

Санкт-Петербургский государственный университет

РОЛДУГИНА Анна Андреевна

Магистерская диссертация

*Моделирование эпидемиологической ситуации, связанной
с распространением коронавирусной инфекции (COVID-19)
в Санкт-Петербурге*

Направление 01.04.02 «Прикладная математика и информатика»
Основная образовательная программа ВМ.5504 «Исследование операций
и системный анализ»

Заведующий кафедрой
д.ф.-м.н., профессор
Петросян Л.А.

Научный руководитель
к.ф.-м.н., доцент
Панкратова Я.Б.

Рецензент
к.т.н.
Асфар С.В.

Санкт-Петербург
2020г.

Содержание

Введение	3
Постановка задачи	5
Обзор литературы	6
Глава 1. Модели распространения эпидемий	8
1.1. Классическая модель SIR	8
1.2. SEIR-модель	9
1.3. Обобщенная SEIR-модель	12
Глава 2. Эпидемиологическая ситуация в Санкт-Петербурге	17
2.1. Автоматический подбор параметров модели	18
2.2. Модель развития эпидемии в Санкт-Петербурге	30
Выводы	35
Заключение	37
Список литературы	38
Приложение	40

Введение

В декабре 2019 года в городе Ухань (провинция Хубэй, Китай) началась вспышка новой коронавирусной инфекции, быстро распространившаяся по Китаю, а затем и по всему миру. В марте 2020 года было объявлено, что вспышка коронавирусной инфекции приняла характер пандемии. Возбудителю инфекции, новому коронавирусу, Всемирной организацией здравоохранения было дано название 2019-nCoV, или SARS-CoV-2, а инфекция была названа COVID-19 («Coronavirus disease 2019»).

Вирус передается воздушно-капельным и контактным путем, имеет достаточно длительный инкубационный период (до 14 дней, в среднем 5-7 дней), вызывает в организме серьезные заболевания, в том числе острую респираторную вирусную инфекцию, пневмонию, дыхательную недостаточность, особенно опасен для людей пожилого возраста и имеющих хронические заболевания. Вакцины против коронавирусной инфекции на данный момент не существует, для лечения используют симптоматическую терапию. По текущим данным, от начала заболевания до выписки пациента проходит 10-20 дней в зависимости от тяжести течения формы новой коронавирусной инфекции.

По данным [12] на середину мая 2020 года в России выявлено более 250 тысяч заболевших COVID-19, среди которых более 2,5 тысяч умерших. В целях предотвращения распространения коронавирусной инфекции в регионах страны проводится комплекс противоэпидемических и профилактических мероприятий.

С начала распространения эпидемии Санкт-Петербург занимает третью позицию по общему числу заразившихся коронавирусом в перечне всех регионов России [13]. Эпидемиологическая ситуация в Санкт-Петербурге

значительно отличается от ситуации в Москве: не такой большой поток иностранных граждан и прибывших из неблагополучных стран россиян, меньшая плотность населения, быстрое введение противоэпидемических мер также повлияли на скорость распространения инфекции в городе. В целях контроля за эпидемиологической ситуацией и оперативного реагирования на ее изменения могут применяться математические методы и модели.

Математическое моделирование позволяет изучить механизмы развития эпидемии. Применение эпидемиологических моделей позволяет спрогнозировать и оценить многие характеристики распространения заболеваний. Адекватные и качественные математические модели позволят оценить такие характеристики развития эпидемии, как общее число заболевших, выздоровевших, умерших и болеющих, длительность эпидемии, время ее окончания при различных параметрах, имитирующих влияние противоэпидемических мер.

Постановка задачи

В условиях борьбы с распространением коронавирусной инфекции крайне важен анализ текущей ситуации в режиме реального времени, основанный на актуальных данных. Моделирование эпидемиологической ситуации с помощью математических методов позволяет оценить текущую ситуацию и спрогнозировать дальнейшее развитие эпидемии. Целью настоящей работы является исследование известных способов построения эпидемиологических моделей, построение модели и прогнозирование эпидемиологической ситуации в соответствии с построенной моделью распространения коронавируса в Санкт-Петербурге.

