УДК 51-7

ГИБРИДНЫЙ ПОДХОД К МОДЕЛИРОВАНИЮ ДИНАМИКИ ИНФЕКЦИОННЫХ ЗАБОЛЕВАНИЙ НА ОСНОВЕ СОВМЕЩЕНИЯ ПОПУЛЯЦИОННЫХ И МУЛЬТИАГЕНТНЫХ МОДЕЛЕЙ

Леоненко В.Н. (Санкт-Петербург)

Введение

Вспышки эпидемий инфекционных заболеваний, таких как грипп и COVID-19, охватывающие практически все страны мира, ежегодно приводят к миллионам случаев тяжелого течения болезней и сотням тысяч случаев смертей в год. Для расширения возможностей органов эпиднадзора и, как следствие, повышения эффективности средств сдерживания эпидемий инфекционных заболеваний и снижения смертности от осложнений, в органах здравоохранения широко используются статистические и математические модели. Среди факторов динамики инфекций важнейшими являются особенности контактов в популяции [1], а также уровни иммунитета и распределение иммунных индивидов в популяции [2]. Последний фактор напрямую связан с первым, так как неоднородности в топологии сетей контактов приводят к неравномерному распределению инфицированных в популяции и, как следствие, разной доле иммунных особей в разных подгруппах населения, что усложняет динамику протекания инфекций и делает невозможной её предсказание при использовании базовых моделей. Например, в детерминированных SEIR- моделях, сформулированных в терминах систем разностных и дифференциальных уравнений, предполагается, что такой показатель, как уровень популяционного иммунитета (в целом по популяции или по отдельным крупным подгруппам), однозначно определяет количество переболевших в результате эпидемии. Однако это противоречит реальным данным, в которых можно наблюдать, что динамика распространения инфекции существенно зависит от вероятностных факторов, присущих начальным стадиям эпидемии [3], и топологии контактной сети. В результате возможны ситуации, когда в некотором городе согласно SEIR-модели эпидемия, безусловно, должна была произойти, потому что был зафиксирован занос инфекции в городскую популяцию, однако в силу благоприятного стечения обстоятельств эпидемия не стартует из-за того, что не сформировалось устойчивых цепочек передачи инфекции (например, инфицированный индивид, прилетевший в город, сразу попал домой и потом был госпитализирован при переходе инфекции в тяжёлую форму, не успев до этого никого инфицировать).

Методикой моделирования, позволяющей учесть неоднородность контактной сети при передаче заболевания, является мультиагентное моделирование. Существует большое количество публикаций по данной теме, например, [4], где описано мультиагентное моделирование сценариев вакцинации в гетерогенных популяциях на основе данных о заболеваемости из социальных сеней. Другими примерами являются многокомпонентная стохастическая модель, воспроизводящая динамику гриппа в отдельных регионах Англии и Уэльса за 14 лет [2], и работы исследовательских групп, использующих мультиагентный подход для прогнозирования динамики гриппа на основе синтетических популяций, например, исследовательские группы Университета Питтсбурга [5], RTI International [6] и Медицинского факультета Университета Уэйк Форест [7]. Эта же концепция активно применялась при моделировании динамики COVID-19 и послужила базой для реализации соответствующих сред моделирования, например, COVASIM [8], [9].

