

САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

**ЮЖАКОВА Алина Сергеевна**

Выпускная квалификационная работа

**Моделирование распространения вирусной инфекции**

Уровень образования: магистратура

Направление *01.04.02 «Прикладная математика и информатика»*

Основная образовательная программа: *ВМ.5691 «Прикладная математика и информатика в задачах медицинской диагностики»*

Научный руководитель:

д-р. физ.-мат. наук, профессор  
кафедры теории управления

Котина Елена Дмитриевна

Рецензент

к.ф.-м.н., доцент

кафедра медицинской биофизики СПбГПМУ

Елизарова Марина Владиславовна

Санкт-Петербург

2021 год

# ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	3
ГЛАВА 1. МОДЕЛИРОВАНИЕ РАСПРОСТРАНЕНИЯ ЭПИДЕМИИ. 6	
1.1. Общие SIR-модели распространения эпидемии.....	7
1.2. Развитие моделирования эпидемий гриппа на базе SIR-модели	13
1.3. Постановка задачи .....	19
1.4. Вариация и градиент функционала .....	21
1.5. Алгоритм определения параметров модели и начальных данных в задаче SIR-моделирования .....	23
ГЛАВА 2. ПРАКТИЧЕСКАЯ РЕАЛИЗАЦИЯ МОДЕЛИ РАСПРОСТРАНЕНИЯ ЭПИДЕМИИ.....	25
2.1. Моделирование распространения вирусной инфекции гриппа через SIR-модель с прямым выбором параметров .....	25
2.2. SIR-модель с выбором параметров на основе оптимизационного алгоритма .....	28
ЗАКЛЮЧЕНИЕ .....	36
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ .....	37

# ВВЕДЕНИЕ

**Актуальность** темы. В современном мире значительное воздействие на распространение болезней оказывают глобализация поездок и торговли, бесплановая урбанизация и такие экологические проблемы, как изменение климата. Для предсказания динамики болезней, оценки угроз и выбора мер по контролю заболеваемости возникает необходимость в математическом моделировании процессов, происходящих во время эпидемий.

Методы моделирования инфекционной заболеваемости активно развиваются с начала XX века. В последние годы число работ на эту тему стремительно растет благодаря развертыванию информационных систем надзора и появлению больших объемов статистики, доступной для анализа. Эпидемиологические прогнозы выполняются для различных сроков и в зависимости от них служат разным целям.

Чтобы предсказать охват и длительность эпидемии, учёные прибегают к моделированию передачи вируса в обществе. Модели могут быть в разной степени детальными. Некоторые из них описывают только заражение и выздоровление: если кто-то переносит инфекцию, то определённая доля людей без иммунитета заразится, доля заразившихся — выздоровеет. Другие модели учитывают дополнительные факторы, такие как иммунитет, приобретённый через вакцинацию. Разумеется, эту последнюю корректировку можно включить в модель, только если вакцина от вируса существует. Поэтому детальность модели напрямую зависит от инфекции, распространение которой она призвана отражать.

Работы по моделированию важных в медицинском отношении и социально значимых процессов предпринимаются давно, а в последние годы активизировались в связи с усложнением эпидемиологической ситуации, что потребовало более рационального использования финансовых средств в медицине и социальной сфере, в том числе и на профилактические мероприятия.

Для анализа и моделирования развития эпидемической ситуации в широко используется в практическом использовании модель SIR, основанная на разделении населения на три группы: S – восприимчивые (Susceptible), I – инфицированные (Infectious) и R – имеющие иммунитет (Removed):  $N = S + I + R$ , где N – общая численность населения.

**Целью** настоящей работы является построение методики выбора начальных данных и параметров системы для моделирования распространения эпидемии вирусной инфекции на примере эпидемии гриппа в Российской Федерации.

Исходя из поставленной цели работы были введены следующие **задачи**:

1. Проанализировать существующие расширения SIR-модели;
2. Разработать алгоритм выбора параметров модели и начальных данных для системы для моделирования распространения вирусной инфекции;
3. Проанализировать распространения вирусной инфекции гриппа на территории Российской Федерации;
4. Провести моделирование распространения вирусной инфекции гриппа на территории Российской Федерации;
5. Решить задачу выбора начальных данных для SIR-модели распространения вирусной инфекции гриппа.

В настоящей работе предлагается определение параметров SIR-модели распространения эпидемии на примере моделирования эпидемии гриппа, а также нахождения начальных условий числа реципиентов, восприимчивых к инфекции, т.к. на первом этапе построения берутся приблизительные данные распространения инфекции по уже зараженным реципиентам, что может давать неточную картину распространения заболевания.

В большинстве изученных работ по теме значение параметра  $I_0$  (число инфицированных в момент начала эпидемии) берется как приблизительно известное, а значение параметра  $S_0$  (начальное число чувствительных) зачастую неопределенно. При этом рассматривается применимость

дальнейшего совершенствования математической модели, т.е. в частности, авторам необходим более надежный и объективный метод задания параметров модели. Данная работа позволяет получить модель распространения эпидемии с выбором коэффициентов модели, а также оценить выбор начальных параметров эпидемии для построения SIR-модели распространения вирусной инфекции.

**Основное содержание работы.** В первой главе содержится общий обзор литературы и постановка задачи. Приводятся постановка задачи и схема предлагаемого решения, а также анализ существующих работ. Во второй главе рассматривается модель распространения эпидемии. Ставится и решается задача моделирования эпидемии распространения гриппа с применением параметров в SIR-модели. В заключении сделаны выводы по работе и рассмотрены направления дальнейших исследований.

# ГЛАВА 1. МОДЕЛИРОВАНИЕ РАСПРОСТРАНЕНИЯ ЭПИДЕМИИ

Острые респираторные заболевания, или ОРВИ являются группой одних из наиболее широко распространенных и известных инфекционных заболеваний человека. Грипп, являющийся одним из самых опасных, имеет способность провоцировать эпидемические вспышки заболевания с повторяемостью, при которых наблюдается значительное превышение среднего сезонного уровня суммарной заболеваемости гриппа и фоновых ОРВИ. Ежегодно в мире отмечают порядка 3-5 миллионов тяжелой формы протекания заболеваний гриппа<sup>1</sup>.

Согласно данным ВОЗ<sup>2</sup>, включающим в статистику также последствия инфарктов и инсультов, то уровень смертности от последствий гриппа составляет 250-500 тыс. индивидов в год. Даже если рассматривать в разрезе эпидемии гриппа, то только порядка 15-20% случаев заболеваемости ОРВИ можно отнести собственно к гриппу, и поставить точный диагноз. Отличие гриппа от иных заболеваний верхних дыхательных путей, которые имеют схожую симптоматику, можно провести только с помощью лабораторных исследований.

Математическая эпидемиология сегодня дает исследователям различные методы планирования мер для сдерживания гриппа. Прогнозирование является необходимой процедурой, которая даст возможность принятия решения о предотвращении или выходе из критической ситуации (в данном случае - эпидемии).

При моделировании эпидемии зачастую применяется долгосрочное прогнозирование на год и более, но достичь высокого качества таких

---

<sup>1</sup> Thompson, W.W. Estimates of US influenza-associated deaths made using four different methods / W.W. Thompson, E. Weintraub, P. Dhankhar, O.Y. Cheng, L. Brammer, M.I. Meltzer et al. // Influenza Other Respi Viruses. — 2009. — №3. — P. 37-49.

<sup>2</sup> Хартл, Г. Смертность от респираторных заболеваний, связанных с сезонным гриппом // Всемирная организация здравоохранения. — 2017.

прогнозов в большинстве случаев невозможно, тем не менее, они необходимы, например, при оценке необходимых объемов производства лекарств и вакцин, оснащении медицинских учреждений и обучении кадров.

Высокий уровень заболеваемости и смертности от гриппа во время эпидемии 2016 года привлекли внимание к изучению особенностей эпидемии гриппа на территории Российской Федерации на примере моделирования распространения эпидемии SIR-модели.

### 1.1. Общие SIR-модели распространения эпидемии

Модель SIR - это своего рода компартментальная модель, описывающая динамику инфекционных заболеваний. Возникает вопрос, почему модель SIR называется «компаратментальной моделью». В этой модели население делится на группы. Ожидается, что каждая группа будет иметь одинаковые характеристики и, получается, что SIR-модель представляет три ветви, сегментированные по модели<sup>3</sup>:

- **Susceptible** – здоровые особи, которые находятся в группе риска и могут подхватить инфекцию;
- **Infectious** – инфицированные особи; являются переносчиками инфекции;
- **Recovered** – «выбывшие» особи, к которым относятся выздоровевшие особи, приобретшие иммунитет к данной болезни, а также умершие.

«Восприимчивые» относятся к группе людей, наиболее уязвимых к заболеванию после контакта с инфицированными людьми. На момент заражения они могут уже быть больными или иметь какое-то хроническое заболевание. Группа «инфицированные» представляет уже зараженных инфекцией людей. Они могут передать болезнь группе восприимчивых людей

---

<sup>3</sup> Бароян, О.В. Прогнозирование эпидемий гриппа в условиях СССР / О.В. Бароян, Л.А. Рвачев // Вопросы вирусологии. — 1978. — № 2. — С. 131–137.

и могут со временем выздороветь. Выздоровевшие пациенты приобретают иммунитет, т.к. они больше не подвержены этому заболеванию.

Модель SIR – это структура, описывающая, как количество людей в каждой группе изменяется в течении времени (рис. 1).