В связи с этим возникает ряд задач:

- проанализировать известные математические модели распространения инфекционных заболеваний;
- построить математическую модель распространения эпидемии коронавируса в Санкт-Петербурге;
- разработать программный продукт для построения модели распространения коронавирусной инфекции в Санкт-Петербурге и прогнозирования эпидемиологической ситуации на реальных статистических данных;
- разработать программный продукт для визуализации полученных результатов в виде таблиц и графиков.

Обзор литературы

Вопрос построения моделей распространения эпидемий исследуется многими учеными. На данный момент описано множество математических методов моделирования распространения инфекционных заболеваний.

Одним из широко используемых типов моделей распространения инфекций являются детерминированные SIR-модели (Susceptible — Infected — Recovered), в основе которых лежит теория Кермака и Маккендрика [8]. Этот вид моделей предполагает разделение популяции (фиксированного числа лиц) на три группы: восприимчивые, инфицированные и выздоровевшие индивиды. Динамика выделенных групп описывается с помощью системы дифференциальных уравнений с непрерывным временем. SIR-модель хорошо подходит для моделирования эпидемий и актуальна до сих пор [10]. Эта модель послужила основой для множества других моделей.

SIR-модель вместе с ее модификациями SI, SIS, SIRS рассматривается в [5]. Описанные модели предполагают возможность повторного инфицирования выздоровевшего индивида. В [6] также предложены модификации исходной SIR-модели: модели MSIR и MSEIR, в которых учитывается возможность приобретения иммунитета к инфекции внутриутробно, от зараженной матери.

В [2] рассматривается еще одна модификация SIR-модели — SEIR, учитывающая наличие инкубационного периода болезни.

В условиях быстрого распространения коронавирусной инфекции многие исследователи изучали поведение эпидемии коронавируса и предлагали различные модификации описанных выше моделей на примерах разных стран, в том числе Китая, США, Италии, Испании [1, 3, 4, 9].

В [7] предлагается еще одна модификация модели SIR, учитываю-

щая возможность нахождения индивида в одном из 7 состояний течения болезни: от восприимчивых и невосприимчивых, заболевших в разных стадиях до момента окончания заболевания (выздоровевших или умерших). В статье с помощью модели SEIQRDP исследуется распространение коронавирусной инфекции в провинциях Китая, в том числе в провинции Хубэй.

В данной работе рассматривается построение модели SEIQRDP на официальных статистических данных [12, 13] Санкт-Петербурга.

Глава 1. Модели распространения эпидемий

Математическое моделирование распространения инфекционных заболеваний представляет собой достаточно сложную задачу из-за большого количества факторов, влияющих на развитие эпидемии. В данной главе описаны и исследованы классическая модель SIR и ее модификаций SEIR и SEIQRDP на возможность их использования для построения модели распространения коронавирусной инфекции в Санкт-Петербурге.

1.1. Классическая модель SIR

Одной из хорошо известных эпидемиологических моделей является модель SIR (Susceptible — Infected — Removed), описанная в 1927 году [8]. Классическая SIR-модель предусматривает деление всей популяции на три класса:

- $S(t)$ (susceptible) – восприимчивые (предрасположенные к заболеванию неинфицированные индивиды);
- $I(t)$ (infected) – инфицированные (болеющие, способные распространить заболевание);
- $R(t)$ (removed) – невосприимчивые (выбывшие из класса инфицированных в случае выздоровления или смерти).

Популяция считается неизменной на протяжении эпидемии, т.е. не учитывает рождаемость населения и смертность от прочих причин:

$$S(t) + I(t) + R(t) = N.$$

Передача инфекции в популяции осуществляется от инфицированных индивидов I к восприимчивым S . Переболевшие индивиды R приобретают иммунитет и не могут быть заражены вторично.

