

УДК 519.673, 57.087.1

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

© 2021 О. И. Криворотко<sup>1,2a</sup>, С. И. Кабанихин<sup>1b</sup>, М. А. Бектемесов<sup>3с</sup>,  
М. И. Сосновская<sup>4d</sup>, А. В. Неверов<sup>4е</sup>

<sup>1</sup> *Институт вычислительной математики и математической геофизики  
Сибирского отделения Российской академии наук, 630090, пр. Академика  
Лаврентьева 6, г. Новосибирск, Россия*

<sup>2</sup> *Институт математики им. С.Л. Соболева Сибирского отделения Российской  
академии наук, 630090, пр. Академика Коптюга 4, г. Новосибирск, Россия*

<sup>3</sup> *Казахский национальный педагогический университет имени Абая, 050010, пр.  
Достык 13, г. Алматы, Республика Казахстан*

<sup>4</sup> *Федеральное государственное автономное образовательное учреждение высшего  
образования «Новосибирский национальный исследовательский государственный  
университет», 630090, ул. Пирогова 2, г. Новосибирск, Россия*

E-mails: <sup>a</sup>krivorotko.olya@mail.ru, <sup>b</sup>ksi52@mail.ru, <sup>c</sup>maktagali@mail.ru,  
<sup>d</sup>m.sosnovskaya@alumni.nsu.ru, <sup>e</sup>a.neverov@g.nsu.ru

В работе построена агентная модель расчета сценариев динамики  $f(t)$  выявленных случаев COVID-19 в Республике Казахстан, в основе которой 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, реализованная на языке программирования Python. Обратная задача в вариационной постановке решается поэтапно: в начале каждого месяца обновляются параметры дня изменения скорости передачи инфекции  $\beta_d(t)$  и значения  $\beta_c(t)$ , на которое оно изменится. В случае резкого изменения ситуации (появление нового штамма, отмена или введение ограничительных мер и т.п.) параметры модели обновляются с учетом дополнительной информации за предыдущий месяц (оперативное усвоение данных). Обратная задача решалась методом стохастической глобальной оптимизации (древовидных оценок Парзена). Показано, что увеличение концентрации агентов в новогодние праздники в общественных местах (магазины, театры, парки) увеличивает количество выявленных случаев COVID-19 (к 15.01.2022 увеличилось в 3.5 раза по сравнению с 01.12.2021). В качестве примера приведены два сценария распространения COVID-19, рассчитанных 12 декабря 2021 года на период до 20 января 2022 года. Сценарий, в котором учитывались новогодние праздники (опубликован 12 декабря 2021 года на сайте <https://covid19-modeling.ru/>), практически совпал с тем, что произошло в реальности (погрешность составила 0.2%).

**Ключевые слова:** агентно-ориентированная модель, COVASIM, коронавирусная инфекция, популяционное моделирование, имитационное моделирование, обработка данных, SARIMA, уточнение параметров, обратная задача, оптимизация, регуляризация, сценарии развития, индекс репродукции.

### 1. ВВЕДЕНИЕ

Агентная модель (АМ) распространения COVID-19 наиболее адекватно работает в регионах с

населением от 100 до 20 млн человек [1, 2, 3] и позволяет:

- учитывать демографическую информацию по конкретной стране (численность и возрастная структура 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 агентной модели каждый действующий агент наделяется признаками (возраст, социальный статус, восприимчивость к заболеванию и т.д.).

Приведем краткий обзор работ по агентному моделированию. Группа американских ученых [3] разработала программный комплекс Covasim [4], основу которого составляет агентный подход моделирования эпидемии с учетом 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 политических решений более чем в десятке стран Африки, Азиатско-Тихоокеанского региона, Европы и Северной Америки.

В работе [5] АМ применялась для описания распространения COVID-19 в Бостонской столичной зоне. 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.

В работе [6] изучалось моделирование распространения COVID-19 в штате Джорджия, США. Авторы 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 [7, 8, 9] в рамках АМ распространения COVID-19 проанализированы противоэпидемические программы в различных регионах Франции и Великобритании, в результате чего получено понимание эффективных мер для разных географических и демографических условий, а также текущих штаммов SARS-CoV-2. В большинстве сценариев высокоэффективного отслеживания контактов и изоляции случаев достаточно для контроля новой вспышки COVID-19 в течение 3 месяцев. Вероятность контроля снижается при длительной задержке от появления симптомов до изоляции, уменьшении числа случаев, выявленных при отслеживании контактов, и увеличении передачи инфекции до появления симптомов.

В работе [10] авторы построили АМ, в котором источники заражения выступали суперраспространителями. Они показали, что сверхраспространение резко усиливает значимость ограничений личных контактов.

Приведем краткий обзор открытых программных комплексов анализа распространения COVID-19:

1. <https://covid19-scenarios.org/>. Реализуется агентно-структурированная по возрастам камерная модель распространения коронавирусной инфекции, основанная на 9 дифференциальных уравнениях с возможностью варьирования параметров модели [11].
2. <https://covid19.biouml.org/>. Группа новосибирских ученых моделируют распространение COVID-19 в Москве, Новосибирской области, Германии, Франции и Италии, в котором участвуют 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 легких (ИВЛ) и других характеристик, необходимых для эффективного управления ситуацией в условиях эпидемии.