Модель SIR дает возможность описания количества людей в каждой группе с помощью обыкновенного дифференциального уравнения. Параметр, который контролирует скорость передачи болезни при контакте, - это  $\beta$ . Данный параметр зависит от вероятности контакта и вероятности передачи болезни. Параметр, который выражает уровень излечения от болезни за определенный период, - это  $\gamma$ . Как только люди выздоравливают, они приобретают иммунитет. Они не могут повторно войти в категорию восприимчивых к болезни<sup>4</sup>.

Система дифференциальных уравнений, описывающих модель, выглядит следующим образом:

$$\begin{aligned}\frac{ds}{dt} &= -\beta IS, \\ \frac{dI}{dt} &= \beta IS - \gamma I, \\ \frac{dR}{dt} &= \gamma I.\end{aligned}\tag{1}$$

Относительно начальных условий, можно отметить, что зачастую изначально в популяции отсутствуют особи с иммунитетом к заболеванию, т. е.  $R(0) = 0$ . Систему уравнений (1) также называют системой Кермака-МакКендрика<sup>5</sup>.

Данная модель не принимает во внимание эффект от влияния естественной смертности или рождаемости, т.к. модель предполагает, что период, выдающийся на заболевание, намного короче, чем продолжительность жизни человека. Это позволяет понять важность знания двух параметров,  $\beta$  и  $\gamma$ . Теперь можно оценить данные значения, которые

---

<sup>4</sup> Бароян, О.В. Моделирование и прогнозирование эпидемий гриппа на территории СССР / О.В. Бароян, Л.А. Рвачев, Ю.Г. Иванников. — М., 1977. — 546 с.

<sup>5</sup> Kermack, W.O. Contributions to the mathematical theory of epidemics —I / W.O. Kermack, A.G. McKendrick // Bulletin of Mathematical Biology. — 1991. — V. 53. — P. 33-55.



привнесут некоторое понимание происходящего в системе. Если  $\gamma$  – параметр, отражающий скорость выздоровления, который можно описать как время проявления от начала первых симптомов болезни до полного выздоровления, то  $D = \frac{1}{\gamma}$ <sup>6</sup>.

Кроме того, можно оценить характер заболевания по силе распространения инфекции  $U_0 = \frac{\beta}{\gamma}$ , который называют индексом репродукции.  $U_0$  – это среднее число людей, которых заразил один человек. Увеличение данного параметра также напрямую влияет на уровень пандемии. Число также используется для оценки общего иммунного порога (НИТ). Если исходная распространенность, умноженная на процент восприимчивых людей, равна 1, это указывает на сбалансированное состояние системы. Соответственно, чем больше  $U_0$ , тем большей дисбаланс внесен в распространение эпидемии.

Численность населения берется постоянной. Если взять предположение о том, что доля людей, которые уже были больны, равна  $p$ , то устойчивое состояние можно сформулировать в следующем виде:

$$U_0(1 - p) = 1 \rightarrow 1 - p = \frac{1}{U_0} \rightarrow p_c = 1 - \frac{1}{U_0}. \quad (2)$$

Следовательно,  $p_c$  – это общий иммунный порог (НИТ), необходимый для остановки распространения инфекционных заболеваний. Повышение иммунного порога и остановка вспышки эпидемии возможны путем вакцинации населения для повышения общего иммунного порога.

Модель SIR изобилует всевозможными улучшениями и модификациями, поскольку считается одной из самых простых моделей эпидемии.

Так, существуют SIR-модели вида SIR-D, в которых фигурируют отдельно параметр D. Усложнение SIR-модели в SIR-D значит, что отдельно

---

<sup>6</sup> Smith, D. The SIR Model for Spread of Disease / D. Smith, L. Moore // Convergence. — 2004. — № 12.

рассматриваются переменные «Recovered» и «Death», т.е. выздоровевших в результате эпидемии реципиентов и умерших.

Развитием модели SIR стали, в частности, следующие модели:

- SIRS — «восприимчивые — инфицированные — выздоровевшие — восприимчивые»: модель описания динамики заболеваний с временным иммунитетом (выздоровевшие индивиды со временем снова становятся восприимчивыми);
- SEIR — «восприимчивые — контактные (Exposed) — инфицированные — выздоровевшие»: модель для описания распространения заболеваний с инкубационным периодом;
- SIS — «восприимчивые — инфицированные — восприимчивые»: модель для распространения заболевания, к которому не вырабатывается иммунитет;
- MSEIR — «наделенные иммунитетом от рождения (Maternally derived immunity) — восприимчивые — контактные — инфицированные — выздоровевшие»: модель, учитывающая иммунитет детей, приобретенный внутриутробно<sup>7</sup>.

Именно в соответствии с моделью SEIR развиваются по-настоящему опасные эпидемии, т.к. длительный инкубационный период может помешать своевременному выявлению болезни. В этом случае есть риск, что заболевание затронет значительное количество особей в популяции.

Заболевание в модели SEIR развивается по схеме «восприимчивые» — «контактные» — «инфицированные» — «выздоровевшие» и описывается системой уравнений:

$$\begin{aligned}\frac{dS}{dt} &= \mu N - \mu S - \beta \frac{I}{N} S, \\ \frac{dE}{dt} &= \beta \frac{I}{N} S - (\mu + a) E, \\ \frac{dI}{dt} &= a E - (\gamma + \mu) I,\end{aligned}\tag{3}$$

---

<sup>7</sup> Menon, A. Modelling and simulation of COVID-19 propagation in a large population with specific reference to India / A. Menon, N. K. Rajendran, A. Chandrachud, G. Setlur // medRxiv. — 2020. — № 5. — P. 25-28.

$$\frac{dR}{dt} = \gamma I - \mu R,$$

где:

- $\mu$  – уровень смертности;
- $\alpha$  – величина, обратная среднему инкубационному периоду заболевания;
- $E(t)$  – численность индивидов – носителей заболевания в момент времени  $t$ .

Модель SEIR включает в себя концепцию так называемого эпидемического перехода. То есть модель ведёт себя радикально по-разному в зависимости от параметра  $U_0$ . Каждый инфицированный человек может заразить некоторое количество здоровых людей.  $U_0$  показывает среднее число людей, которых один инфицированный может заразить в период болезни, то есть до полного выздоровления.

Если  $U_0$  меньше единицы, то наблюдается постепенное затухание эпидемии, но если  $U_0$  больше единицы, то заболевание распространяется в геометрической прогрессии. Как именно работает  $U_0$  зависит от конкретного рассматриваемого вируса, процента населения, которое становится невосприимчивым к нему (будь то с помощью вакцины или путём выявления вируса), и от мер, принимаемых населением для подавления эпидемии (например, карантин и социальное дистанцирование)<sup>8</sup>.

Модель SIS («восприимчивые — инфицированные — восприимчивые») применяется для анализа распространения заболеваний, к которым не вырабатывается иммунитет, к примеру, можно привести здесь грипп и ОРВИ. Модель SIS описывается следующей системой уравнений<sup>9</sup>:

---

<sup>8</sup> Mwalili, S. SEIR model for COVID-19 dynamics incorporating the environment and social distancing / S. Mwalili, M. Kimathi, V. Ojiambo, D. Gathungu, R. Mbogo // BMC Research Notes. — № 52(2020) — P. 1-5.

<sup>9</sup> Moehlis, J.M. Seasonal Epidemic Models: An SIS model // Mathematics Department, Macquarie University. — 2002. — №10(14).

$$\begin{aligned}\frac{dS}{dt} &= -\frac{\beta SI}{N} + \gamma I, \\ \frac{dI}{dt} &= \frac{\beta SI}{N} - \gamma I.\end{aligned}\tag{4}$$

Совместно первое и второе уравнение означают, что количество здоровых и больных реципиентов в сумме не меняется, а число заражений пропорционально числу контактов между инфицированными и восприимчивыми.

Модель MSEIR построена для заболеваний, которые включают инкубационный период, а также является моделью, учитывающей иммунитет детей, приобретенный внутриутробно. Данная модель является одной из самых сложных для анализа в силу наличия большого числа независимых параметров<sup>10</sup>.

Эта система уравнений отличается от ранее рассмотренных моделей тем, что в ней учитывается рождение детей, вероятность заражения которых со временем увеличивается, поскольку они теряют иммунитет, приобретенный в утробе матери. Данные зависимости описаны в первых двух уравнениях системы.

Приобретенный внутриутробный иммунитет может быть не у всех рожденных младенцев, но сто процентов малышей можно охватить вакцинацией. Введение этого параметра в математическую модель приводит к качественному изменению картины развития эпидемий<sup>11</sup>.

---

<sup>10</sup> Jason, B. How to Create an ARIMA Model for Time Series Forecasting in Python // Time Series. — 2020.

<sup>11</sup> Ковалев, С.В. Мультиагентное моделирование распространения эпидемий / С.В. Ковалев, Рюмин Н.Н., Ковалева О.А., Сидляр М.Ю., Хромова Т.А. // Вестник технологического университета. — 2021. — Т.24. — №1. — С. 91-97.

## 1.2. Развитие моделирования эпидемий гриппа на базе SIR-модели

Количество работ, посвященных моделированию инфекционной заболеваемости, быстро растет по мере появления статистических данных, позволяющих проводить анализ.

Все рассмотренные в обзоре литературы методы проиллюстрированы на примере моделирования заболеваемости гриппом.