SIR-модель без учета рождаемости и смертности определяется системой дифференциальных уравнений с непрерывным временем:

$$\begin{cases} \frac{dS(t)}{dt} = -\beta \frac{S(t)I(t)}{N}, \\ \frac{dI(t)}{dt} = \beta \frac{S(t)I(t)}{N} - \delta I(t), \\ \frac{dR(t)}{dt} = \delta I(t). \end{cases}$$

где $\beta \in [0; \infty)$ – скорость распространения инфекции в популяции, а $\delta \in [0; \infty)$ – величина, обратная времени болезни (лечения), которую можно интерпретировать как скорость выздоровления.

На данный момент об эпидемии коронавируса известно, что течение болезни происходит с наличием инкубационного периода. Это значит, что восприимчивые не заболевают сразу после контакта с инфицированными, т.е. присутствует промежуточный класс, в котором находятся индивиды при переходе из класса восприимчивых в инфицированные. Наличие промежуточного звена не позволяет использовать модель SIR для построения модели распространения коронавирусной инфекции.

1.2. SEIR-модель

Инфекционные заболевания могут иметь скрытый (латентный, инкубационный) период заболевания - период от момента заражения до проявления первых симптомов.

О коронавирусной инфекции известно о наличии такого инкубационного периода. Длительность его составляет 1-14 дней, в среднем 5-7 дней. Поэтому для построения необходимо использовать модель, учитывающую способность вируса находиться в латентном состоянии.

Одной из моделей, учитывающих эту особенность развития инфекции, является модель SEIR (Susceptible — Exposed — Infected — Removed) [2]. Популяция в данной модели делится на 4 непересекающихся класса:

- $S(t)$ (susceptible) – восприимчивые (предрасположенные к заболеванию неинфицированные индивиды);
- $E(t)$ (exposed) – латентные (инфицированные в инкубационном периоде, еще не заразные);
- $I(t)$ (infected) – инфицированные (болеющие, распространяющие заболевание);
- $R(t)$ (removed) – невосприимчивые (выбывшие из класса инфицированных в случае выздоровления или смерти).

Популяцию также будем считать фиксированной во время вспышки эпидемии, поэтому в любой момент времени $t \in [0; \infty)$ будет выполняться равенство:

$$S(t) + E(t) + I(t) + R(t) = N.$$

SEIR-модель без учета изменения популяции может быть описана следующей системой дифференциальных уравнений:

$$\begin{cases} \frac{dS(t)}{dt} = -\beta \frac{S(t)I(t)}{N}, \\ \frac{dE(t)}{dt} = \beta \frac{S(t)I(t)}{N} - \gamma E(t), \\ \frac{dI(t)}{dt} = \gamma E(t) - \delta I(t), \\ \frac{dR(t)}{dt} = \delta I(t). \end{cases}$$

где коэффициент β – скорость распространения инфекции, δ – скорость выздоровления, γ – величина, обратная продолжительности инкубационного периода.

SEIR-модель по сравнению с SIR учитывает наличие инкубационного периода, что может быть использовано для построения простой модели распространения коронавирусной инфекции в Санкт-Петербурге. Однако рассмотренных в SEIR-модели четырех классов недостаточно для моделирования более узких групп, таких как умершие и выздоровевшие. В рассмотренной SEIR-модели два этих показателя объединены в один класс, что делает невозможным построение модели распространения эпидемии коронавируса в Санкт-Петербурге.

Еще одним важным недостатком SEIR-модели для построения модели коронавирусной инфекции является наличие определенных групп людей, перенесших новый вирус бессимптомно. Данная группа уже переболела, но ее нельзя отнести к выздоровевшим, так как факта болезни этой группы зафиксировано не было.

