

УДК 519.876.5

## ИНДИВИДУУМ-ОРИЕНТИРОВАННОЕ МОДЕЛИРОВАНИЕ ЭПИДЕМИЧЕСКИХ ОРВИ В ГОРОДАХ РФ: МЕТОДЫ РЕАЛИЗАЦИИ И ОЦЕНКИ ПРИМЕНИМОСТИ

© 2025 А. И. Корзин<sup>1,a</sup>, Н. А. Чичкова<sup>1,b</sup>, Т. И. Капарулин<sup>1,c</sup>,  
В. Н. Леоненко<sup>1,2,d</sup>

<sup>1</sup> Университет ИТМО,

Кронверкский просп., 49, лит. А., г. Санкт-Петербург 197101, Россия,

<sup>2</sup> НИИ гриппа им. Смородинойцев Минздрава РФ,

ул. проф. Попова, 15/17, г. Санкт-Петербург 197022, Россия

E-mails: <sup>a</sup>corzin.an@gmail.com, <sup>b</sup>nachichkova@itmo.ru, <sup>c</sup>kaparulinti@mail.ru,  
<sup>d</sup>vnleonenco@itmo.ru

Поступила в редакцию 28.03.2024 г.; после доработки 30.10.2025 г.;  
принята к публикации 10.12.2025 г.

Индивидуум-ориентированные модели эпидемических вспышек завоевали широкую популярность среди исследователей общественного здравоохранения благодаря их способности описывать эпидемический процесс с высокой детализацией. Основным недостатком таких моделей является необходимость подготовки большого количества детализированных данных в качестве входа модели, а также выделения мощных вычислительных ресурсов для выполнения расчётов. Как следствие, из-за недостаточности данных и больших временных затрат при идентификации таких моделей их объяснительная и предсказательная сила может быть поставлена под вопрос. В настоящем исследовании предлагается метод разработки индивидуум-ориентированных моделей эпидемических вспышек ОРВИ, которые могут быть удовлетворительно откалиброваны на имеющиеся данные демографической и эпидемической статистики в городах РФ и позволяют получить результаты моделирования за разумное время за счёт применения процедуры сэмпинга. Прозрачность модельной структуры, применение модульного принципа на уровне алгоритма (разделение демографической и эпидемической составляющих), а также открытость кода программы позволяет обеспечить принцип воспроизводимости результатов моделирования, независимую проверку результатов и потенциальное их переиспользование со стороны исследователей коллективов, занимающихся смежными темами. Оценка применимости метода произведена на примере моделирования вспышек гриппа в Самаре и Челябинске.

**Ключевые слова:** синтетические популяции, индивидуум-ориентированные модели, эпидемиология, грипп, COVID-19.

DOI: 10.33048/SIBJIM.2025.28.408

### 1. ВВЕДЕНИЕ

#### 1.1. Постановка проблемы

Ежегодные всплески заболеваемости острыми респираторными инфекциями, такими как грипп и COVID-19, являются серьёзной социально-экономической проблемой, поскольку они приводят к повышению смертности населения и финансовым потерям из-за временной нетрудоспособности заболевших. Помимо этого, как показала пандемия коронавируса SARS-CoV-2, отдельные вспышки заболеваемости большой интенсивности могут приводить к перегруженности больниц и нарушению привычного рабочего графика и образа жизни людей. Эти последствия крайне негативны для экономики и здоровья населения. Для оценки эффективности

мер борьбы с эпидемическими вспышками органами здравоохранения используются математические и статистические модели, с помощью которых производится ретроспективный анализ и прогнозирование динамики острых респираторных вирусных инфекций.

Проблемами моделирования распространения эпидемических ОРВИ занимается большое количество научных коллективов по всему миру, с концентрацией большинства центров исследований в США и в Европе. Количество исследований по моделированию ОРВИ ощутимо растёт каждый раз при начале циркуляции в человеческих популяциях новых штаммов пандемического гриппа, например, A(H1N1)pdm09, а также других ОРВИ, таких как различные варианты SARS-CoV. Одним из старейших и наиболее распространённых типов математических моделей являются компартментальные модели, основанные на системах разностных и обыкновенных дифференциальных уравнений (SIR-модели и их модификации, логистическое уравнение). Основным плюсом компартментальных моделей является их простота и, как следствие, высокая скорость получения результатов моделирования и лёгкость их интерпретации. Между тем, ряд задач, связанных с детальным анализом эпидемических вспышек (на уровне контактов отдельных индивидов) не может быть решён с помощью компартментальных моделей. В силу этого, а также вследствие увеличения доступных компьютерных мощностей, с начала 2000-х гг. всё большую популярность приобретают стохастические мультиагентные (индивидуум-ориентированные) модели. Обзор по пространственным подходам к моделированию COVID-19 может быть найдён в [1], мультиагентные модели для широкого класса инфекций рассмотрены в [2].

Достоинством мультиагентных моделей является возможность получить динамическую картину пространственного распространения эпидемии, а также отслеживать динамику вспышки эпидемии с точностью до одного индивида, отвечая на вопрос «кто кого инфицировал и когда». Все это позволяет обнаружить неочевидные зависимости между характером течения эпидемии и демографическими параметрами. В частности, индивидуум-ориентированные модели дают возможность детального воспроизведения сети контактов населения города, которая оказывает значительное влияние на динамику заболеваемости. Структура контактов отдельного индивида зависит от количества людей, которых он встречает за день, и мест, которые он посещает. Построение точной сети контактов для каждого конкретного города позволяет количественно оценить связь между структурой мобильности людей и темпами роста числа случаев заболевания во время вспышки эпидемии.

Несмотря на то что в теории мультиагентные модели дают исследователям почти неограниченные возможности в плане детализации моделирования, на практике их применение сталкивается с множественными ограничениями, такими как недоступность данных для калибровки, повышенная неопределённость результатов моделирования из-за слишком большого количества параметров, значительные временные затраты для вычислительных экспериментов по сравнению с SIR-моделями (как следствие, затруднение в проведении анализа чувствительности и полноценной калибровки — в целом, решения любых задач, связанных с множественным повторным запуском моделей). В частности, большинство реальных практически значимых результатов, связанных с моделированием динамики COVID-19, были получены с помощью SIR-моделей [3], [4] или даже более простых моделей, таких как дифференциальные уравнения типа Ферхюльста — Пирла, несмотря на наличие у многих коллективов готовых решений на основе высокопроизводительных мультиагентных модельных комплексов, в частности, готовившихся для ответа на пандемии новых штаммов гриппа [5].

Пандемия COVID-19 и опыт использования мультиагентных моделей для прогнозирования динамики этого заболевания привела к широкой дискуссии в научной среде на тему, возможно ли в принципе опираться на результаты таких моделей для планирования мер сдерживания заболеваемости. В частности, исследователями в дискуссии [6] отмечалось, что детализированные модели, основанные на плохих данных, могут демонстрировать некорректные результаты, при этом вызывая доверие за счёт большого количества якобы учтённых

в них факторов. При этом некорректность результатов будет сложно обнаружить при отсутствии кода модели в открытом доступе или хотя бы детального описания алгоритма модели, что является частым случаем. Отечественными исследователями отмечается, что недостатком мультиагентных моделей «может стать избыточная подробность, которая усложняет общую задачу моделирования и приводит к созданию чрезмерно сложных моделей с избыточными параметрами, которые не вносят существенного вклада в результаты моделирования», а также тот факт, что «в большинстве случаев нет возможности критически осмыслить модель ввиду наличия общего, часто поверхностного описания её устройства, приводимого в публикации, и отсутствия доступа к исходному коду модели» [2]. Таким образом, несмотря на наличие большого количества мультиагентных моделей, описанных в научных публикациях, лишь немногие из них могут быть использованы для решения задач реального эпидемического надзора.

## 1.2. Существующие решения

Анализ публикаций, посвящённых проблеме, позволяет сформулировать следующие требования к мультиагентным моделям, результаты работы которых призваны использоваться в принятии решений органами здравоохранения [7]:

— Допущения, сделанные в модели, а также лежащие в её основе механизмы, должны быть детально изложены.

— Должна быть обеспечена максимальная прозрачность модели путём предоставления более подробной информации о базовых механизмах модели и сделанных предположениях, оптимально — благодаря использованию стандартного протокола (например, ODD), а также путём предоставления кода моделирующей программы для загрузки.

— Модель должна быть валидирована, при этом сама процедура валидации должна быть подробно описана. Даже при отсутствии необходимых данных валидация может быть проведена — в виде систематического тестирования модели, сравнения её с другими моделями или экспертной оценки.

— Отдельное внимание должно быть уделено структуре имитируемой популяции. В случае если характеристики или поведение искусственной популяции не отражают характеристики или поведение реальной популяции адекватным образом, это приведёт к ограниченной применимости результатов моделирования к изучению реальной популяции.

В табл. 1 приведён ряд комплексов для мультиагентного моделирования эпидемических ОРВИ, получивших наиболее широкое распространение и доступных для использования.