Одним из основных недостатков мультиагентного подхода, который, наряду с потребностью в большом количестве детальных данных для калибровки моделей, серьезно ограничивает его применимость, связан с большой вычислительной трудоёмкостью соответствующих модельных программ. Даже в простейшем случае, когда цель состоит в том, чтобы рассчитать динамику заболеваемости для одной эпидемической вспышки в масштабе города с заранее известным набором значений параметров, требуется несколько запусков моделирования для устранения стохастической неопределенности. В результате эксперимент может длиться от нескольких часов до нескольких дней, в зависимости от используемой модели и доступных вычислительных ресурсов. Очевидно, что в этих условиях задачи, требующие большого количества повторяющихся запусков с различными значениями параметров, такие как калибровка модели по данным или анализ неопределенности/чувствительности, не могут быть выполнены в разумные сроки. Существуют различные методы преодоления этого препятствия, а именно методы, связанные с предварительной модификацией данных (например, использование репрезентативной выборки индивидуумов, а не всей совокупности в симуляциях), оптимизацией алгоритмов и параллельными вычислениями (в частности, реализацией GPGPU-совместимых фреймворков), а также упрощением некоторых процессов в динамике рассматриваемой системы (например, путем обучения нейронных сетей на выходных данных мультиагентных моделей для воспроизведения траекторий заболеваемости и распространенности заболеваний без фактического моделирования). Среди последней группы подходов можно назвать гибридное моделирование динамики заболеваемости [10]. Указанный метод основан на идее о том, что в некоторых наборах условий различия в результатах работы детализированных мультиагентных моделей и значительно более простых популяционных моделей могут быть пренебрежимо малыми [11], что в этих специальных условиях позволяет заменить первый подход вторым без потери точности. В результате подобной замены получается искомая гибридная модель с динамическим переключением между двумя подходами, которая в известной степени совмещает преимущества популяционного и мультиагентного подходов (высокая степень детализации при малом количестве заболевших на начальном этапе эпидемической вспышки, когда эта детализация действительно важна для корректного описания происходящего, и низкая детализация с высокой скоростью моделирования, когда инфекционных индивидов становится настолько много, что слежение за отдельными цепями передачи инфекции теряет смысл). Основной недостаток гибридного подхода вытекает из его достоинства – необходимость использования двух методов моделирования для описания протекания одной эпидемической вспышки ставит вопрос о соответствии первого модельного описания системы второму, т.е. необходимо обоснование того, что при изменении типа описания системы сама описываемая система не меняет своих свойств. Требуется обеспечить плавность переключения между двумя моделями и удостовериться в том, что технический момент, связанный с переключениями, не влияет на корректность отображения эпидемической динамики.

В настоящей работе обсуждается подход к гибридному моделированию распространения эпидемических ОРВИ в городской популяции, проблемы его реализации в условиях калибровки моделей на данные РФ и особенности применения подхода для моделирования распространения сезонного гриппа.

Материалы и методы

Исходная концепция

Общая концепция разработанной автором гибридной модели основывается на постулатах, предложенных в работе [10]. Концепция исходит из предпосылки, что при значительном увеличении числа активных агентов в мультиагентной модели к наблюдениям могут быть применены закон больших чисел и центральная предельная теорема. Таким образом, появляется возможность агрегировать поведение подобных агентов и моделировать их поведение с помощью приближений среднего поля. Применение подхода было проиллюстрировано авторами для простой SEIR-модели, в которой процесс передачи инфекции был основан на классическом для математической эпидемиологии законе действующих масс. Согласно этому закону каждый инфицированный агент имеет равный шанс встретиться с любым другим агентом и передать ему инфекцию. В мультиагентной модели, которая аппроксимировалась описанной выше SEIR-моделью, считалось, что человек находится в одном из следующих четырех состояний: (1) восприимчивый (может заразиться болезнью при контакте с инфицированным человеком), (2) подвергшийся воздействию (заразился, но находится в латентном состоянии без проявления симптомов), (3) инфекционный (проявляющий симптомы и способный заражать других) и (4) выздоровевший (приобрел постоянный иммунитет и не может заразить других). Поскольку взаимодействие между восприимчивыми агентами не имеет отношения к эпидемическому процессу, в мультиагентной модели отслеживались контакты подвергшихся воздействию и инфекционных агентов, сохраняя восприимчивых агентов неактивными. Когда инфицированный человек передает инфекцию восприимчивому человеку, активируется соответствующий восприимчивый агент. Этот алгоритм позволяет поддерживать низкое количество активных агентов в начале эпидемического процесса и, таким образом, снижает вычислительную нагрузку. По мере увеличения количества активированных агентов увеличивается и объем необходимых вычислительных ресурсов (например, оперативной памяти).

