

Министерство образования и науки Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования «Московский физико-технический институт
(государственный университет)»

Физтех-школа биологической и медицинской физики
Кафедра системной и синтетической биологии

Исследование роли транспортных потоков в развитии эпидемии на основе компьютерной агентной модели

Выпускная квалификационная работа бакалавра

Выполнил:
студент Б06-101 группы,
Клочков Константин Александрович.

Научный руководитель:
к.б.н., Манолов Александр Иванович.

Долгопрудный, 2025

Содержание

1 Список сокращений	2
2 Введение	3
3 Обзор литературы	4
3.1 Виды эпидемиологических моделей	4
3.1.1 Детерминистические компартментные модели	4
3.1.2 Детерминистические модели среднего поля	5
3.1.3 Стохастические модели	6
3.1.4 Агентные модели	7
3.2 Подходы к изучению транспортных потоков в эпидемиологии	7
3.2.1 Математические модели	8
3.2.2 Модели на графах	9
3.2.3 Агентные модели	9
3.3 Агентная модель Covasim	11
3.4 Дисперсионный анализ чувствительности математических моделей	13
3.5 Калибровка моделей	14
4 Материалы и методы	14
4.1 Используемые программные пакеты	14
4.2 Вычисление индексов Соболя	14
4.3 Исследуемая агентная модель	16
5 Результаты	16
5.1 Город и сATEллиты	16
5.2 Базовые потоки	19
5.3 Индексы Соболя	21
6 Выводы	21
7 Благодарности	21
8 Список литературы	21

1 Список сокращений

Аббревиатуры

ABM Agent-based model (от англ. агентная модель). [7](#), [9](#), [10](#)

Covasim COVID-19 Agent-based Simulator. [10](#), [11](#)

COVID-19 Coronavirus Disease 2019. [11](#)

EBCM edge-based compartmental models (от англ. компартментные модели на ребрах графа). [6](#)

MP message passing (от англ. обмен сообщениями). [6](#)

SEIR susceptible → exposed → infectious → recovered (от англ. восприимчивый → подверженный → инфицированный → восстановившийся). [10](#)

SEIS susceptible → exposed → infectious → susceptible (от англ. восприимчивый → подверженный → инфицированный → восприимчивый). [9](#)

SIQS susceptible → infectious → quarantined → susceptible (от англ. восприимчивый → инфицированный → изолированный → восприимчивый). [8](#)

SIR susceptible → infectious → recovered (от англ. восприимчивый → инфицированный → восстановившийся). [4](#), [5](#), [9](#)

SIRP susceptible → infectious → recovered → partially immune (от англ. восприимчивый → инфицированный → восстановившийся → частично иммунный). [8](#)

SIS susceptible → infectious → susceptible (от англ. восприимчивый → инфицированный → восприимчивый). [4–6](#), [8](#)

ЗППП заболевания, передающиеся половым путем. [5](#)

2 Введение

3 Обзор литературы

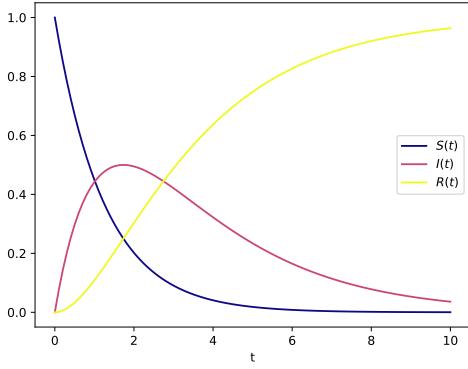
3.1 Виды эпидемиологических моделей

Математические модели стали важными инструментами анализа распространения инфекционных заболеваний и способов борьбы с ними. В процессе создания модели уточняются предположения и параметры, а само моделирование позволяет оценить критические значения параметров модели, приводящие к развитию эпидемии, индекс репродукции инфекции R_0 , среднее число контактов между людьми σ . Математические модели и компьютерное моделирование являются полезными экспериментальными инструментами для проверки теорий, оценки количественных гипотез, ответов на конкретные вопросы, определения чувствительности к изменениям значений параметров и оценки ключевых параметров по данным. Понимание особенностей передачи инфекционных заболеваний в сообществах, регионах и странах может привести к улучшению подходов к снижению распространения этих заболеваний. Математические модели используются при сравнении, планировании, реализации, оценке и оптимизации различных методов программ по выявлению, профилактике, лечению и контролю. Моделирование в эпидемиологии может способствовать разработке и анализу эпидемиологических обследований, предлагать важнейшие данные которые необходимо собрать, выявить тенденции, сделать общие прогнозы и оценить неопределенность прогнозов [1–3].

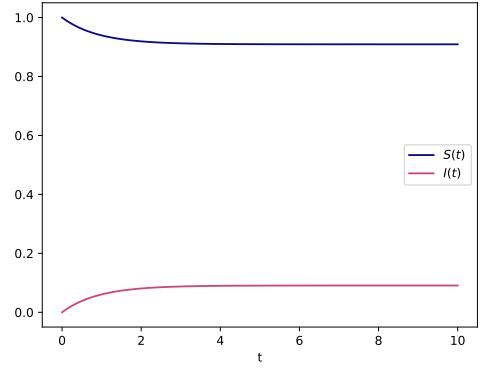
3.1.1 Детерминистические компартментные модели

Подавляющее большинство эпидемиологических моделей основано на компартментализации людей в зависимости от их состояния [4–7]. Базовые модели описывают лишь три компартмента: восприимчивых, инфицированных и восстановившихся. В таком случае происходит пренебрежение многими деталями развития эпидемии, тем не менее такие модели все еще активно применяются. Основным из используемых типов является **susceptible → infectious → recovered** (от англ. восприимчивый → инфицированный → восстановившийся) (**SIR**) модель, описываемая системой дифференциальных уравнений

$$\begin{cases} \frac{dS}{dt} = bN - \lambda S - dS, \\ \frac{dI}{dt} = \lambda S - gI - dI, \\ \frac{dR}{dt} = gI - dR, \end{cases} \quad (1)$$



(a) SIR модель



(b) SIS модель

Рис. 1: Численные решения задачи Коши компартментных моделей с начальными условиями $S(0) = 1$, $I(0) = 0$, $R(0) = 0$ и параметрами $b = d = 0$, $g = 0.4$, $\lambda = 0.8$, $N = 1$

и susceptible → infectious → susceptible (от англ. восприимчивый → инфицированный → восприимчивый) (SIS) модель

$$\begin{cases} \frac{dS}{dt} = gI - \lambda S, \\ \frac{dI}{dt} = \lambda S - gI. \end{cases} \quad (2)$$

Переменные S , I , R обозначают количества восприимчивых, инфицированных, восстановившихся людей соответственно; N — размер популяции; b , d , g — коэффициенты рождаемости, смертности и восстановления (доля рождающихся, умирающих, восстанавливющихся людей в единицу времени); λ — сила инфекции (доля восприимчивых людей, инфицируемых в единицу времени).

SIR модели применяются для описания инфекционных заболеваний, формирующих очень стойкий или пожизненный иммунитет, например, кори, коклюша [4; 5; 7—9]. SIS модели используют преимущественно при описании распространения ЗППП, реинфицирование которыми вполне возможно, например, хламидиоза, гонореи [4; 10; 11].

Также активно применяются модифицированные модели с уточненной компартментализацией для более сложного протекания болезни [4; 12; 13] и с более проработанной структурой популяции [10; 14; 15].

3.1.2 Детерминистические модели среднего поля

Компартментные модели строятся в предположении, что каждый человек взаимодействует с каждым. На самом деле в большой популяции у каждого индивида есть свой ограниченный круг общения, и именно им определяется путь распространения инфекции, причем число контактов очень сильно меняется от человека к человеку [16]. Для более реалистичного описания таких социальных взаимодействий распространение эпидемии моделируют на графе контактов (или сети контактов), в котором каждому человеку соответствует узел, а каждой социальной связи — ребро графа [17].

В связи с тем, что прямой анализ стохастического поведения эпидемии на графе является сложным, его описывают детерминистическим приближением среднего поля, анализируя некоторые характерные средние характеристики моделей [17]. Другими словами, в таких моделях для упрощения математического описания вводятся различные приближения: попарные модели рассматривают состояния пар узлов, соединенных ребром [18; 19]; *message passing* (от англ. обмен сообщениями) (MP) модели рассматривают каждый узел и все возможные пути передачи инфекции в этот узел независимо [20]; модели средней степени вершины графа учитывают звездчатые структуры — узлы с их соседями с учетом их состояния [21]; *edge-based compartmental models* (от англ. компартментные модели на ребрах графа) (EBCM) оценивают вероятность случайного узла остаться восприимчивым [22]. Все эти модели можно построить на одном и том же графе контактов, исходя из разных предположений, но, на самом деле, часть из них оказывается эквивалентна друг другу [17].

Эти модели лучше описывают социальное устройство популяции, однако они рассматривают марковский процесс, то есть исходят из предположения, что состояние графа на следующем шаге определяется лишь его состоянием на текущем шаге и никак не зависит от предыдущих шагов. Эти предположения часто оказываются неверны, так как инфицирование или выздоровление происходит не случайно в любой момент времени, а после конкретного инкубационного периода или периода болезни соответственно [4; 19; 23].

3.1.3 Стохастические модели

Результатами применения стохастических моделей также являются временные зависимости некоторых глобальных величин [24; 25], однако, в отличие от детерминистических моделей, стохастический подход допускает временные флюктуации как на глобальном уровне, так и на локальном — флюктуации на уровне агентов в зависимости от их контактов [26].