Основным препятствием к использованию указанных моделей для городов РФ является необходимость обеспечения для них входных статистических данных заданных форматов. В исследовании по динамике гриппа в Нью-Йорке авторы создают синтетическую популяцию в масштабе города, объединяя данные микропереписи (PUMS Microsurvey) и отраслевые реестры — это исследование одним из первых опиралось на синтетические популяции стандарта RTI. Данные популяции созданы для всей территории США, ряда стран Европы и Юго-Восточной Азии и предоставлены желающим в открытом доступе. В частности, они являются стандартным видом входных данных для модельного комплекса FRED. Опыт работы авторов с FRED показал его применимость к задачам моделирования динамики гриппа в Санкт-Петербурге [15] — при условии подготовки популяций в формате RTI. В то же время критической проблемой оказалось закрытие доступа к новым версиям комплекса и документации к ней, что не позволило продолжить применение модели для моделирования динамики ОРВИ в городах РФ. OpenABM создаёт свои синтетические домохозяйства и возрастные структуры на основе таблиц панельных исследований, которые в случае РФ могут быть конвертированы из ранее собранных популяций стандарта RTI или сгенерированы заново из данных Росстата. MATSIM+EPISIM опираются на чрезвычайно подробные модели поведения населения города, которые требуют большого количества данных, по большей части недоступных для РФ,

Таблица 1

## Комплексы мультиагентного моделирования ОРВИ

Модель	Описание	Население	Социальная структура
FRED (Framework for Reconstructing Epidemic Dynamics, 2013) [8]	Открытая агентная модель для США на основе синтетической популяции. Создана на базе моделей гриппа MIDAS (Университет Питтсбурга и Университет Карнеги – Меллона).	Синтетические популяции для всех округов США (~300 млн агентов), пространственные регионы (сетка 1 км <sup>2</sup> ).	Домохозяйства, школы (в т. ч. классы), рабочие места (офисы), жилые районы; шаг 1 день; SEIR с поддержкой нескольких штаммов ОРВИ.
Модель динамики гриппа в синтетических популяциях [9]	Количественная оценка роли поездок в метро Нью-Йорка в распространении гриппа; сравнение мер, ориентированных на транспорт, с менее конкретными мерами контроля.	7,85 млн синтетических агентов (пять районов города).	Домохозяйства, школы, рабочие места, локальные сообщества, стагус пользования общественным транспортом (да/нет); калибровка по гриппу 1957–1958 гг.
Covasim (2021) [10]	Высокопроизводительная агентная модель для сравнения стратегий сдерживания COVID-19	Обычно 100 тыс.–1 млн агентов; демография стран (данные ООН) или синтетические сети SynthPops.	Многослойные сети (домохозяйства, школы, рабочие места, локальные сообщества, госпитали); возрастная структура; инфекционность на основе вирусной нагрузки.
OpenABM–Covid19 (2021)[11]	Британская агентная модель для оценки мер сдерживания и отслеживания контактов (contact tracing).	По умолчанию 1 млн городских агентов; метапопуляции масштабируются до десятков миллионов индивидов.	Три сети: домохозяйства («малый мир»), профессия (работа/школа), случайные ежедневные контакты; возраст по десятилетиям.
MATSIM+EPISIM [12]	Интегрирует поведение в транспорте (MATSIM) и распространение эпидемии (EPISIM [13]) для количественной оценки риска COVID-19 по видам активности применительно к Монреалю.	~1 млн агентов (25% сэмплирование); пространственно явные ежедневные поездки и активности.	Контакты реконструируются по совместному расположению в общественных местах и транспорте; вероятность заражения пропорциональна длительности контакта и трансмиссивности вируса; учёт сезонности.
REINA (2020) [14]	Финская национальная АБМ для сравнения контрольных мер по сдерживанию COVID-19.	1,6 млн агентов (регион Хельсинки) с возрастной структурой и учётом максимальной нагрузки больниц.	Случайные контакты, зависящие от возраста; ведётся история контактов для отслеживания.

и значительного времени на их предобработку. REINA использует агрегированные демографические данные (возрастное распределение) без синтеза домохозяйств, что делает эту модель наиболее приспособленной для быстрого развёртывания, но ограничивает детальность генерируемых ей выходных данных. Вторым после REINA по скорости применения для данных РФ можно считать Covasim, который допускает генерацию синтетических сети контактов на основе уже существующих общедоступных источников (данные ООН, SynthPops) и позволяет с минимальными модификациями производить моделирование в предположении регулярных структур контактов населения (сети Барабаши — Альберта или Уотса — Строгаца). Однако эксперименты с этим модельным комплексом применительно к популяции Санкт-Петербурга показали, что использование стандартных топологий контактов ухудшает результаты моделирования по сравнению с синтетической популяцией. В то же время использование собственных синтетических популяций в Covasim вместо сгенерированных контактных сетей на основе вышеупомянутых стандартных моделей значительно замедляет время выполнения экспериментов и уменьшает вероятность получения качественной калибровки на данные заболеваемости [16].

### 1.3. Задачи исследования

Целью настоящей работы является разработка и демонстрация метода индивидуум-ориентированного моделирования эпидемий ОРВИ, направленного на практическое использование в эпидемическом надзоре РФ и в максимальной степени опирающегося на вышеуказанные принципы. В задачи работы входит:

- 1) Разработка метода генерации данных для модели для произвольного города РФ,
- 2) Разработка алгоритмов численного моделирования на основе методов Монте-Карло,
- 3) Разработка алгоритма калибровки модели и оценка его эффективности,
- 4) Иллюстрация разработанных методов на примере моделирования сезонного гриппа в Челябинске и Самаре.

Решаемые задачи являются частью более общей проблемы в области создания и использования синтетических популяций и мультиагентных моделей, которая заключается в поиске оптимальных методов параметризации моделей и баланса между детализацией и интерпретируемостью моделей для достижения максимальной точности и достоверности результатов [2]. Прозрачность модельной структуры, применение модульного принципа на уровне алгоритма (разделение демографической и эпидемической составляющих), использование открытых демографических данных, а также доступность кода программы для скачивания (репозиторий с кодом доступен по приведённой ссылке: [17]) позволяет обеспечить принцип воспроизводимости результатов моделирования, независимую проверку результатов и потенциальное их переиспользование со стороны исследовательских коллективов, занимающихся смежными темами.

Для обеспечения соответствия озвученным принципам предложенный модельный подход реализован согласно следующим требованиям:

— Максимальное использование готовых решений, как в части моделей и программ, так и в части стандартов данных. При генерации синтетических популяций авторы опираются на стандарт компании RTI [18], который даёт возможность использовать любые открытые и проприетарные модели, совместимые с этим стандартом.

— Ускоренная подготовка данных на основе стандартизированной информации, имеющейся для городов РФ, в противовес «штучному» сбору данных для каждого конкретного модельного случая, который практиковался в более ранних работах [19]. Для этого внедрено максимальное использование открытых данных (таких, как данные карт OSM, переписи населения, статистика Росстата, сервис Яндекс.Аудитории) и верифицированных алгоритмов для их обработки, включая алгоритмы распределения индивидов по домохозяйствам [20].

— Ускорение работы моделирующей программы. Одна из возможностей – использование моделей с разумно достаточным уровнем детализации. Для достижения этой цели в используемых моделях, в частности, отсутствует реализация динамических контактов (нет расчёта вероятности встреч индивидов в реальном времени согласно графикам перемещения), не различаются выходные и будние дни при оценке вероятностей заражения в различных местах, ограничено число мест заражения.

## 2. МЕТОДЫ

### 2.1. Генерация данных для моделирования

В качестве входных демографических данных для мультиагентной модели используется синтетическая популяция — это искусственно созданная популяция, которая по своим агрегированным характеристикам статистически приближена к реальной. Генерация синтетических популяций может быть выполнена в автоматическом и ручном режиме. В то время как ручная генерация данных обладает большей точностью, автоматический сбор информации ускоряет процесс подготовки данных к моделированию. Популяция составляется из записей индивидов с данными о возрасте, поле, месте жительства и месте работы/обучения. Для подачи на вход моделирующей программе необходимо сгенерировать эти данные и скомпоновать в виде нескольких входных файлов. В настоящей работе процесс подготовки синтетической популяции Санкт-Петербурга 2010 года, продемонстрированный ранее в [19], был изменён и обобщён для применения к произвольному городу. По умолчанию генерация входных демографических данных для модели производится в автоматическом режиме с использованием открытой платформы «Цифровая урбанистика» [21] — в этом случае алгоритм моделирования популяции города сокращается до получения данных о домах, объектах инфраструктуры из платформы и моделировании рабочих мест. Такой подход, однако, доступен лишь при наличии уже загруженных данных о городе на платформе. Платформа позволяет получать данные о половозрастном составе жителей домов, геолоцированные данные с объектами инфраструктуры с восстановленными значениями вместимости школ, а также геолоцированными данными о жилых домах городов, представленных на платформе. Данные о половозрастном составе жителей домов моделируются с помощью библиотеки с открытым исходным кодом «Population-restorator» [22], используя метод расселения жителей по жилым домам в городах, имея данные об общем количестве жителей в городе и половозрастное распределение людей в регионе [20]. Также на платформе возможно моделирование численности и половозрастного состава населения с использованием данных о миграции и исторических данных о численности населения [23]. Модель реализована в виде библиотеки с открытым исходным кодом ITMO-2 [24]. В случае недоступности данных на платформе возможна их ручная генерация с последующим сохранением для дальнейшего переиспользования в рамках «Цифровой урбанистики» (в том числе и для задач, не связанных с анализом распространения эпидемических вспышек). Описание шагов алгоритма ручной генерации приведено ниже.

#### **Получение набора данных со всеми зданиями в городе.**

Для создания базы данных жилых домов в городе был использован сервис Open Street Map (OSM), который предоставляет географические данные в открытом доступе. Для получения набора данных из OSM используется сервис Overpass-Turbo, выбранный в силу своей популярности. Данный инструмент позволяет выгружать наборы данных в формате .geojson, что удобно для дальнейшей работы с пространственными данными. Процесс получения данных из OSM происходит с помощью комбинации ключа и тега для идентификации объектов.