3. <https://www.anylogic.com/healthcare/>, Богота, Колумбия – на основе агентной модели с учетом географических особенностей города (расположение школ, медицинских учреждений, общественных мест) и социальной дистанции между агентами с возможностью варьирования параметров.
4. <https://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 [12].
5. <https://github.com/InstituteForDiseaseModeling/covasim>, США – на основе агентной модели со случайной структурой с возможностью идентификации параметров для конкретного региона [3]. Данный программный комплекс послужил основой создания собственного программного комплекса (см. п. 6), а также был адаптирован для получения результатов для Республики Казахстан в рамках данной работы.
6. <https://covid19-modeling.ru/>, Новосибирская область – на базе комбинации камерной и агентной моделей с возможностью построения сценариев развития на основе решения обратных задач [13, 14].

В силу новизны и сложности заболевания COVID-19 параметры большинства математических моделей, как правило, неизвестны, что затрудняет адаптировать существующие программные решения для анализа ситуации в конкретном регионе. Основные проблемы моделирования распространения COVID-19:

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  $\beta(t)$ , вероятность появления тяжелых 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 удаленный рабочий режим, закрытие школ, предприятий, районов и городов), появление новых штаммов (альфа, бета, гамма, дельта, омикрон и другие).

Указанные проблемы приводят к необходимости рассматривать и решать обратные задачи, а именно задачи идентификации неизвестных эпидемиологических параметров в конкретном регионе с учетом мутации вируса и различных вмешательств (политические, фармацевтические) по дополнительным измерениям эпидемиологических характеристик в данный момент времени (количество проведенных ПЦР-тестов, выявленных, госпитализированных, критических и умерших от COVID-19 случаев). Ввиду того, что исследуемые обратные задачи некорректны (решение может быть неединственным и/или неустойчивым), в работе используется регуляризация решения обратных задач. Регуляризация осуществляется с использованием ограничений на искомые параметры на каждой итерации, которые получены из оценок Всемирной организации здравоохранения (ВОЗ) и анализа идентифицируемости [15, 16].

Особенность представленной нами работы заключается в адаптации агентной модели к Республике Казахстан с учетом распределения населения, городов, ограничительных мер и типа статистических данных. Особенностью рассмотренной модели для Республики Казахстан является структура данных. В данной работе предложен алгоритм поэтапного уточнения эпидемиологических параметров агентной модели распространения COVID-19 в случае неполных

и неточных данных на примере Республики Казахстан, а также алгоритм построения сценариев развития эпидемии в регионе на основе комбинации методов машинного обучения и решения обратных задач.

Для достоверного моделирования тех или иных факторов распространения COVID-19 в основу расчета 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 зашумленных статистических данных в работе применяются методы машинного обучения и регрессионного анализа обработки и анализа временных рядов (см. Раздел 2). Исходя из особенностей данных для Республики Казахстан, в Разделе 3 строится агентная модель, основная цель которой – составление сценариев развития и оценка влияния различных вмешательств на эпидемию, формулируются прямая и обратная задачи для АМ. В Разделе 4 описан алгоритм решения обратной задачи на основе метода древовидных оценок Парзена минимизации абсолютного целевого функционала и усвоения данных. В Разделе 5 построены и проанализированы сценарии распространения COVID-19 в Республике Казахстан с учетом ограничительных мер с 13 декабря 2021 года по 20 января 2022 года. Показано, что увеличение концентрации агентов в новогодние праздники в общественных местах (магазины, театры, парки) увеличивает количество выявленных случаев COVID-19 (к 15.01.2022 увеличилось в 3.5 раза по сравнению с 01.12.2021). В разделе 6 приведено заключение.

## 2. АНАЛИЗ ДАННЫХ ПО РЕСПУБЛИКЕ КАЗАХСТАН

В данном разделе представлены основные демографические и эпидемиологические данные по Республике Казахстан, которые были использованы в последующем построении агентной модели распространения COVID-19, описанной в Разделе 3.

### 2.1. Демографические данные

Статистические данные о численности населения по возрастным категориям (возрастные группы разбиты на периоды по 10 лет) были предоставлены соавтором статьи М.А. Бектемесовым (Таблица 1). Такого рода данные используются при инициации искусственной популяции агентной модели (см. Раздел 3.2).

Возрастная категория (лет)	Количество, человек
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

Таблица 1: Распределение населения по возрастам в группах по 10 лет в Республике Казахстан при общей численности популяции 18 879 552 человек на 01.10.2021.

### 2.2. Эпидемиологические данные

Данные были собраны из открытых источников [ourworldindata.org](https://ourworldindata.org) (OWD) и [worldhealthorg.shinyapps.io/covid/](https://worldhealthorg.shinyapps.io/covid/) (WHO), а также из локальных источников Республики

Казахстан [www.coronavirus2020.kz/ru](http://www.coronavirus2020.kz/ru) (CV2020) и [kt.kz](http://kt.kz) (КТ) при помощи программы для автоматизированного сбора.

Для построения и валидации агентной модели были собраны следующие эпидемиологические данные о распространённости COVID-19 в день  $t$ :

- количество проведенных ПЦР тестов  $T(t)$  (см. Рис. 1) – OWD,
- количество диагностированных случаев COVID-19 (пациентов с положительным ПЦР тестом)  $f(t)$  (см. 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, подключенных к аппарату ИВЛ  $C(t)$  – КТ, CV2020,
- число умерших от COVID-19  $D(t)$  – OWD, WHO, CV2020, КТ.

Схема работы программы:

1. Задается 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 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.

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

Статистические данные  $T(t)$  о количестве проведенных ПЦР тестов в Республике Казахстан имеют пропуски с 06.08.2021, а также свойственную недельную сезонность, поэтому для построения агентной модели был использован прогноз временного ряда  $T(t)$  с помощью сезонной авторегрессионной модели SARIMA являющейся модификацией модели ARIMA (AutoRegressive Integrated Moving Average) [18], которая описывает одномерные временные ряды с сезонной компонентой [19].