Одной из наиболее ранних и широко используемых в настоящее время работ, посвященных моделированию эпидемии гриппа, является работа У. Кермака (W.O. Kermack) и А. МакКендрика (A. G. McKendrick) «A Contribution to the Mathematical Theory of Epidemics», написанная в 1927 г. Основные принципы моделирования эпидемии, которые были изложены в этой работе, продолжают успешно применяться и в наше время. Динамика эпидемического процесса в данной работе описывается с помощью системы дифференциальных уравнений, решения которых характеризуют динамику изменения численности в различных подгруппах населения<sup>12</sup>.

С одной стороны, грипп и ОРВИ занимают ведущее место среди инфекционных заболеваний человечества и, несмотря на все принимаемые противоэпидемические меры, их заболеваемость не имеет тенденции к снижению как в России, так и за рубежом. Поэтому многие модели были предложены специально для анализа распространения гриппа.

С другой стороны, методы прогнозирования гриппа подходят напрямую или могут быть легко адаптированы для других инфекций, передающихся воздушно-капельным путем и при бытовом контакте. Однако значительная

---

<sup>12</sup> Kermack, W.O. Contributions to the mathematical theory of epidemics —I / W.O. Kermack, A.G. McKendrick // Bulletin of Mathematical Biology. — 1991. — V. 53. — P. 33-55.

часть методов применима для прогнозирования даже неинфекционной заболеваемости<sup>13</sup>.

Существующие способы прогнозирования заболеваемости можно условно разделить на несколько категорий:

- статистические методы;
- методы на основе машинного обучения и прецедентов;
- методы на базе фильтрации;
- математическое моделирование распространения инфекции.

Это очень условное разделение, поскольку ряд способов находится на стыке введенных категорий, а некоторые приемы даже могут быть представлены как частный случай других методов. Тем не менее все они имеют самостоятельное значение и особенности практического применения.

Для большинства инфекций, а также для ОРВИ, можно отметить характерный сезонный рост заболеваемости, и часто аналогичные значения показателей повторяются из года в год. Выявленная сезонность учитывается при любом прогнозе от месяца. В простейшем случае отдельно рассматриваются данные за определенный календарный период (неделя, месяц, год). Пусть  $T$  — период сезонности, тогда прогноз  $\hat{y}_t, \hat{y}_{t+1}, \dots, \hat{y}_{t-1+T}$  может быть рассчитан для каждого участка процесса заболеваемости на основе наборов известных значений  $\{y_{t-T}, y_{t-2T}, y_{t-3T}, \dots\}, \dots, \{y_{t-T}, y_{t-2T}, y_{t-3T}, \dots\}$  соответственно. Таким образом, обычным способом получения оценок ожидаемой заболеваемости является простое усреднение:

$$\hat{y}_t = \sum_{j=1}^n y_{t-jT} / n, \quad (5)$$

где  $n$  — количество доступных наблюдений. Применение данного подхода широко используется при расчете уровня обычной заболеваемости.

---

<sup>13</sup> Кондратьев, М.А. Методы прогнозирования и модели распространения заболеваний // Компьютерные исследования и моделирование. — 2013. — Т. 5. — № 5. — С. 863–882.

Классический регрессионный анализ, вероятно, является самым популярным методом прогнозирования заболеваемости. Проблема регрессии состоит в том, чтобы найти оценки неизвестных параметров и сформировать функциональную связь между заболеваемостью и факторами, которые ее определяют. Регрессионные модели можно разделить на две группы<sup>14</sup>: неадаптивные модели, для оценки параметров которых используются все доступные данные, и адаптивные модели, значения параметров которых рассчитываются на основе скользящего окна наблюдения.

Вид регрессионной зависимости для неадаптивной модели, конечно же, выбирается на основе свойств анализируемого временного ряда. В некоторых случаях можно использовать полиномиальные или степенные функции<sup>15</sup>, но зачастую от модели требуется учитывать сезонный характер заболеваемости. Для этих целей используют предложенную Серфлингом циклическую функцию вида<sup>16</sup>:

$$\hat{y}_t = \sum_{j=0}^v a_j t^j + \sum_{j=1}^k (\beta_{2j-1} \sin \theta_j + \beta_{2j} \cos \theta_j), \quad (6)$$

где  $\hat{y}_t$  — оценка заболеваемости в момент времени  $t$ ,  $a_j$  и  $\beta_j$  — параметры регрессии, степень полинома  $v$  обычно равна единице, а  $\theta_j$  — линейная функция времени  $t$ . Ее следует выбирать по результатам спектрального анализа<sup>17</sup>. Обычно используют  $\theta_j = 2\pi j t / T$ , где  $T$  — период сезонности, например, 12 месяцев или 52 недели. Число гармоник  $k$  редко превышает две.

Неадаптивные регрессионные модели предназначены для учета всей предыстории заболеваемости на анализируемой территории. Для их

---

<sup>14</sup> Shmueli, G. Statistical Challenges Facing Early Outbreak Detection in Biosurveillance / G. Shmueli, H. Burkom // *Technometrics*. — 2010. — V. 52, №. 1. — P 39-51.

<sup>15</sup> Кизим, Н.А., Доровской А.В. Анализ и прогнозирование тенденций заболеваемости населения Украины и Харьковской области / Н.А. Кизим, А.В. Доровской // *Проблемы экономики*. — 2010. — № 3.

<sup>16</sup> Serfling, R.E. Methods for current statistical analysis of excess pneumonia-influenza deaths // *Public Health Rep.* — 1963. — №78 (6). — P. 494-506.

<sup>17</sup> Sumi, A. MEM spectral analysis for predicting influenza epidemics in Japan / A. Sumi, K. Kamo // *Environ Health Prev Med.* — 2012. — № 17(2). P. 98-108.

построения используются все доступные данные или, по крайней мере, наблюдения последних лет с аналогичными характеристиками. Итак, если свойства процесса заболеваемости изменились, скорее всего, устаревшие данные уже не помогут уточнить прогноз. Модели адаптивной регрессии представляют собой ограниченный промежуток временного ряда и должны быть достаточно чувствительными, чтобы реагировать на текущие тенденции заболеваемости.

Регрессионный анализ - это базовый подход к прогнозированию временных рядов любого характера, легко реализуемый с использованием любой системы компьютерной математики. Неадаптивные модели позволяют получать прогнозы заболеваемости на любой период, но такие модели игнорируют локальные колебания эпидемических показателей и плохо подходят для краткосрочного прогнозирования.

Напротив, адаптивные модели предназначены для создания прогнозов на несколько недель вперед. При достаточно большой ширине скользящего окна можно рассчитать среднесрочные оценки заболеваемости, но добиться их высокого качества проблематично. Использование описанных выше регрессионных форм предполагает, что остатки модели (ошибки в сопоставлении модели с наблюдениями) являются независимыми случайными величинами, которые имеют одинаковый закон распределения с нулевым математическим ожиданием и постоянной дисперсией. Распространенной проблемой является наличие значительной автокорреляции остатков, которая обычно указывает на то, что модель может быть дополнена, а прогноз может быть уточнен, например, с использованием авторегрессионных моделей<sup>18</sup>.

---

<sup>18</sup> Кондратьев, М. А. Методы прогнозирования и модели распространения заболеваний // Компьютерные исследования и моделирование 2013 Т. 5 № 5 С. 863–882.



Чаще всего в методологии Бокса и Дженкинса используются авторегрессионные модели, в которых измерения временных рядов представлены как<sup>19</sup>:

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}. \quad (7)$$

Здесь  $\mu$  — «уровень» случайного процесса (константа, подлежащая определению),  $\phi_j$  и  $\theta_j$  — неизвестные параметры, а  $\varepsilon_j$  — случайные импульсы невязок модели (распределенные нормально с нулевым математическим ожиданием и постоянной дисперсией  $\sigma_\varepsilon^2$ , которую оценивают по наблюдениям). Вышеприведенное выражение называется смешанной авторегрессионной моделью - скользящей средней (АРСС) порядка  $(p, q)$ . АРСС описывают только стационарные временные ряды.

Также применяется модель авторегрессии — проинтегрированного скользящего среднего (АРПСС) порядка  $(p, d, q)$ <sup>20</sup>:

$$(1 - \sum_{j=1}^p \phi_j L^j)(1 - L)^d y_t = \mu + (1 - \sum_{j=1}^q \theta_j L^j), \quad (8)$$

где  $d$  — порядок разности, который требуется для достижения стационарности (обычно не превышает двух).

АРСС зачастую используется для краткосрочного прогнозирования заболеваемости, но чаще требуется трансформация анализируемых временных рядов. АРПСС подходит как для краткосрочного, так и для долгосрочного прогнозирования, алгоритмы оценки параметров модели реализованы во всех специализированных пакетах статистических программ.

Многие временные ряды не могут строго удовлетворять требованию стационарности, что влечет за собой неточности в применении АРСС. Кроме

---

<sup>19</sup> Бокс, Дж. Анализ временных рядов прогноз и управление / Дж. Бокс, Г. Дженкинс. — М.: Мир, 1974. — 200 с.

<sup>20</sup> Box George, E.P. Time series analysis, forecasting and control / E.P. Box George, G.M. Jenkins // Holden-day, Inc. — 1976.

того, для оценки параметров АРПСС требуется большое количество наблюдений.

При небольшом количестве исходных данных предпочтительнее использовать другую технику - целочисленную авторегрессию, которую можно отнести к активно развивающимся методам прогнозирования. Существуют и другие варианты моделей авторегрессии, например, обобщенные модели авторегрессии - скользящее среднее<sup>21</sup>.