#### **Проверка атрибута назначения здания.**

В полученный набор зданий добавляется атрибут `is_living`, характеризующий тип здания (жилое или нежилое). Для его создания и корректировки используется программное обеспечение QGIS [25], предназначенное для работы с геопространственными данными. Атрибут

принимает значение `true`, если тип здания в поле `building` указан как один из следующих: общежитие, многоквартирный дом, частный дом. Для каждого города необходима ручная проверка и выставление значений данного атрибута.

#### **Проверка атрибута этажности для каждого жилого дома.**

Для корректного расселения людей по жилым домам необходима проверка атрибута `floors`, отвечающего за количество этажей в доме. Это значение используется в вычислении жилой площади здания. Для его проверки используется программное обеспечение QGIS и сравнение набора данных с данными сервисов 2GIS [26] и «Яндекс.Карты» [27], а также ручное внесение атрибута для каждого здания. Такая проверка требует больших временных затрат. Для города с населением в миллион человек проверка занимает до 80 человеко-часов.

#### **Добавление пропущенных зданий.**

Недостающие жилые здания добавляются при помощи ручной проверки с наложением карт географических сервисов поверх карты с полученными ранее домами. Для новых домов вручную заполняются атрибуты `is_living` и `floors`. Время ручной обработки данных зависит от качества исходных данных сервиса OSM, а также от общего количества жилых домов. Подготовка дополненного набора данных с недостающими зданиями для города населением миллион человек занимает до 120 человеко-часов.

#### **Расселение людей по жилым зданиям.**

Половозрастная структура населения формируется согласно данным Росстата. При расселении используется подход, описанный в работе [20]. Алгоритм распределяет заданное количество жителей в городе по зданиям пропорционально величине жилой площади. Для каждого жилого здания определяется координата центроида для дальнейшего использования в качестве координаты всех домохозяйств популяции, находящихся в этом доме.

#### **Получение данных о школах.**

Для получения данных о школах используется сервис OSM с запросом по соответствующим ключам и тегам. Проблема со сбором информации о школах заключается в том, что под понятие «школа» попадают сразу несколько типов заведений: языковая школа, автошкола и другие подобные учреждения. В силу этого для повышения адекватности данных предпочтительна их дополнительная ручная корректировка.

#### **Данные о рабочих местах.**

Данные о рабочих местах для синтетической популяции были получены при помощи открытого сервиса «Яндекс.Аудитории», позволяющего получать количество рабочих мест на заданном ограниченном участке площади в городе. Для удобства сбора данных территория города была разделена на ячейки в виде шестиугольников. Для каждой из ячеек были получены значения количества рабочих мест. В большинстве случаев при выгрузке данных из данного сервиса оценка количества рабочих мест превышает официальное количество трудозанятых, поэтому после получения данных из сервиса количество рабочих мест нормируется, исходя из количества трудозанятых в городе согласно данным Росстата. Нормировочный коэффициент рассчитывается по формуле

$$k_{\text{norm}} = \frac{N_{\text{wp}}^{(\text{of})}}{N_{\text{wp}}},$$

где  $N_{\text{wp}}^{(\text{of})}$  — количество трудозанятых согласно данным Росстата,  $N_{\text{wp}}$  — суммарное количество рабочих мест по городу, полученное при помощи сервиса «Яндекс.Аудитории». Тогда итоговый размер рабочего места  $n_{\text{wp}_i}$  вычисляется по формуле

$$n_{\text{wp}_i} = k_{\text{norm}} \cdot n_{\text{wp}_i}^{(0)},$$

где  $n_{\text{wp}_i}^{(0)}$  — размер  $i$ -го рабочего места согласно [28].

### Разделение домов на квартиры.

Население каждого дома предполагается разделённым на квартиры. Количество жильцов в каждой квартире генерируется случайным образом согласно распределению Пуассона (см. использование такого подхода в [29]) с функцией вероятности  $p(k) = \frac{\mu^k}{k!} e^{-\mu}$ , где  $k$  — итоговое количество людей в квартире,  $\mu$  — параметр распределения. Значение параметра  $\mu = 2.57$  выбиралось согласно [19] и соответствует предполагаемому среднему размеру домохозяйства. Суммарное число людей во всех квартирах совпадает с предполагаемым числом жителей дома.

### Присваивание рабочих мест и школ.

Следующим этапом добавляются рабочие места. Предполагается, что все индивиды в возрасте от 7 до 17 лет ежедневно посещают школу, а взрослые трудоспособного возраста (с 18 до 60 для мужчин и с 18 до 55 для женщин) ежедневно посещают рабочие места. Выбор школы и места работы зависит от расстояния до места жительства человека. В данной работе максимальное расстояние до школы и работы было задано 5 км и 15 км соответственно. Если свободного места в школе (на работе) в заданном радиусе не находилось, то присваивалось ближайшее доступное место, невзирая на расстояние. Результатом работы данного этапа алгоритма является список индивидов, где каждой записи приписан пол, возраст, идентификатор домохозяйства, идентификатор школы и координата рабочего места без привязки к конкретному офису.

### Разделение рабочих мест на офисы.

Рабочие места, координаты которых соответствуют центрам шестиугольников, полученных из сервиса «Яндекс.Аудитории», разделяются на более мелкие части — офисы. Размер рабочего места, т. е. максимальное количество работников, генерируется случайным образом согласно распределению Пуассона с параметром  $\mu = 8.5$ .

## 2.2. Сэмплинг данных синтетической популяции

Мультиагентное моделирование трудозатратно и занимает много времени по сравнению с компарментальными моделями. Возможным решением проблемы является использование вместо полноразмерной синтетической популяции её масштабной копии. Такой подход в научной литературе называется сэмплингом (масштабированием) или разукрупнением (downscaling). К примеру, в работе [30] авторы запускают мультиагентную модель распространения COVID-19 для популяции в округе Форд, штат Канзас, США, производят таким образом оценку параметров, и в дальнейшем эти параметры используются для проведения всех экспериментов для уменьшенной версии Нью-Йорка, США. Моделирование города Нью-Йорка на период 120 дней выполнено при предположении численности населения в 10 000 человек; при этом входные параметры, специфичные для Нью-Йорка, приспособлены для меньшей численности населения, чтобы обеспечить соответствующее масштабирование.

Хотя даунскейлинг (разукрупнение) стал обычной практикой, лишь в ограниченном числе исследований изучалось его влияние на результаты моделирования [31]. Например, в работе [32], посвящённой транспортному моделированию, указывается: «В примере для агломерации Мюнхена 5% агентов и 50 итераций воспроизвели вероятностное распределение путевого времени аналогично полученному при 100% агентов и 500 итераций (но в 50 раз быстрее). Соответственно, в данном случае масштабные коэффициенты около 5% кажутся разумными. Однако требования конкретных сценариев мер контроля могут повлиять на наиболее подходящий масштабный коэффициент и необходимую плотность транспортной сети в подвыборке для каждого приложения. При тестировании сценариев, где результаты анализируются только на высокоагрегированном уровне, меньшие коэффициенты масштабирования могут быть приемлемыми. Однако анализ транспортных потоков по одному конкретному транспортному коридору вероятно, потребует 100% выборки». В случае подготовки сэмплированной популяции

для моделирования распространения заболеваемости аналогично важно сохранять структуру исходных связей агентов, которая в нашем случае соответствует сети контактов индивидов.

### 2.3. Мультиагентная модель

Для моделирования динамики эпидемии в настоящей работе была использована актуальная версия мультиагентной модели, впервые представленной в [19]:

#### Описание.

Модель предназначена для моделирования динамики изменения состояний индивидов синтетической популяции в течение заданного периода времени. Время модели дискретное с шагом в один день. Генерация переходов между состояниями производится имитационным моделированием с помощью методов Монте-Карло. Входными данными являются параметры заболевания и синтетическая популяция, задаваемая согласно стандарту RTI [18]. Выходными данными является количество индивидов синтетической популяции в каждом из состояний, относящихся к заболеванию, в момент времени  $t \in \overline{0, T}$ , а также списки индексов индивидов, заражённых определённым штаммом вируса в момент времени  $t$ . Поскольку каждому индивиду сопоставлена координата домохозяйства, на основе этих данных возможно построение карт распространения заболеваемости в городе и их отображение в геоинформационных системах.

#### Население.

Население представляется синтетической популяцией, состоящей из нескольких текстовых файлов. Файл представляет собой список, каждая строка которого представляет собой отдельную запись, соответствующую определённой сущности – человеку, домохозяйству, месту работы или школе:

– households.txt. Местоположение и описательные характеристики каждого домохозяйства. Записи о домохозяйствах в файле «households.txt» ссылаются на записи об отдельных лицах в таблице «people.txt».

– people.txt. Записи о каждом человеке с указанием его возраста и пола, принадлежности к школе и рабочему месту. Эти синтетические записи о человеке связаны с файлом «households.txt» через поле `sp_hh_id`, содержащее индекс домохозяйства.

– schools.txt. Записи по каждой школе, включая её почтовый индекс, максимальную вместимость и координаты.

– workplaces.txt. Записи по каждому рабочему месту с указанием его координат и размера.

Характерные размеры популяций составляют от нескольких сот тысяч до нескольких миллионов индивидов.