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

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

Здесь  $\epsilon(n)$  – стационарный временной ряд белого шума,  $c, a_i, b_j$  – параметры модели,  $\Delta^d$  – 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.),  $n$  – временной параметр (день).

Алгоритм прогнозирования временного ряда:

1. Применяем преобразование Бокса-Кокса [20] для уменьшения дисперсии.

2. Вычисляем сезонную разность (сдвиг на 7 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 [21].
5. Передаем соответствующие проделанным действиям параметры в модель ARIMA(1, 2, 2) и подбираем остальные на основе минимизации информационного критерия Акаике. В качестве данных передается ряд из п. 1.
6. Применяем обратное преобразование Бокса-Кокса к полученной модели с настроенными гиперпараметрами.

Прогноз временного ряда  $T(t)$  количества ежедневных ПЦР-тестов в Республике Казахстан с 06.08.2021 по 20.01.2022 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 [22]. Таким образом, мы избавляемся от возможной неустойчивости и неинтерпретируемости результатов модели ввиду резких скачков в данных (серая линия на рис. 1 и 2). Сглаженные данные, которые были использованы при построении модели, представлены черной линией на рис. 1 и 2.

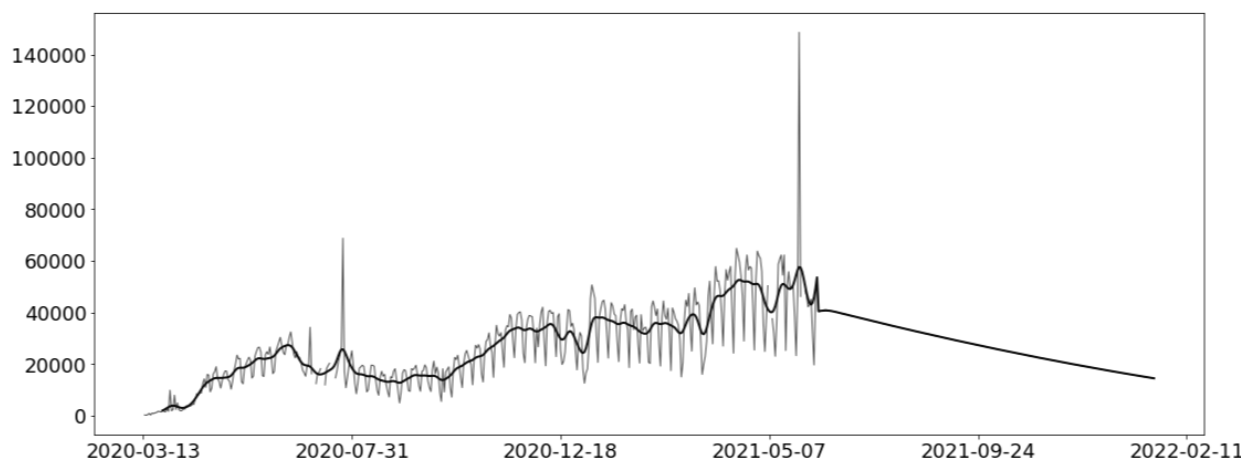


Рис. 1: Статистические данные о количестве проведенных ПЦР тестов  $T(t)$  в Республике Казахстан, используемые в построении модели. Серой линией обозначены необработанные данные, полученные с открытого источника [ourworldindata.org](https://ourworldindata.org), черной линией – данные после обработки (заполнение пропусков, экстраполяция данных и сглаживание), которые были переданы в модель.

### 3. АГЕНТНАЯ МОДЕЛЬ

Агентное моделирование основано на исследовании динамики развития заболевания путем 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 подразделе 3.1. Оно включает в себя инициацию популяции (подраздел 3.2), правила распространения заболевания (подраздел 3.3) и тестирования агентов (подраздел 3.4). В разделе 3.5 приведена постановка прямой задачи для АМ, а в разделе 3.6 – постановка обратной задачи для АМ.

#### 3.1. Общее описание модели



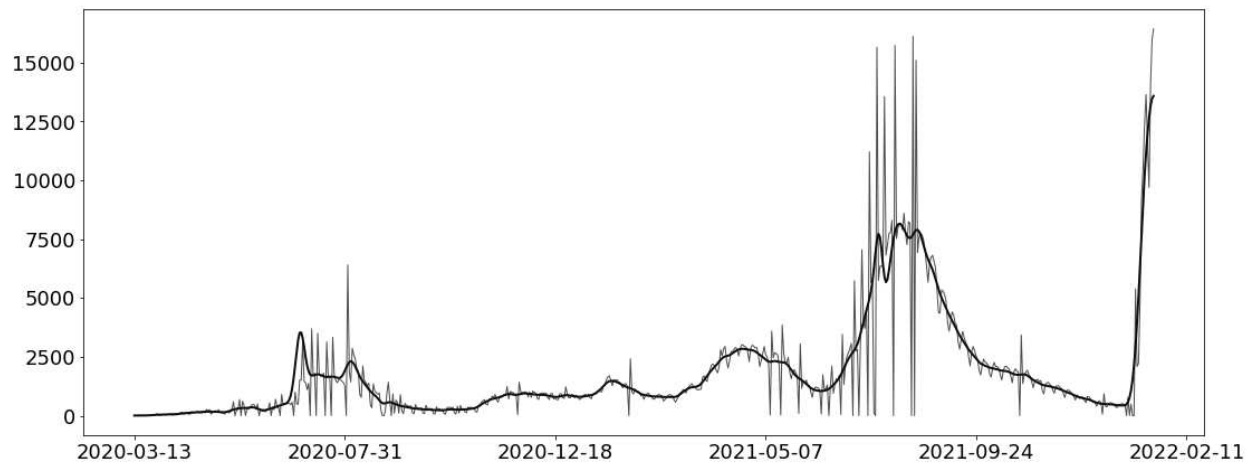


Рис. 2: Статистические данные по выявленным случаям COVID-19  $f(t)$  в Республике Казахстан, используемые в построении модели. Серой линией обозначены необработанные данные, полученные с открытого источника [ourworldindata.org](https://ourworldindata.org), черной линией – данные после обработки (заполнение пропусков, экстраполяция данных и сглаживание), которые были переданы в модель.