Ключевым моментом, связанным с работой гибридного модельного комплекса, является условие переключения между моделями. В описываемой концепции гибридного моделирования предполагается, что условие переключения связано с достигнутым абсолютным количеством инфицированных людей в популяции. Когда оно достигает определенного значения, модель останавливалась и производилась оценка состояний, в которых агенты находились к этому моменту. Затем посчитанные значения использовались в качестве начальных условий для популяционной SEIR-модели, и указанная модель стартует с начального момента времени, равного моменту остановки мультиагентной модели. И наоборот, когда число инфицированных людей становится небольшим при затухании эпидемического процесса в популяции, модельный комплекс переключается в мультиагентный режим, чтобы избежать артефактов, которые в такой ситуации могут быть вызваны использованием популяционной модели. Изначально для переключения авторами был выбран порог в 1000 инфицированных индивидов, далее в экспериментах с моделью авторами был изучен ряд более низких порогов переключения, соответствующих 50, 100 и 200 индивидов. Было выяснено, что, когда порог составляет около 100 новых зараженных агентов, комплекс слишком рано переключается на настройку на основе уравнений и очень точно следует траектории «чистой» популяционной модели (т. е. модели вне гибридной схемы). Переход на модель, основанную на уравнениях, на более позднем этапе, например, пороговое значение 200, делает модель более близкой к модели с переключением в 1000 случаях, что привело авторов к выводу полагать, что дальнейшее повышение порога не приведет к принципиально другим результатам.

Авторы отмечают, что поскольку пороговое значение может быть разным для городов разного размера, а одно и то же значение уровня переключения не подходит для всех сценариев, необходима разработка правила для динамической установки порогового значения. В частности, один из вариантов связан с отслеживанием оценки скорости передачи заболевания в модели. Когда число инфицированных людей достаточно велико, оценка данного параметра стабилизируется вокруг теоретического значения, соответствующего популяционной SEIR-модели. В силу этого момент такой стабилизации подходит для переключения между компонентами гибридной системы.

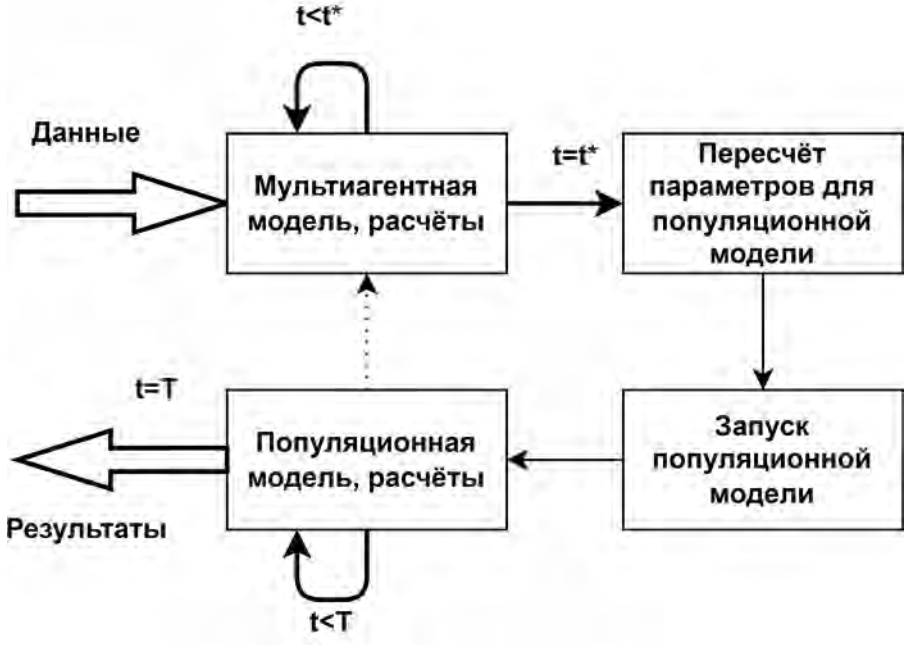

Рис. 1. Схема работы гибридного модельного комплекса

Гибридная модель ко-циркуляции штаммов гриппа в Санкт-Петербурге

В качестве развития описанной выше концепции для решения конкретной прикладной задачи автором была предложена гибридная модель динамики гриппа с учётом различий в циркулирующих штаммах. Полное техническое описание модели доступно в [12], далее в тексте приводится его краткое обобщение. Принципиальная схема работы комплекса представлена на рис. 1.

