Sym) с вероятностью p_{sym} через t_{sym} дней или остаться бессимптомным больным (A) с вероятностью $1 - p_{\text{sym}}$ через t_{inc} после заражения. Бессимптомные больные вылечиваются через t_{rec1} дней и переходят в группу (R). Инфицированные с симптомами (Sym) могут развить тяжёлую степень заболевания и быть госпитализированными (H) с вероятностью p_{sev} или остаться больным лёгкой степени тяжести (M) с вероятностью $1 - p_{\text{sev}}$ через t_{inf} дней после попадания в группу (Sym). Больные лёгкой степени тяжести вылечиваются через t_{rec2} дней и переходят в группу (R). Госпитализированные (H) могут перейти развить критическое состояние (C), т. е. нуждаться в подключении аппарата ИВЛ, с вероятностью p_{crit} через t_{hosp} дней после госпитализации или выздороветь с вероятностью $1 - p_{\text{crit}}$ через t_{rec2} дней. Больные в критическом состоянии (C) умирают с вероятностью p_{death} через t_{crit} дней или вылечиваются с вероятностью $1 - p_{\text{death}}$ через t_{rec2} дней. Все перечисленные вероятности можно записать в следующем виде:

$$p(S \rightarrow E) = \beta, \quad p(E \rightarrow \text{Sym}) = \frac{p_{\text{sym}}}{t_{\text{sym}}}, \quad p(E \rightarrow A) = \frac{1 - p_{\text{sym}}}{t_{\text{inc}}},$$

$$p(\text{Sym} \rightarrow H) = \frac{p_{\text{sev}}}{t_{\text{inf}}}, \quad p(\text{Sym} \rightarrow M) = \frac{1 - p_{\text{sev}}}{t_{\text{inf}}}, \quad p(H \rightarrow C) = \frac{p_{\text{crit}}}{t_{\text{hosp}}},$$

$$p(H \rightarrow R) = \frac{1 - p_{\text{crit}}}{t_{\text{rec2}}}, \quad p(M \rightarrow R) = \frac{1}{t_{\text{rec1}}}, \quad p(A \rightarrow R) = \frac{1}{t_{\text{rec1}}},$$

$$p(C \rightarrow D) = \frac{p_{\text{death}}}{t_{\text{crit}}}, \quad p(C \rightarrow R) = \frac{1 - p_{\text{death}}}{t_{\text{rec2}}}.$$

2.4. Тестирование агентов в АМ. Тестирование агентов проводится в количестве, соответствующем ежедневным статистическим данным в Республике Казахстан (см. п. 1.2). Шанс быть протестированным на COVID-19 зависит от эпидемиологического состояния агента (восприимчивый, инфицированный с симптомами, госпитализированные и т. п.). На каждом шаге моделирования тесты распределяются среди всей популяции (исключая умерших). Положительный результат могут получить агенты, статус которых обведён в рамку на рис. 3 (инфицированные бессимптомные и с симптомами, госпитализированные, больные лёгкой степени тяжести и критические случаи). В случае положительного теста на COVID-19 агенты попадают в статистику ежедневно выявленных. В модели предполагается, что вероятность тестирования агентов с симптомами выше, чем у бессимптомных больных. Данное соотношение шансов контролируется параметром p , который восстанавливается в ходе решения обратной задачи (см. п. 2.6).

2.5. Постановка прямой задачи для АМ. Прямая задача агентного моделирования состоит в определении количества инфицированных (в том числе $f(t)$ выявленных случаев в результате ПЦР тестирования), госпитализированных, умерших и других состояний агента, которые учитываются в АМ. В прямой задаче все входные параметры модели считаются известными. В этом случае агентная модель позволяет вычислить значения вектора

$$\vec{X}(t) = (S(t), E(t), A(t), \text{Sym}(t), M(t), H(t), C(t), R(t), D(t))$$

на следующий день, т. е. $\vec{X}(t+1)$. Поскольку многие параметры вектора

$$\vec{q}(t) = (E(0), \beta, p, \beta_d(i), \beta_c(i)), \quad i = 1, \dots, N,$$

АМ распространения COVID-19 неизвестны, необходимо сформулировать и решить обратную задачу, используя дополнительную информацию в текущий день t . Здесь $E(0)$ — начальное количество инфицированных, β — параметр передачи вируса, p — параметр тестирования, $\beta_d(i)$ — дни изменения параметра β , $\beta_c(i)$ — значения, на которые изменяется параметр β в дни $\beta_d(i)$, i соответствует месяцу изменения параметра контагиозности β , т. е. $i + 1 = t + 30$.

2.6. Постановка обратной задачи для АМ. В данной работе на каждом временном этапе (равном 30 дням; подробнее см. п. 3.2) решается своя обратная задача. Обратная задача 1 состоит в восстановлении

вектора параметров

$$\vec{q}(0) = (E(0), \beta, p, \beta_d(1), \beta_c(1))$$

по дополнительной информации о количестве ежедневно выявленных случаев $f(t)$, t измеряется в днях.