#### Инициализация модели.

Перед началом моделирования в популяции случайным образом выбираются восприимчивые индивиды в количестве

$$S(0) = [\alpha \cdot N(0)],$$

где  $\alpha \in [0, 1]$  — доля восприимчивых в популяции на момент начала моделирования,  $N(0)$  — общее количество людей в популяции в начальный момент времени. Поскольку при моделировании краткосрочных эпидемий демографическими и миграционными процессами можно пренебречь, полагаем  $N(t) \equiv N(0)$ . Затем среди общего списка восприимчивых индивидов выбирается фиксированное начальное количество инфицированных  $I_0$ , т. е. происходит случайный выбор их порядковых номеров.

#### Модель контактов.

Каждый агент в модели может взаимодействовать с заранее определённым количеством людей с заданной вероятностью контакта. Согласно модели, индивиды могут взаимодействовать и передавать инфекцию внутри домохозяйств, на рабочих местах или в школах. Для

каждого индивида в синтетической популяции моделью добавляется три индикатора, относящихся к его инфекционному статусу:

- *isImmune* — индикатор, показывающий, имеет ли индивид иммунитет к данной инфекции (с учётом штамма в мультиштабном варианте моделирования);
- *isSick* — индикатор, показывающий, болеет ли человек в текущий день моделирования (с учётом штамма);
- *infectionDay* — число, равное количеству дней, прошедших с начала заболевания.

Для каждого индивида возможно нахождение в одном из трёх состояний: восприимчивый ( $S$ ), инфицированный ( $I$ ), выздоровевший ( $R$ ), что соответствует нотации классической компартментальной SIR-модели. Для каждого дня моделирования определяются места встреч восприимчивых и инфицированных индивидов. Вероятность заражения человека вычисляется по формуле

$$p_{\text{inf}}^{(\text{pl})} = 1 - \prod_{i=1}^{N^{(\text{pl})}-1} (1 - \lambda \cdot p_{\text{cnt}}^{(\text{pl})} \cdot g(\tau_i)),$$

где  $\lambda$  — трансмиссивность,  $p_{\text{cnt}}^{(\text{pl})}$  — вероятность события, что человек встретит случайно выбранного человека в общем месте их пребывания,  $g(\tau_i)$  — функция доли инфекционных в зависимости от времени  $\tau_i$ , прошедшего с начала инфицирования (*infectionDay*) для конкретного индивида (отвечает за изменение заразности индивида с течением времени после момента инфицирования). Верхний индекс (pl) указывает на рассмотрение процесса заражения в конкретном месте (школа, дом или рабочее место). Сумма вероятности заражения индивида в заданном месте pl и вероятности того, что он не будет заражён, составляет единицу, что отражено в формуле. Внутри оператора произведения каждый множитель отражает вероятность отсутствия заражения рассматриваемого индивида в результате эффективного контакта с любым из индивидов  $i \in 1, N^{(\text{pl})} - 1$ , находящихся в данном месте. Вероятность  $p_{\text{cnt}}^{(\text{pl})}$  вычисляется по следующей формуле:

$$p_{\text{cnt}}^{(\text{pl})} = \min \left\{ \frac{c^{(\text{pl})}}{N^{(\text{pl})}(t) - 1}, 1 \right\},$$

где  $c^{(\text{pl})}$  — среднее число контактов в месте пребывания индивида (отдельно для дома, школы и офиса),  $N^{(\text{pl})}(t)$  — общее количество людей в таком месте.

Количество инфицированных в месте pl людей определяется согласно биномиальному распределению:

$$n_{\text{inf}}^{(\text{pl})} \sim \text{Bin}(S(t), p_{\text{inf}}^{(\text{pl})}),$$

Процесс заражения инициируется последовательно в домах, на рабочих местах и в школах. После окончания расчёта всех переходов индивидов между состояниями номер текущего дня увеличивается на единицу и процессы заражения повторяются снова.

#### Модель протекания заболевания.

Для модельного комплекса реализовано две модели протекания заболевания:

- Детерминированная модель, связанная с функцией  $g(\tau_i)$  доли инфекционных в зависимости от времени  $\tau_i$ , прошедшего с начала инфицирования (*infectionDay*). С помощью этой функции задаётся длина периода инфекционности, длина периода, когда индивид не заразен, и момент наступления выздоровления (в двух последних случаях  $g(\tau_i) = 0$ ). Функция считается одинаковой для всех индивидов.

- Стохастическая модель, связанная с генерацией длины инкубационного  $\tau_{\text{inc}}$  и инфекционного периода  $\tau_{\text{inf}}$  индивидуально для каждого заражённого. Данная модель, в отличие от предыдущей, не позволяет использовать в явном виде функции инфекционности, зависящие от времени. Её преимуществом является включение стохастического эффекта протекания заболевания и возможность учёта влияния индивидуальных характеристик (например, пола

и возраста) на длину инкубационного периода и периода инфекционности. Пример калибровки мультиагентной модели со стохастической моделью протекания заболевания на данные заболеваемости COVID-19 приведён в [16].

В экспериментах данной работы использовалась детерминированная модель.

### Контрольные меры.

Карантинные меры в модели имитируются уменьшением среднего числа контактов  $c^{(pl)}$  в разных категориях мест. Использование индивидуальных средств защиты моделируется путём изменения значения параметра трансмиссивности вирусов  $\lambda$ . Вакцинация имитируется присвоением статуса *isImmune* перед началом моделирования индивидам синтетической популяции. Количество защищённых индивидов фиксированной возрастной группы  $a$  вычисляется по следующей формуле:

$$N_{\text{immune}}^{(a)} = VE(a) \cdot VR(a) \cdot N_{\text{total}}^{(a)},$$

где  $N_{\text{total}}^{(a)}$  — общая численность возрастной группы,  $VR(a)$  — охват вакцинацией возрастной группы,  $VE(a)$  — вакцинная эффективность для данной группы.

Пример применения модели для оценки эффективности вакцинации с разным охватом групп риска приведён в [33].

## 3. ЧИСЛЕННЫЕ ЭКСПЕРИМЕНТЫ

Программа, реализующая мультиагентную модель, написана на языке Python. Моделирование выполнялось с использованием параллельных вычислений на основе библиотеки `multiprocessing`. Результатами моделирования является количество восприимчивых, заболевших и выздоровевших индивидов в каждый из дней моделирования, а также списки идентификаторов заражённых индивидов и мест заражения, что позволяет восстанавливать пространственную картину распространения инфекции. Значения параметров модели указаны в табл. 2.

Таблица 2

Параметры модели

Параметр модели	Описание	Значение
$\alpha$	Доля восприимчивого населения	0.78
$\lambda$	Трансмиссивность	0.15
$I_0$	Начальное количество инфицированных	10
$T_{\text{mod}}$	Количество дней моделирования	120
$N_{\text{runs}}$	Количество запусков	10

### 3.1. Сравнение динамики эпидемического процесса в разных городах РФ

В качестве исследуемых городов были рассмотрены два близких по населению города: Самара и Челябинск. При помощи вышеописанного алгоритма генерации синтетических популяций были подготовлены входные данные для моделирования. Результаты работы алгоритма по оценке количества жилых зданий на каждом из этапов обработки данных приведены в табл.2. Разница в итоговом количестве заселённых зданий предположительно связана с различием в процентном соотношении многоэтажных жилых зданий и частного сектора в рассматриваемых городах, что влияет на результаты работы алгоритма расселения [20]. В табл. 3 приведены основные количественные характеристики синтетических популяций для Самары и Челябинска.

Таблица 3

Оценки количества жилых зданий в процессе генерации

Город	Самара	Челябинск
OSM, общее количество зданий	61305	73367
OSM, кол-во жилых зданий	8880	29972
Фонд развития территорий, общее кол-во	11613	6775
После корректировки, общее кол-во	61305	74176
После корректировки, кол-во жилых зданий	40552	40421
Количество заселённых зданий	32932	23478

Таблица 4

Основные характеристики синтетической популяции Самары и Челябинска

Город	Население	Жилые дома	Школы	Уровень занятости
Самара	1158000	32931	219	61.9%
Челябинск	1189000	23478	113	64.07%

В качестве исходной информации для анализа эпидпроцессов в исследуемых городах рассматривались данные о заболеваемости эпидемических сезонов с 2010 по 2021 годы, предоставленные ФГБУ «НИИ гриппа им. А. А. Смородинцева» Минздрава России. На рис. 1 приведены графики количества новых случаев инфицирования ОРВИ и гриппом в Челябинске и Самаре за вычетом сезонных уровней ОРВИ.