В рамках данного исследования была реализована стохастическая АМ для Республики Казахстан. Основным инструментом для создания модели является библиотека COVASIM [3], реализованная на языке программирования Python (открытый код доступен по ссылке [4]) и созданная для исследования агентных моделей COVID-19 с нетривиальными структурами. Общий алгоритм выглядит следующим образом: загружаются все необходимые параметры и статистические данные, создается искусственная популяция с учетом распределения по возрастам в регионе. Далее агенты соединяются в контактные сети, представляющие собой полносвязные графы. Затем начинается цикл по времени: на каждом шаге (временной интервал равен одному дню) обновляется эпидемиологический статус агента как суперпозиция вероятностей с учетом его структуры контактов и введенных ограничительных мер (самоизоляция, закрытие общественных мест, ношение масок и т.д.).

### 3.2. 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 лет, 10–19, ..., 80+) согласно статистическим данным Республики Казахстан (Таблица 1).
- Социальный статус (рабочий, студент, ребенок, 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, представляющие собой полносвязные графы, степень которых определяется пуассоновской случайной величиной с различными параметрами  $\lambda$ :

- для домохозяйств – средний размер семьи, равный  $\lambda = 3.496$  человек [23];
- 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.

### 3.3. Распространение заболевания

В рамках модели предполагается, что вирус передается между агентами, соединенными ребром графа. Заражение при близком контакте описывается кусочно-постоянным параметром  $\beta(t)$ , который в зависимости от структуры контакта умножается на соответствующую константу  $w_\beta$  (для домохозяйств  $w_\beta = 3$ , 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 (рис. 3):  $S$  – восприимчивые к заражению,  $E$  – зараженные незаразные,  $A$  – бессимптомные больные,  $Sym$  – больные с симптомами,  $M$  – больные легкой степени тяжести,  $H$  – госпитализированные,  $C$  – больные в критическом состоянии (нуждаются в реанимации),  $R$  – вылечившиеся,  $D$  – умершие. Переход из одной стадии заболевания в другую контролируется параметрами, зависящими от возраста (т.е. чем старше агент, тем он более уязвим, см. Таблицу 2).

Возраст Пар-р	0-9	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80+
$p_{sym}$	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9
$p_{sev}$	$5 \cdot 10^{-4}$	0.0016	0.0072	0.0208	0.0343	0.0765	0.133	0.207	0.246
$p_{crit}$	$3 \cdot 10^{-5}$	$8 \cdot 10^{-5}$	$4 \cdot 10^{-4}$	0.001	0.0022	0.0093	0.0364	0.089	0.174
$p_{death}$	$2 \cdot 10^{-5}$	$2 \cdot 10^{-5}$	$10^{-4}$	$3 \cdot 10^{-4}$	0.0098	0.0026	0.008	0.024	0.082

Таблица 2: Вероятности перехода между состояниями заболевания в зависимости от возраста. Здесь  $p_{sym}$  – вероятность проявлять симптомы после заражения;  $p_{sev}$  – вероятность перехода больного с симптомами в тяжелое состояние (нуждается в госпитализации);  $p_{crit}$  – вероятность перехода больного из тяжелого состояния в критическое (нуждается в реанимации);  $p_{death}$  – вероятность смерти для больного, находящегося в реанимации.

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 Таблице 3).

Таким образом, восприимчивый к заражению агент ( $S$ ) при контакте с зараженными агентами, соединенными ребром графа, переходит в стадию зараженного незаразного ( $E$ ) с вероятностью  $\beta$  в момент времени  $t$ . Затем агент в состоянии зараженного незаразного ( $E$ ) может перейти в состояние инфицированного с симптомами ( $Sym$ ) с вероятностью  $p_{sym}$  через  $t_{sym}$  дней или остаться бессимптомным больным с вероятностью  $1 - p_{sym}$  через  $t_{inc}$  после заражения. Бессимптомные больные вылечиваются через  $t_{rec1}$  дней и переходят в группу ( $R$ ). Инфицированные с симптомами ( $Sym$ ) могут развить тяжелую степень заболевания и быть госпитализированными ( $H$ ) с вероятностью  $p_{sev}$  или остаться больным легкой степени тяжести ( $M$ ) с вероятностью  $1 - p_{sev}$  через  $t_{inf}$  дней после попадания в группу ( $Sym$ ). Больные легкой степени тяжести вылечиваются через  $t_{rec2}$  дней и переходят в группу ( $R$ ). Госпитализированные ( $H$ ) могут перейти развить критическое состояние ( $C$ ), т.е. нуждаются в подключении аппарата ИВЛ, с вероятностью  $p_{crit}$  через  $t_{hosp}$  дней после госпитализации или выздороветь с вероятностью  $1 - p_{crit}$  через  $t_{rec2}$  дней. Больные в критическом состоянии ( $C$ ) умирают с вероятностью  $p_{death}$  через  $t_{crit}$  дней или вылечиваются с вероятностью  $1 - p_{death}$  через  $t_{rec2}$  дней.