В модели предполагается, что изменчивость вируса (появление новых штаммов, фармацевтические и социальные меры) происходит не чаще, чем раз в месяц. Ввиду этого решается обратная задача 2, которая состоит в восстановлении вектора параметров ежемесячно

$$\vec{q}(t + 30) = (\beta_d(i), \beta_c(i)), \quad i + 1 = t + 30,$$

по дополнительной информации о количестве ежедневно выявленных случаев $f(t)$. Здесь i соответствует месяцу моделирования.

Решения обратных задач восстановления вектора $\vec{q}(t)$ сводились к решению задачи минимизации целевого функционала

$$J(\vec{q}) = \sum_{t_i=1}^T \frac{|f_d(t_i) - f_m(t_i, \vec{q})|}{M_{\text{diag}}}. \quad (1)$$

Здесь $f_d(t_i)$, $f_m(t_i, \vec{q})$ — сглаженные данные и результат моделирования ежедневно выявленных случаев соответственно, T — количество дней моделирования, $M_{\text{diag}} = \max_{t_i} \{f_d(t_i)\}$ — нормирующий член.

В работах [28, 29] для исследуемой модели авторы провели анализ чувствительности неизвестных параметров к измерениям на основе методов дифференциальной алгебры и байесовского подхода. Показано, что параметр β , отвечающий за передачу вируса, наиболее чувствителен к измерениям. С помощью методов анализа чувствительности область изменения параметра β удалось уменьшить в два раза благодаря добавлению дополнительной информации об эпидемии (а именно, к измерениям о количестве выявленных случаев и умерших была добавлена информация о критических случаях).

3. Алгоритм решения обратной задачи

В ходе решения обратной задачи вектор неизвестных параметров \vec{q} восстанавливался с помощью применения пакета Optuna [30], в основе которого лежит метод древовидных оценок Парзена (Tree-structured Parzen estimators или коротко ТРЕ), а также подхода усвоения данных для поэтапного восстановления параметров агентной модели (п. 3.2).

3.1. Метод древовидных оценок Парзена. Идея метода состоит в следующем: вычисляются вероятности $p(\vec{q} \mid J(\vec{q}))$ и $p(J(\vec{q}))$ для определения области параметров для минимизации функционала J . Для этого пространство значений параметров $\mathcal{D}_K = \{q_k, J(q_k) \mid k = 1, \dots, K\}$

разделяется на 2 подмножества $\mathcal{D}_{K_l}^l$ и $\mathcal{D}_{K_g}^g$ такие, что $\mathcal{D}_{K_l}^l$ содержит квантиль уровня γ наименьших значений функционала в точках из $\mathcal{D}_K(J_\gamma)$, т. е. $P(J < J_\gamma) = \gamma$. Подмножество $\mathcal{D}_{K_g}^g$ содержит все остальные точки из \mathcal{D}_K . Далее с помощью метода парзеновского окна оцениваются плотности распределения $l(x)$ и $g(x)$, которые получены из $\mathcal{D}_{K_l}^l$ и $\mathcal{D}_{K_g}^g$ соответственно. Таким образом, с помощью $l(x)$ можно получить область точек, в которых функционал достигает наименьших значений. Тем самым вероятность $p(\vec{q}_{K+1}|J(\vec{q}))$ определяется следующим образом:

$$p(\vec{q}_{K+1}|J(\vec{q})) = \begin{cases} l(\vec{q}), & J(\vec{q}_{K+1}) < J^\gamma, \\ g(\vec{q}), & J(\vec{q}_{K+1}) \geq J^\gamma. \end{cases}$$

Затем генерируется набор векторов согласно плотности $l(x)$. Из них выбирается вектор \vec{q}^* , на котором достигается максимальное ожидаемое улучшение $EI(x)$, выражающееся по формуле

$$EI(\vec{q}) = \left(\gamma + \frac{g(\vec{q})}{l(\vec{q})}(1 - \gamma) \right)^{-1}.$$

Сходимость по вероятности статистических методов отражена в общей теореме в [31].

Критерием останковки в представленном алгоритме выбрано ограничение на количество итераций $\max_iter = 100$. Схема алгоритма выглядит следующим образом.

Алгоритм 1. Алгоритм древовидных оценок Парзена

Вход: Значения параметров γ , n_{samp} и \max_iter

- 1: **Инициализация:** пространство значений неизвестных параметров $\mathcal{D}_{\text{init}} = \{\vec{q}_k, J(\vec{q}_k), k = 1, \dots, n_{\text{init}}\}$
 - 2: **for** $m = 0, \dots, \max_iter$ **do**
 - 3: Разделить $\mathcal{D}_{n_{\text{init}}+m}$ для генерации пространств $\mathcal{D}_{m_g}^g, \mathcal{D}_{m_l}^l$
 - 4: Получить оценку плотности $l(\vec{q})$ для наборов параметров из $\mathcal{D}_{n_{\text{init}}+m_l}^l$
 - 5: Получить оценку плотности $g(\vec{q})$ для наборов параметров из $\mathcal{D}_{n_{\text{init}}+m_g}^g$
 - 6: Сгенерировать $\vec{q}^s = \{\vec{q}_k^s \mid k = 1, \dots, n_{\text{samp}}\}$, где $\vec{q}_k^s \sim l(\vec{q})$
 - 7: Выбрать $\vec{q}_{m+1} = \underset{\vec{q} \in \vec{q}^s}{\operatorname{argmax}} EI(\vec{q})$
 - 8: Вычислить $J(\vec{q}_{m+1})$
 - 9: $\mathcal{D}_{n_{\text{init}}+m} \leftarrow \mathcal{D}_{n_{\text{init}}+m+1}$
-

Подробнее с методом древовидных оценок Парзена можно ознакомиться в [32].