Дополнительно к указанным выше требованиям необходимо учитывать изменчивость условий распространения коронавируса, которую можно описать с помощью коэффициентов модели. В реальных условиях на эти параметры сильно влияют многие факторы. В частности, принимаемые решения, такие как социальное дистанцирование, ношение масок, ограничения работы предприятий и образовательных организаций - все это позволяют снизить коэффициент распространения инфекции, снизить нагрузку на медицинские организации, что благотворительно повлияет на качество ухода за больными, вследствие чего повысится скорость выздоровления, уменьшится летальность заболеваний, связанных с эпидемией.

С учетом вышесказанного, рассмотрим еще одну модель, удовлетворя-

ющую необходимым условиям.

1.3. Обобщенная SEIR-модель

Рассмотрим модель SEIQRDP (Susceptible — Exposed — Infected — Quarantined — Removed — Death — Insusceptible), описанную китайскими учеными в [7]. Модель представляет собой обобщение классической SEIR-модели и включает в себя новые классы, учитывает длительность инкубационного периода и выздоровления, коэффициенты распространения инфекции.

Обобщенная SEIR-модель делит всю популяцию на 7 различных классов, определяющих состояние индивида в момент времени t :

- $S(t)$ (susceptible) – восприимчивые (предрасположенные к заболеванию неинфицированные индивиды);
- $P(t)$ (insusceptible) – невосприимчивые (устойчивые к заболеванию, или перенесшие его бессимптомно);
- $E(t)$ (exposed) – латентные (инфицированные в инкубационном периоде (бессимптомные), еще не заразные);
- $I(t)$ (infected) – инфицированные (инфицированные в инкубационном периоде (бессимптомные), распространяющие заболевание);
- $Q(t)$ (quarantined) – карантинные (инфицированные, распространяющие заболевание, болеющие и находящиеся в карантине);
- $R(t)$ (recovered) – выздоровевшие;
- $D(t)$ (death) – умершие.

Популяция считается фиксированной во время вспышки, т.е. не учитывается рождаемость и смертность от прочих факторов. Поэтому в любой момент времени t популяция равна N , т.е. выполняется равенство:

$$S(t) + P(t) + E(t) + I(t) + Q(t) + R(t) + D(t) = N.$$

Согласно модели, в начальный момент времени вся популяция делится на две группы: восприимчивые S и невосприимчивые P . При занесении вируса в популяцию восприимчивый индивид S может стать латентным E после контакта с вирусом или другими инфицированными I . В латентном состоянии E болезнь развивается и индивид переходит к классу инфицированных I , а затем в класс болеющих Q . Для лиц из класса болеющих Q возможны два исхода: выздоровление R или смерть D .

Схематичное изображение перехода между выделенными классами представлено на рисунке 1.

Переход между классами осуществляется с помощью следующих переменных:

- α – уровень защищенности ;
- β – скорость заражения;
- γ – величина, обратная среднему времени инкубационного периода;
- δ – величина, обратная среднему времени лечения (карантина);
- λ – скорость выздоровления;
- κ – уровень смертности.

Каждая из указанных переменных должна принадлежать интервалу $[0; +\infty)$. Эти параметры модели позволяют учитывать влияние мер по

борьбе с коронавирусной инфекцией: например, введение режима самоизоляции, карантина, ношение защитных средств в виде перчаток и масок, ограничение контактов болеющих, закрытие для посещений мест с большим скоплением людей приводит к снижению количества потенциально восприимчивых людей за счет ограничения возможных мест и способов заражения конкретного человека. Вследствие этого, повышается уровень защищенности населения α и снижается скорость передачи вируса между людьми β . Конечно, в зависимости от обеспеченности рассматриваемой территории необходимыми для лечения больных лекарствами, медицинским персоналом, инфекционными койками в стационарах, принятые меры также влияют на скорость выздоровления λ и уровень смертности κ .