Параметр	Описание	Распределение
$t_{inc}$	Количество дней с момента контакта до того, как агент станет заразным.	LogN(4.6, 4.8) [24]
$t_{sym}$	Количество дней с момента, когда агент стал заражен, до проявления симптомов.	LogN(1, 0.9) [24]
$t_{rec1}$	Продолжительность болезни для бессимптомных и легких случаев.	LogN(8, 2) [25]
$t_{rec2}$	Продолжительность болезни для тяжелых и критических случаев.	LogN(14, 2.4) [26]
$t_{inf}$	Количество дней, за которое агент переходит из легкого состояния в тяжелое.	LogN(6.6, 4.9) [24]
$t_{hosp}$	Количество дней, за которое агент переходит из тяжелого состояния в критическое.	LogN(3, 7.4) [27]
$t_{crit}$	Длительность пребывания агента в критическом состоянии.	LogN(6.2, 1.7) [26]

Таблица 3: Распределения параметров продолжительности стадий заболевания болезни в днях.

### 3.4. Тестирование агентов

Тестирование агентов проводится согласно ежедневным статистическим данным о количестве проведенных тестов в Республике Казахстан (см. Раздел 2.2). Шанс быть протестированным на COVID-19 зависит от эпидемиологического статуса агента. На каждом шаге моделирования тесты распределяются среди популяции, и положительный результат могут получить агенты, статус которых обведен в рамку на рис. 3 (инфицированные бессимптомные и с симптомами, госпитализированные, больные легкой 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. Раздел 3.6).

### 3.5. Постановка прямой задачи для агентной модели

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  $\vec{X}(t) = (S(t), E(t), A(t), Sym(t), M(t), H(t), C(t), R(t), D(t))$  на следующий день, т.е.  $\vec{X}(t+1)$ . Поскольку многие параметры АМ распространения COVID-19 неизвестны

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

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

### 3.6. Постановка и решение обратной задачи

В данной работе на каждом временном этапе решается своя обратная задача. Обратная задача 1 состоит в восстановлении вектора параметров

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

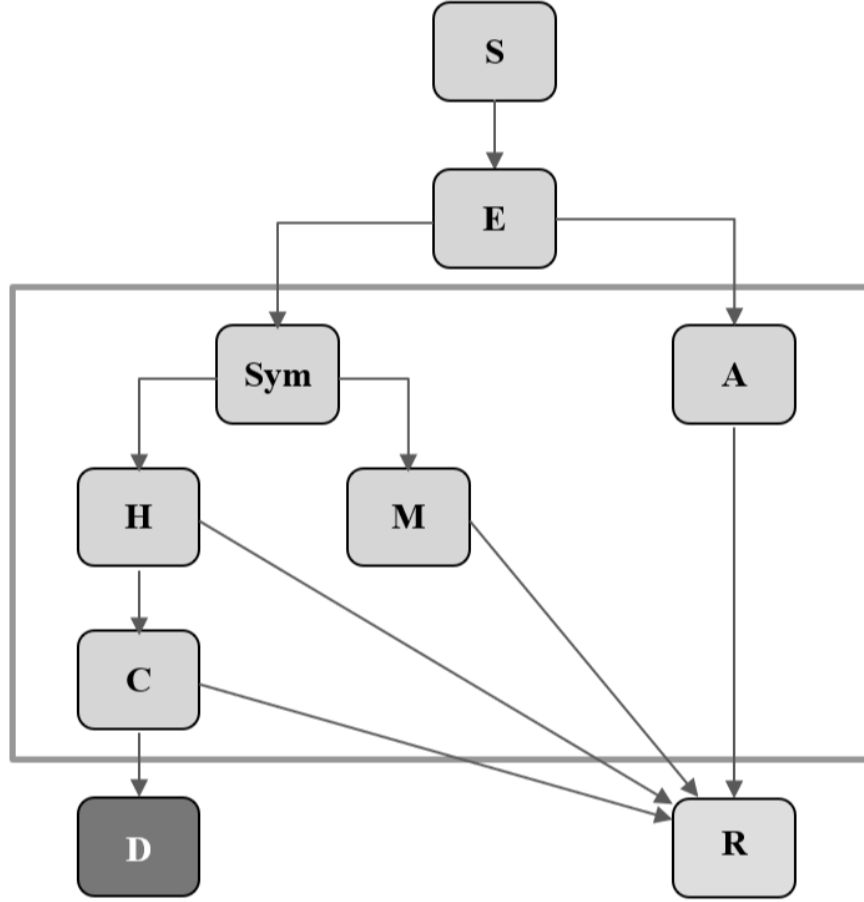


Рис. 3: Диаграмма состояний агентов в Covasim. В рамку обведены те состояния, находясь в которых агент имеет возможность получить положительный тест на COVID-19.

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

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

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

В работах [15, 16] для исследуемой модели авторы провели анализ чувствительности неизвестных параметров к измерениям на основе методов дифференциальной алгебры и байесовского подхода. Показано, что параметр  $\beta$ , отвечающий за передачу вируса, наиболее чувствителен к измерениям. С помощью методов анализа чувствительности область изменения параметра  $\beta$  удалось уменьшить в два раза благодаря добавлению дополнительной информации об эпидемии (а именно, к измерениям о количестве выявленных и умерших случаях была добавлена информация о критических случаях).

#### 4. АЛГОРИТМ РЕШЕНИЯ ОБРАТНОЙ ЗАДАЧИ

В ходе решения обратной задачи вектор неизвестных параметров  $\vec{q}$  восстанавливался с помощью применения пакета Optuna [28], в основе которого лежит метод древовидных оценок Парзена (Tree-structured Parzen estimators или коротко TPE), а также подхода усвоения данных для поэтапного восстановления параметров агентной модели (см. Раздел 4.2).