Общая идея. Исходной моделью, использованной в качестве основы для гибридного комплекса, являлась мультиагентная модель ко-циркуляции штаммов гриппа в синтетической популяции, которая подробно описана в [13]. Она имеет дискретное время с шагом моделирования, равным одному дню. Эпидемический процесс инициируется путем случайного присвоения инфекционного статуса некоторым особям в синтетической популяции в начале моделирования. Выходные данные модели включают генерацию пространственных распределений случаев заболеваемости, расчет общей заболеваемости и распространенности в исследуемой популяции, а также оценку уровней коллективного иммунитета в популяции после вспышки. При необходимости возможна генерация дополнительных данных, таких как идентификаторы места заражения (школа, рабочее место, дом и его непосредственные окрестности) для каждого случая заболеваемости, что позволяет оценить вклад контактов в каждом типе общественных мест в распространение инфекции.

Население. Параметры, связанные с населением, используемые для модели, организованы в виде синтетической популяции Санкт-Петербурга за 2010 год. Население включает 40 213 домашних хозяйств с совокупным числом жителей 4 865

118 человек. Жилые дома рассматриваются как совокупность отдельных жилищ, и человек может контактировать только с теми людьми, с которыми он живет в одном помещении (например, с членами семьи). Предполагалось, что все молодые люди в возрасте от 7 до 17 лет посещают школы, а взрослые трудоспособного возраста (от 18 до 55 лет для мужчин и от 18 до 60 лет для женщин) могут работать. Рабочие места разделены на небольшие отсеки, внутри которых происходят ежедневные контакты. Средний размер рабочего места был выбран равным среднесуточному количеству контактов на рабочем месте. Размеры конкретных «отсеков» рабочих мест генерировались случайным образом, предполагая распределение Пуассона с соответствующим средним значением. Рабочие места для взрослых и школы для детей школьного возраста выбирались случайным образом исходя из имеющихся вакантных мест в определенном радиусе от домохозяйства (исходя из общих соображений, мы взяли 15 км и 5 км соответственно, что представляется адекватным для Санкт-Петербурга). Если в пределах заданного радиуса не было вакансий в школах/рабочих мест, ближайшая вакансия назначалась независимо от расстояния. Оставшиеся лица, не имеющие школы, распределялись по ближайшим школам без учета школьной вместимости, в то время как оставшиеся лица, не имеющие работы, считались безработными.

Контакты. Предполагалось, что существуют следующие модели повседневной деятельности в зависимости от свойств конкретного индивида популяции: пребывание в домашнем хозяйстве в течение всего дня (дети-дошкольники, пенсионеры, безработные), посещение школы (учащиеся), посещение рабочего места (работающие взрослые). Таким образом, в каждый день у одного человека имеется от 1 до 2 мест потенциальных контактов, которые заданы изначально при запуске модели и не меняются с течением времени. Роль общественного транспорта в распространении инфекции не рассматривается. Выходные дни не рассматриваются отдельно, т. е. поведение индивидов одинаково в течение всех дней.

Начало заболевания и выздоровление. Интенсивность так называемых «эффективных контактов» (т. е. контактов между восприимчивым и инфицированным человеком, которые приводят к новым случаям заражения) в конкретном общественном месте зависит от среднего числа контактов на человека в день и трансмиссивности вируса. Было принято упрощающее предположение, что трансмиссивность вирусов гриппа не зависит от штамма. Если разные штаммы одновременно передаются человеку в месте контакта, один из них выбирается случайным образом как вызывающий инфекцию. Каждый агент в популяции потенциально взаимодействует с другими агентами, если они посещают одну и ту же школу (для школьников), рабочее место (для взрослых трудоспособного возраста) или живут в одном домохозяйстве. Заразность каждого человека зависит от дня заражения. Предполагалось, что доля заразных особей во времени не зависит от штамма. Люди, выздоровевшие от болезни, считаются невосприимчивыми к конкретному штамму гриппа, который ее вызвал, до конца моделирования. Перекрестный иммунитет не учитывается, т.е. упомянутые выздоровевшие лица не приобретают иммунитета к другим штаммам гриппа.