3.2. Поэтапное восстановление неизвестных параметров. Параметр β предполагается кусочно постоянным. Соответственно, чем продолжительнее рассматриваемый период моделирования, тем больше число неизвестных параметров. Однако каждый из параметров $\beta_d(i)$ и $\beta_c(i)$, $i = 1, \dots, N$, где $N = 18$ — количество месяцев моделирования, зависит только от данных на конкретном подпериоде моделирования. Каждый промежуток последовательно калибровался один за другим (метод усвоения данных), и параметры, восстановленные на предыдущем шаге, использовались в последующем запуске алгоритма оптимизации (алгоритм описан в [33]). Таким образом, на первом промежутке 13.03.2020 г. — 12.04.2020 г. ($i = 1$) использовался вектор неизвестных параметров (начальные условия $t = 0$ соответствуют дате 13.03.2020 г.)

$$\vec{q}(0) = (E(0), \beta, p, \beta_d(1), \beta_c(1)),$$

а на всех последующих промежутках —

$$\vec{q}(i) = (\beta_d(i), \beta_c(i)), \quad i = 2, \dots, N,$$

где N — количество месяцев моделирования.

Регуляризация решения обратной задачи состоит в использовании ограничений на искомые параметры, полученные при анализе чувствительности на основе используемых измерений [28, 29].

4. Численные результаты

Представим результаты математического моделирования распространения COVID-19 в Республике Казахстан. Как было показано в разд. 3.2,

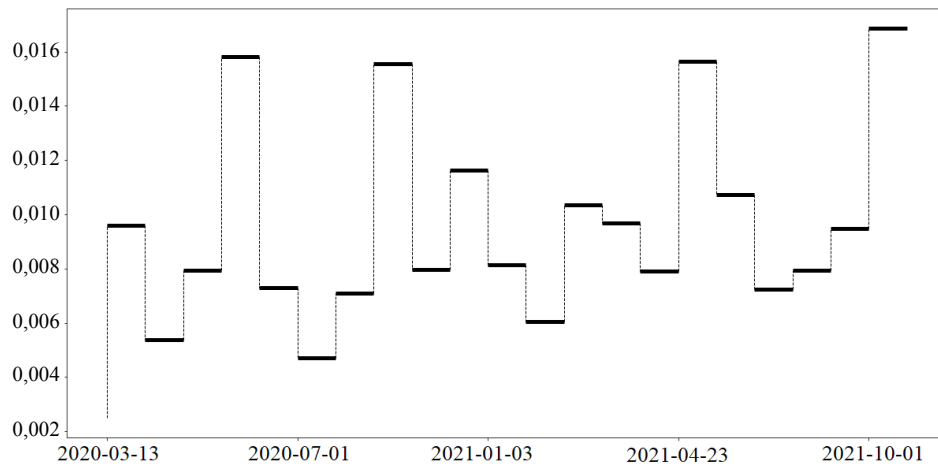


Рис. 4. График изменения параметра заразности $\beta_c(i)$ и дней его изменения $\beta_d(i)$ в РК с 13.03.2020 г. по 12.12.2021 г.

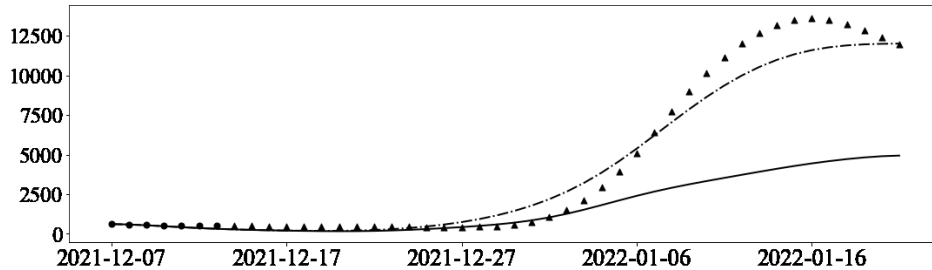


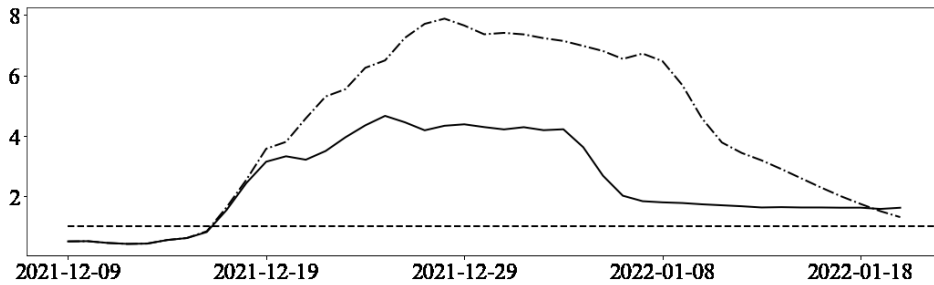
Рис. 5. Моделирование двух сценариев распространения ежедневно выявленных случаев в результате ПЦР-тестирования в РК

на первом этапе восстанавливаются количество бессимптомных инфицированных $E(0)$, скорость передачи вируса от инфицированного к восприимчивому агенту β и значение $\beta_c(1)$, на которое изменится параметр β в день $\beta_d(1)$, а также шанс быть протестированным p . Ежемесячно на основе решения обратной задачи обновлялись параметры дня $\beta_d(i)$ и значения $\beta_c(i)$ изменения скорости передачи вируса в РК (результат восстановления $\beta_d(i)$ и $\beta_c(i)$ приведён на рис. 4).