Основой таких моделей, как и с моделями среднего поля, часто является граф взаимодействий, часто описываемый матрицей контактов A [27]. В отличие от компартментных моделей, матрица контактов может содержать в себе крайне сложные социальные структуры, что позволяет наблюдать более широкий спектр эпидемиологических явлений [28]. Распространение инфекции задается матрицей перехода \hat{T} , матричные элементы $T_{\mu\nu}$ которой равны вероятностям перехода из состояния графа ν в μ [29]. Эта разница между компартментными и стохастическими моделями приводит к различию между временными зависимостями эпидемиологических статистик. К примеру, оказывается, что, хоть SIS модель демонстрирует схожее с аналогичной стохастической поведение при различных соотношениях между силой инфекции и коэффициентом восстановления, на ранних этапах развития эпидемии, когда среднее число инфицированных сильно меньше размера популяции, ключевым фактором являются именно флюктуации числа инфицированных. И именно такому режиму соответствуют санитарные меры по

сдерживанию эпидемий на ранних стадиях [26].

3.1.4 Агентные модели

Agent-based model (от англ. агентная модель) (АВМ) применяются при моделировании распространения инфекционных заболеваний уже более 40 лет [30; 31], однако из-за недостатка данных и вычислительных мощностей раскрывать свой потенциал они начали совсем недавно [32].

Устройство АВМ позволяет напрямую описывать сложные социальные и физические системы. Хоть для воспроизведения социальной структуры конкретной страны или города может потребоваться очень много данных, АВМ может отразить неоднородность социальных взаимодействий и динамические изменения в структуре сети [33].

Такой подход очень полезен для простой реализации учета иммунологической истории индивидов, вариации климата между регионами, социальных структур, что позволяет проводить более тонкую калибровку. Также АВМ позволяют рассматривать распространение инфекции в маленьких группах людей [33].

Построение АВМ требует, во-первых, воссоздания искусственной популяции, соответствующей демографическим данным в регионе, и, во-вторых, создания сети публичных объектов, таких как школ, рабочих мест, где агенты могут взаимодействовать друг с другом. Главной сложностью при построении АВМ является тот факт, что требуемая информация часто скрыта при агрегировании федеральными службами государственной статистики. В связи с этим возникает необходимость в разработке методов, позволяющих воссоздать исходные данные из агрегатов, или моделей, приближающих искомые данные [33].

3.2 Подходы к изучению транспортных потоков в эпидемиологии

Транспорт играет важнейшую роль в жизни людей и в то же время активно способствует распространению эпидемий. В связи с этим введение ограничений на транспортные потоки является важной эпидемиологической мерой [34], а понимание влияния этих мер позволит человечеству эффективнее бороться с распространением инфекционных заболеваний.

В имеющихся исследованиях по этой теме транспорт описывается преимущественно математическими моделями, моделями на графах и агентными моделями. Статистические (регрессионные и авторегрессионные модели на пространственных данных) же для этих целей практически не применяются в связи с необходимостью иметь огромное множество данных о мобильности людей, в отсутствие которых качество результатов работы таких моделей крайне ограничено. Клеточными автоматами симулировать перемещение индивидов также затруднительно, так что их для этих целей тоже применяют редко [34].

3.2.1 Математические модели

Одним из преимуществ математических моделей является тот факт, что для их калибровки требуются малые объемы данных. По этой причине было предпринято множество попыток адаптирования существующих математических моделей под описание системы нескольких городов. Хоть эти модели и предполагают, что города обладают одинаковыми демографическими параметрами, что перемещения происходят мгновенно, некоторые игнорируют возможность инфицирования при перемещении, из них можно получить важные выводы [34].

Дж. Хайман и др. (2003) [35] для описания распространения гриппа в США в присутствие авиаотранспорта каждый исследуемый город описали моделью типа **susceptible → infectious → recovered → partially immune** (от англ. восприимчивый → инфицированный → восстановившийся → частично иммунный) (**SIRP**) — добавили компартмент, содержащий частично иммунных агентов, в который попадает часть восстановившихся. Перемещения людей ввели, определив постоянную матрицу миграции M , содержащую долю перемещающихся людей в единицу времени, внеся в производные $\frac{d}{dt}(S_k, I_k, R_k, P_k)$ слагаемое $\sum_{j=1}^n \left[m_{jk} \frac{(S_j, I_j, R_j, P_j)}{N_j} - m_{kj} \frac{(S_k, I_k, R_k, P_k)}{N_k} \right]$. В результате авторы выявили, что наиболее эффективно замедляет эпидемию сокращение продолжительности инфекционной стадии, сокращение числа контактов r , уменьшение вероятности передачи β . Этот подход полностью игнорирует возможность инфицирования при перемещении, что некорректно особенно при рассмотрении быстро распространяющихся заболеваний как грипп или атипичная пневмония, ведь во время поездок люди в течение долгого времени находятся в очень близком контакте друг с другом [34].

Ванг и Жао (2004) [36] на совокупности **SIS** на двух примерах проиллюстрировали важность перемещения агентов в развитии эпидемии: первый показал, что при большом индексе репродукции инфекции одного города пассажиропотоки могут усилить распространение болезни, если же пассажиропотоки большие при малых числах репродукции во всех населенных пунктах, то наличие транспорта может снизить распространение. Второй пример продемонстрировал, что инфекция из-за транспортных потоков может стать эндемичной, даже если в каждом изолированном городе она должна исчезнуть.

Такеучи и др. (2006) [37] предложили аналогичный подход для системы из двух **SIS** и ввели слагаемое, учитывающее инфицирование агентов в процессе перемещения между населенными пунктами, считая, что частота заболевания туристов $\gamma(\alpha S_j)(\alpha I_j)/(\alpha S_j + \alpha I_j)$, где γ — трансмиссивность инфекции в процессе перемещения, α — доля перемещающихся агентов в единицу времени. Анализ этой **SIS** привел авторов к выводу, что инфицирования, связанные с транспортными потоками, приводят к повышению как абсолютного, так и относительного числа больных, что подтверждает необходимость вводить транспортные ограничения сразу при обнаружении вспышки инфекции [37].

Лиу и Такеучи (2006) [38] рассматривали возможность тестирования инфицированных агентов и помещения их в карантинный компартмент в системе двух **susceptible**

\rightarrow infectious \rightarrow quarantined \rightarrow susceptible (от англ. восприимчивый \rightarrow инфицированный \rightarrow изолированный \rightarrow восприимчивый) (SIQS), учитывая туризм аналогично авторам упомянутым выше. Они пришли к тому, что входной скрининг с последующим карантином инфицированных крайне полезен, так как всегда может привести к завершению эпидемии, даже если инфекция эндемична в каждом из рассматриваемых городов [38].

Ван и Цуй (2007) [39] исследовали возможности контроля распространения инфекционных заболеваний, изучая модель susceptible \rightarrow exposed \rightarrow infectious \rightarrow susceptible (от англ. восприимчивый \rightarrow подверженный \rightarrow инфицированный \rightarrow восприимчивый) (SEIS). Оказалось, что даже если запретить инфекционным агентам перемещаться, инфицированные в инкубационной фазе все равно приносят болезнь в города. Помимо этого они подтвердили результаты, полученные ранее Вангом и Жао [36] и Такеучи и др. [37], о том, что наличие транспортных потоков может качественно изменить ход эпидемии.

Лиу и Жоу (2007) [40] при анализе SIR вывели, что при базовом репродуктивном числе $R_{0\gamma} \leq 1$ болезнь исчезает со временем, и это состояние глобально асимптотически устойчиво. Если же $R_{0\gamma} > 1$ существует локально асимптотически устойчивое эндемическое равновесие, то есть болезнь остается в популяции. Также, как и у Ванг и Жао [36], авторы получили, что при наличии транспорта инфекция может стать эндемичной, даже если в каждом отдельном городе в отсутствие перемещений агентов она должна была исчезнуть. Аналогичный результат получили и Ли и Жоу (2009) [41], используя систему дифференциальных уравнений с запаздыванием при учете латентного периода в SIR.

3.2.2 Модели на графах

3.2.3 Агентные модели

В АВМ каждому агенту могут быть присвоены свои атрибуты, при этом каждый принимает индивидуальные решения, что позволяет отразить все детали межличностных взаимодействий. В связи с развитием систем общественного транспорта им пользуется все больше людей, что значительно усложняет сеть контактов. Поэтому моделирование транспортных систем очень важно в эпидемиологических АВМ.

Симоэс (2006) [42] для описания эпидемии паротита в Португалии 1996 года разработала АВМ, состоящую из susceptible \rightarrow exposed \rightarrow infectious \rightarrow recovered (от англ. восприимчивый \rightarrow подверженный \rightarrow инфицированный \rightarrow восстановившийся) (SEIR) и транспортной модели: она базировалась на делении Португалии на регионы, между которыми вводились 4 типа перемещений, связанные с различными типами социальных активностей: движения в пределах квартала; региона; между соседними регионами и между удаленными регионами. Перемещение агента на каждом шаге по времени определялось взвешенной суммой всевозможных перемещений между регионами с весами, равными вероятностям этих перемещений. Такой подход учитывает пространственное распределение популяции и ее перемещения, что позволило получить результаты моде-

лирования эпидемии паротита близкие к реальным.