С учетом введенных классов и параметров, обобщенная SEIR-модель

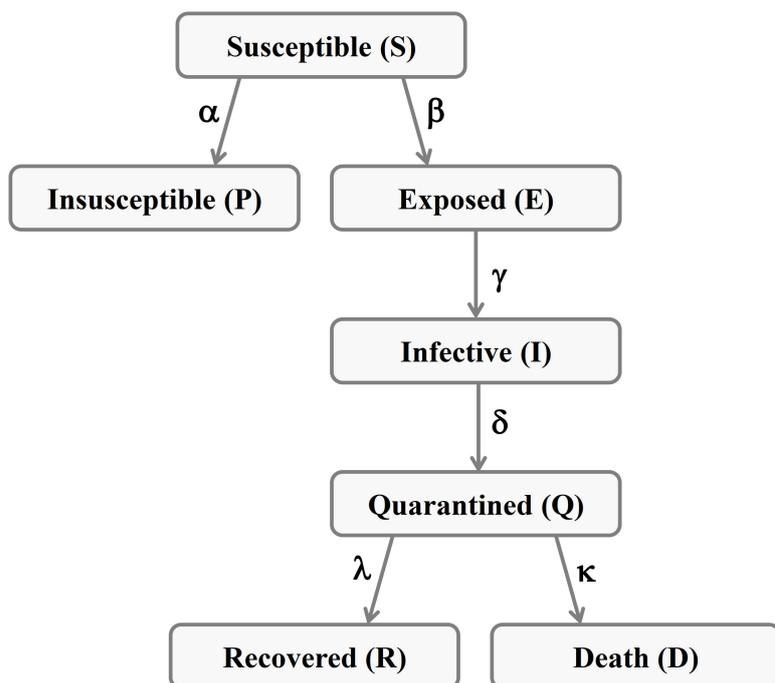


Рис. 1. Схема обобщенной SEIR-модели

описывается системой дифференциальных уравнений (1):

$$\left\{ \begin{array}{l} \frac{dS(t)}{dt} = -\beta \frac{S(t)I(t)}{N} - \alpha S(t), \\ \frac{dE(t)}{dt} = \beta \frac{S(t)I(t)}{N} - \gamma E(t), \\ \frac{dI(t)}{dt} = \gamma E(t) - \delta I(t), \\ \frac{dQ(t)}{dt} = \delta I(t) - \lambda Q(t) - \kappa Q(t), \\ \frac{dR(t)}{dt} = \lambda Q(t), \\ \frac{dD(t)}{dt} = \kappa Q(t), \\ \frac{dP(t)}{dt} = \alpha S(t). \end{array} \right. \quad (1)$$

Модель SEIQRDP (обобщенная SEIR-модель) позволяет учитывать классы невосприимчивых к болезни, умерших и выздоровевших, необходимые для моделирования эпидемиологической ситуации, связанной с распространением коронавируса в Санкт-Петербурге. Теперь для дополнительного учета влияния изменчивых внешних факторов будем задавать параметры λ (скорость выздоровления) и κ (уровень смертности) в виде функций времени t следующим образом [4]:

$$\lambda(t) = \lambda_0(1 - e^{-\lambda_1 t}), \quad (2)$$

$$\kappa(t) = \kappa_0 e^{-\kappa_1 t}, \quad (3)$$

где $\lambda_0, \lambda_1, \kappa_0, \kappa_1$ - неотрицательные константы.

Выбранные для моделирования распространения коронавирусной инфекции в Санкт-Петербурге функции (2) и (3) позволяют учитывать выявленные (в том числе на фактических значениях других эпидемий) закономерности:

- со временем скорость выздоровления больных увеличивается, что связано с повышением уровня знаний ученых о новом коронавирусе, получением врачами практического опыта лечения (поскольку на данный момент не существует вакцины, а лечение симптоматическое, врачи вынуждены применять новые, опробованные на первых пациентах методы и комбинации лекарств);
- смертность со временем будет снижаться, в том числе из-за повышения качества ухода и лечения пациентов, изобретения и использования новых лекарств и вакцин.

Модернизированную таким образом обобщенную SEIR-модель, которая позволяет учесть новые группы населения и динамические параметры, будем использовать для построения модели распространения коронавирусной инфекции в Санкт-Петербурге в следующей главе.