Другой подход к моделированию временных рядов, учитывающий структуру корреляции данных, - это динамические байесовские сети. Байесовские сети в настоящее время быстро набирают популярность в различных областях знаний и применяются для решения проблемы прогнозирования заболеваемости, в основном в простой форме скрытых марковских моделей<sup>22</sup>.

Основная идея СММ состоит в том, чтобы сравнить каждую случайную величину  $Y_t$  с ненаблюдаемой случайной величиной  $S_t$ , которая определяет условное распределение  $Y_t$ . Задав форму закона распределения, связывающего  $Y_t$  и  $S_t$ , можно оценить его параметры по наблюдаемым значениям  $y_t$ <sup>23</sup>.

Важным представителем методов, основанных на машинном обучении, являются искусственные нейронные сети (ИНС). ИНС сегодня - очень перспективный подход, широко распространенный не только в зарубежных, но и в отечественных работах. Методология ИНС реализована в большинстве систем компьютерной математики и статистических пакетов<sup>24</sup>.

Любой временной ряд заболеваемости можно рассматривать как случайный процесс, состоящий из сигнала, отражающего реальную

---

<sup>21</sup> Jason Brownlee. How to Create an ARIMA Model for Time Series Forecasting in Python // Time Series, December 10, 2020 <https://machinelearningmastery.com/arma-for-time-series-forecasting-with-python/>

<sup>22</sup> Sebastiani, P. A Bayesian dynamic model for influenza surveillance / P. Sebastiani, K.D. Mandl, P. Szolovits, I. Kohane, M.F. Ramoni // Statistics in Medicine. — 25(11). — P. 1803-16.

<sup>23</sup> Le Strat, Y. Monitoring epidemiologic surveillance data using hidden Markov models / Y. Le Strat, F. Carrat // Statistics in Medicine. — 1999. — Vol. 18, № 24. — P. 3463-3478.

<sup>24</sup> Bai Y., Jin Z. Prediction of SARS epidemic by BP neural networks with online prediction strategy // Chaos, Solitons and Fractals. — 2005. — Vol. 26, № 2. — P. 559-569.

эпидемическую ситуацию, и высокочастотного шума. Фильтрация шума позволяет уточнить прогноз и может выполняться как при предварительной обработке исходных данных, так и в составе самого алгоритма прогнозирования.

Комбинации методов прогнозирования разнообразны и используются авторами по своему усмотрению в зависимости от решаемой задачи. В целом использование смешанных методов прогнозирования целесообразно, если другие подходы не позволяют достичь требуемой точности прогноза.

На сегодняшний день только для среднесрочного и долгосрочного прогнозирования используются точечные оценки и модели неадаптивной регрессии. Остальные модели, такие как адаптивная регрессия, авторегрессия, байесовские и искусственные нейронные сети, используются как для краткосрочного прогнозирования, так и для среднесрочного и долгосрочного.

Не существует универсального метода прогнозирования заболеваемости - лучший подход следует выбирать, сравнивая результаты, полученные с использованием различных методик на основе эмпирических данных. Часто бывает сложно отдать предпочтение одному методу прогнозирования - несколько подходов дают сопоставимые по качеству результаты.

### **1.3. Постановка задачи**

В SIR-модели распространения эпидемии Кермака-МакКендрика, как правило, используются параметры  $\gamma$  и  $\beta$ , которые задаются исходя из экспериментальных данных и известных значений начальных данных о конкретном заболевании.

Задача моделирования распространения вирусной инфекции в работе состоит в определении данных параметров, а также нахождения начальных условий числа реципиентов, восприимчивых к инфекции, т.к. на первом этапе построения берутся приблизительные данные распространения инфекции по

уже зараженным реципиентам, что никак не учитывает критерий восприимчивости.

В большинстве изученных работ по теме значение параметра  $I_0$  (число инфицированных в момент начала эпидемии) берется как приблизительно известное, а значение параметра  $S_0$  (начальное число чувствительных) зачастую неопределенно. При этом рассматривается применимость дальнейшего совершенствования математической модели, т.е. в частности, авторам необходим более надежный и объективный метод задания параметров модели.

Будем рассматривать модель SIR:

$$\begin{aligned} \frac{dS}{dt} &= -\beta IS, \\ \frac{dI}{dt} &= \beta IS - \gamma I, \\ \frac{dR}{dt} &= \gamma I. \end{aligned} \quad (9)$$

Можем записать первый интеграл системы  $S + I + R = N$ , где  $N$  – общая численность населения.

Начальные условия для системы (9) имеют вид:

$$S(0) = S_0, I(0) = I_0, R(0) = N - S_0 - I_0. \quad (10)$$

Принимая во внимание первый интеграл и то, что  $R$  не входит в первые два уравнения, можем ограничиться рассмотрением первых двух уравнений, найдя решение которых, можно будет определить и  $R$  так, что  $R(t) = N - S(t) - I(t)$ .

Далее для удобства изложения, введем следующие обозначения:

$$\begin{aligned} S &= x_1 - \text{восприимчивые,} \\ I &= x_2 - \text{инфицированные,} \\ x &= (x_1, x_2)^*, \end{aligned}$$

и, где  $u = (u_1, u_2)^* = (\beta, \gamma)^*$  – вектор, компоненты которого являются искомыми параметрами системы,

$$\begin{aligned} f_1 &= u_1 x_1 x_2, \\ f_2 &= u_1 x_1 x_2 - u_2 x_2, \\ f &= (f_1, f_2)^*. \end{aligned} \tag{11}$$

Таким образом, система дифференциальных уравнений SIR-модели примет следующий вид:

$$\dot{x} = f(t, x, u), \tag{13}$$

$$x(0) = x_0. \tag{14}$$

Здесь  $u$  – вектор (допустимое управление),  $n \in U \subset R^2$ ,  $U$  – компактное выпуклое множество и  $x(t) = x(t, x_0, u)$  – соответствующая ему траектория системы (13), с начальными условиями (14),  $x_0 \in M_0$ .

Будем рассматривать задачу определения параметров системы (13) - вектора  $u$ , а также определения начальных условий (14).

На траекториях системы (13) зададим функционал следующего вида:

$$J(u, x_0) = \int_0^T \varphi(t, x(t)) dt. \tag{15}$$

В качестве функции  $\varphi$  в формуле (15) рассматривается функция вида  $\varphi = (x_2(t) - \hat{x}_2(t))^2$ , где  $\hat{x}_2(t)$  – изменение числа инфицированных реципиентов, зависящее от времени на интервале  $t \in [0, T]$ .

Будем рассматривать задачу нахождения минимума функционала (15) по всем допустимым  $u$  и начальным данным.

#### 1.4. Вариация и градиент функционала

При решении задач оптимального управления сложными системами часто используются численные методы оптимизации. Наиболее популярными среди них являются градиентные методы минимизации функционала.

В настоящей работе рассматривается задача минимизации интегрального функционала (15) для нахождения параметров системы SIR и выбора начальных данных модели.

С этой целью выпишем вариацию функционала (15). Можно показать, что вариация функционала по  $u \in U$  и начальным данным  $x_0(t) \in N$  имеет следующий вид<sup>25</sup>:

$$\delta J = - \int_0^T \psi^* \frac{\partial f}{\partial u} \Delta u dt - \psi^*(0) \delta x_0. \quad (16)$$

Здесь функция  $\psi$  удовлетворяет следующему уравнению:

$$\dot{\psi} = - \left( \frac{\partial f}{\partial x} \right)^* \psi + \frac{\partial \varphi}{\partial x}. \quad (17)$$

с конечным условием  $\psi(T) = 0$ .

Перейдя от общего вида к (13), можем записать, таким образом  $\frac{\partial f}{\partial x}$ :

$$\frac{\partial f}{\partial x} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{pmatrix} = \begin{pmatrix} -u_1 x_2 & -u_1 x_1 \\ u_1 x_2 & u_1 x_1 - u_2 \end{pmatrix}.$$

Запишем для нашего случая систему уравнений (17):

$$\begin{aligned} \dot{\psi}_1 &= u_1 x_2 \psi_1 - u_1 x_2 \psi_2, \\ \dot{\psi}_2 &= u_1 x_1 \psi_1 - (u_1 x_1 + u_2) \psi_2 + 2(x_2 - \hat{x}_2), \end{aligned} \quad (18)$$

с конечным условием

$$\psi_1(T) = 0, \psi_2(T) = 0. \quad (19)$$

Запишем выражение для градиента рассматриваемого функционала (15) по параметрам  $u_1, u_2$  и  $x_0$ .

---

<sup>25</sup> Овсянников, Д.А. Моделирование и оптимизация динамики пучков заряженных частиц. — Л.: Изд-во Ленингр. ун-та, 1990. — 312 с.

$$\begin{aligned}\frac{\partial J}{\partial u_1} &= - \int_0^T (-\psi_1 x_1 x_2 + \psi_2 x_1 x_2) dt, \\ \frac{\partial J}{\partial u_2} &= - \int_0^T (-\psi_2 x_2 + 2(x_2 - \hat{x}_2)) dt, \\ \frac{\partial J}{\partial x_{1,0}} &= -\psi_1(0).\end{aligned}\tag{20}$$

Решение поставленной задачи состоит в минимизации значения функционала (15).

### **1.5. Алгоритм определения параметров модели и начальных данных в задаче SIR-моделирования**

Рассмотрим построение алгоритма минимизации значения интегрального функционала (15) на основе полученного выражения градиента функционала (20). Далее для удобства записи обозначим искомый вектор  $(u_1, u_2, x_{1,0})^*$  за  $V$ , т.е.

$$V = (u_1, u_2, x_{1,0})^*.$$

Был разработан итерационный алгоритм в задаче SIR-моделирования, который позволит более точно подобрать начальные условия и параметры SIR модели на основе реальных статистических данных.