Популяционная модель. В качестве упрощенной замены мультиагентной модели используется SEIR-модель, основанная на детерминированной системе разностных уравнений с временным шагом, равным одному дню. Подробное описание модели можно найти в [14]. По аналогии с мультиагентной моделью мы рассматриваем ко-циркуляцию трех штаммов гриппа, A(H1N1)pdm09, A(H3N2) и B. Возможность последовательного заражения несколькими штаммами в течение одного сезона не рассматривается, поэтому лица, выздоровевшие от гриппа, вызванного любым из циркулирующих штаммов, считаются иммунными. Однако этот предполагаемый

перекрестный иммунитет между штаммами вируса не переносится на следующий эпидемический сезон.

Алгоритм переключения. Поскольку время вычисления одного шага алгоритма (дня моделирования) растет очень быстро с увеличением числа инфицированных в популяции, важно как можно раньше аппроксимировать процесс заражения более простой моделью, начиная с момента, когда количество инфицированных становится большим. Это условие вступает в противоречие с необходимостью моделирования начального этапа эпидемии с помощью детальной модели, поскольку ранее переключение, как было упомянуто в предыдущем разделе данной публикации, приводит к совпадению гибридной модели с результатами работы чисто популяционной модели. В этом же разделе были упомянуты два типа условий переключения, предложенных авторами работы [10]:

- Переход на популяционную модель при достижении определенного количества инфицированных особей в популяции;
- Переход на популяционную модель, когда эффективное репродуктивное число инфекции R_t стабилизируется (разница между соответствующими значениями для последующих дней моделирования становится ниже определенного порога).

Оба условия переключения легко реализуются в случае модели динамики заболевания, вызванной одним вирусом, однако они требуют адаптации для задачи ко-циркуляции. Типичная ситуация в моделировании ко-циркуляции вирусов соответствует высокой распространенности инфекции, вызванной одним штаммом, и низкой распространенности, вызванной другими штаммами. В результате, если сделать переключение на основе суммарного по всем штаммам числа инфицированных или «средневзвешенного» репродуктивного числа, динамика заболевания вирусом с низкой распространенностью будет искажена из-за слишком раннего переключения. С другой стороны, если переключение должно быть выполнено, когда пороговое значение достигнуто для числа заражений самым редким штаммом из рассматриваемых, это приведет к полному отсутствию переключения или к позднему переключению, потому что незначительные эпидемии с распространенностью, никогда не превышающей порога, характерны практически для каждого модельного запуска. В силу этой проблемы, для реализованной автором гибридной модели было выбрано условие переключения, зависящее от времени, т. е. момент переключения привязан к конкретному модельному дню t^*.

Вычислительные эксперименты

В вычислительных экспериментах, проводимых с гибридной моделью, значения параметров, общие для двух модельных компонент (размер популяции, доля переболевших тем или иным штаммом в предыдущем сезоне, вирулентность инфекций, скорость потери иммунитета), выставлялись равными. Однако часть параметров в силу особенностей реализации моделей оказалось невозможным отождествить друг другу. Естественным образом они оказались связаны с детализацией описания индивидов и сетей их контактов в мультиагентной модели. Например, сложную систему контактов с различными их интенсивностями в зависимости от свойств индивидов в популяционной модели заменяет одно значение, соответствующее среднему числу контактов в популяции. Особенную проблему вызвала интерпретация параметра μ, который в популяционной модели отражает долю населения, по какой-то причине выключенной из эпидемического процесса [14] и должен соответствовать косвенному иммунитету в мультиагентной модели (например, защищённость людей, которые не имеют социальных связей или связаны исключительно с иммунными индивидами, сами не являясь иммунными). Значение $\mu = 0.9$ в популяционной компоненте комплекса было установлено на основе предыдущего опыта калибровки популяционных моделей.

На рисунках 2 и 3 приведены в качестве примера полученные траектории заболеваемости гриппом B при переключении модели в моменты времени t*=5 и t*=15. На графиках для случая t*=15 хорошо виден сам момент переключения, что свидетельствует о том, что сопоставление подмоделей простым выравниванием значений параметров не позволяет осуществить плавный переход. Таким образом, текущая структура алгоритма переключения может потребовать модификации для обеспечения правильной связи подмоделей. Это наблюдение согласуется с выводами, продемонстрированными другими исследовательскими группами. Можно заметить, что переключение в день 5 приводит к гладким траекториям, но это связано с незначительной численностью инфицированных на этом модельном шаге и эффектом масштаба. Кроме того, при таком раннем переключении траектория динамики гибридного комплекса предположительно совпадает с чисто популяционной моделью, что может быть показано экспериментально и согласуется с результатом уже цитированной работы [10]. Таким образом, мультиагентная компонента перестаёт играть какую-либо значимую роль, и итоговая модель излишне огрубляется, теряя возможность описать стохастическую динамику в начальный период эпидемической вспышки.