При построении сценариев распространения ежедневно выявленных случаев COVID-19 предполагается, что в регионе сохранится средний уровень тестирования населения $\mathcal{T}(t)$. Количество ПЦР тестов в день t рассчитан по регрессионной модели SARIMA (см. разд. 1).

4.1. Сценарии распространения COVID-19 в РК. На рис. 5 представлен результат моделирования среднего количества ежедневно выявленных случаев в результате ПЦР-тестирования в РК с прогнозом на 45 дней (точками изображены реальные данные с 13.03.2020 г. по 12.12.2021 г., которые участвуют в решении обратной задачи, а треугольниками — данные с 13.12.2021 г. по 20.01.2022 г., которые использовались для проверки прогнозирования). При построении прогноза были учтены 2 вида сценариев:

- базовый сценарий (сплошная линия), в котором не учитывалось повышенное скопление людей во время новогодних праздников, а учитывалось только увеличение мобильности граждан в период подготовки к каникулам (с 20.12.2021 г. по 30.12.2021 г.);
- повышенная мобильность граждан на новогодние праздники (пунктирная линия) — учитывалась повышенная передача вируса в общественных местах в период с 2.01.2022 г. по 6.01.2022 г. Характеризуется увеличением значения $\beta_c(t)$ передачи вируса в 2,5 раза на период предновогодних праздников 20–30.12.2021 г.: достигает значения 0,039, затем увеличивается до 0,548 в период 1–10.01.2022 г., после чего спадает до 0,03.

Рис. 6. Значение индекса репродукции $\mathcal{R}(t)$

По результатам валидации модели можно сделать вывод, что учёт повышенной мобильности граждан демонстрирует более точное соответствие реальным данным. Значения параметра β во втором сценарии (пунктирная линия) говорят о том, что в период новогодних праздников передача вируса увеличилась в 3,5 раза по сравнению с началом декабря. Так, на 20.01.2022 г. число выявленных случаев COVID-19 составляло 12 032 человека, число ожидаемых выявленных случаев в Республике Казахстан по базовому сценарию АМ (сплошная линия), рассчитанному 12.12.2021 г., равнялось 4 939 человек (погрешность составила 59%), а с учётом повышенной мобильности граждан в общественных местах — 12 007 человек (погрешность составила 0,2%).

4.2. Базовый индекс репродукции вируса в РК. Основным показателем распространения эпидемии является базовый индекс репродукции вируса $R(t)$, который характеризует среднее количество людей, которые заражаются от активно инфицированных в полностью неиммунизированном окружении при отсутствии специальных эпидемиологических мер. В данной работе мы использовали выражение для базового индекса репродукции, предложенное авторами работы [3]:

$$\mathcal{R}(t) = \frac{I_N(t) \cdot d}{I_C(t)}. \quad (2)$$

Здесь $I_N(t)$ — количество новых заражённых в день t , $I_C(t)$ — текущее количество заражённых в день t , d — средняя продолжительность заболевания в днях. Если $\mathcal{R}(t) < 1$, то считается, что эпидемия перестанет распространяться, в противном случае она растёт. На рис. 6 представлен график $\mathcal{R}(t)$ для Республики Казахстан для двух рассматриваемых выше сценариев. Результаты показывают рост количества выявленных случаев COVID-19 в Республике Казахстан и высокую нагрузку на систему здравоохранения в период с 19.12.2021 г. по 18.01.2022 г. для сценария

с увеличением мобильности граждан в общественных местах ввиду новогодних праздников (штрих-пунктирная линия), после чего количество новых выявленных случаев сокращается до размеров базового сценария развития (сплошная линия). Горизонтальной штриховой линией обозначено пороговое значение $\mathcal{R}(t) = 1$.

Заключение

Разработана агентная модель распространения COVID-19 в РК, основанная на программном пакете Covasim, реализованном на языке Python, и включающая в себя инициацию популяции на основе демографических данных страны, правила распространения заболевания и тестирования агентов в зависимости от возраста и эпидемиологического статуса. На первом этапе происходит сбор, обработка и анализ неполных данных методами регрессионного анализа и машинного обучения. На втором этапе уточняются эпидемиологические параметры агентной модели (скорости передачи инфекции, тестирования, начальное количество инфицированных агентов) по дополнительной информации о количестве выявленных случаев COVID-19 в РК. Для этого разработан алгоритм усвоения данных, в рамках которого неизвестные параметры идентифицируются ежемесячно по дополнительной информации о количестве ежедневно выявленных случаев COVID-19 на основе метода глобальной оптимизации древовидных оценок Парзена. На третьем этапе учитываются нефармацевтические вмешательства в процесс распространения эпидемии с целью построения наиболее реалистичных сценариев распространения COVID-19.