Глава 2. Эпидемиологическая ситуация в Санкт-Петербурге

В Санкт-Петербурге первые случаи заболевания коронавирусной инфекцией были выявлены 3 марта согласно официальным источникам [12] и [13]. Также организован и проводится комплекс противоэпидемических и профилактических мероприятий, включающий в себя введение режима самоизоляции, наблюдение за прибывшими из неблагополучных стран, ограничение контактов, тестирование и лечение людей с симптомами острых респираторных инфекций (ОРВИ) и непосредственно симптомами коронавирусной инфекции, приостановлена, ограничена или переведена на удаленный режим деятельность многих предприятий и организаций, организован переход работников и обучающихся образовательных организаций на дистанционное образование. Предпринятые меры помогли немного сдержать всплеск заражения коронавирусом, однако несоблюдение рекомендаций уже в апреле-мае 2020 года или слишком ранняя отмена ограничительных мер могут привести к повторному росту заболеваемости. В целях анализа и прогнозирования эпидемиологической ситуации в Санкт-Петербурге была программно реализована обобщенная SEIR-модель в среде MATLAB на основе [7] и [14].

Основными показателями для построения модели и прогнозирования эпидемиологической ситуации, наглядно показывающими развитие эпидемии, являются количество выявленных заболевших, выздоровевших, умерших и количество активных случаев (болеющих в данный момент времени).

2.1. Автоматический подбор параметров модели

На основе рассмотренной в предыдущей главе обобщенной SEIR-модели построим математическую модель распространения коронавирусной инфекции в Санкт-Петербурге без учета рождаемости и смертности. В таком случае численность популяции, в данном случае – численность населения Санкт-Петербурга, обозначим $N = 5\,393\,000$ человек (по состоянию на 1 января 2020 года [11]).

В качестве основных показателей развития эпидемиологической ситуации будем рассматривать следующие группы:

- R – выздоровевшие, соответствует классу $R(t)$ обобщенной SEIR-модели;
- D – умершие от коронавируса, соответствует классу $D(t)$ обобщенной SEIR-модели;
- Q – болеющие, соответствует классу $Q(t)$ обобщенной SEIR-модели.

Дополнительно к этим трем группам будем рассматривать C (Confirmed) – общее число подтвержденных случаев заболевания COVID-19 в Санкт-Петербурге.

Данные по рассматриваемым четырем группам предоставляются официальными источниками [12, 13] и будут использованы в качестве фактических значений.

Поскольку обобщенная SEIR-модель не учитывает при построении группу C , будем находить численные значения этой группы по следующему равенству:

$$C = Q + R + D.$$

Классы S , P , E и I не будут входить в общее число подтвержденных случаев, что следует из описания этих классов в обобщенной SEIR-модели.

Система дифференциальных уравнений (1) выбранной модели не является линейной, поэтому не разрешима аналитически. Для решения данной системы уравнений с известными параметрами обобщенной SEIR-модели α , β , γ , δ , λ_0 , λ_1 , κ_0 и κ_1 в программной реализации в среде MATLAB используются численные методы решения (формулы Рунге-Кутты четвертого порядка).

На данный момент известны приблизительные значения двух величин, используемых для определения параметров: средняя длительность инкубационного периода и средняя продолжительность лечения. Известно, что длительность инкубационного периода составляет 1-14 дней, а средние значения – 5-7 дней. Зная, что параметр γ задан как величина, обратная среднему времени инкубационного периода, можно считать параметр γ принадлежащим рассчитанному интервалу $[0, 14; 0, 2]$. Аналогичным образом можно сузить границы на параметр δ , исходя из средних значений длительности лечения в 10-20 дней: будем предполагать δ принадлежащим интервалу $[0, 05; 0, 1]$. Параметры γ и δ , тем не менее, нельзя считать ограниченными только в заданных интервалах (скорее, эти интервалы можно назвать рекомендуемыми к использованию в построении модели), так как эпидемия коронавирусной инфекции еще не полностью изучена.