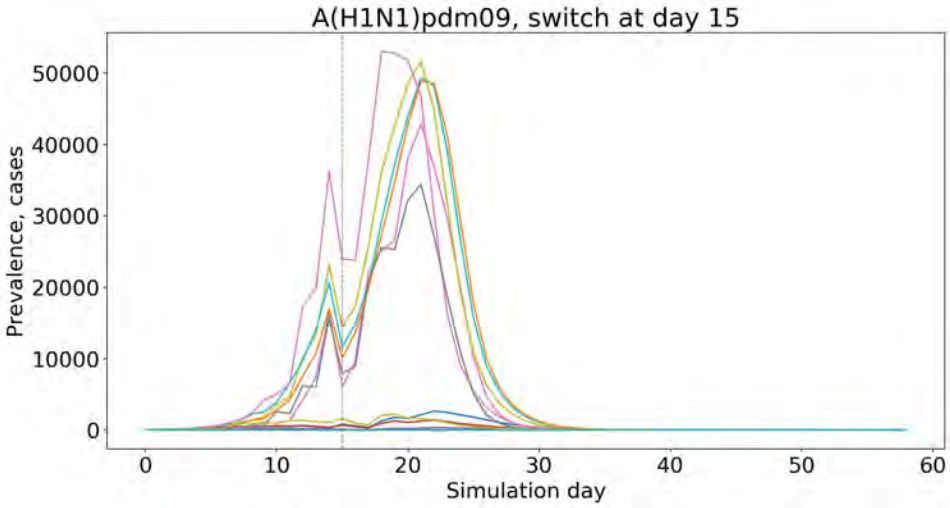

Рис. 2. Траектории заболеваемости гриппом A(H1N1)pdm09 при переключении модели в момент t*=15

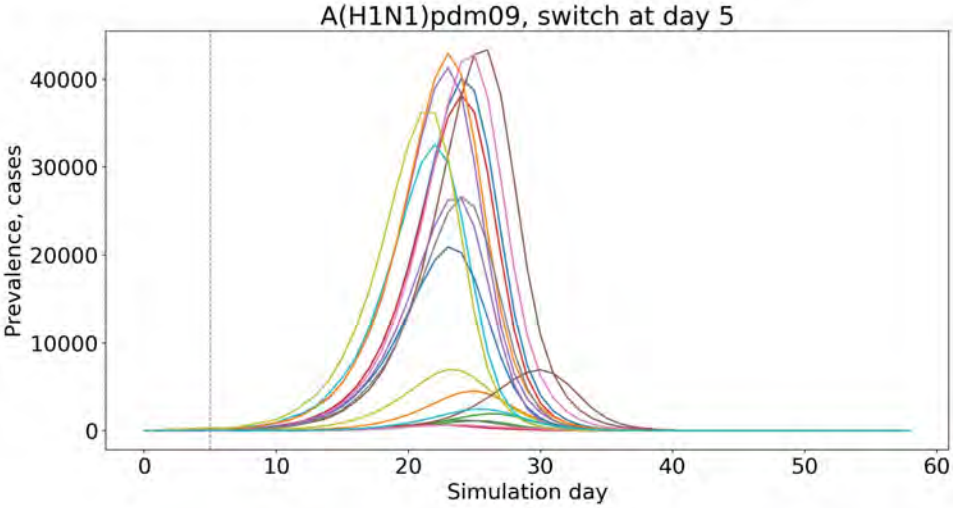

Рис. 3. Траектории заболеваемости гриппом A(H1N1)pdm09 при переключении модели в момент t*=5

Продемонстрированная прерывистость траекторий заболеваемости в момент переключения потребовала пересмотра реализации алгоритма переключения. В частности, в настоящий момент предлагаются альтернативные формулы для пересчёта значений параметров, эффективность которых будет установлена многомасштабными численными экспериментами. После устранения проблем с отсутствием гладкости в моменты t=t* и дополнительной верификации комплекса, связанной со сравнением максимальной и суммарной заболеваемостей при разных моментах переключения, гибридный комплекс будет откалиброван на данных заболеваемости в Санкт-Петербурге в 2010-2022 гг. с использованием синтетических популяций, получаемых с помощью индивидуум-ориентированной демографической модели [15].