##### 4.1. Метод древовидных оценок Парзена

Идея метода состоит в следующем: вычисляются вероятности  $p(\vec{q}|J(\vec{q}))$  и  $p(J(\vec{q}))$  для определения области параметров для минимизации функционала  $J$ . Для этого пространство значений параметров  $\mathcal{D}_K = \{q_k, J(q_k) : k = 1, \dots, K\}$  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  $\gamma$  наименьших значений функционала в точках из  $\mathcal{D}_K(J_\gamma)$ , т.е.  $P(J < J_\gamma) = \gamma$ .  $\mathcal{D}_{K_g}^g$  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. Т.е. 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 такой вектор  $\vec{q}^*$ , который доставляет максимальное ожидаемое улучшение  $EI(x)$ , выражающееся по формуле

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

Сходимость по вероятности статистических методов отражена в общей теореме в работе [29].

Критерием остановки в данном алгоритме являлось ограничение на количество итераций  $max\_iter = 100$ . Схема алгоритма выглядит следующим образом:

Подробнее с методом древовидных оценок Парзена можно ознакомиться в работе [30].

##### 4.2. Поэтапное восстановление неизвестных параметров

В модели предполагается, что параметр  $\beta$  является кусочно-постоянным. Соответственно, чем продолжительнее рассматриваемый период моделирования, тем больше количество 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, и параметры, восстановленные на предыдущем шаге, использовались в последующем запуске алгоритма оптимизации. Таким образом, на первом промежутке 13.03.2020 – 12.04.2020 ( $i = 1$ ) использовался вектор неизвестных параметров (начальные условия  $t = 0$  соответствуют дате 13.03.2020)

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

**Алгоритм 1** древовидных оценок Парзена**Вход:** Значения параметров  $\gamma$ ,  $n_{samp}$  и  $max\_iter$ 

- 1: **Инициализация:** пространство значений неизвестных параметров  $\mathcal{D}_{init} = \{\vec{q}_k, J(\vec{q}_k), k = 1, \dots, n_{init}\}$
- 2: **Цикл от  $m=0$  до  $max\_iter$  выполнять**
- 3:   Разделить  $\mathcal{D}_{n_{init}+m}$  для генерации пространств  $\mathcal{D}_{m_g}^g, \mathcal{D}_{m_l}^l$
- 4:   Получить оценку плотности  $l(\vec{q})$  для наборов параметров из  $\mathcal{D}_{n_{init}+m_l}^l$
- 5:   Получить оценку плотности  $g(\vec{q})$  для наборов параметров из  $\mathcal{D}_{n_{init}+m_g}^g$
- 6:   Сгенерировать  $\vec{q}^s = \{\vec{q}_k^s : k = 1, \dots, n_{samp}\}$ , 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:    $\mathcal{D}_{n_{init}+m} \leftarrow \mathcal{D}_{n_{init}+m+1}$
- 10: **Конец цикла**

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 осуществляется с использованием ограничений на искомые параметры, полученные при анализе чувствительности на основе используемых измерений обратной задачи [15, 16].

## 5. ЧИСЛЕННЫЕ РЕЗУЛЬТАТЫ

В данном разделе представлен результат математического моделирования распространения COVID-19 в Республике Казахстан с восстановленным вектором неизвестных параметров  $\vec{q}$  (см. Разделы 3.6 и 4.2), а также сценарии распространения эпидемии при вариации нефармацевтических вмешательств. При построении сценариев распространения ежедневно выявленных случаев предполагается, что в регионе сохранится прежний уровень тестирования населения  $T(t)$ , полученный с помощью регрессионной модели SARIMA (см. Раздел 2).

На рис. 4 представлен результат моделирования среднего количества ежедневно выявленных случаев в результате ПЦР-тестирования в Республике Казахстан с прогнозом на 45 дней (показан период моделирования с 07.12.2021 по 20.01.2022, а прогноз построен с 13.12.2021 по 20.01.2022). При построении прогноза были учтены 2 вида сценариев:

- базовый сценарий (сплошная линия), в котором не учитывалось повышенное скопление людей во время новогодних праздников, учитываются только увеличение мобильности граждан в период подготовки к каникулам (с 20.12.2021 по 30.12.2021);
- повышенная мобильность граждан в период новогодних праздников (пунктирная линия) – учитывается повышенная передача вируса в общественных местах в период с 02.01.2022 по 06.01.2022.

Точками изображены реальные данные 13.03.2020 – 12.12.2021, которые участвуют в решении обратной задачи, а треугольниками – данные 13.12.2021 – 20.01.2022, которые использовались для проверки прогнозирования.

По результатам валидации модели можно сделать вывод, что учет повышенной мобильности граждан демонстрирует более точное соответствие реальным данным. Таким образом, по значениям параметра  $\beta$  во втором сценарии (пунктирная линия) можно сделать вывод, что в