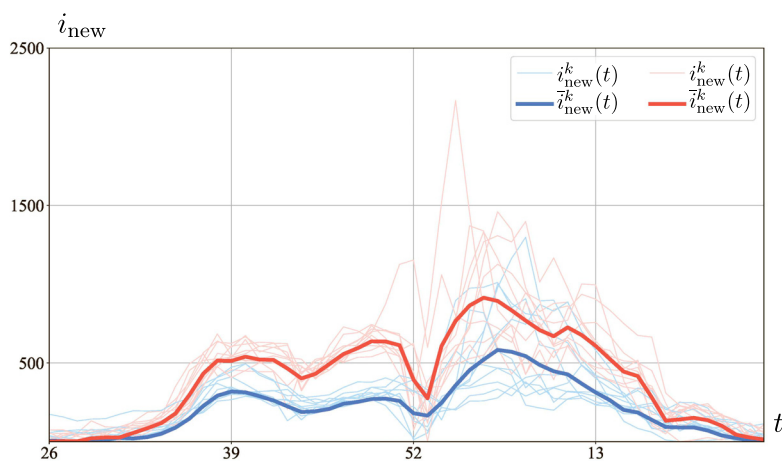


Рис. 1. Количество случаев инфицирования ОРВИ и гриппом для Самары (синий цвет) и Челябинска (красный цвет) с 2010 по 2021 годы;  $i_{\text{new}}^k(t)$  — количество новых случаев инфицирования за вычетом сезонных уровней ОРВИ в эпидемический сезон  $k$  на неделю  $t$ ,  $\bar{i}_{\text{new}}^k(t)$  — среднее значение  $i_{\text{new}}^k(t)$  по всем приведенным эпидемическим сезонам

Корректировка по сезонным уровням вводилась с целью точнее оценить эпидемическую заболеваемость гриппом без использования данных лабораторной диагностики (см. [34]). Как видно из графиков, для Челябинска наблюдается превышение числа заболевших в большинстве эпидемических сезонов по сравнению с аналогичными показателями для Самары. Это различие может быть вызвано рядом отдельных факторов или их сочетанием, такими как

особенности структуры сети контактов, климатические факторы, особенности работы городских систем учёта заболеваемости.

Результатом каждого запуска модели является динамика заболеваемости гриппом в течение 150 дней с изначальным количеством инфицированных в 10 человек. Значение параметра трансмиссивности  $\lambda$  равно 0.15. График относительного количества заболеваемости приведён на рис. 2.

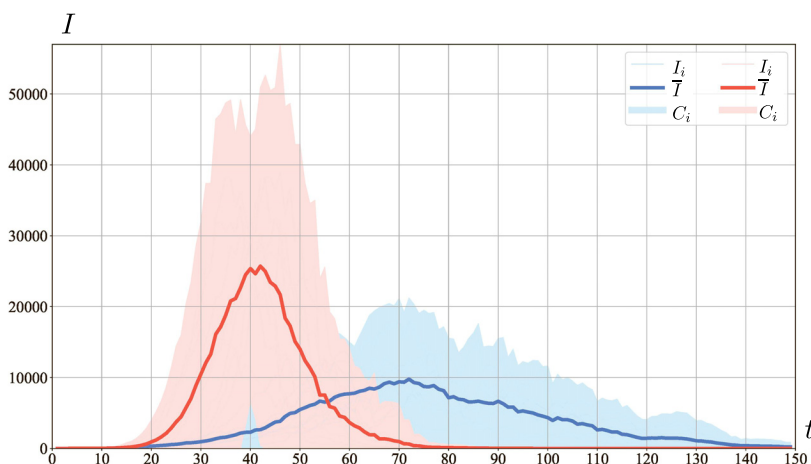


Рис. 2. Количество инфицированных гриппом индивидов для Самары (синий цвет) и Челябинска (красный цвет) согласно результатам моделирования; на горизонтальной оси  $t$  отложены дни; доверительные интервалы  $C_i$  рассчитаны по правилу  $3\sigma$

Для каждого из дней моделирования вычислено усреднённое значение заболеваемости  $\bar{I}(t)$ :

$$\bar{I}(t) = \frac{1}{N_l} \sum_{i=1}^{N_l} I_i(t),$$

где  $I_i(t)$  — значение количества инфицированных на 100 тысяч человек в день моделирования  $t$  для  $i$ -го номера запуска моделирования,  $N_l$  — количество запусков моделирования.

Как видно из рис. 2, формы кривых заболеваемости гриппом для Самары и Челябинска существенно отличаются друг от друга. В сравнении с кривой заболеваемости для Самары, кривая для Челябинска демонстрирует более резкий рост и большую высоту пика, что свидетельствует о более интенсивном протекании эпидемии. Эпидемические параметры модели для двух городов были заданы одинаковыми, в силу этого различие в кривых заболеваемости, очевидно, связано с различиями в структуре сетей контактов, определяемой синтетическими популяциями.

### 3.2. Моделирование на сэмплированных данных

В данной серии численных экспериментов производилась оценка качества результатов работы модели на сэмплированных данных. Параметром алгоритма сэмплирования является процент  $r$  сохранённых домохозяйств по отношению к исходной популяции. Алгоритмом производится случайный выбор  $r$  (%) квартир в каждом из районов и формируется популяция на основе людей, проживающих в этих квартирах. Поскольку распределение населения по рабочим местам и школам остаётся аналогичным изначальной популяции, такой подход может привести к понижению плотности людей в этих местах и как следствие уменьшению числа контактов. При малых значениях  $r$  возможно полное отсутствие людей в некоторых офисах и

школах и увеличение кластеризации сети контактов. Поэтому важным является оценка доли населения, которая позволяет проводить моделирование с достаточной достоверностью.

Для мегаполисов с населением свыше миллиона человек расчётный эксперимент с использованием мультиагентной модели занимает до нескольких часов, поэтому в качестве входных данных логично использовать сэмплированные популяции вместо полноразмерных. На рис. 3 приведены результаты моделирования на синтетических популяциях для Самары и Челябинска.

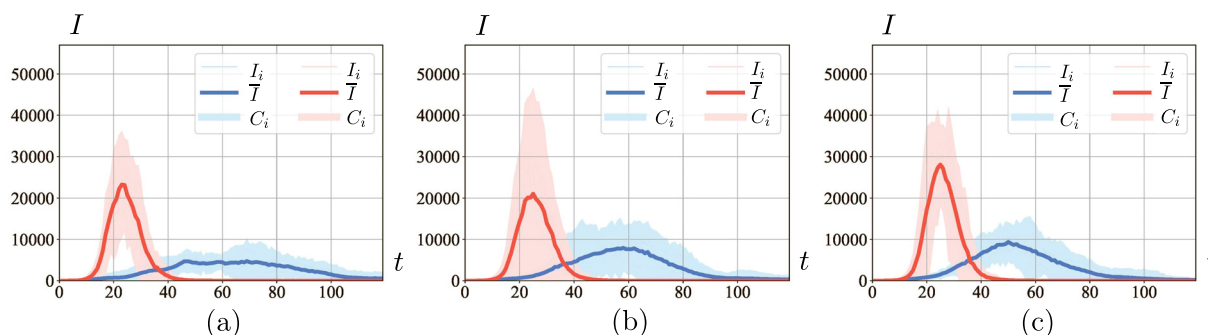


Рис. 3. Количество инфицированных индивидов в сэмплированных популяциях Самары (синий цвет) и Челябинска (красный цвет): (a)  $r = 25\%$ , (b)  $50\%$ , (c)  $75\%$ ; на горизонтальной оси  $t$  отложены дни; доверительные интервалы  $C_i$  рассчитаны по правилу  $3\sigma$

При помощи алгоритма сэмплинга были составлены синтетические популяции с параметром  $r = 25; 50; 75\%$ . На каждой выборке была запущена программа моделирования с десятью параллельно вычисляемыми реализациями. Время расчётов, а также пиковые значения в относительных величинах для различных экспериментов приведены в табл. 5 и на рис. 4.

Таблица 5

Сравнение времени моделирования и средних пиковых значений в относительных величинах.

$r, \%$	Самара		Челябинск	
	$\tau(r)/\tau(100)$	$\bar{I}_{\max}(r)/\bar{I}_{\max}(100)$	$\tau(r)/\tau(100)$	$\bar{I}_{\max}(r)/\bar{I}_{\max}(100) \cdot 0.2$
25	0.24	0.67	0.2	0.86
50	0.42	0.98	0.32	0.9
75	0.89	1.1	0.41	1.02
100	1 (17 889 с.)	1 (9 627 чел.)	1 (27 937 с.)	1 (29 014 чел.)

Полное время моделирование для Самары и Челябинска ( $\tau(r)$  при  $r = 100\%$ ) составило 17 889 и 27 937 секунд соответственно. Значение  $\bar{I}_{\max}$  было усреднено по пиковым значениям отдельных запусков и для полных популяций Самары и Челябинска составило 9 627 и 29 014 человек.

По приведённым результатам видно, что использование сэмплированных популяций в мультиагентной модели позволило сократить время моделирования. Для  $r = 25\%$  моделирование каждого из городов не превосходит двух часов. Кривые заболеваемости для сэмплированных популяций, приведённые на рис. 4, сохраняют своё расположение друг относительно друга, а пиковые значения и вид кривой заболеваемости варьируются в зависимости от параметра  $r$  по-разному для каждого города.