При реализации модели использовались данные о количестве выявленных случаев с 13.03.2020 г. по 12.12.2021 г. в Республике Казахстан. Показано, что увеличение плотности агентов в новогодние праздники в общественных местах (магазины, театры, парки) увеличивает количество выявленных случаев COVID-19 (к 15.01.2022 г. увеличилось в 3,5 раза по сравнению с 1.12.2021 г.). В качестве примера приведены два сценария распространения COVID-19, рассчитанных 12.12.2021 г. на период до 20.01.2022 г. Сценарий, в котором учитывались новогодние праздники (опубликован 12.12.2021 г. на сайте covid19-modeling.ru), практически совпал с тем, что произошло в реальности (погрешность составила 0,2% или 25 человек). Таким образом, математическое моделирование позволяет получить качественное и количественное совпадение прогноза эпидемиологической ситуации с действительностью.

Особенность моделирования распространения коронавирусной инфекции в Республике Казахстан состоит в том, что необходимо учитывать концентрацию населения в крупных городах: Алматы (1 993 067 человек), Астана (1 199 083 человека), Шымкент (1 090 160 человек) и др., —

в которых проживает более 11 млн человек, а в сёлах — более 7 млн человек. Для более детального моделирования сценариев распространения COVID-19 необходимо учитывать транспортные потоки между крупнейшими городами, а также транспортные потоки на уровне город-область. Для получения более детальных сценариев распространения COVID-19 необходимо комбинировать агентные и SIR-модели, как это сделано в Новосибирской области [2, 15, 16].

Вклад каждого автора в работу следующий:

- О. И. Криворотько и С. И. Кабанихин — постановка прямой и обратной задач, формулировка алгоритмов решения и анализ результатов расчётов, координация работ;
- М. А. Бектемесов — предоставление данных по Республике Казахстан;
- М. И. Сосновская — реализация агентной модели и алгоритмов решения прямой и обратной задач;
- А. В. Неверов — обработка данных, разработка программы для вычисления на кластере ССКЦ СО РАН.

ЛИТЕРАТУРА

1. Адарченко В. А., Бабань С. А., Брагин А. А. [и др.]. Моделирование развития эпидемии коронавируса по дифференциальной и статистической моделям. Снежинск, 2020. (Препринт РФЯЦ — ВНИИТФ; № 264).
2. Криворотько О. И., Кабанихин С. И. Математические модели распространения COVID-19. Новосибирск, 2022. (Препринт Ин-та математики им. С. Л. Соболева; № 300).
3. Kerr C. C., Stuart R. M., Mistry D. [et al.]. Covasim: An agent-based model of COVID-19 dynamics and interventions // PLoS Comput. Biol. 2021. V. 17, No. 7, ID e1009149. 32 p.
4. Laubenbacher R., Hinkelmann F., Oremland M. Agent-based models and optimal control in biology: A discrete approach // Mathematical concepts and methods in modern biology: Using modern discrete models. Ch. 5. San Diego, CA: Acad. Press, 2013. P. 143–178.
5. Влад А. И., Санникова Т. Е., Романюха А. А. Моделирование распространения респираторных вирусных инфекций в городе: мультиагентный подход // Мат. биология и биоинформатика. 2020. Т. 15, № 2. С. 338–356.
6. Aleta A., Martín-Corral D., Pastore y Piontti A. [et al.]. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19 // Nat. Hum. Behav. 2020. V. 4, No. 9. P. 964–971.
7. Lau M. S. Y., Grenfell B., Thomas M., Bryan M., Nelson K., Loman B. Characterizing superspreading events and age-specific infectiousness of SARS-CoV-2 transmission in Georgia, USA // PNAS. 2020. V. 117, No. 36. P. 22430–22435.