Основная задача алгоритма состоит в минимизации значения интегрального функционала  $J(V^k)$ , здесь  $k$  - номер итерации. Предлагается следующий алгоритм:

- 1) Задаем начальное приближение параметров  $u$ , где  $u = (u_1, u_2)$  – и начальных данных  $x(0) = x_0$ . Обозначили их за вектор  $V^k$ , полагая  $k = 0$ ;
- 2) Интегрируем полученную систему (13), начав итерационный процесс с заданным начальным приближением;
- 3) Вычисляем интегральный функционал  $J(V)$  по формуле (15).

4) Вычисляем вспомогательную векторную функцию  $\psi(t)$  путем решения системы уравнений (18) с конечными условиями (19).

5) Вычисляем градиент функционала  $gradJ(V^k)$  по следующим формулам:

$$\begin{aligned}\frac{\partial J}{\partial u_1} &= -\int_0^T (-\psi_1 x_1 x_2 + \psi_2 x_1 x_2) dt, \\ \frac{\partial J}{\partial u_2} &= -\int_0^T (-\psi_2 x_2 + 2(x_2 - \hat{x}_2)) dt, \\ \frac{\partial J}{\partial x_{1,0}} &= -\psi_1(0).\end{aligned}$$

6) Получаем значение  $V^{k+1}$  методом градиентного спуска  $V^{k+1} = V^k - \bar{\mu} gradJ(V^k)$ . Для нахождения  $\bar{\mu}$  находится минимум функции  $Z(\mu)$  такой, что

$$\min_{\mu \in [0, a]} Z(\mu) = J(V^k - \mu gradJ(V^k)) = Z(\bar{\mu}).$$

7) Проверяем условия остановки алгоритма. Алгоритм останавливается, когда либо достигается заданное число итераций  $n$  либо  $|J(V^{k+1}) - J(V^k)| < \varepsilon$ ,  $\varepsilon$  - заданное значение.

Если после проверки ни одно из поставленных условий не выполняется, возвращаемся к шагу 2. Повторяем заданный алгоритм, получая новые  $u_1, u_2$  и начальные значения для нахождения наилучшего результата.



## ГЛАВА 2. ПРАКТИЧЕСКАЯ РЕАЛИЗАЦИЯ МОДЕЛИ РАСПРОСТРАНЕНИЯ ЭПИДЕМИИ

### 2.1. Моделирование распространения вирусной инфекции гриппа через SIR-модель с прямым выбором параметров

В большинстве изученных работ по теме значение параметра  $I_0$  (число инфицированных в момент начала эпидемии) берется как приблизительно известное, а значение параметра  $S_0$  (начальное число чувствительных) зачастую неопределенно. При этом рассматривается применимость дальнейшего совершенствования математической модели, т.е. в частности, авторам необходим более надежный и объективный метод задания параметров модели.

Рассматривается модель SIR. Дифференциальные уравнения, описывающие модель выглядят следующим образом:

$$\begin{aligned}\frac{dS}{dt} &= -\beta IS, \\ \frac{dI}{dt} &= \beta IS - \gamma I, \\ \frac{dR}{dt} &= \gamma I.\end{aligned}$$

Первоначально рассматривается метод задания параметров SIR-модели, который выглядит следующим образом:

```
U0 = 3
gamma = 1./14
beta = gamma*U0
условие начала эпидемии: beta*S_0/gamma =>1
```

Выбор  $U_0$  относится к постановке задачи. Правильный выбор величины  $U_0$  в некоторой мере обуславливает величину заболеваемости. Так, было проведено построение модели распространения вирусной инфекции гриппа через SIR-модель.

Население взято из расчета количества заболевших на 100 000 человек. Построение ведется по неделям, массив данных составляет вектор из 21

значения, в котором возвращается значение функции на каждый 7 день. Как видно из рисунка 2.1 период эпидемии составил почти 21 неделю, а количество заболевших не превышает пятой части населения.

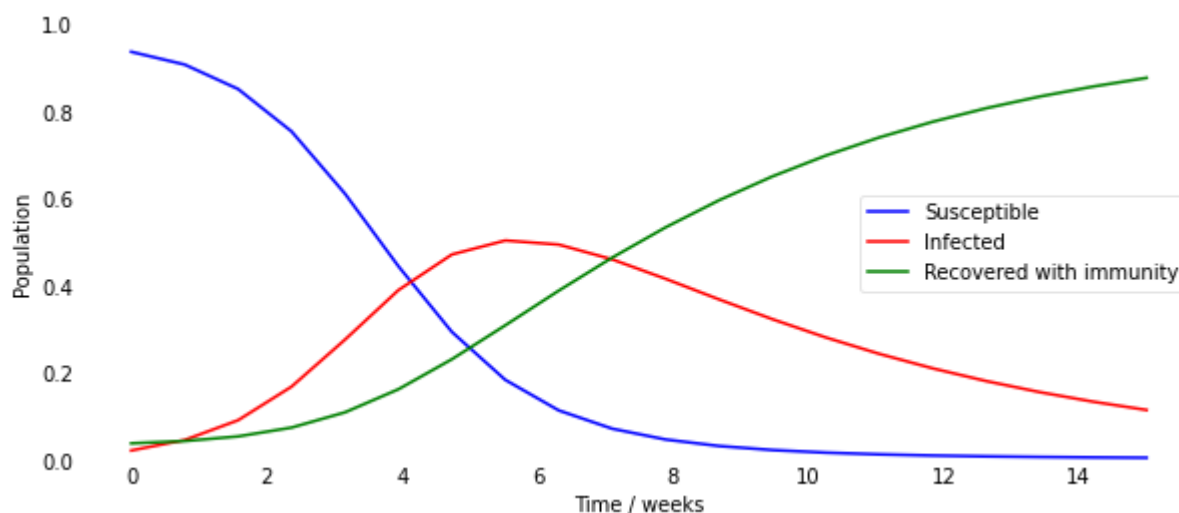


Рис. 2.1. SIR-модель с обычным выбором параметров, где  $U_0 = 3$ ,  $\gamma = 1./14$ ,  $\beta = \gamma * U_0$

Было построено решение системы SIR для различных значений параметров  $\gamma, \beta$  и одинаковых начальных данных. Вариативность параметров при одинаковых  $S_0, I_0$  показана на рисунке 2.2.

Как видно из рисунка, чем меньше величина параметров  $\gamma, \beta$ , тем дольше будет длиться эпидемия. При этом, если оба параметра имеют небольшую величину, то эпидемия не случится вовсе.

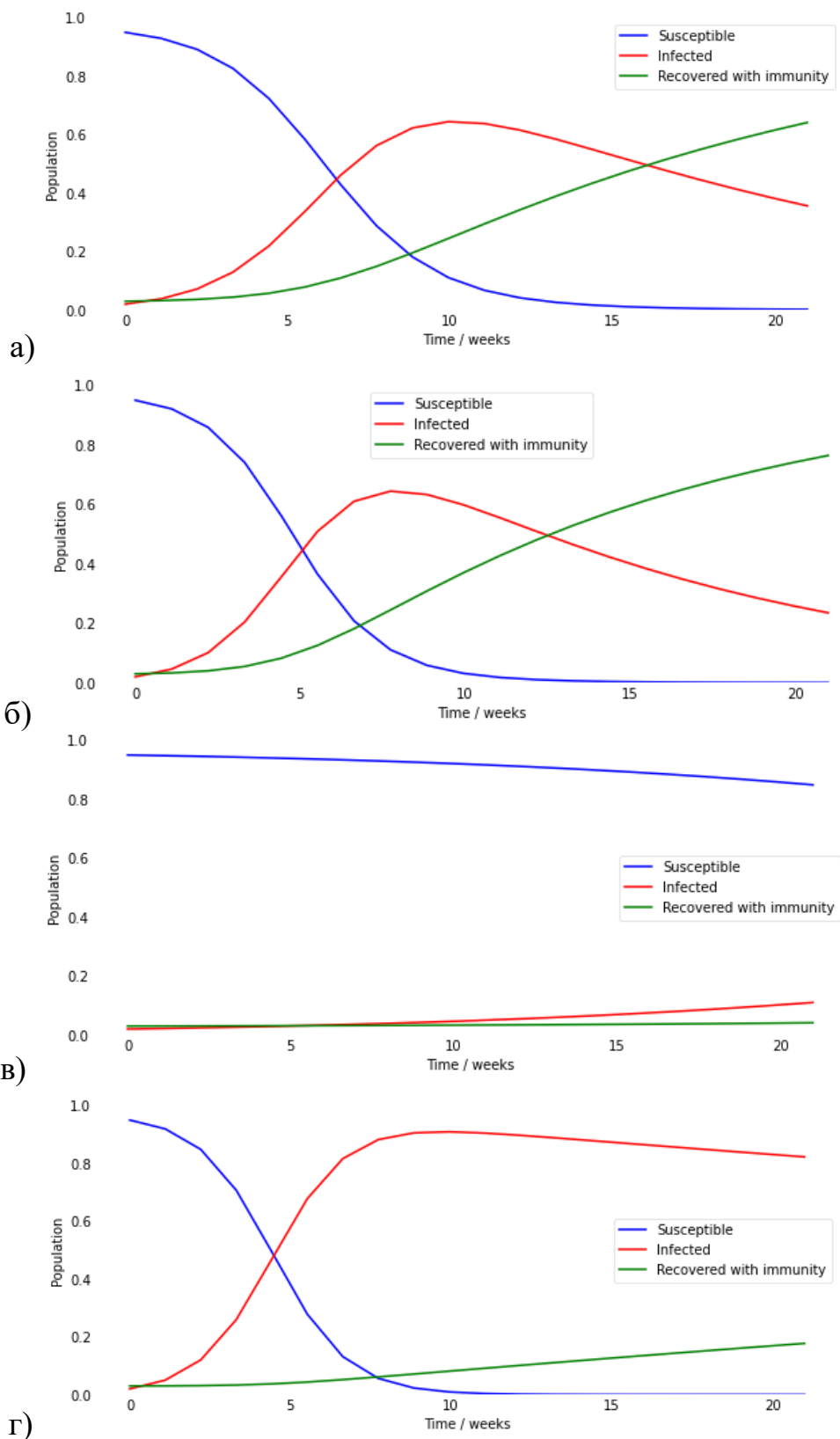


Рис. 2.2. Построение SIR-модели разными заданными параметрами  $\gamma, \beta$ :  
 а) 0.1, 0.9 б) 0.9, 0.1 в) 0.1, 0.1 г) 0.9, 0.9