Раковски и др. (2010) [33] при построении АВМ распространения гриппа в Польше использовали простые правила транспортировки для имитации поездок людей и позволили в модели инфекции передаваться непосредственно в транспорте. Авторы выбрали 38 крупнейших городов Польши и на основе их построили граф транспортной сети, в котором веса ребер были пропорциональны расстоянию между городами и обратно пропорциональны размерам их популяций. С помощью алгоритма Дейкстры находился кратчайший маршрут между любыми двумя городами, а список промежуточных городов на этом пути заносился в матрицу S . На каждом шаге модели выбирались случайные агенты, отправляемые в другой город; каждому агенту назначался город прибытия (вероятность выбора города пропорциональна его населению); определялись города, в которых делались пересадки; формировались группы путешественников размером до 10 человек. Эта модель, хоть и не учитывает точные маршруты, позволяет оценить их исключительно из карты распределения плотности населения, что сильно снижает требования к объемам необходимых для моделирования данных.

Фриас-Мартинез и др. (2011) [43] первыми использовали реальные данные сотовых операторов в АВМ для решения задачи моделирования транспортных потоков, ранее применялись только результаты опросов и переписей населения, которые не дают требуемого временного разрешения. Они оценили покрытия вышек операторов сотовой связи и на основе имеющихся данных создали модель перемещения пользователей, которая оценивала положение людей в каждый момент времени, и модель социальных связей, которая определяла наиболее близкие взаимоотношения. Помимо этого каждый агент подчинялся модели заболевания SEIR. Так авторы выяснили, что ограничение на передвижение людей в 2009 году в Мексике позволило снизить пиковое число зараженных H1N1 гриппом на 10% и отложить момент наступления этого пика на 2 дня.

Крукс и др. (2014) [44] моделировали вспышки холеры в лагере для беженцев Дадааб в Кении с помощью АВМ. Распорядок дня беженцев была внесена в поведение агентов и могла исполняться в специализированных для этого местах: школе, религиозном центре, магазине и так далее. Для этого карта Дадааба была оцифрована и поделена на участки по видам возможной там деятельности. Также каждый агент отличался личными характеристиками (пол, возраст), социальными связями (размер семьи, круга общения), уровнем иммунитета (симптоматический, асимптоматический), целями и приоритетами. На каждом шаге днем агент принимает решение, остаться там, где он находится, или переместиться, чтобы закрыть свои потребности (в воде, еде, школе), а ночью возвращается домой. Само распространение холеры определяется SEIR, где инфицированные могут быть как симптоматическими, так и асимптоматическими, а само заражение происходит из источников воды, а источники воды становятся заразными после того, как туда попали фекалии инфицированных. Благодаря полученной модели авторы смогли рассмотреть два сценария развития холеры: радиально от источника

загрязненной воды и через дождевые стоки, причем результаты качественно совпадают с реальными данными, например, для лагеря Дагахлей. Потенциально такая модель может стать частью системы раннего предупреждения вспышек холеры.

3.3 Агентная модель Covasim

В этой работе в дальнейшем речь будет идти о модели [COVID-19 Agent-based Simulator \(Covasim\)](#) и ее модернизации. [Covasim](#) — АВМ с открытым кодом, содержащая информацию о возрастной структуре и размере популяции, реалистичную систему передачи инфекции в разных социальных слоях (домашние хозяйства, школы, рабочие места, случайные контакты и др.), течение болезни, зависящее от возраста, иммунитет, определяемый уровнем антител. Также в данной модели реализовано множество различных эпидемиологических мер: социального дистанцирования, ношения масок, вакцинирования, тестирования и других.

Каждый запуск [Covasim](#) состоит из нескольких этапов. Сначала создается объект симуляции и загружаются все необходимые параметры. Затем в соответствии с данными о распределении возрастов создаются агенты и объединяются в сети контактов, параметры которой берутся из демографических данных (например, средний размер семьи, число случайных контактов и др.). Затем в цикле на каждом шаге по времени: производится масштабирование популяции (если необходимо для повышения производительности); обновляются состояния агентов, в том числе связанные с развитием инфекции; инфицируются случайные агенты; применяются эпидемиологические меры, вычисляются вероятности дальнейшего инфицирования по сети контактов; вычисляются результирующие метрики.

В [Covasim](#) реализована довольно сложная модель протекания болезни, в которой каждый агент может пребывать в одном из 9 состояний: восприимчивый, подвергшийся воздействию (инфицированный, но пока не инфекционный), инфекционный (разделяется по тяжести симптомов на пресимптоматический, бессимптомный, легкий, тяжелый, критический), восстановившийся и умерший. Длительности нахождения в каждой фазе заболевания определяются для каждого агента отдельно из логнормального распределения с параметрами, взятыми из исследований [Coronavirus Disease 2019 \(COVID-19\)](#) [45–54]. Вероятности приобретения симптомов, развития тяжелой, критической фазы или смерти различны для агентов в зависимости от их возраста, и эти данные также взяты из исследований [COVID-19](#) [52; 54–57]. В зависимости от слоя, которому принадлежит связь, вероятность передачи инфекции домножается на 0.05 для домохозяйства, на 0.01 для школ и рабочих мест, на 0.005 для случайных контактов, что согласуется с литературными данными [58; 59].

В зависимости от объемов имеющихся демографических данных для запуска [Covasim](#) можно использовать 3 способа генерации синтетической популяции: случайные сети, SynthPops и гибридные сети.

В отсутствие данных применяются случайные сети: всем агентам из Пуассонов-

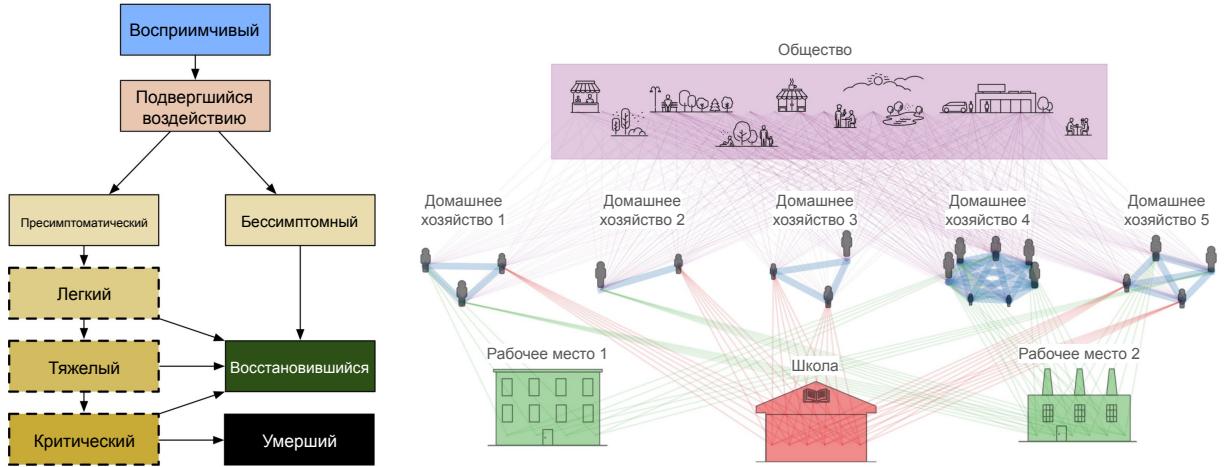


Рис. 2: Структура модели развития инфекции (слева) и иллюстрация сети контактов в Covasim. Источник: [60]

ского распределения определяются числа контактов, после чего в соответствии с этими значениями формируются произвольные связи между людьми.

SynthPops применяется при наличии информации о вероятностях взаимодействий между возрастными группами в домохозяйствах, школах, на рабочих местах, на улице; о распределении численностей людей в школах, числах набора школьников разных возрастов, отношении числа преподавателей к числу школьников; о распределении численностей рабочих мест и числах трудоустроенных людей разных возрастов; о распределении размеров домохозяйств, возрастов и полов членов семей.

При формировании связей в домохозяйствах SynthPops сначала выбирает размеры семей из их распределения, после чего за каждой семьей закрепляет главу семейства. Остальные члены домохозяйств набираются в соответствии с вероятностями контактов в семьях и распределением возрастов в них.

Контакты между учениками в школах формируются аналогично, собирая агентов по семьям (дети из одних домохозяйств в одних школах), после чего выбираются агенты-работники учебного заведения и в зависимости от размера школы либо создаются контакты между всеми школьниками и работниками, либо каждому работнику определяется число связей с учениками, которые затем формируются.

Связи на рабочих местах формируются аналогично из агентов случайных домохозяйств. Также после определения чисел случайных контактов они добавляются в SynthPops.

Для создания гибридной сети необходимы данные о распределении возрастов и размеров семей, однако они уже загружены в Covasim для каждой страны. Гибридный подход не учитывает распределение возрастов в домохозяйствах, все дети (все агенты от 6 до 22 лет) объединяются в школы, взрослые (агенты возрастом между 22 и 65) в рабочие места с предопределенными для них числами контактов в этих слоях из Пуассоновских распределений. Это позволяет воссоздать некоторую популяционную гетерогенность в случае недостатка демографических данных.

3.4 Дисперсионный анализ чувствительности математических моделей

В данной работе анализ чувствительности модели будет проводиться с помощью дисперсионных методов в связи с её стохастичностью и их информативностью, несмотря на большие вычислительные затраты: такие методы никак не зависят от природы модели и позволяют оценить влияние изменений любых комбинаций параметров [61].