8. **Kucharski A. J., Klepac P., Conlan A. J. K.** [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.* 2020. Vol. 20, No. 10. P. 1151–1160.
9. **Hoertel N., Blachier M., Blanco C.** [et al.]. A stochastic agent-based model of the SARS-CoV-2 epidemic in France // *Nat. Med.* 2020. V. 26, No. 9. P. 1417–1421.
10. **Hellewell J., Abbott S., Gimma A.** [et al.]. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts // *Lancet Glob. Health.* 2020. V. 8, No. 4. P. e488–e496.
11. **Nielsen B. F., Simonsen L., Sneppen K.** COVID-19 superspreading suggests mitigation by social network modulation // *Phys. Rev. Lett.* 2021. V. 126, No. 11, ID 118301. 6 p.
12. COVID-19 agent-based simulator. Bellevue, WA: Inst. Disease Modeling, 2022. Available at github.com/InstituteForDiseaseModeling/covasim/ (accessed Jan. 9, 2023).
13. **Noll N. B., Aksamentov I., Druelle V.** [et al.]. COVID-19 scenarios: An interactive tool to explore the spread and associated morbidity and mortality of SARS-CoV-2. Oyster Bay, NY: Cold Spring Harbor Lab., 2020. (Prepr. Server Health Sci. medRxiv). Available at medrxiv.org/content/10.1101/2020.05.05.20091363 (accessed Jan. 9, 2023).
14. **Tuomisto J. T., Yrjölä J., Kolehmainen M., Bonsdorff J., Pekkanen J., Tikkanen T.** An agent-based epidemic model REINA for COVID-19 to identify destructive policies. Oyster Bay, NY: Cold Spring Harbor Lab., 2020. (Prepr. Server Health Sci. medRxiv). Available at medrxiv.org/content/10.1101/2020.04.09.20047498 (accessed Jan. 9, 2023).
15. **Криворотко О. И., Кабанихин С. И., Зятков Н. Ю., Приходько А. Ю., Прохoshин Н. М., Шишленин М. А.** Математическое моделирование и прогнозирование COVID-19 в Москве и Новосибирской области // *Сиб. журн. вычисл. математики.* 2020. Т. 23, № 4. С. 395–414.
16. **Krivorotko O. I., Sosnovskaia M. I., Vashchenko I. A., Kerr C. C., Lesnic D.** Agent-based modeling of COVID-19 outbreaks for New York state and UK: Parameter identification algorithm // *Infect. Dis. Model.* 2022. V. 7, No. 1. P. 30–44.
17. API reference — pandas 1.5.2 documentation. Austin, TX: NumFOCUS, 2022. Available at pandas.pydata.org/docs/reference/ (accessed Jan. 9, 2023).
18. **Бокс Дж., Дженкинс Г.** Анализ временных рядов: Прогноз и управление. Вып. 1. М.: Мир, 1974. 406 с.
19. **Dabral P. P., Murry M. Z.** Modelling and forecasting of rainfall time series using SARIMA // *Environ. Process.* 2017. V. 4, No. 2. P. 399–419.
20. **Dickey D. A., Fuller W. A.** Distribution of the estimators for autoregressive time series with a unit root // *J. Am. Stat. Assoc.* 1979. V. 74, No. 366a. P. 427–431.
21. SciPy API — SciPy v1.10.0 manual. Austin, TX: Enthought, 2022. Available at docs.scipy.org/doc/scipy/reference/ (accessed Jan. 9, 2023).

22. Household: Size & composition, 2022. New York: United Nations, 2022. Available at population.un.org/Household/#/countries/840 (accessed Jan. 9, 2023).
23. SynthPops — create synthetic populations for COVID-19 epidemic analyses. Bellevue, WA: Inst. Disease Modeling, 2021. Available at github.com/InstituteForDiseaseModeling/synthpops/ (accessed Jan. 9, 2023).
24. Lauer S. A., Grantz K. H., Bi Q. [et al.]. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application // *Ann. Intern. Med.* 2020. V. 172, No. 9. P. 577–582.
25. Wölfel R., Corman V. M., Guggemos W. [et al.]. Virological assessment of hospitalized patients with COVID-2019 // *Nature*. 2020. V. 581, No. 7809. P. 465–469.
26. Verity R., Okell L. C., Dorigatti I. [et al.]. Estimates of the severity of coronavirus disease 2019: A model-based analysis // *Lancet Infect. Dis.* 2020. V. 20, No. 6. P. 669–677.
27. Wang D., Hu B., Hu C. [et al.]. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China // *JAMA*. 2020. V. 323, No. 11. P. 1061–1069.
28. Криворотько О. И., Кабанихин С. И., Сосновская М. И., Андорная Д. В. Анализ чувствительности и идентифицируемости математических моделей распространения эпидемии COVID-19 // *Вавиловский журн. генетики и селекции*. 2021. Т. 25, № 1. С. 82–91.
29. Krivorotko O. I., Sosnovskaia M. I., Vashchenko I. A. Agent-based mathematical model of COVID-19 spread in Novosibirsk region: Identifiability, optimization and forecasting // *J. Inverse Ill-Posed Probl.* [in print].
30. Optuna — a hyperparameter optimization framework. Tokyo: Preferred Networks, 2018. Available at optuna.org (accessed Jan. 9, 2023).
31. Жиглявский А. А., Жилинскас А. Г. Методы поиска глобального экстремума. М.: Наука, 1991.
32. Bergstra J., Bardenet R., Bengio Y., Kégl B. Algorithms for hyperparameter optimization // *Advances in neural information processing systems* 24. 25th Annual Conf. (Granada, Spain, Dec. 12–15, 2011). V. 3. Red Hook, NY: Curran Associates, 2012. P. 2546–2554.
33. Daza-Torres M. L., Capistrán M. A., Capella A., Andrés Christen J. Bayesian sequential data assimilation for COVID-19 forecasting // *Epidemics*. 2022. V. 39, ID 100564. 10 p.

Криворотько Ольга Игоревна
 Кабанихин Сергей Игоревич
 Бектемесов Мактагали Абдимажитович
 Сосновская Мария Игоревна
 Неверов Андрей Вячеславович

Статья поступила
 4 июля 2022 г.
 После доработки —
 27 сентября 2022 г.
 Принята к публикации
 28 сентября 2022 г.