Заключение

В данной статье представлены подходы к организации структуры гибридного комплекса моделирования, который потенциально позволяет обеспечить хороший компромисс между уровнем детализации моделирования и скоростью вычислений. Показаны проблемы, связанные с необходимостью гладкого перехода от подмодели к подмодели и с связанным с этим требованием корректного пересчёта параметров подмоделей. Обсуждены варианты решения проблем, связанные с изменением формул пересчёта и алгоритма переключения. По мнению автора, указанные шаги сделают описанную структуру гибридного моделирования ценным инструментом для краткосрочного прогнозирования распространенности гриппа и COVID-19.

Работа выполнена при поддержке Российского научного фонда (грант РНФ 22-71-10067).

Литература

1. **Ajelli M., Litvinova M.** Estimating contact patterns relevant to the spread of infectious diseases in Russia // Journal of Theoretical Biology. – 2017. – 419, 1 7.
2. **Baguelin M. et al.** Assessing optimal target populations for influenza vaccination programmes: an evidence synthesis and modelling study // PLOS medicine. – 2013. – 10(10), e1001527.
3. **Drake J.M. et al.** The statistics of epidemic transitions // PLOS Computational Biology. – 15(5), 1 14 (05 2019)
4. **Lee B.Y. et al.** A computer simulation of vaccine prioritization, allocation, and rationing during the 2009 h1n1 influenza pandemic // Vaccine. – 2010. – 28(31). – P. 4875 4879.
5. **Kumar S. et al.** Policies to reduce influenza in the workplace: impact assessments using an agent-based model // American journal of public health. – 2013. – 103(8). – P. 1406-1411.
6. **Cooley P. et al.** The role of subway travel in an influenza epidemic: a New York City simulation // Journal of Urban Health. – 2011. – 88(5). – P. 982.
7. **Guo D. et al.** Multiscale modeling for the transmission of influenza and the evaluation of interventions toward it // Scientific Reports. – 2015. – 5(1). – 1 9
8. **Kerr C.C. et al.** COVASIM: an agent-based model of COVID-19 dynamics and interventions // PLOS Computational Biology. – 2021. – 17(7). – e1009149.
9. **Krivorotko O. et al.** Agent-based modeling of COVID-19 outbreaks for New York State and UK: Parameter identification algorithm // Infectious Disease Modelling. – 2022. – 7(1). – P. 30-44.

10. **Bobashev G.V., Goedecke D.M., Yu F., Epstein J.M.** A hybrid epidemic model: combining the advantages of agent-based and equation-based approaches // 2007 winter simulation conference. – 2007. – Pp. 1532-1537. IEEE.

11. **Rahmandad H., Sterman J.** Heterogeneity and network structure in the dynamics of diffusion: Comparing agent-based and differential equation models // Management Science. – 2008. – 54(5). – P. 998-1014.

12. **Leonenko V.** A Hybrid Modeling Framework for City-Scale Dynamics of Multi-strain Influenza Epidemics // International Conference on Computational Science. – Cham: Springer International Publishing, 2022. – С. 164-177.

13. **Leonenko V., Arzamastsev S., Bobashev G.** Contact patterns and influenza outbreaks in Russian cities: A proof-of-concept study via agent-based modeling // Journal of Computational Science. – 2020. – 44. – 101156.

14. **Leonenko V.N.** Herd immunity levels and multi-strain influenza epidemics in russia: a modelling study // Russian Journal of Numerical Analysis and Mathematical Modelling. – 2021. – 36(5). – P. 279-291.

15. **Arzamastsev S. A., Leonenko V. N.** A demographic microsimulation model for the long-term evolution of synthetic populations in Saint-Petersburg // Математическая биология и биоинформатика. – 2020. – С. 157-161.