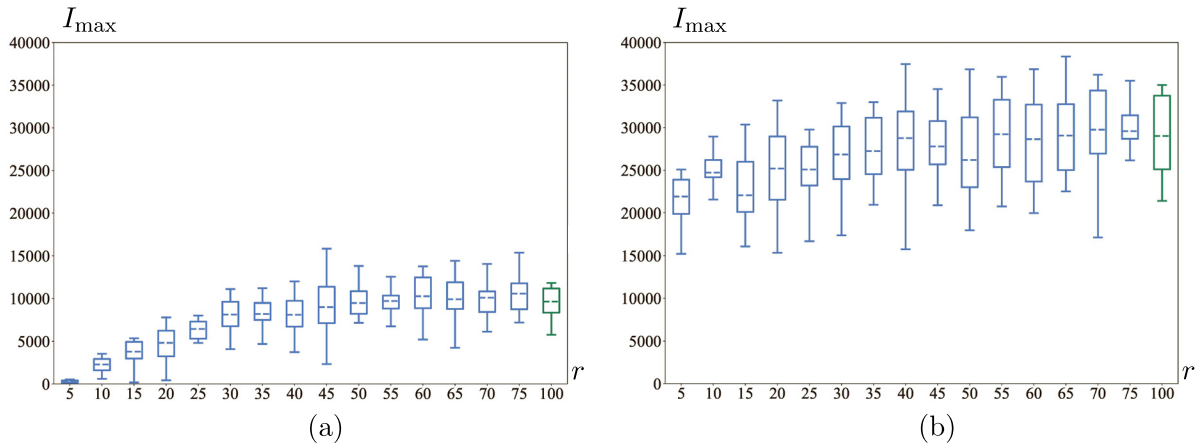


Рис. 4. Распределение значений пиковой заболеваемости для различного процента  $r$  сэмплинга населения: Самара (а) и Челябинск (б)

### 3.3. Калибровка моделей на данные

Подбор модельных кривых, соответствующих данным, выполняется полным перебором значений параметров доли изначально восприимчивых  $\alpha$  и трансмиссивности инфекции  $\lambda$  с фиксированным шагом (так называемый grid search или parameter sweep). Наиболее близкие к данным заболеваемости модельные кривые определяются подсчётом значения  $R^2$ .

Пример поиска модельных кривых, соответствующих данным заболеваемости гриппом в Самаре, приведён на рис. 5.

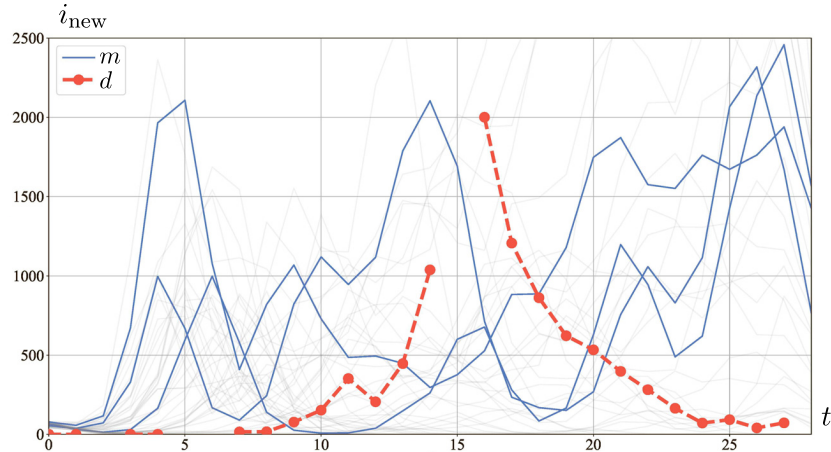


Рис. 5. Сопоставление модельных кривых с данными заболеваемости гриппом А(Н3N1) в Самаре в 2023–2024 гг.;  $m$  — модельные кривые,  $d$  — данные заболеваемости

Использовались следующие значения параметров:  $\alpha_{\min} = 0.8$ ,  $\alpha_{\max} = 0.92$ ,  $\alpha_{\text{step}} = 0.02$ ,  $\lambda_{\min} = 0.05$ ,  $\lambda_{\max} = 0.15$ ,  $\lambda_{\text{step}} = 0.01$ . Пример использования авторами данного алгоритма также приведён в работе [16] для прогнозирования COVID-19 в синтетической популяции Санкт-Петербурга.

## ЗАКЛЮЧЕНИЕ

В данной работе авторами приведён набор методов, позволяющий оперативно подготовить данные и произвести мультиагентное моделирование процессов распространения заболеваемости в городах РФ. Целью настоящей работы была демонстрация принципиальной

возможности применения детализированного моделирования динамики инфекций в практических целях. В то время как в большинстве публикуемых исследований упор делается на описание реализации и применения конкретной мультиагентной модели, авторам представляется более существенным акцентировать внимание на решение проблемы оперативного сбора и обработки детальных демографических данных программным путём в автоматическом и полуавтоматическом режиме, а также обеспечения удовлетворительной скорости получения результатов работы модели при множественных запусках, без которых реалистичное прогнозирование детальными моделями невозможно. Представленные в настоящем исследовании результаты, в частности, методы генерации синтетических популяций на основе данных открытой инструментальной платформы [21] и методы масштабирования популяций, значительно ускоряют получение результатов моделирования и дают возможность использовать мультиагентное моделирование в составе реальных систем поддержки принятия решений. Основным предполагаемым примером применения является поддержка принятия решений в экстремальных ситуациях (таких, как пандемии вирусов нового типа), в которых ключевым фактором является учёт стохастичности процесса и индивидуальное поведение отдельных агентов системы (например, суперраспространителей).

Рассмотренный метод разработки и реализации моделей имеет следующие ограничения:

— Несовершенство автоматического алгоритма генерации синтетических популяций на основе платформы [21]. Одной из особенностей представляемого подхода является максимальное переиспользование существующих данных, и этот момент имеет недостатки, связанные с неполным соответствием свойств имеющихся данных задачам моделирования эпидемических вспышек. В частности, в случае городов с платформы [21] уровень детализации популяций соответствует решаемым урбанистическим задачам, но недостаточен для формирования достоверных сетей контактов населения, например, в части разделения домов на квартиры, которое в ходе данного исследования пришлось реализовывать дополнительно. Данная проблема решается адаптацией платформы под задачи эпидемиологии либо ручной доработкой популяций — в зависимости от наличия финансовых и временных ресурсов.

— Данные о рабочих местах слишком общие (нет деления по отраслям), требуют дополнительных вычислений (не являются частью открытой инструментальной платформы) и ограничены списком городов, для которых есть информация «Яндекс.Аудитории».

— Отсутствие в модели реализации заражения в транспорте, которое может является весомым фактором передачи респираторных инфекций. Хотя согласно ряду оценок (например, в [9]) контакты в транспорте имеют незначительное влияние на процесс распространения гриппа, в случае коронавирусной инфекции есть результаты, подтверждающие значимость роли транспорта [35]. Данное упрощение модели вызвано историческими причинами (использование модельных гипотез из комплекса FRED, где передача инфекции в транспорте также не реализована), а также отсутствием доступных стандартизованных наборов данных по использованию общественного транспорта в городах РФ, на основе которых можно было бы оценить соответствующие параметры модели. При необходимости предполагается быстрая реализация модели заражения в транспорте с помощью добавления дополнительного источника инфицирования индивидов по принципу спонтанной активации инфекции (переход восприимчивого индивида в инфицированные с вероятностью, зависящей от общего числа инфицированных в популяции) или путём образования дальних связей между парами индивидов (генерация определённого количества случайных индивидов в качестве «соседей» данного выбранного индивида в качестве нового типа общественного места «транспорт»), которая позволит учесть этот фактор без значительного усложнения модели.

— Неточности в популяциях, возникающие после процедуры сэмплирования и более подробно описанные ранее. Для уменьшения неточностей планируется введение оптимизационного алгоритма с оценкой оптимального процента сэмплирования, аналогично ранее прово-

дившемуся подобному анализу для гибридной модели [36]. Сам гибридный метод моделирования также предполагается принять для практического использования, что даёт два варианта ускорения работы модели: через использование сэмпированных популяций (уровень данных) и применение гибридного подхода (уровень модельного комплекса).

— На данный момент нет адекватной оценки количества запусков модели, требуемых для получения репрезентативных траекторий распространённости. Планируется проведение такой оценки в случае для разной структуры контактов (т. е. для разных городов) на основе известных методов [37].

— Ограниченная валидация результатов моделирования, что является общей проблемой подхода [38].

— Неясно, даёт ли использование столь детальных контактных сетей преимущество в объясняющей и предсказательной силе моделей по сравнению с менее детальными моделями. Погрешности в данных и большая неопределённость, вызванная отсутствием валидации корректности расселения людей и назначения рабочих мест, может сводить на нет потенциальные эффекты от детализации.

В качестве дальнейшего направления работ предполагается изучить использование упрощённых сетей контактов на основе стандартных топологий вместо индивидуальной генерации связей между индивидами в результате их распределения по местам контактов. Такое упрощение модели, по оценкам авторов, позволит значительно ускорить работу моделирующей программы и тем самым ещё более упростить все задачи, связанные с многократным запуском модели. Между тем, вопрос, будут ли упрощённые сети контактов адекватно воспроизводить динамику контактов реального населения и передачу инфекции в результате этих контактов, требует дополнительного исследования.

## ФИНАНСИРОВАНИЕ РАБОТЫ

Исследование выполнено при финансовой поддержке Российского научного фонда (проект № 22-71-10067). Других источников финансирования проведения или руководства данным конкретным исследованием не было.

## КОНФЛИКТ ИНТЕРЕСОВ

Авторы данной работы заявляют, что у них нет конфликта интересов.