Пусть задана квадратично интегрируемая функция f на Ω^k , единичном k -мерном гиперкубе,

$$\Omega^k = (X | 0 \leq x_i \leq 1; i = 1, \dots, k),$$

Илья Меерович Соболь предложил раскладывать эту функцию в сумму по возрастающим размерностям:

$$f = f_0 + \sum_i f_i + \sum_i \sum_{j > i} f_{ij} + \dots + f_{12\dots k},$$

где каждое слагаемое также является квадратично интегрируемым на области определения и является функцией лишь коэффициентов в его индексе, то есть $f_i = f_i(X_i)$, $f_{ij} = f_{ij}(X_i, X_j)$ и так далее. Соболь показал, что слагаемые суммы можно последовательно вычислить из результатов расчетов модели Y как

$$\begin{aligned} f_0 &= E(Y) \\ f_i &= E(Y|X_i) - E(Y) \\ f_{ij} &= E(Y|X_i, X_j) - f_i - f_j - E(Y) \end{aligned}$$

Оказывается, что дисперсия условного матожидания может быть использована как показатель чувствительности, на самом деле дисперсии слагаемых в приведенном выше разложении и есть искомые меры. К примеру, $V(f_i(X_i))$ есть $V[E(Y|X_i)]$, откуда индекс чувствительности первого порядка

$$S_i = \frac{V[E(Y|X_i)]}{V(Y)}.$$

Индекс первого порядка представляет собой основной вклад варьирования каждого параметра в дисперсию модели и был описан разными учеными как мера важности [62–66] и коэффициент корреляции [67].

Тогда же Соболь предложил аналогичное определение [68], основанное на корреляции между результатами модели Y и условным матожиданием $E(Y|X_i)$

$$S_i = \text{Corr}(Y, E(Y|X_i)).$$

С учетом того, что $V_i = V(f_i(X_i)) = V[E(Y|X_i)]$ выражение

$$V_{ij} = V(f_{ij}(X_i, X_j)) = V(E(Y|X_i, X_j)) - V(E(Y|X_i)) - V(E(Y|X_j))$$

позволяет оценить эффект совместного варьирования (X_i, X_j) на результат Y , который называют индексом чувствительности второго порядка [69]. Аналогичные формулы могут быть выписаны для индексов высших порядков, основываясь на ANalysis Of VAriance high-dimensional model representation (от англ. анализ дисперсии по многомерному представлению) (ANOVA-HDMR) разложении

$$V(Y) = \sum_i V_i + \sum_i \sum_{j>i} V_{ij} + \dots + V_{123\dots k}.$$

Деля обе части равенства на $V(Y)$, получаем

$$\sum_i S_i + \sum_i \sum_{j>i} S_{ij} + \dots + S_{123\dots k} = 1.$$

Также возможно оценить вклад варьирования параметра и всех, связанных с ним, взаимодействий вычислением полного индекса Соболя

$$S_{T_i} = \frac{E[V(Y|\mathbf{X}_{\sim i})]}{V(Y)} = 1 - \frac{V[E(Y|\mathbf{X}_{\sim i})]}{V(Y)}.$$

3.5 Калибровка моделей

4 Материалы и методы

4.1 Используемые программные пакеты

4.2 Вычисление индексов Соболя

Вычисление индексов Соболя производилось с помощью модуля SALib [70; 71] по следующему алгоритму [61]:

- Создается матрица случайных параметров k размера $(N, 2k)$, определяющая матрицы A и B . Параметры лучше брать, используя последовательности псевдослучайных чисел для более равномерного заполнения пространства параметров [72; 73]

$$A = \begin{bmatrix} x_1^{(1)} & x_2^{(1)} & \dots & x_i^{(1)} & \dots & x_k^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \dots & x_i^{(2)} & \dots & x_k^{(2)} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_1^{(N-1)} & x_2^{(N-1)} & \dots & x_i^{(N-1)} & \dots & x_k^{(N-1)} \\ x_1^{(N)} & x_2^{(N)} & \dots & x_i^{(N)} & \dots & x_k^{(N)} \end{bmatrix},$$

$$B = \begin{bmatrix} x_{k+1}^{(1)} & x_{k+2}^{(1)} & \cdots & x_{k+i}^{(1)} & \cdots & x_{2k}^{(1)} \\ x_{k+1}^{(2)} & x_{k+2}^{(2)} & \cdots & x_{k+i}^{(2)} & \cdots & x_{2k}^{(2)} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{k+1}^{(N-1)} & x_{k+2}^{(N-1)} & \cdots & x_{k+i}^{(N-1)} & \cdots & x_{2k}^{(N-1)} \\ x_{k+1}^{(N)} & x_{k+2}^{(N)} & \cdots & x_{k+i}^{(N)} & \cdots & x_{2k}^{(N)} \end{bmatrix}.$$

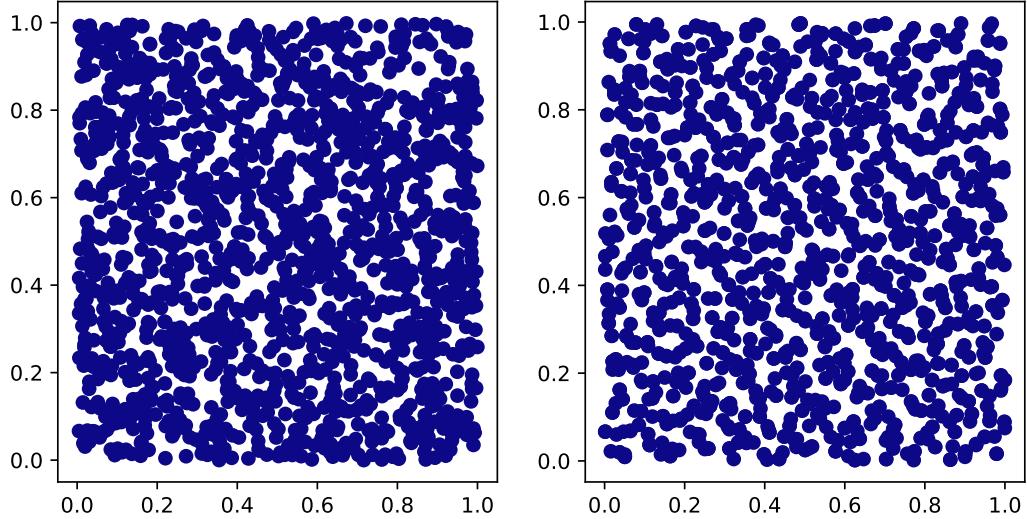


Рис. 3: Наборы точек из двумерного пространства, взятые из равномерного случайного распределения (слева) и последовательности псевдослучайных чисел Соболя (справа)

- Столбцы матрицы C_i есть столбцы B за исключением i -го, который берется из A :

$$C_i = \begin{bmatrix} x_{k+1}^{(1)} & x_{k+2}^{(1)} & \cdots & x_i^{(1)} & \cdots & x_{2k}^{(1)} \\ x_{k+1}^{(2)} & x_{k+2}^{(2)} & \cdots & x_i^{(2)} & \cdots & x_{2k}^{(2)} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{k+1}^{(N-1)} & x_{k+2}^{(N-1)} & \cdots & x_i^{(N-1)} & \cdots & x_{2k}^{(N-1)} \\ x_{k+1}^{(N)} & x_{k+2}^{(N)} & \cdots & x_i^{(N)} & \cdots & x_{2k}^{(N)} \end{bmatrix}.$$

- Вычисляется вывод модели при всех наборах параметров из матриц A, B, C_i , то есть получаются три вектора размерности $N \times 1$:

$$y_A = f(A) \quad y_B = f(B) \quad y_{C_i} = f(C_i).$$

- Вычисляется индекс Соболя первого порядка

$$S_i = \frac{V[E(Y|X_i)]}{V(Y)} = \frac{y_A \cdot y_{C_i} - f_0^2}{y_A \cdot y_A - f_0^2} = \frac{(1/N) \sum_{j=1}^N y_A^{(j)} y_{C_i}^{(j)} - f_0^2}{(1/N) \sum_{j=1}^N (y_A^{(j)})^2 - f_0^2},$$

где

$$f_0^2 = \left(\frac{1}{N} \sum_{j=1}^N y_A^{(j)} \right)^2.$$

Аналогично полный индекс Соболя

$$S_{T_i} = 1 - \frac{V[E(Y|\mathbf{X}_{\sim i})]}{V(Y)} = 1 - \frac{y_B \cdot y_{C_i} - f_0^2}{y_A \cdot y_A - f_0^2} = 1 - \frac{(1/N) \sum_{j=1}^N y_B^{(j)} y_{C_i}^{(j)} - f_0^2}{(1/N) \sum_{j=1}^N (y_A^{(j)})^2 - f_0^2}.$$

Преимуществом такого подхода является тот факт, что для вычисления индексов по k параметрам требуется только $N(k + 2)$ запусков симуляции в отличие от полного перебора N^2 точек пространства параметров. Метод позволяет вычислить только полные индексы и индексы первого порядка, но они являются самыми информативными, поэтому мы ограничимся ими в данной работе [61].

4.3 Исследуемая агентная модель

5 Результаты

5.1 Город и сателлиты

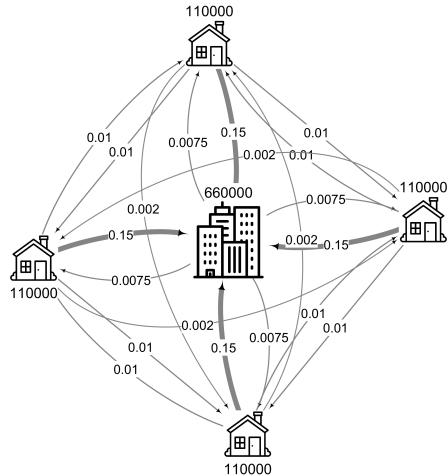


Рис. 4: Представление рассматриваемой модели транспортных потоков в системе большой город и сателлиты в виде графа (слева).