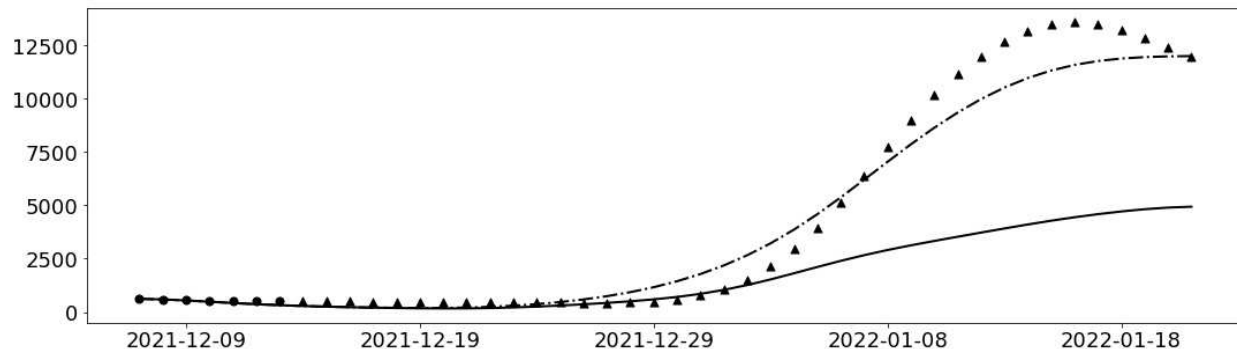


Рис. 4: Моделирование двух сценариев распространения ежедневно выявленных случаев в результате ПЦР-тестирования в Республике Казахстан (точками изображены реальные данные, участвовавшие в восстановлении неизвестных параметров 13.03.2020 – 12.12.2021, треугольники – данные для валидации модели 13.12.2021 – 20.01.2022): базовый сценарий (сплошная линия) и увеличение агентов в общественных местах в новогодние праздники (пунктирная линия).

период новогодних праздников передача вируса увеличилась в 3.5 раза по сравнению с началом декабря. Так, на 20 января 2022 года количество выявленных случаев COVID-19 составляло 12032 человека, количество ожидаемых выявленных случаев в Республике Казахстан по базовому сценарию АМ (сплошная линия), рассчитанному 12.12.2021, равнялось 4939 человек (погрешность составила 59%), а с учетом повышенной мобильности граждан в общественных местах (пунктирная линия) – 12007 человек (погрешность составила 0.2%).

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 зараженных в день  $t$ ,  $I_C(t)$  – текущее количество зараженных в день  $t$ ,  $d$  – средняя продолжительность заболевания в днях. Если  $\mathcal{R}(t) < 1$ , тогда считается, что эпидемия перестает распространяться, в противном случае она растет. На рис. 5 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 19.12.2021–18.01.2022 для сценария с увеличением мобильности граждан в общественных местах ввиду новогодних праздников (пунктирная с точками линия), после чего количество новых выявленных случаев сокращается до размеров базового сценария развития (сплошная линия).

## 6. ЗАКЛЮЧЕНИЕ

Разработан алгоритм численного моделирования и построения сценариев распространения эпидемии COVID-19 в Республике Казахстан на основе статистических данных об эпидемии на базе агентной модели. Выбор агентной модели обоснован возможностью учитывать индивидуальные взаимодействия агентов, возрастные исходы заболеваемости, политические и фармацевтические вмешательства. Модель основывается на программном пакете COVASIM,

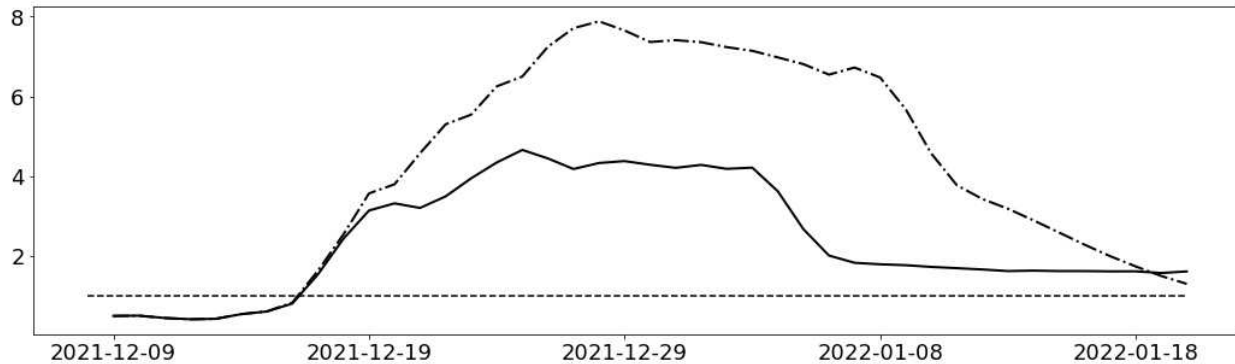


Рис. 5: Значение индекса репродукции  $\mathcal{R}(t)$  для базового сценария (сплошная линия) и для сценария с повышенной мобильностью граждан в общественных местах в новогодние праздники (пунктирная с точками линия). Пунктирной линией обозначено пороговое значение  $\mathcal{R}(t) = 1$ .

реализованным на языке программирования Python, и включает 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 (прогноз сезонных временных рядов ожидаемых ПЦР-тестов) неполных данных методами регрессионного анализа и машинного обучения. На втором этапе на основе статистических данных уточняются эпидемиологические параметры агентной модели (скорости передачи инфекции, тестирования, начальное количество инфицированных агентов). Для этого разработан регуляризованный алгоритм численного решения обратной задачи, 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. Регуляризация решения обратной задачи осуществляется с использованием ограничений на искомые параметры при каждом шаге итерации. На третьем этапе учитываются нефармацевтические вмешательства в процесс распространения эпидемии с целью построения наиболее реалистичных сценариев развития.

При реализации модели использовались данные о количестве выявленных случаев с 13.03.2020 по 12.12.2021 в Республике Казахстан. Показано, что увеличение концентрации агентов в новогодние праздники в общественных местах (магазины, театры, парки) увеличивает количество выявленных случаев COVID-19 (к 15.01.2022 увеличилось в 3.5 раза по сравнению с 01.12.2021). В качестве примера приведены два сценария распространения COVID-19, рассчитанных 12 декабря 2021 года на период до 20 января 2022 года. Сценарий, в котором учитывались новогодние праздники (опубликован 12 декабря 2021 года на сайте <https://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 эпидемиологической ситуации с действительностью.