## ЛИТЕРАТУРА

1. Wang P., Zheng X., Liu H. Simulation and forecasting models of COVID-19 taking into account spatio-temporal dynamic characteristics // Review. *Frontiers in Public Health*. 2022, V. 10. Article number 1033432.
2. Маслова И.И., Манолов А.И., Глущенко О.Е., Козлов И.Е., Цуркис В.И., Попов Н.С., Самойлов А.Е., Лукашев А.Н., Ильина Е.Н. Ограничения в создании искусственных популяций в агентном моделировании эпидемий: систематический обзор // *Журн. микробиологии, эпидемиологии и иммунобиологии*. 2024. Т. 101, No 4. С. 4.
3. Friedman J., Liu P., Troeger C.E. et al. Predictive performance of international COVID-19 mortality forecasting models // *Nature Communications*. 2021. V. 12, N 1. P. 2609.
4. Ioannidis J. P., Cripps S., Tanner M. Forecasting for COVID-19 has failed // *Intern. J. Forecasting*. 2022. V. 38, N 2. P. 423–438.
5. Stroud P., Del Valle S., Sydoriak S. et al. Spatial dynamics of pandemic influenza in a massive artificial society // *J. Artificial Societies and Social Simulation*. 2007. V. 10, N 4. P. 9.
6. JASSS-Covid19-Thread // Review of Artificial Societies and Social Simulation; <https://rofasss.org/tag/JASSS-Covid19-Thread/>

7. Lorig F., Johansson E., Davidsson P. Agent-based social simulation of the covid-19 pandemic: A systematic review // J. Artif. Soc. Soc. Simul. 2021. 24(3); DOI: 10.18564/jasss.4601
8. Grefenstette J.J., Brown S.T., Rosenfeld R. et al. FRED (A Framework for Reconstructing Epidemic Dynamics): An open-source software system for modeling infectious diseases and control strategies using census-based populations // BMC Public Health. 2013. V. 13, N 1. P. 940.
9. Cooley P., Brown S., Cajka J. et al. The role of subway travel in an influenza epidemic: a New York City simulation // J. Urban Health. 2011. V. 88. P. 982–995.
10. Kerr C.C., Stuart R.M., Mistry D., Abeysuriya R.G., Rosenfeld K., Hart G.R., Nunez R.C., Cohen J.A., Selvaraj P., Hagedorn B., George L. Covasim: an agent-based model of COVID-19 dynamics and interventions // PLOS Computational Biology. 2021. V. 17, N 7. Article number e1009149.
11. Hinch R., Probert W.J., Nurtay A., Kendall M., Wymant C., Hall M., Lythgoe K., Bulas Cruz A., Zhao L., Stewart A., Ferretti L. OpenABM-Covid19—An agent-based model for non-pharmaceutical interventions against COVID-19 including contact tracing // PLoS Computational Biology. 2021. V. 17, N 7. Article number e1009146.
12. Manout O., Ciari F. Assessing the role of daily activities and mobility in the spread of COVID-19 in Montreal with an agent-based approach // Frontiers in Built Environment. 2021. V. 7. Article number 654279.
13. Mniszewski S.M., Del Valle S.Y., Stroud P.D., Riese J.M., Sydorik S.J. EpiSimS simulation of a multi-component strategy for pandemic influenza // Proceed. of the 2008 Spring Simulation Multiconf. 2008, April. 2008. P. 556–563.
14. Tuomisto J.T., Yrjola J., Kolehmainen M., Bonsdorff J., Pekkanen J., Tikkanen T. An agent-based epidemic model REINA for COVID-19 to identify destructive policies // MedRxiv. P. 2020-04; DOI:10.1101/2020.04.09.20047498
15. Leonenko V., Lobachev A., Bobashev G. Spatial modeling of influenza outbreaks in Saint Petersburg using synthetic populations // InInternat. Conf. Comput. Sci., 2019 Jun 8. Cham: Springer International Publishing, 2019. P. 492–505.
16. Korzin A.I., Kaparulin T.I., Leonenko V.N. Assessing the Applicability of the Multiagent Modeling Approach to the Epidemic Surveillance of COVID-19 in Russian Cities // IEEE Internat. Multi-Conf. Engrg., Computer and Information Sci. (SIBIRCON), 2024, September. 2004. P. 237–242. IEEE.
17. URL: [https://github.com/vnleonenko/Multiagent\\_ARI](https://github.com/vnleonenko/Multiagent_ARI)
18. Wheaton W.D., Cajka J.C., Chasteen B.M. et al. Synthesized population databases: A US geospatial database for agent-based models // Methods report (RTI Press). 2009. V. 2009, N 10. P. 905.
19. Leonenko V., Arzamastsev S., Bobashev G. Contact patterns and influenza outbreaks in Russian cities: A proof-of-concept study via agent-based modeling // J. Comput. Sci. 2020. V. 44. Article number 101156.
20. Kontsevich G., Sokol A., Bogomolov Y. et al. Modeling the citizens' settlement in residential buildings // Procedia Comput. Sci. 2022. V. 212. P. 51–63.
21. Платформа «Цифровая урбанистика»; URL: <https://dc.idu.actcognitive.org/>
22. Библиотека «population-restorator»; URL: <https://github.com/kanootoko/population-restorator/>
23. Zakharov K., Aghajanyan A., Kovantsev A., Boukhanovsky A. Forecasting Population Migration in Small Settlements Using Generative Models under Conditions of Data Scarcity // Smart Cities. 2024. V. 7. P. 2495–2513; <https://doi.org/10.3390/smartcities7050097>
24. Библиотека «ITMO-2» // URL: <https://github.com/aghajanyan/ITMO-2/>
25. Геоинформационная система QGIS // URL: <https://qgis.org/ru/site/>
26. Электронный справочник карт городов 2GIS // URL: <https://2gis.ru>
27. Яндекс.Карты // <https://yandex.ru/maps>
28. Яндекс.Аудитории // URL: <https://audience.yandex.ru/>
29. Jennings V., Lloyd-Smith B., Ironmonger D. Household size and the Poisson distribution // J. Australian Population Association. 1999. V. 16. N 1. P. 65–84.

30. *Shamil M.S., Farheen F., Ibtehad N. et al.* An agent-based modeling of COVID-19: validation, analysis, and recommendations // *Cognitive Computation*. 2021. P. 1–12.
31. *Kagho G.O., Meli J., Walser D., Balac M.* Effects of population sampling on agent-based transport simulation of on-demand services // *Procedia Comput. Sci.* 2022. V. 201. P. 305–312.
32. *Llorca C., Moeckel R.* Effects of scaling down the population for agent-based traffic simulations // *Procedia Comput. Sci.* 2019. N 151. P. 782–787.
33. *Korzin A.I., Kaparulin T.I., Leonenko V.N.* Assessing the Effect of Influenza Vaccination Strategies Using Multi-agent Modeling // In *2024 IEEE 3rd International Conference on Problems of Informatics, Electronics and Radio Engineering (PIERE)*, 2024, November. IEEE. 2024. P. 1000–1003.
34. *Leonenko V.N., Ivanov S.V.* Fitting the SEIR model of seasonal influenza outbreak to the incidence data for Russian cities // *Russian J. Numerical Analysis and Math. Modelling*. 2016. V. 31. N 5. P. 267–279.
35. *Harris J.E.* Critical role of the subways in the initial spread of SARS-CoV-2 in New York City // *Frontiers in Public Health*. 2021. V. 9. Article number 754767.
36. *Leonenko V.* A Hybrid Modeling Framework for City-Scale Dynamics of Multi-strain Influenza Epidemics // *Proc. Internat. Conf. Comput. Sci.* 2022. V. 13352. P. 164–177; DOI: 10.1007/978-3-031-08757-8\_16
37. *Lee J.S., Filatova T., Ligmann-Zielinska A., Hassani-Mahmooei B., Stonedahl F., Lorscheid I., Voinov A., Polhill J.G., Sun Z., Parker D.C.* The complexities of agent-based modeling output analysis // *J. Artif. Soc. Soc. Simul.* 2015. 18(4); DOI:10.18564/jasss.2897
38. *Castro B.M., Reis M.D.M., Salles R.M.* Multi-agent simulation model updating and forecasting for the evaluation of COVID-19 transmission // *Scientific Reports*. 2022. V. 12, N 1. Article number 22091.

UDC 519.876.5

**INDIVIDUAL-ORIENTED MODELING OF EPIDEMIC ARI IN CITIES  
OF THE RUSSIAN FEDERATION: METHODS OF IMPLEMENTATION  
AND APPLICABILITY ASSESSMENT**

© 2025 A. I. Korzin<sup>1a</sup>, N. A. Chichkova<sup>1b</sup>, T. I. Kaparulin<sup>1c</sup>,  
V. N. Leonenko<sup>1,2d</sup>

<sup>1</sup>*ITMO University,  
Kronverkskii pr., 49, lit. A., Sankt-Peterburg 197101, Russia,*  
<sup>2</sup>*Smorodintsev Research Institute Institute of Influenza  
Ministry of Health of the Russian Federation,  
Prof. Popov st., 15/17, Sankt-Peterburg 197022, Russia*

E-mails: <sup>a</sup>corzin.an@gmail.com, <sup>b</sup>nachichkova@itmo.ru, <sup>c</sup>kaparulinti@mail.ru,  
<sup>d</sup>vnleonenko@itmo.ru

Received 28.03.2024, revised 30.10.2025, accepted 10.12.2025

**Abstract.** Individual-based models of epidemic outbreaks have gained wide popularity among public health researchers due to their ability to describe the epidemic process with high detail. The main disadvantage of such models is the need to prepare a large amount of detailed data as an input to the model, as well as to allocate powerful computing resources to perform calculations. As a consequence, due to the lack of data and difficulties in performing the identification procedure of such models, their explanatory and predictive power can be called into question. In this study, a method is proposed for developing individual-based models of ARVI epidemic outbreaks that can be satisfactorily calibrated on the available demographic and epidemic statistics data in the cities of the Russian Federation in a reasonable time, taking into account the actual availability of data and the degree of detail of the infection transmission patterns in a heterogeneous population at the level of individuals. By using the sampling procedure, an increase in the speed of the modeling program is achieved. The transparency of the model structure, the use of the modular principle at the algorithm level (separation of the demographic and epidemic components), and the openness of the program code allow us to ensure the principle of reproducibility of the modeling results, independent verification of the results, and their potential reuse by research teams working on related topics. The applicability of the methods is assessed using the example of modeling influenza outbreaks in Samara and Chelyabinsk.