SIMULATION OF COVID-19 PROPAGATION SCENARIOS
IN THE REPUBLIC OF KAZAKHSTAN
BASED ON REGULARIZATION OF AGENT MODEL

O. I. Krivorotko^{1,2,3,a}, *S. I. Kabanikhin*^{2,3,b}, *M. A. Bektemesov*^{4,c},
M. I. Sosnovskaya^{3,d}, and *A. V. Neverov*^{1,3,e}

¹ Institute of Computational Mathematics and Mathematical Geophysics,
6 Acad. Lavrentiev Avenue, 630090 Novosibirsk, Russia

² Sobolev Institute of Mathematics,
4 Acad. Koptug Avenue, 630090 Novosibirsk, Russia

³ Novosibirsk State University,
2 Pirogov Street, 630090 Novosibirsk, Russia

⁴ Abai Kazakh National Pedagogical University,
13 Dostyk Avenue, 050010 Almaty, Kazakhstan

E-mail: ^a*krivorotko.olya@mail.ru*, ^b*ksi52@mail.ru*, ^c*maktagali@mail.ru*,
^d*m.sosnovskaya@alumni.nsu.ru*, ^e*a.neverov@g.nsu.ru*

Abstract. An algorithm for modeling scenarios for new diagnosed cases of COVID-19 in the Republic of Kazakhstan is proposed. The algorithm is based on the treatment of incomplete epidemiological data and the inverse problem solving for the agent-based model (ABM) using a set of available epidemiological data. The main tool for building the ABM is the open library Covasim. In the event of a sudden change in the situation (appearance of a new strain, removal or introduction of restrictive measures, etc.), the model parameters are updated with additional information for the previous month (data assimilation). The inverse problem was solved by tree Parzen estimates optimization. As an example, two scenarios of COVID-19 propagation are given, calculated on December 12, 2021 for the period up to January 20, 2022. The scenario,

This research is supported by the Council for Grants under the President of the Russian Federation (Project MK–4994.2021.1.1), Russian Foundation for Basic Research (Project 21–51–10003) and Ministry of Education and Science of the Republic of Kazakhstan (Project AP09260317).

English version: Journal of Applied and Industrial Mathematics **17** (1) (2023).

which took into account the New Year holidays (published on December 12, 2021 on `covid19-modeling.ru`), almost coincided with what happened in reality (the error was 0,2%). Tab. 3, illustr. 6, bibliogr. 33.

Keywords: agent oriented model, COVID-19, inverse problem, optimization, regularization, scenario, index of virus reproduction.

REFERENCES

1. **V. A. Adarchenko, S. A. Baban', A. A. Bragin**, [et al.]. Modelling the coronavirus epidemic development with the use of differential and statistical models (Snezhinsk, 2020) (Prepr. RFYaTs—VNIITF, No. 264) [Russian].
2. **O. I. Krivorotko and S. I. Kabanikhin**, Mathematical models for the spread of COVID-19 (Novosibirsk, 2022) (Prepr. Inst. Mat. Sobolev., No. 300) [Russian].
3. **C. C. Kerr, R. M. Stuart, D. Mistry**, [et al.]. Covasim: An agent-based model of COVID-19 dynamics and interventions, *PLoS Comput. Biol.* **17** (7), ID e1009149, 32 p. (2021).
4. **R. Laubenbacher, F. Hinkelmann**, and **M. Oremland**, Agent-based models and optimal control in biology: A discrete approach, in *Mathematical Concepts and Methods in Modern Biology: Using Modern Discrete Models*, Ch. 5 (Acad. Press, San Diego, CA, 2013), pp. 143–178.
5. **A. I. Vlad, T. E. Sannikova**, and **A. A. Romanyukha**, Modelling the spread of respiratory viral infections in a city: Multi-agent approach, *Mat. Biol. Bioinform.* **15** (2), 338–356 (2020) [Russian].
6. **A. Aleta, D. Martín-Corral, A. Pastore y Piontti**, [et al.]. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19, *Nat. Hum. Behav.* **4** (9), 964–971 (2020).
7. **M. S. Y. Lau, B. Grenfell, M. Thomas, M. Bryan, K. Nelson**, and **B. Lopman**, Characterizing superspreading events and age-specific infectiousness of SARS-CoV-2 transmission in Georgia, USA, *PNAS* **117** (36), 22430–22435 (2020).
8. **A. J. Kucharski, P. Klepac, A. J. K. Conlan**, [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** (10), 1151–1160 (2020).
9. **N. Hoertel, M. Blachier, C. Blanco**, [et al.]. A stochastic agent-based model of the SARS-CoV-2 epidemic in France, *Nat. Med.* **26** (9), 1417–1421 (2020).
10. **J. Hellewell, S. Abbott, A. Gimma**, [et al.]. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts, *Lancet Glob. Health.* **8** (4), e488–e496 (2020).
11. **B. F. Nielsen, L. Simonsen**, and **K. Sneppen**, COVID-19 superspreading suggests mitigation by social network modulation, *Phys. Rev. Lett.* **126** (11), ID 118301, 6 p. (2021).