## 2.2. SIR-модель с выбором параметров на основе оптимизационного алгоритма

Были получены данные по распространению гриппа в Российской Федерации за 2015-2018 гг. (рис. 2.3.). Данные брались за период в 21 неделю, после первого зафиксированного роста заболеваемости гриппа. Отображение данных идет в расчете на 100 тысяч человек.

Как видно из рис. 2.2. Наибольший пик распространения эпидемии пришелся на 2015-2016 гг., в который было зарегистрировано порядка 14 тысяч заболевших в расчете на 100 тысяч человек по России.

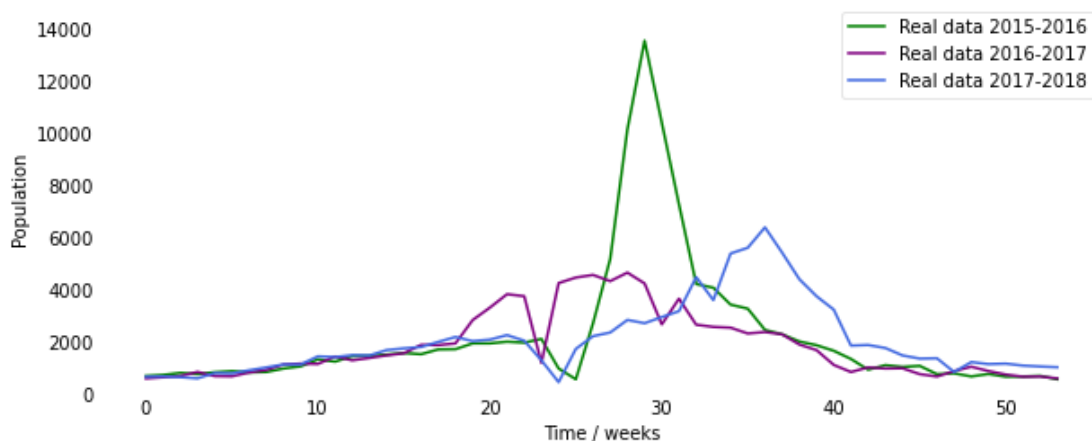


Рис. 2.3. Построение кривых распространения эпидемии гриппа на территории Российской Федерации по реальным статистическим данным, где зеленым цветом показаны данные за 2015-16 гг, фиолетовым за 2016-2017 гг., синим за 2017-2018 гг.

После этого были построены графики статистических данных по распространению эпидемии гриппа за 21 неделю и кривой «Infected» с выбором начальных условий и подбором параметров  $\gamma$  и  $\beta$  для каждой из полученных кривых статистических данных (рис. 2.4 а,б,в). Начальные приближения и параметры системы были взяты приближенно.

Отталкиваясь от полученных начальных данных и выбранных параметров  $\gamma$  и  $\beta$  для каждой из полученных кривых статистических данных в дальнейшем применен алгоритм, основанный на градиентном методе для более точного выбора начальных данных и параметров  $\gamma$  и  $\beta$ .

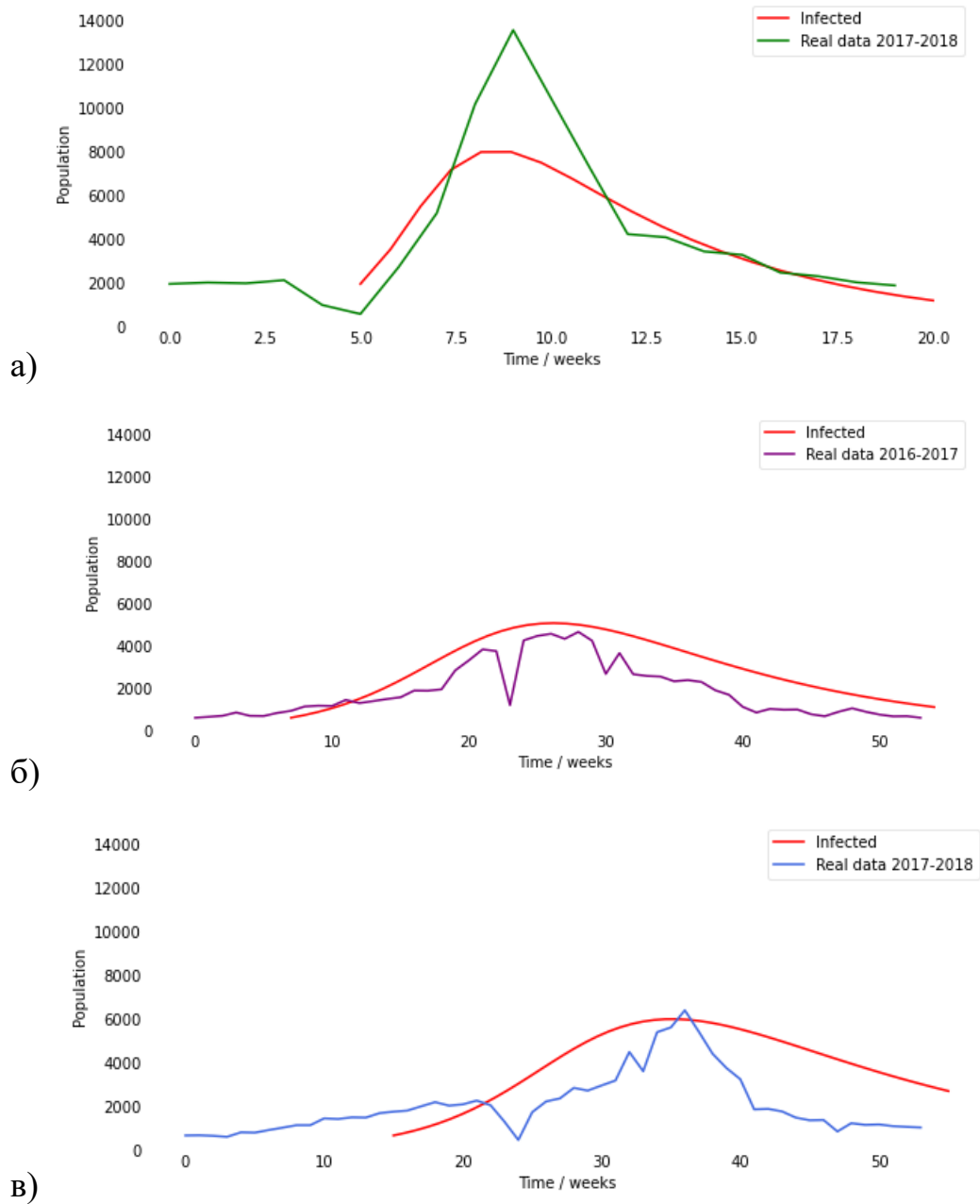


Рисунок 2.3. График статистических данных по распространению эпидемии гриппа за 21 неделю и кривой «Infected» с выбором начальных условий и подбором параметров  $\gamma$  и  $\beta$  для каждой из полученных кривых статистических данных, где берутся параметры: а)  $I_0, R_0 = 680, 0, S_0 = N - I_0 -$

$R_0, \gamma = 0.2 \beta = 1.2$ ; б)  $I_0, R_0 = 582, 0, S_0 = N - I_0 - R_0, \gamma = 0.09 \beta = 0.3$ ; в)  $I_0, S_0 = 647, 13753, R_0 = N - S_0 - I_0, \gamma = 0.07 \beta = 0.28$ .

Для начала была проведена реализация алгоритма для соотношения со статистическими данными распространения эпидемии гриппа в Российской Федерации за 2015-2016 гг.

Действуя исходя из алгоритма, полученного в параграфе 1.5 задаем начальное приближение параметров  $u$ . Пусть  $u_1 = 0.2, u_2 = 1.2$  исходя из параметров, подобранных для статистических данных по распространению эпидемии гриппа в 2015.2016 гг. в Российской Федерации за 20 недель. Начальные условия нормируются так, что  $N = 1$ :  $x_1 = 0.938014221, x_2 = 0.022901853$ .