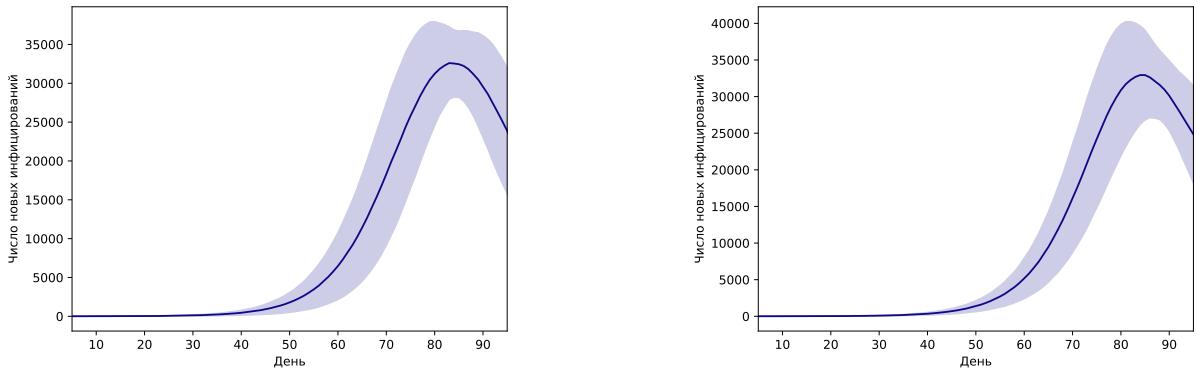


Рис. 5: Суммарное число новых ежедневных инфицирований в системе при моделировании в отсутствие ограничительных мер при начале эпидемии в большом городе (слева) и в спутнике (справа).

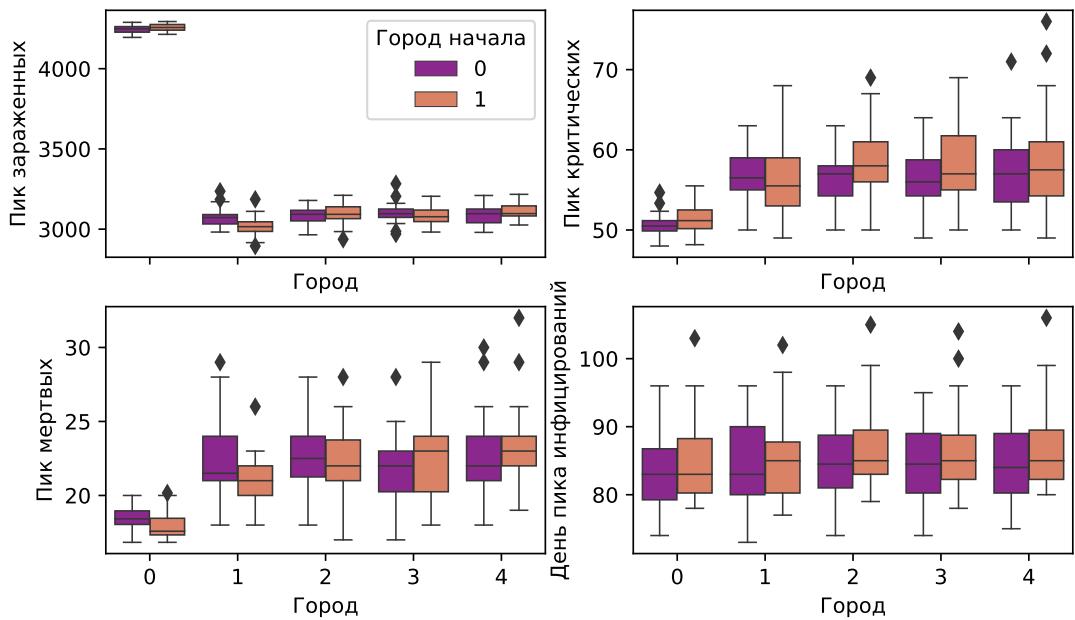


Рис. 6: Исследуемые метрики в различных городах системы в зависимости от места начала эпидемии (ограничительные меры отсутствуют)

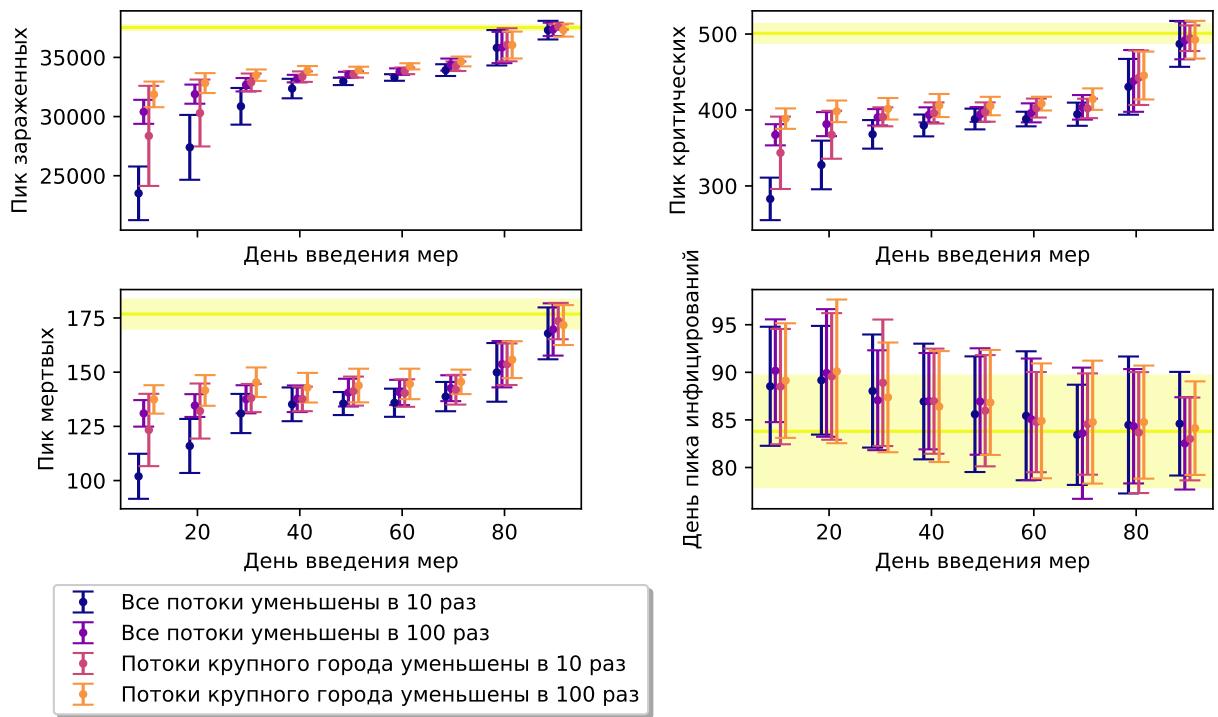


Рис. 7: Зависимости исследуемых метрик системы от дня введения ограничений на транспортные потоки и величины этих ограничений при начале эпидемии в центральном городе.

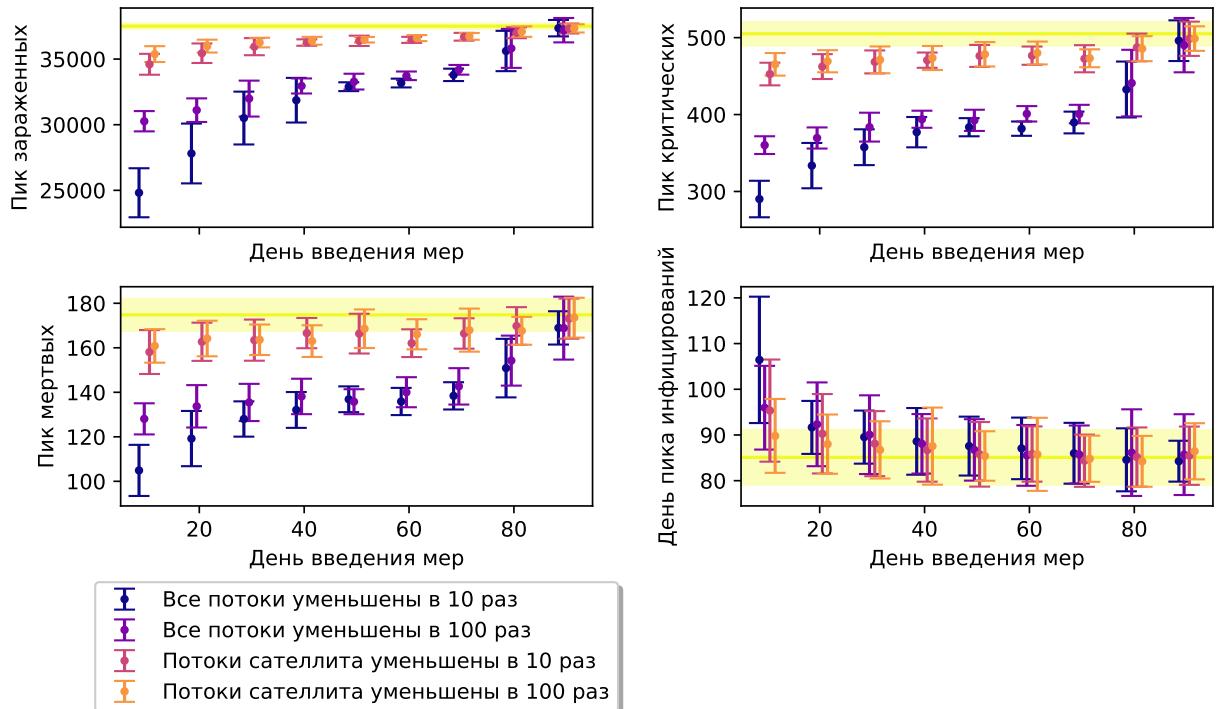


Рис. 8: Зависимости исследуемых метрик системы от дня введения ограничений на транспортные потоки и величины этих ограничений при начале эпидемии в город-сателлите.