12. COVID-19 Agent-Based Simulator (Inst. Disease Modeling, Bellevue, WA, 2022). Available at github.com/InstituteForDiseaseModeling/covasim/ (accessed Jan. 9, 2023).
13. **N. B. Noll, I. Aksamentov, V. Druelle**, [et al.]. *COVID-19 scenarios: An interactive tool to explore the spread and associated morbidity and mortality of SARS-CoV-2* (Cold Spring Harbor Lab., Oyster Bay, NY, 2020). (Prepr. Server Health Sci. medRxiv). Available at medrxiv.org/content/10.1101/2020.05.05.20091363 (accessed Jan. 9, 2023).
14. **J. T. Tuomisto, J. Yrjölä, M. Kolehmainen, J. Bonsdorff, J. Pekkanen**, and **T. Tikkanen**, *An agent-based epidemic model REINA for COVID-19 to identify destructive policies* (Cold Spring Harbor Lab., Oyster Bay, NY, 2020). (Prepr. Server Health Sci. medRxiv). Available at medrxiv.org/content/10.1101/2020.04.09.20047498 (accessed Jan. 9, 2023).
15. **O. I. Krivorotko, S. I. Kabanikhin, N. Yu. Ziatkov, A. Yu. Prikhodko, N. M. Prokhoshin**, and **M. A. Shishlenin**, Mathematical modeling and forecasting of COVID-19 in Moscow and Novosibirsk region, *Sib. Zh. Vychisl. Mat.* **23** (4), 395–414 (2020) [Russian] [*Num. Anal. Appl.* **13** (4), 332–348 (2020)].
16. **O. I. Krivorotko, M. I. Sosnovskaia, I. A. Vashchenko, C. C. Kerr**, and **D. Lesnic**, Agent-based modeling of COVID-19 outbreaks for New York state and UK: Parameter identification algorithm, *Infect. Dis. Model.* **7** (1), 30–44 (2022).
17. API Reference — Pandas 1.5.2 Documentation (NumFOCUS, Austin, TX, 2022). Available at pandas.pydata.org/docs/reference/ (accessed Jan. 9, 2023).
18. **G. E. P. Box** and **G. M. Jenkins**, *Time Series Analysis: Forecasting and Control* (Holden-Day, San Francisco, 1970; Mir, Moscow, 1974 [Russian]).
19. **P. P. Dabral** and **M. Z. Murry**, Modelling and forecasting of rainfall time series using SARIMA, *Environ. Process.* **4** (2), 399–419 (2017).
20. **D. A. Dickey** and **W. A. Fuller**, Distribution of the estimators for autoregressive time series with a unit root, *J. Am. Stat. Assoc.* **74** (366a), 427–431 (1979).
21. SciPy API — SciPy v1.10.0 Manual (Enthought, Austin, TX, 2022). Available at docs.scipy.org/doc/scipy/reference/ (accessed Jan. 9, 2023).
22. Household: Size & Composition, 2022 (United Nations, New York, 2022). Available at population.un.org/Household/#/countries/840 (accessed Jan. 9, 2023).
23. SynthPops — create synthetic populations for COVID-19 epidemic analyses (Inst. Disease Modeling, Bellevue, WA, 2021). Available at github.com/InstituteForDiseaseModeling/synthpops/ (accessed Jan. 9, 2023).
24. **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), 577–582 (2020).
25. **R. Wölfel, V. M. Corman, W. Guggemos**, [et al.]. Virological assessment of hospitalized patients with COVID-2019, *Nature* **581** (7809), 465–469 (2020).

26. **R. Verity, L. C. Okell, I. Dorigatti, [et al.]**. Estimates of the severity of coronavirus disease 2019: A model-based analysis, *Lancet Infect. Dis.* **20** (6), 669–677 (2020).
27. **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), 1061–1069 (2020).
28. **O. I. Krivorotko, S. I. Kabanikhin, M. I. Sosnovskaya, and D. V. Andornaya**, Sensitivity and identifiability analysis of COVID-19 pandemic models, *Vavilov. Zh. Genet. Sel.* **25** (1), 82–91 (2021) [Russian].
29. **O. I. Krivorotko, M. I. Sosnovskaya, and I. A. Vashchenko**, Agent-based mathematical model of COVID-19 spread in Novosibirsk region: Identifiability, optimization and forecasting, *J. Inverse Ill-Posed Probl.* [in print].
30. Optuna — A Hyperparameter Optimization Framework (Preferred Networks, Tokyo, 2018). Available at optuna.org (accessed Jan. 9, 2023).
31. **A. A. Zhiglyavskii and A. G. Zhilinskii**, *Methods for Global Extremum Search* (Moscow, Nauka, 1991).
32. **J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl**, Algorithms for hyperparameter optimization, in *Advances in Neural Information Processing Systems 24* (25th Annual Conf., Granada, Spain, Dec. 12–15, 2011), Vol. 3 (Curran Associates, Red Hook, NY, 2012), pp. 2546–2554.
33. **M. L. Daza-Torres, M. A. Capistrán, A. Capella, and J. Andrés Christen**, Bayesian sequential data assimilation for COVID-19 forecasting, *Epidemics* **39**, ID 100564, 10 p. (2022).

Olga I. Krivorotko
 Sergey I. Kabanikhin
 Maktagali A. Bektemesov
 Mariya I. Sosnovskaya
 Andrey V. Neverov

Received July 4, 2022
 Revised September 27, 2022
 Accepted September 28, 2022