Код программы реализован на программном языке Python с применением функции `spi.solve_ivp`:

```
sol = spi.solve_ivp(lambda y, t: x1x2_model(y, t, N, u1, u2), [0,T],
[x1_0, x2_0], method = 'LSODA',
t_eval=[0, 7, 14, 21, 28, 35, 42, 49, 56, 63,
70, 77, 84, 91, 98, 105, 112, 119, 126, 133, 140])
```

Найдено решение системы (13) в округлении до 10 степени значимости. График построения найденного решения выглядит следующим образом (рис 2.4):

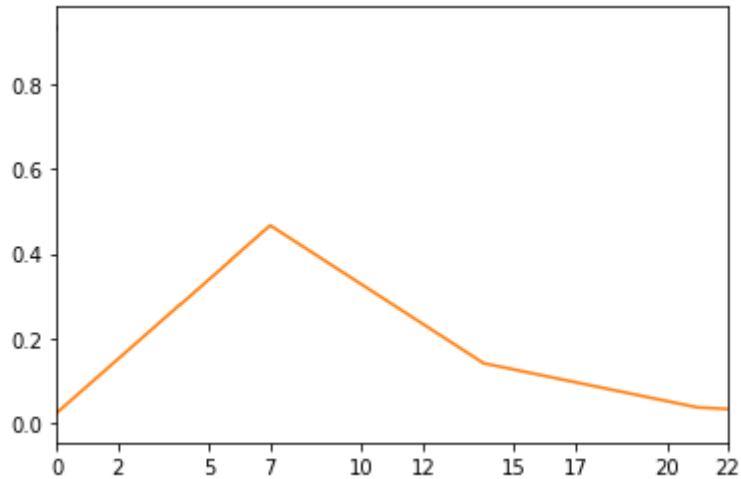


Рис 2.4. Визуализация  $x_2$  из решения системы (13) с заданными параметрами  $u_1 = 0.2$ ,  $u_2 = 1.2$  и начальными значениями  $x_1 = 0.938014221$ ,  $x_2 = 0.022901853$

Вычисляем функционал  $J(u, x_0)$  по формуле (21) приближенными методами. Получается, что значение функционала после нормирования данных равно  $J(u, x_0) = 0.290146797$ .

Вычисляем вспомогательную векторную функцию  $\psi(t)$  путем решения системы уравнений (18) с конечными условиями (19).

Код программы для вычисления системы реализован также на языке Python с применением встроенного алгоритма `fsolve`:

```
for i in range(21):
    set_param(i)
    x, y = fsolve(equations, (f1f2rez[i]))
```

Найдено решение системы (18) с конечными условиями (19) с округлением до 10 степени значимости. Параметры системы  $u_1, u_2, x_1, x_2, \hat{x}_2$  берутся как вектора данных, конечные условия равны нулю.

Вычисляем градиент функционала  $gradJ(V^k)$  по следующим формулам:

$$\frac{\partial J}{\partial u_1} = - \int_0^T (-\psi_1 x_1 x_2 + \psi_2 x_1 x_2) dt,$$

$$\frac{\partial J}{\partial u_2} = - \int_0^T (-\psi_2 x_2 + 2(x_2 - \hat{x}_2)) dt,$$

$$\frac{\partial J}{\partial x_{1,0}} = -\psi_1(0).$$

Получаем, что значение градиента по нормированным данным  $\frac{\partial J}{\partial u_1} = 0.0000334232$ ,  $\frac{\partial J}{\partial u_2} = 0.6201257678$ ,  $\frac{\partial J}{\partial x_{1,0}} = -0.025322287$ .

Т.е.  $gradJ(V^0) = (0.0000334232, 0.6201257678, -0.025322287)$ .

Получаем значение  $V^{k+1}$  методом градиентного спуска  $V^{k+1} = V^k - \mu gradJ(V^k)$ . Проверяем условия остановки алгоритма. Алгоритм останавливается, т.к. достигнута заданная точность  $|J(V^{k+1}) - J(V^k)| < \varepsilon$ ,  $\varepsilon = 0.01$ .

Получается, что итоговый вектор, найденный для улучшения начальных данных и параметров SIR-модели после реализации алгоритма, выглядит следующим образом:

$$V^{k+n} = (1.22, 0.24021898, 0.93801422, 0.02290185).$$

Рисунок 2.5 отображает результат построения SIR-модели по начальным данным, полученным в результате работы алгоритма и реальным данным распространения эпидемии гриппа на территории Российской Федерации в 2015-2016 гг.

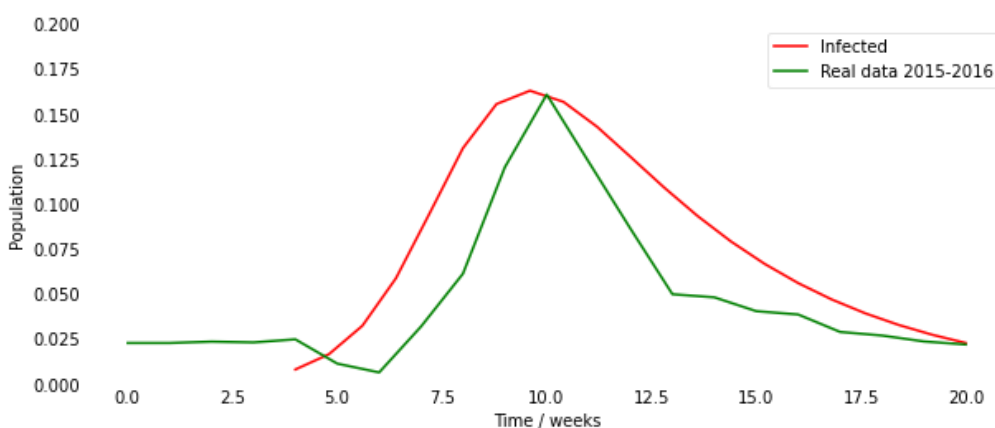


Рис. 2.5. Построение кривой «Infected» SIR-модели с заданными параметрами  $\beta = 0.24021898$ ,  $\gamma = 1.22$ , начальными значениями  $S_0 =$



0.938014221,  $I_0 = 0.022901853$  и статистических данных распространения эпидемии гриппа в Российской Федерации за 2015-2016 гг.

После получения в результате работы алгоритма улучшенных начальных данных и параметров модели для статистических данных по распространению эпидемии гриппа в Российской Федерации за 2015-2016 год были получены начальные данные и параметры за 2016-2017 гг. и 2017-2018 гг. Результаты построения моделей представлены на рис 2.6 а, б.

Как видно из результатов, рис. 2.6. показывает большее приближение кривой «Infected» к действительным статистическим данным распространения эпидемии, нежели выбор начальных данных и параметров на рис. 2.3.

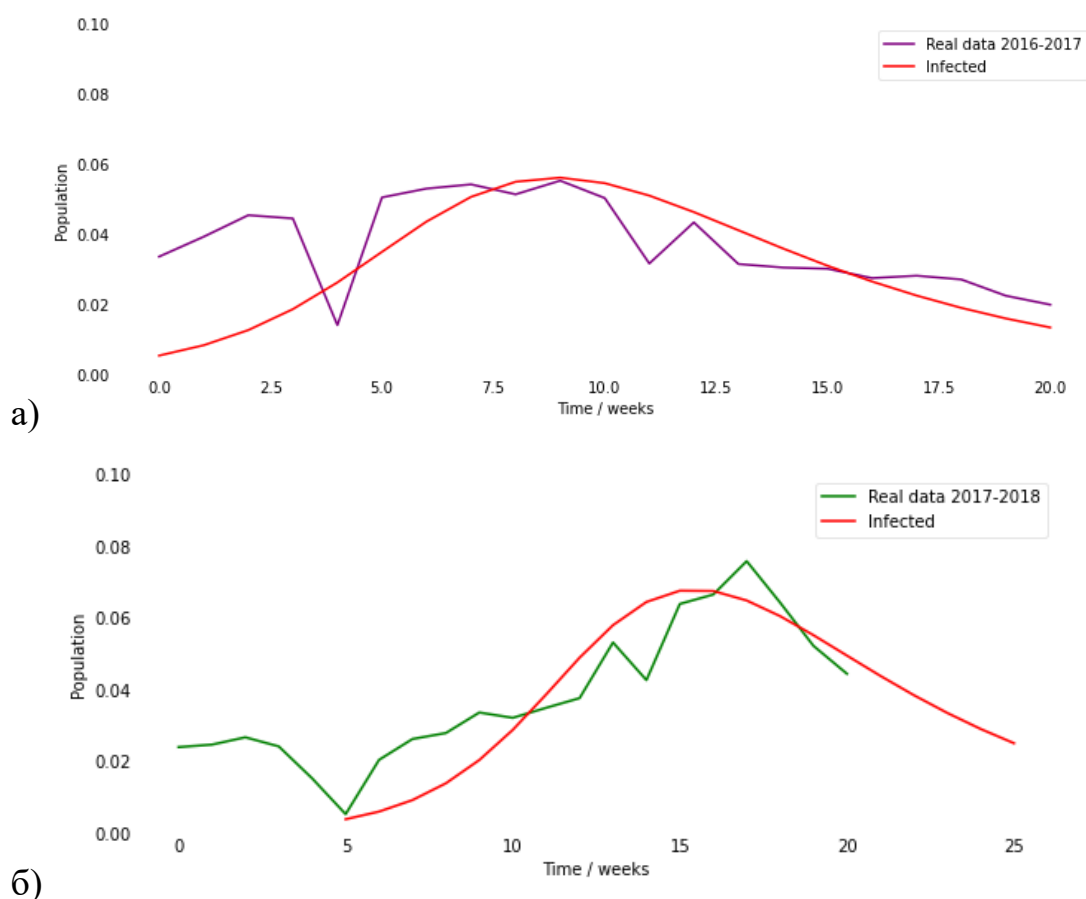


Рис. 2.7. Построение кривой «Infected» SIR-модели с заданными параметрами и начальными значениями: а)  $\gamma = 0.09, \beta = 0.29430084, I_0 = 0.02864886, S_0 = 0.92292553$

Для анализа точности нахождения начальных данных и параметров был совершен тестовой прогон алгоритма. Для этого были заданы случайным образом заданы параметры  $\gamma = 0.08, \beta = 0.5$  и начальные данные  $I_0 = 0.21, S_0 = 0.7584$ . После чего проведено построение SIR-модели. На следующем этапе взяты значения точек кривой «Infected» для создания тестовой выборки. После чего данные были занесены в алгоритм и изменены начальные условия и параметры. Результат вывода алгоритма можно увидеть на рис. 2.8. После улучшения точности данных алгоритмом были получены параметры  $\gamma = 0.0794568, \beta = 0.49723655$  и начальные данные  $I_0 = 0.22, S_0 = 0.7488$ .