**Keywords:** synthetic populations, individual-based models, epidemiology, influenza, COVID-19.

DOI: 10.33048/SIBJIM.2025.28.408

## REFERENCES

1. Wang P., Zheng X., Liu H. Simulation and forecasting models of COVID-19 taking into account spatio-temporal dynamic characteristics. *Review. Frontiers in Public Health*, 2022, Vol. 10, Article number 1033432.
2. Maslova I.I., Manolov A.I., Glushchenko O.E., Kozlov I.E., Tsurkis V.I., Popov N.S., Samoilov A.E., Lukashov A.N., Il'ina E.N. Ogranicheniya v sozdanii iskusstvennykh populyatsii v agentnom modelirovanii epidemii: sistemicheskii obzor [Limitations in the creation of artificial populations in agent-based epidemic modeling: a systematic review]. *Zhurn. mikrobiologii, epidemiologii i immunobiologii* [J. Microbiology, Epidemiology and Immunobiology], 2024, Vol. 101, No. 4, pp. 4.

3. Friedman J., Liu P., Troeger C.E. et al. Predictive performance of international COVID-19 mortality forecasting models. *Nature Communications*, 2021, Vol. 12, No. 1, pp. 2609.
4. Ioannidis J. P., Cripps S., Tanner M. Forecasting for COVID-19 has failed. *Intern. J. Forecasting*, 2022, Vol. 38, No. 2, pp. 423–438.
5. Stroud P., Del Valle S., Sydoriak S. et al. Spatial dynamics of pandemic influenza in a massive artificial society. *J. Artificial Societies and Social Simulation*, 2007, Vol. 10, No. 4, pp. 9.
6. JASSS-Covid19-Thread // Review of Artificial Societies and Social Simulation; <https://rofasss.org/tag/JASSS-Covid19-Thread/>
7. Lorig F., Johansson E., Davidsson P. Agent-based social simulation of the covid-19 pandemic: A systematic review // *J. Artif. Soc. Soc. Simul.*, 2021, Vol. 24, No. 3, Article number 5; DOI: 10.18564/jasss.4601
8. Grefenstette J.J., Brown S.T., Rosenfeld R. et al. FRED (A Framework for Reconstructing Epidemic Dynamics): An open-source software system for modeling infectious diseases and control strategies using census-based populations. *BMC Public Health*, 2013, Vol. 13, No. 1, pp. 940.
9. Cooley P., Brown S., Cajka J. et al. The role of subway travel in an influenza epidemic: a New York City simulation. *J. Urban Health*, 2011, Vol. 88, pp. 982–995.
10. Kerr C.C., Stuart R.M., Mistry D., Abeyseriya R.G., Rosenfeld K., Hart G.R., Nunez R.C., Cohen J.A., Selvaraj P., Hagedorn B., George L. Covasim: an agent-based model of COVID-19 dynamics and interventions. *PLoS Computational Biology*, 2021, Vol. 17, No. 7, Article number e1009149.
11. Hinch R., Probert W.J., Nurtay A., Kendall M., Wymant C., Hall M., Lythgoe K., Bulas Cruz A., Zhao L., Stewart A., Ferretti L. OpenABM-Covid19—An agent-based model for non-pharmaceutical interventions against COVID-19 including contact tracing. *PLoS Computational Biology*, 2021, Vol. 17, No. 7, Article number e1009146.
12. Manout O., Ciari F. Assessing the role of daily activities and mobility in the spread of COVID-19 in Montreal with an agent-based approach. *Frontiers in Built Environment* 2021, Vol. 7, Article number 654279.
13. Mniszewski S.M., Del Valle S.Y., Stroud P.D., Riese J.M., Sydoriak S.J. EpiSimS simulation of a multi-component strategy for pandemic influenza. *Proceed. of the 2008 Spring Simulation Multiconf.*, 2008, April, pp. 556–563.
14. Tuomisto J.T., Yrjola J., Kolehmainen M., Bonsdorff J., Pekkanen J., Tikkanen T. An agent-based epidemic model REINA for COVID-19 to identify destructive policies. *MedRxiv.*, pp.2020-04; DOI:10.1101/2020.04.09.20047498
15. Leonenko V, Lobachev A, Bobashev G. Spatial modeling of influenza outbreaks in Saint Petersburg using synthetic populations. *In Internat. Conf. Comput. Sci.*, 2019 Jun 8, Cham: Springer International Publishing, 2019, pp. 492–505.
16. Korzin A.I., Kaparulin T.I., Leonenko V.N. Assessing the Applicability of the Multiagent Modeling Approach to the Epidemic Surveillance of COVID-19 in Russian Cities. *IEEE Internat. Multi-Conf. Engrg., Computer and Information Sci. (SIBIRCON)*, 2024, September, pp. 237–242.
17. URL: [https://github.com/vnleonenko/Multiagent\\_ARI](https://github.com/vnleonenko/Multiagent_ARI)
18. Wheaton W.D., Cajka J.C., Chasteen B.M. et al. Synthesized population databases: A US geospatial database for agent-based models. *Methods report*, 2009, Vol. 2009, No. 10, pp. 905.
19. Leonenko V., Arzamastsev S., Bobashev G. Contact patterns and influenza outbreaks in Russian cities: A proof-of-concept study via agent-based modeling. *J. Comput. Sci.*, 2020, Vol. 44. Article number 101156.
20. Kontsevich G., Sokol A., Bogomolov Y. et al. Modeling the citizens' settlement in residential buildings. *Procedia Comput. Sci.* 2022, Vol. 212, pp. 51–63.
21. Platforma «Tsifrovaya urbanistika» [The «Digital Urbanism» platform]; URL: <https://dc.idu.actcognitive.org/>
22. Biblioteka «population-restorator» [Library «population-restorator»]; URL: <https://github.com/kanootoko/population-restorator/>

23. Zakharov K., Aghajanyan A., Kovantsev A., Boukhanovsky A. Forecasting Population Migration in Small Settlements Using Generative Models under Conditions of Data Scarcity. *Smart Cities*, 2024, Vol. 7, pp. 2495–2513; <https://doi.org/10.3390/smartcities7050097>
24. Biblioteka «ITMO-2» [Library «ITMO-2»]; URL: <https://github.com/aghajanyan/ITMO-2/>
25. Geoinformatsionnaya sistema QGIS [Geographic information system QGIS]; URL: <https://qgis.org/ru/site/>
26. Elektronnyi spravochnik kart gorodov 2GIS [Electronic directory of city maps 2GIS]; URL: <https://2gis.ru>
27. Yandex.Karty [Yandex.Maps]; <https://yandex.ru/maps>
28. Yandex.Auditorii [Yandex.Audiences]; URL: <https://audience.yandex.ru/>
29. Jennings V., Lloyd-Smith B., Ironmonger D. Household size and the Poisson distribution. *J. Australian Population Association*, 1999, Vol. 16, No. 1, pp. 65–84.
30. Shamil M.S., Farheen F., Ibtehad N. et al. An agent-based modeling of COVID-19: validation, analysis, and recommendations. *Cognitive Computation*, 2021, pp. 1–12.
31. Kagho G.O., Meli J., Walser D., Balac M. Effects of population sampling on agent-based transport simulation of on-demand services. *Procedia Comput. Sci.*, 2022, Vol. 201, pp. 305–312.
32. Llorca C., Moeckel R. Effects of scaling down the population for agent-based traffic simulations. *Procedia Comput. Sci.*, 2019, No. 151, pp. 782–787.
33. Korzin A.I., Kaparulin T.I., Leonenko V.N. Assessing the Effect of Influenza Vaccination Strategies Using Multi-agent Modeling. IEEE 3rd International Conference on Problems of Informatics, Electronics and Radio Engineering (PIERE), 2024, November, 2024, pp. 1000–1003.
34. Leonenko V.N., Ivanov S.V. Fitting the SEIR model of seasonal influenza outbreak to the incidence data for Russian cities. *Russian J. Numerical Analysis and Math. Modelling.*, 2016, Vol. 31, No. 5, pp. 267–279.
35. Harris J.E. Critical role of the subways in the initial spread of SARS-CoV-2 in New York City. *Frontiers in Public Health*, 2021, Vol. 9, Article number 754767.
36. Leonenko V. A Hybrid Modeling Framework for City-Scale Dynamics of Multi-strain Influenza Epidemics. *Proc. Internat. Conf. Comput. Sci.*, 2022, Vol. 13352, pp. 164–177; DOI: 10.1007/978-3-031-08757-8\_16
37. Lee J.S., Filatova T., Ligmann-Zielinska A., Hassani-Mahmooei B., Stonedahl F., Lorscheid I., Voinov A., Polhill J.G., Sun Z., Parker D.C. The complexities of agent-based modeling output analysis. *J. Artif. Soc. Soc. Simul.*, 2015, 18(4); DOI:10.18564/jasss.2897
38. Castro B.M., Reis M.D.M., Salles R.M. Multi-agent simulation model updating and forecasting for the evaluation of COVID-19 transmission. *Scientific Reports*, 2022, Vol. 12, No. 1, Article number 22091.