[74]

5.2 Базовые потоки

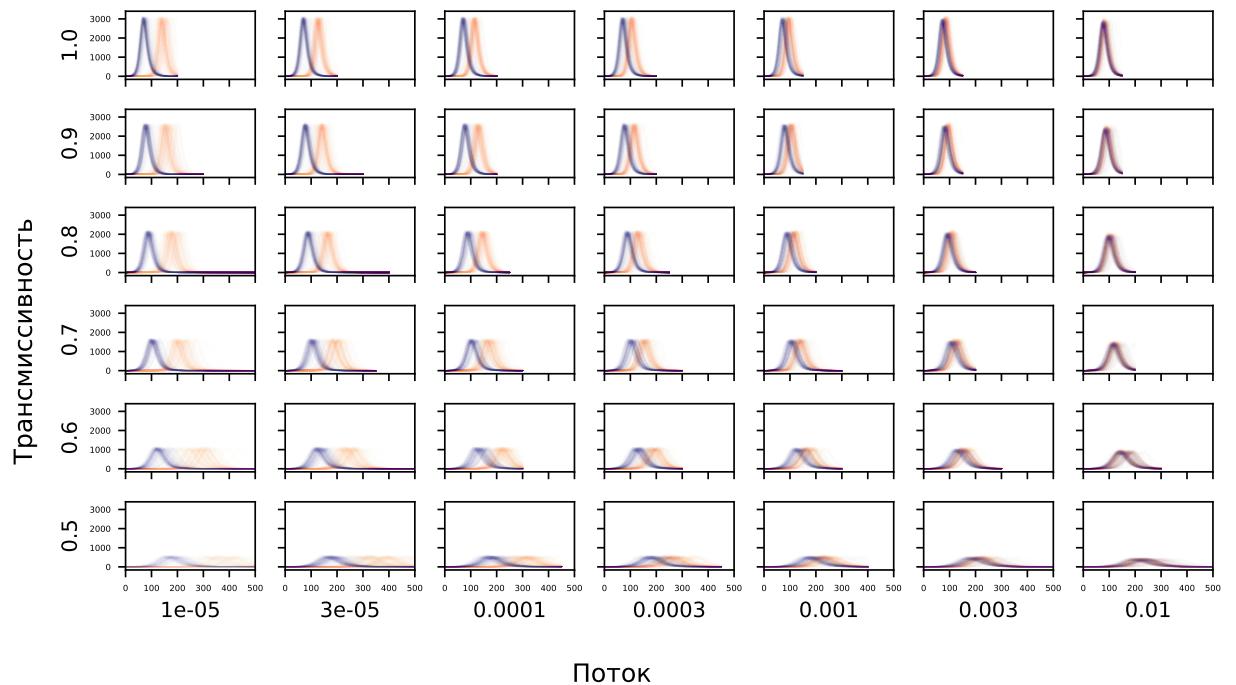


Рис. 9: Эпидемиологические кривые по результатам моделирования двух городов при разных трансмиссивностях инфекции и пассажиропотоках (на каждом графике 150 экспериментов).

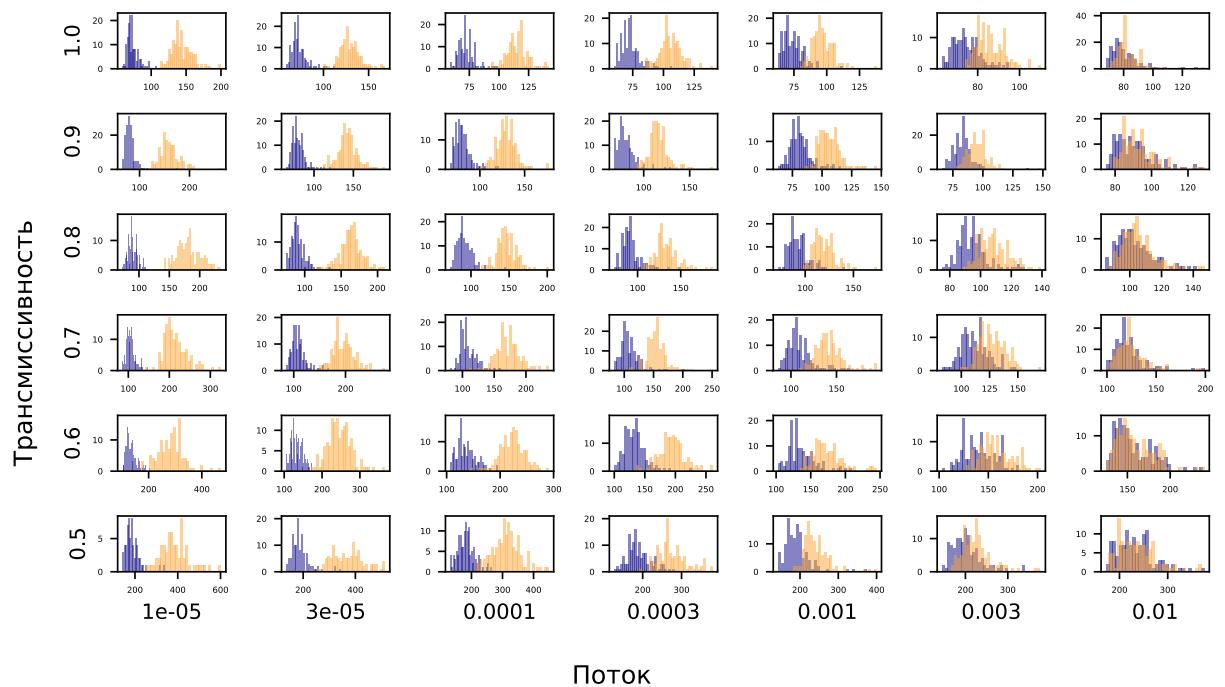


Рис. 10: Гистограммы распределения дня пика инфицирований для двух городов (синяя соответствует городу начала эпидемии, оранжевая — второму городу).

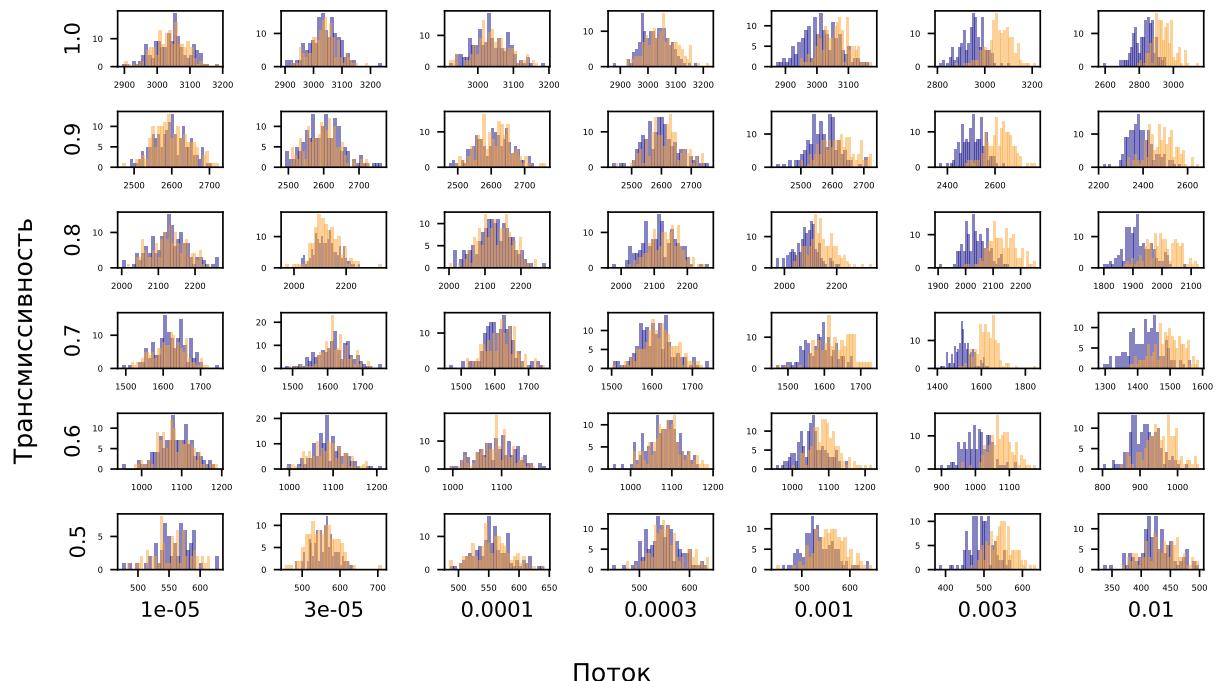


Рис. 11: Гистограммы распределения пикового числа инфицирований для двух городов (синяя соответствует городу начала эпидемии, оранжевая — второму городу).