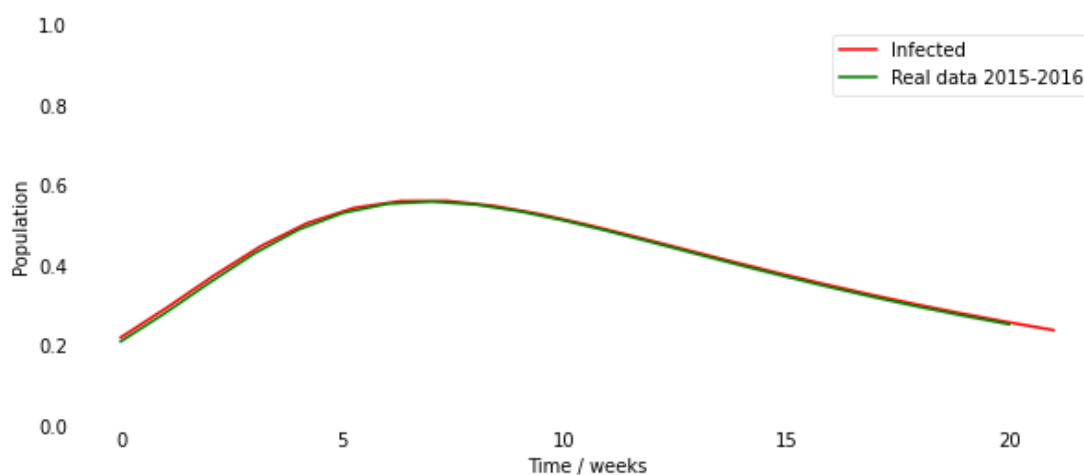


Рис. 2.8. Тестовое построение кривой «Infected» SIR-модели и выборки для проверки работы алгоритма

По итогу, можно говорить о том, что цель работы достигнута. Введенный алгоритм позволяет получать улучшенные параметры SIR-модели и начальные данные для моделирования распространения эпидемии.



## ЗАКЛЮЧЕНИЕ

В результате работы была достигнута главная цель исследования, а именно построение методики выбора начальных данных и параметров системы для моделирования распространения эпидемии вирусной инфекции на примере эпидемии гриппа в Российской Федерации.

В ходе работы были проанализированы существующие расширения SIR-модели, а также другие методы моделирования и анализа распространения заболеваний, основанные на статистических преобразованиях, машинном обучении и методе прецедентов, на базе фильтрации.

Разработан алгоритм выбора параметров модели и начальных данных для системы для моделирования распространения вирусной инфекции. Данный алгоритм основан на градиентном методе, использующем аналитическое представление градиента функционала.

Был проведен анализ распространения вирусной инфекции гриппа на территории Российской Федерации за 2015-2016 гг., 2016-2017 гг., 2017-2018 гг., а также проведено моделирование SIR-модели распространения эпидемии, начальные данные и параметры которой в дальнейшем были улучшены с помощью разработанного алгоритма.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Bai Y., Jin Z. Prediction of SARS epidemic by BP neural networks with online prediction strategy // *Chaos, Solitons and Fractals*. — 2005. — Vol. 26, № 2. — P. 559–569.
2. Box George, E.P. Time series analysis, forecasting and control / E.P. Box George, G.M. Jenkins // Holden-day, Inc. — 1976.
3. Hethcote, W.H. The Mathematics of Infectious Diseases // *Society for Industrial and Applied Mathematics*. — 2000. — V. 42, № 4. — P. 599-653.
4. Jason, B. How to Create an ARIMA Model for Time Series Forecasting in Python // *Time Series*. — 2020. — URL: <https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python> (data views: 25.04.2021).
5. Kermack, W.O. Contributions to the mathematical theory of epidemics —I / W.O. Kermack, A.G. McKendrick // *Bulletin of Mathematical Biology*. — 1991. — V. 53. — P. 33-55.
6. Le Strat, Y. Monitoring epidemiologic surveillance data using hidden Markov models / Y. Le Strat, F. Carrat // *Statistics in Medicine*. — 1999. — Vol. 18, № 24. — P. 3463–3478.
7. Menon, A. Modelling and simulation of COVID-19 propagation in a large population with specific reference to India / A. Menon, N. K. Rajendran, A. Chandrachud, G. Setlur // *medRxiv*. — 2020. — № 5. — P. 25-28.
8. Moehlis, J.M. Seasonal Epidemic Models: An SIS model // *Mathematics Department, Macquarie University*. — 2002. — №10(14). — URL: <https://sites.me.ucsb.edu/~moehlis/APC591/tutorials/tutorial6/tutorial6.html> (date views: 12.03.2021).
9. Mwalili, S. SEIR model for COVID-19 dynamics incorporating the environment and social distancing / S. Mwalili, M. Kimathi, V. Ojiambo, D. Gathungu, R. Mbogo // *BMC Research Notes*. — № 52(2020) — P. 1-5.

10. Sebastiani, P. A Bayesian dynamic model for influenza surveillance / P. Sebastiani, K.D. Mandl, P. Szolovits, I. Kohane, M.F. Ramoni // *Statistics in Medicine*. — 25(11). — P. 1803-16.
11. Serfling, R.E. Methods for current statistical analysis of excess pneumonia-influenza deaths // *Public Health Rep.* — 1963. — №78 (6). — P. 494-506.
12. Shmueli, G. Statistical Challenges Facing Early Outbreak Detection in Biosurveillance / G. Shmueli, H. Burkom // *Technometrics*. — 2010. — V. 52, №. 1. — P 39-51.
13. Smith, D. The SIR Model for Spread of Disease / D. Smith, L. Moore // *Convergence*. — 2004. — № 12. — URL: <https://www.maa.org/press/periodicals/loci/joma/the-sir-model-for-spread-of-disease-the-differential-equation-model> (date views: 29.03.2021).
14. Sumi, A. MEM spectral analysis for predicting influenza epidemics in Japan / A. Sumi, K. Kamo // *Environ Health Prev Med*. — 2012. — № 17(2). P. 98-108.
15. Thompson, W.W. Estimates of US influenza-associated deaths made using four different methods / W.W. Thompson, E. Weintraub, P. Dhankhar, O.Y. Cheng, L. Brammer, M.I. Meltzer et al. // *Influenza Other Respi Viruses*. — 2009. — №3. — P. 37-49.
16. Бароян, О.В. Моделирование и прогнозирование эпидемий гриппа на территории СССР / О.В. Бароян, Л.А. Рвачев, Ю.Г. Иванников. — М., 1977. — 546 с.
17. Бароян, О.В. Прогнозирование эпидемий гриппа в условиях СССР / О.В. Бароян, Л.А. Рвачев // *Вопросы вирусологии*. — 1978. — № 2. — С. 131–137.
18. Бокс, Дж. Анализ временных рядов прогноз и управление / Дж. Бокс, Г. Дженкинс. — М.: Мир, 1974. — 200 с.
19. Кизим, Н.А., Доровской А.В. Анализ и прогнозирование тенденций заболеваемости населения Украины и Харьковской области / Н.А.

Кизим, А.В. Доровской // Проблемы экономики. — 2010. — № 3. — URL: <https://cyberleninka.ru/article/n/analiz-i-prognozirovanie-tendentsiy-zabolevaemosti-naseleniya-ukrainy-i-harkovskoy-oblasti> (дата обращения: 21.04.2021).

20. Ковалев, С.В. Мультиагентное моделирование распространения эпидемий /

21. С.В. Ковалев, Рюмин Н.Н., Ковалева О.А., Сидляр М.Ю., Хромова Т.А. // Вестник технологического университета. — 2021. — Т.24. — №1. — С. 91-97.

22. Кондратьев, М.А. Методы прогнозирования и модели распространения заболеваний // Компьютерные исследования и моделирование. — 2013. — Т. 5. — № 5. — С. 863–882.

23. Овсянников, Д.А. Моделирование и оптимизация динамики пучков заряженных частиц. — Л.: Изд-во Ленингр. ун-та, 1990. — 312 с.

24. Хартл, Г. Смертность от респираторных заболеваний, связанных с сезонным гриппом // Всемирная организация здравоохранения. — 2017. — URL: <https://www.who.int/ru/news/item/14-12-2017-up-to-650-000-people-die-of-respiratory-diseases-linked-to-seasonal-flu-each-year> (дата обращения: 14.04.2021).