Особенность моделирования распространения коронавирусной инфекции в Республике Казахстан заключается в необходимости учета концентрации населения в крупных городах: Алматы (1 993 067 человек), Нур-Султан (1 199 083 человека), Шымкент (1 090 160 человек) и другие, в которых проживает более 11 млн человек, а в селах – 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, 13, 14].



Вклад каждого автора в работу следующий:

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

## СПИСОК ЛИТЕРАТУРЫ

1. Адарченко В. А., Бабань С. А., Брагин А. А., Гребенкин К. Ф. и др. Моделирование развития эпидемии коронавируса по дифференциальной и статистической моделям. Препринт № 264. Снежинск: РФЯЦ-ВНИИТФ, 2020.
2. Криворотько О. И., Кабанихин С. И. Математические модели распространения COVID-19. Препринт № 300. Новосибирск: Институт математики им. С.Л. Соболева, 2022.
3. Kerr C. C., Stuart R. M., Mistry D., Abeysuriya R. G., Rosenfeld K., Hart G. R., et al. Covasim: An agent-based model of COVID-19 dynamics and interventions // PLOS Computational Biology. 2021. V. 17, № 7. P. e1009149.
4. COVID-19 Agent-based Simulator <https://github.com/InstituteForDiseaseModeling/covasim>
5. Aleta A., Martin-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, № 9. P. 964–971.
6. Lau M.S.Y., Grenfell B., Thomas M., Bryan M. et al. Characterizing superspreading events and age-specific infectiousness of SARS-CoV-2 transmission in Georgia, USA // Proc. Natl. Acad. Sci. USA. 2020. V. 117, № 36. P. 22430–22435.
7. Kucharski A. J., Klepac P., Conlan A. J. K., Kissler S. M. 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. V. 20, № 10. P. 1151–1160.
8. Hoertel N., Blachier M., Blanco C., Olsson M. et al. A stochastic agent-based model of the SARS-CoV-2 epidemic in France // Nat. Med. 2020. V. 26, № 9. P. 1417–1421.
9. Hellewell J., Abbott S., Gimma A., Bosse N. I. et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts // Lancet. Glob. Health. 2020. V. 8, № 4. P. e488–e496.
10. Nielsen B. F., Sneppen K. COVID-19 superspreading suggests mitigation by social network modulation // MedRxiv. 2020. V. 2020.09.15.20195008.
11. Noll N. B., Aksamentov I., Druelle V., Badenhorst A. et al. COVID-19 scenarios: an interactive tool to explore the spread and associated morbidity and mortality of SARS-CoV-2 // MedRxiv. 2020. DOI: 10.1101/2020.05.05.20091363.
12. Tuomisto J. T., Yrjölä J., Kolehmainen M., Bonsdorff J. et al. An agent-based epidemic model REINA for COVID-19 to identify destructive policies // MedRxiv. 2020. DOI: 10.1101/2020.04.09.20047498
13. Криворотько О.И., Кабанихин С.И., Зятков Н.Ю., Приходько А.Ю. и др. Математическое моделирование и прогнозирование COVID-19 в Москве и Новосибирской области // Сибирский журнал вычислительной математики. 2020. Т. 23, № 4. С. 395–414.
14. Krivorotko O., Sosnovskaia M., Vashchenko I., Kerr C., Lesnic D. Agent-based modeling of COVID-19 outbreaks for New York state and UK: parameter identification algorithm // Infectious Disease Modelling. 2022. V. 7. P. 30–44.

15. Криворотько О. И., Кабанихин С. И., Сосновская М. И., Андорная Д. В. Анализ чувствительности и идентифицируемости математических моделей распространения эпидемии COVID-19 // Вавиловский журнал генетики и селекции. 2021. Т. 25, № 1. С. 82–91.
16. Krivorotko O., Sosnovskaia M., Vashchenko I. Agent-based mathematical model of COVID-19 spread in Novosibirsk region: identifiability, optimization and forecasting // Journal of Inverse and Ill-Posed Problems. 2022. In print.
17. Интерполяционная функция <https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.interpolate.html>
18. Бокс Д., Дженкинс Г. Анализ временных рядов: прогноз и управление. Выпуск 1. М.: Мир, 1974.
19. Dabral P. P., Murry M. Z. Modelling and forecasting of rainfall time series using SARIMA // Environ. Process. 2017. V. 4. P. 399–419.
20. Box G. E. P., Cox D. R. An analysis of transformations // J. Roy. Statist. Soc. Ser. B. 1964. V. 26. P. 211–252.
21. Dickey D. A., Fuller W. A. Distribution of the estimators for autoregressive time series with a unit root // Journal of the American Statistical Association. 1979. V. 74. P. 427–431.
22. Гауссовский фильтр [https://docs.scipy.org/doc/scipy/reference/generated/scipy.ndimage.gaussian\\_filter.html](https://docs.scipy.org/doc/scipy/reference/generated/scipy.ndimage.gaussian_filter.html)
23. Household Size, 2019, UN. <https://population.un.org/Household/#/countries/840>
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, № 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. 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 // The Lancet Infectious Diseases. 2020. V. 20, № 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, № 11. P. 1061–1069.
28. OPTUNA: hyperparameter optimization framework <https://optuna.org/>
29. Жиглявский А.А., Жилинская А.Г. Методы поиска глобального экстремума. М.: Наука, 1991.
30. Bergstra J., Bardenet R., Kégl B., Bengio Y. Algorithms for Hyper-Parameter Optimization // Advances in Neural Information Processing Systems. 2011. V. 24. P. 2546–2554.