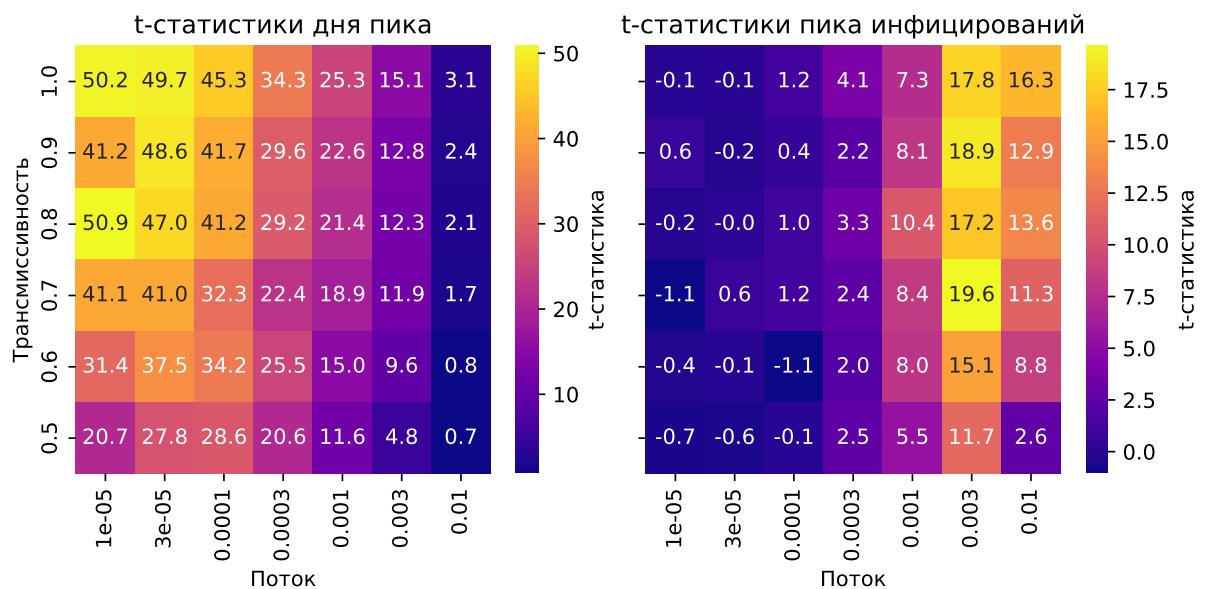


Рис. 12: Тепловые карты t-статистик при сравнении метрик для двух городов (слева — день пика инфицирований, справа — пиковое число инфицирований).

5.3 Индексы Соболя

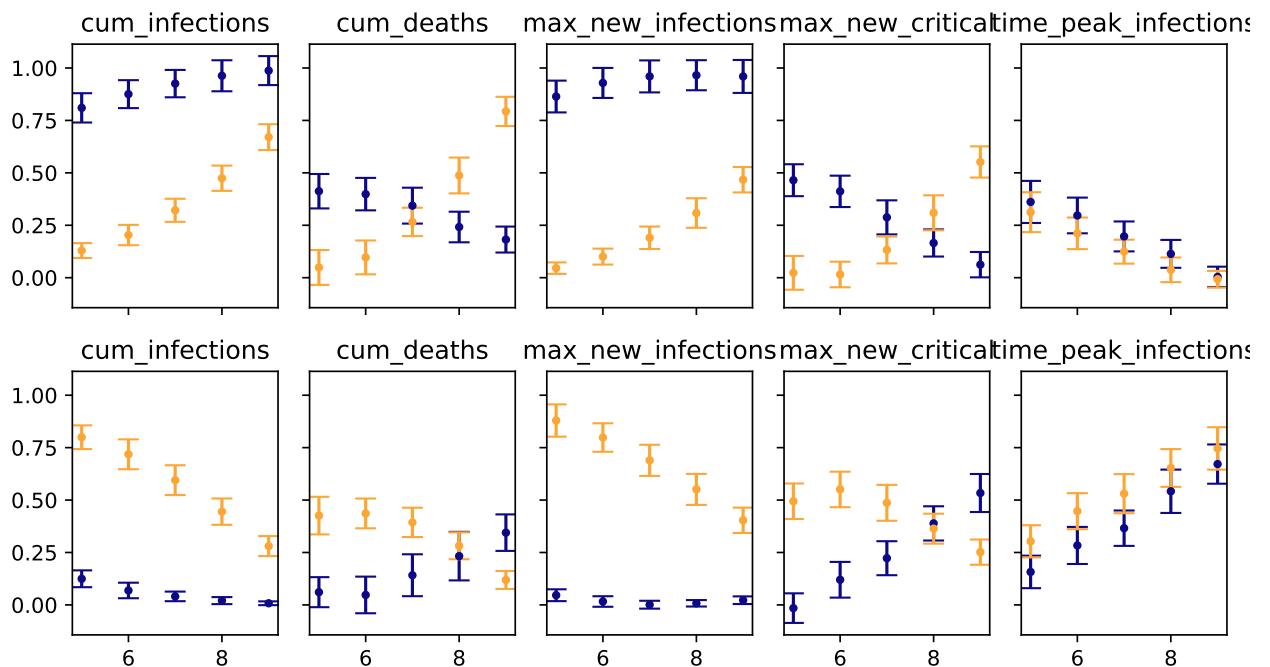


Рис. 13

6 Выводы

7 Благодарности

8 Список литературы

Список литературы

1. Hethcote H. W. The mathematics of infectious diseases // SIAM review. — 2000. — Т. 42, № 4. — С. 599—653.
2. Hethcote H. W. Three basic epidemiological models // Applied mathematical ecology. — Springer, 1989. — С. 119—144.
3. Hethcote H., Van Ark J. Transmission and AIDS in the United States, Lect // Notes Biomath. — 1992. — Т. 95.
4. Keeling M. J., Eames K. T. Networks and epidemic models // Journal of The Royal Society Interface. — 2005. — Т. 2, № 4. — С. 295—307.
5. Kermack W. O., McKendrick A. G. A contribution to the mathematical theory of epidemics // Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character. — 1927. — Т. 115, № 772. — С. 700—721.

6. *Bailey N. T.* The mathematical theory of epidemics // (No Title). — 1957.
7. *Anderson R. M., Robert M. May.* Infectious diseases of humans: dynamics and control. — 1992.
8. *Grenfell B.* Chance and chaos in measles dynamics // Journal of the Royal Statistical Society: Series B (Methodological). — 1992. — Т. 54, № 2. — С. 383—398.
9. *Rohani P., Earn D. J., Grenfell B. T.* Impact of immunisation on pertussis transmission in England and Wales // The Lancet. — 2000. — Т. 355, № 9200. — С. 285—286.
10. *Hethcote H., Yorke J.* Springer Lecture Notes in Biomathematics // Lecture Notes in Biomathematics. — 1984.
11. *Garnett G. P., Anderson R. M.* Sexually transmitted diseases and sexual behavior: insights from mathematical models // Journal of Infectious Diseases. — 1996. — Т. 174, Supplement_2. — S150—S161.
12. *Anderson R. M.* The epidemiology of HIV infection: variable incubation plus infectious periods and heterogeneity in sexual activity // Journal of the Royal Statistical Society: Series A (Statistics in Society). — 1988. — Т. 151, № 1. — С. 66—93.
13. *Grenfell B. T., Bjørnstad O. N., Kappey J.* Travelling waves and spatial hierarchies in measles epidemics // Nature. — 2001. — Т. 414, № 6865. — С. 716—723.
14. *Ghani A. C., Swinton J., Garnett G. P.* The role of sexual partnership networks in the epidemiology of gonorrhea // Sexually transmitted diseases. — 1997. — Т. 24, № 1. — С. 45—56.
15. *Keeling M. J.* Modelling the persistence of measles // Trends in microbiology. — 1997. — Т. 5, № 12. — С. 513—518.
16. *Pastor-Satorras R., Vespignani A.* Epidemic spreading in scale-free networks // Physical review letters. — 2001. — Т. 86, № 14. — С. 3200.
17. Mean-field models for non-Markovian epidemics on networks / N. Sherborne [и др.] // Journal of mathematical biology. — 2018. — Т. 76. — С. 755—778.
18. *Keeling M. J.* The effects of local spatial structure on epidemiological invasions // Proceedings of the Royal Society of London. Series B: Biological Sciences. — 1999. — Т. 266, № 1421. — С. 859—867.
19. *House T., Keeling M. J.* Insights from unifying modern approximations to infections on networks // Journal of The Royal Society Interface. — 2011. — Т. 8, № 54. — С. 67—73.
20. *Karrer B., Newman M. E.* Message passing approach for general epidemic models // Physical Review E—Statistical, Nonlinear, and Soft Matter Physics. — 2010. — Т. 82, № 1. — С. 016101.
21. Effective degree network disease models / J. Lindquist [и др.] // Journal of mathematical biology. — 2011. — Т. 62. — С. 143—164.

22. *Miller J. C., Slim A. C., Volz E. M.* Edge-based compartmental modelling for infectious disease spread // *Journal of the Royal Society Interface*. — 2012. — Т. 9, № 70. — С. 890—906.
23. *Volz E.* SIR dynamics in random networks with heterogeneous connectivity // *Journal of mathematical biology*. — 2008. — Т. 56. — С. 293—310.
24. *Aiélo O. E., Silva M. A. da.* New approach to dynamical Monte Carlo methods: application to an epidemic model // *Physica A: Statistical Mechanics and its Applications*. — 2003. — Т. 327, № 3/4. — С. 525—534.
25. *Haas V. J., Caliri A., Da Silva M.* Temporal duration and event size distribution at the epidemic threshold // *Journal of Biological Physics*. — 1999. — Т. 25. — С. 309—324.
26. Efficient method for comprehensive computation of agent-level epidemic dissemination in networks / G. M. Nakamura [и др.] // *Scientific reports*. — 2017. — Т. 7, № 1. — С. 40885.
27. *Albert R., Barabási A.-L.* Statistical mechanics of complex networks // *Reviews of modern physics*. — 2002. — Т. 74, № 1. — С. 47.
28. Epidemic processes in complex networks / R. Pastor-Satorras [и др.] // *Reviews of modern physics*. — 2015. — Т. 87, № 3. — С. 925—979.
29. *Van Kampen N. G.* Stochastic processes in physics and chemistry. Т. 1. — Elsevier, 1992.
30. Herd immunity: basic concept and relevance to public health immunization practices / J. P. Fox [и др.] // *American journal of epidemiology*. — 1971. — Т. 94, № 3. — С. 179—189.
31. An influmza simulation model for immunization studies / L. R. Elveback [и др.] // *American journal of epidemiology*. — 1976. — Т. 103, № 2. — С. 152—165.
32. *Koopman J.* Controlling smallpox // *Science*. — 2002. — Т. 298, № 5597. — С. 1342—1344.
33. Influenza epidemic spread simulation for Poland—a large scale, individual based model study / F. Rakowski [и др.] // *Physica A: Statistical Mechanics and its Applications*. — 2010. — Т. 389, № 16. — С. 3149—3165.
34. *Li J., Xiang T., He L.* Modeling epidemic spread in transportation networks: A review // *Journal of Traffic and Transportation Engineering (English Edition)*. — 2021. — Т. 8, № 2. — С. 139—152.
35. *Hyman J. M., LaForce T.* Modeling the spread of influenza among cities // *Bioterrorism: Mathematical modeling applications in homeland security*. — SIAM, 2003. — С. 211—236.
36. *Wang W., Zhao X.-Q.* An epidemic model in a patchy environment // *Mathematical biosciences*. — 2004. — Т. 190, № 1. — С. 97—112.

37. Spreading disease with transport-related infection / Y. Takeuchi, Y. Saito [и др.] // Journal of theoretical biology. — 2006. — Т. 239, № 3. — С. 376—390.
38. Liu X., Takeuchi Y. Spread of disease with transport-related infection and entry screening // Journal of Theoretical Biology. — 2006. — Т. 242, № 2. — С. 517—528.
39. An SEIS epidemic model with transport-related infection / H. Wan [и др.] // Journal of theoretical biology. — 2007. — Т. 247, № 3. — С. 507—524.
40. Liu J., Zhou Y. Global stability of an SIRS epidemic model with transport-related infection // Chaos, Solitons & Fractals. — 2009. — Т. 40, № 1. — С. 145—158.
41. Li J., Zou X. Dynamics of an epidemic model with non-local infections for diseases with latency over a patchy environment // Journal of mathematical biology. — 2010. — Т. 60, № 5. — С. 645—686.
42. Simoes J. Modelling a mumps outbreak through spatially explicit agents // Potentials of Complexity Science for Business, Governments, and the Media. — 2006.
43. Frias-Martinez E., Williamson G., Frias-Martinez V. An agent-based model of epidemic spread using human mobility and social network information // 2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing. — IEEE. 2011. — С. 57—64.
44. Crooks A. T., Hailegiorgis A. B. An agent-based modeling approach applied to the spread of cholera // Environmental Modelling & Software. — 2014. — Т. 62. — С. 164—177.
45. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application / S. A. Lauer [и др.] // Annals of internal medicine. — 2020. — Т. 172, № 9. — С. 577—582.
46. Serial interval of COVID-19 among publicly reported confirmed cases / Z. Du [и др.] // Emerging infectious diseases. — 2020. — Т. 26, № 6. — С. 1341.
47. Nishiura H., Linton N. M., Akhmetzhanov A. R. Serial interval of novel coronavirus (COVID-19) infections // International journal of infectious diseases. — 2020. — Т. 93. — С. 284—286.
48. Investigation of three clusters of COVID-19 in Singapore: implications for surveillance and response measures / R. Pung [и др.] // The Lancet. — 2020. — Т. 395, № 10229. — С. 1039—1046.
49. Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data / N. M. Linton [и др.] // Journal of clinical medicine. — 2020. — Т. 9, № 2. — С. 538.
50. Temporal dynamics in viral shedding and transmissibility of COVID-19 / X. He [и др.] // Nature medicine. — 2020. — Т. 26, № 5. — С. 672—675.

51. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China / D. Wang [и др.] // jama. — 2020. — Т. 323, № 11. — С. 1061—1069.
52. Clinical progression of patients with COVID-19 in Shanghai, China / J. Chen [и др.] // Journal of infection. — 2020. — Т. 80, № 5. — e1—e6.
53. Estimates of the severity of coronavirus disease 2019: a model-based analysis / R. Verity [и др.] // The Lancet infectious diseases. — 2020. — Т. 20, № 6. — С. 669—677.
54. Virological assessment of hospitalized patients with COVID-2019 / R. Wölfel [и др.] // Nature. — 2020. — Т. 581, № 7809. — С. 465—469.
55. Age-specific mortality and immunity patterns of SARS-CoV-2 / M. O'Driscoll [и др.] // Nature. — 2021. — Т. 590, № 7844. — С. 140—145.
56. Report 34: COVID-19 Infection Fatality Ratio: Estimates from Seroprevalence / L. W. Baguelin [и др.] // See paper. — 2020. — Т. 201029.
57. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Imperial College COVID-19 Response Team / N. M. Ferguson [и др.] // Imperial College COVID-19 Response Team. — 2020. — Т. 20, № 10.25561. — С. 77482.
58. Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China / J. Zhang [и др.] // Science. — 2020. — Т. 368, № 6498. — С. 1481—1486.
59. *Lader D., Short S., Gershuny J.* The time use survey, 2005 // Office for National Statistics, London. — 2006.
60. Covasim: an agent-based model of COVID-19 dynamics and interventions / C. C. Kerr [и др.] // PLOS Computational Biology. — 2021. — Т. 17, № 7. — e1009149.
61. Global sensitivity analysis: the primer / A. Saltelli [и др.]. — John Wiley & Sons, 2008.
62. *Hora S., Iman R.* A comparison of maximum/bounding and Bayesian/Monte Carlo for fault tree uncertainty analysis // Report SAND85-2839, Sandia Laboratories. — 1986.
63. *Ishigami T., Homma T.* An importance quantification technique in uncertainty analysis for computer models // [1990] Proceedings. First international symposium on uncertainty modeling and analysis. — IEEE. 1990. — С. 398—403.
64. *Iman R. L., Hora S. C.* A robust measure of uncertainty importance for use in fault tree system analysis // Risk analysis. — 1990. — Т. 10, № 3. — С. 401—406.
65. *Saltelli A., Andres T., Homma T.* Sensitivity analysis of model output: An investigation of new techniques // Computational statistics & data analysis. — 1993. — Т. 15, № 2. — С. 211—238.
66. *Homma T., Saltelli A.* Importance measures in global sensitivity analysis of nonlinear models // Reliability Engineering & System Safety. — 1996. — Т. 52, № 1. — С. 1—17.

67. *McKay M.* Variance-based methods for assessing uncertainty importance in NUREG-1150 analyses // Los Alamos National Laboratory LA-UR-96-2695. — 1996.
68. *Sobol' I. M.* On the “freezing” of inessential variables // Vestnik Moskovskogo Universiteta. Seriya 1. Matematika. Mekhanika. — 1996. — № 6. — C. 92—94.
69. Statistics for experimenters. T. 664 / G. E. Box, W. H. Hunter, S. Hunter [и др.]. — John Wiley, sons New York, 1978.
70. *Iwanaga T., Usher W., Herman J.* Toward SALib 2.0: Advancing the accessibility and interpretability of global sensitivity analyses // Socio-Environmental Systems Modelling. — 2022. — Май. — Т. 4. — С. 18155. — DOI: [10.18174/sesmo.18155](https://doi.org/10.18174/sesmo.18155). — URL: <https://sesmo.org/article/view/18155>.
71. *Herman J., Usher W.* SALib: An open-source Python library for Sensitivity Analysis // The Journal of Open Source Software. — 2017. — Янв. — Т. 2, № 9. — DOI: [10.21105/joss.00097](https://doi.org/10.21105/joss.00097). — URL: <https://doi.org/10.21105/joss.00097>.
72. *Sobol' I. M.* The distribution of points in a cube and the approximate evaluation of integrals // USSR Computational mathematics and mathematical physics. — 1967. — Т. 7. — С. 86—112.
73. *Sobol' I.* Uniformly distributed sequences with additional uniformity properties // USSR Computational mathematics and mathematical physics. — 1976. — Т. 16, № 5. — С. 1332—1337.
74. *Makhrova A. G., Kirillov P. L., Bochkarev A. N.* Work commuting of the population in the Moscow agglomeration: estimating commuting flows using mobile operator data // Regional Research of Russia. — 2017. — Т. 7. — С. 36—44.