Остальные параметры в настоящий момент не могут быть рассчитаны с помощью других значений и показателей. Для того, чтобы определить значения всех параметров обобщенной SEIR-модели в среде MATLAB разработан механизм автоматического расчета параметров с помощью метода наименьших квадратов. Полученные значения позволяют построить приближение модельных значений к фактическим.

Построение модели с автоматически рассчитанными параметрами во многих случаях позволяет получить близкие к фактическим модельные

значения рассматриваемых показателей эпидемиологической ситуации в Санкт-Петербурге. При необходимости параметры могут быть изменены в целях улучшения качества построенной модели или отдельных ее показателей. С помощью корректировки показателей также можно получить различные сценарии развития эпидемиологической ситуации и построить для них прогнозные значения на требуемый период.

Стоит однако учитывать, что в условиях быстро изменяющихся внешних факторов эпидемиологическая ситуация также будет изменяться. Поэтому построенная на актуальных значениях обобщенная SEIR-модель с автоматическим расчетом параметров позволяет построить качественные прогнозы только на краткосрочном периоде. Построенные прогнозы позволяют в целом оценить текущую ситуацию и проанализировать необходимость и возможность дополнительного введения или частичного снятия противоэпидемических мер для предотвращения ухудшения эпидемиологической ситуации в Санкт-Петербурге.

Рассмотрим несколько примеров построения обобщенной SEIR-модели с автоматическим подбором параметров.

Пример 1.

Построим обобщенную SEIR-модель с автоматическим подбором параметров для Санкт-Петербурга. В качестве ретроспективных данных будем использовать значения с 3 марта до 1 мая 2020 года. Ограниченный таким образом ряд содержит 60 значений.

В результате применения автоматического расчета параметров получим следующие значения:

$$\begin{aligned} \alpha &= 0,0319; & \beta &= 1,1453; \\ \gamma &= 0,1346; & \delta &= 0,1035; \\ \lambda_0 &= 0,0483; & \lambda_1 &= 0,0113; \\ \kappa_0 &= 1,0000; & \kappa_1 &= 0,1729. \end{aligned}$$

Значения α и β положительны, что соответствует естественным ограничениям на параметры. Из значений параметров γ и δ можно вычислить среднее время инкубационного периода и карантина: они равны 7,43 и 9,66 дней соответственно. Полученные значения выходят за границы принятых средних значений, однако близки к ним. Параметры $\lambda(t)$ и $\kappa(t)$ согласно (2) и (3) быстро возрастают и убывают соответственно.

Построим модель по полученным параметрам с прогнозными значениями до 31 августа. На рисунке 2 представлены 4 рассматриваемые группы: заболевшие, болеющие, выздоровевшие и умершие. Точками на графике показаны фактические значения.

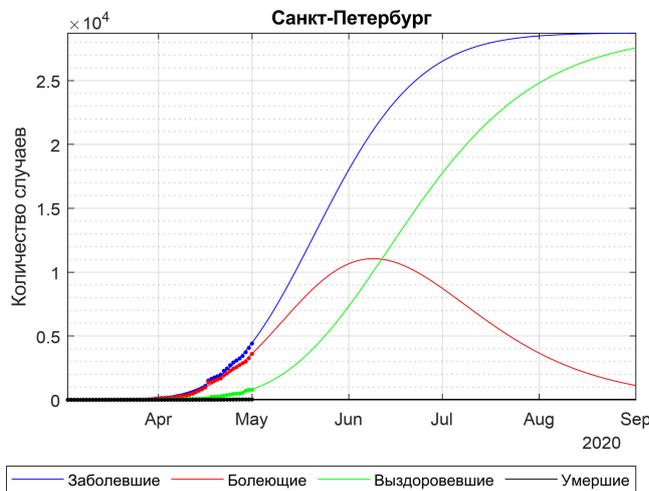


Рис. 2. Моделирование развития коронавирусной инфекции (пример 1)

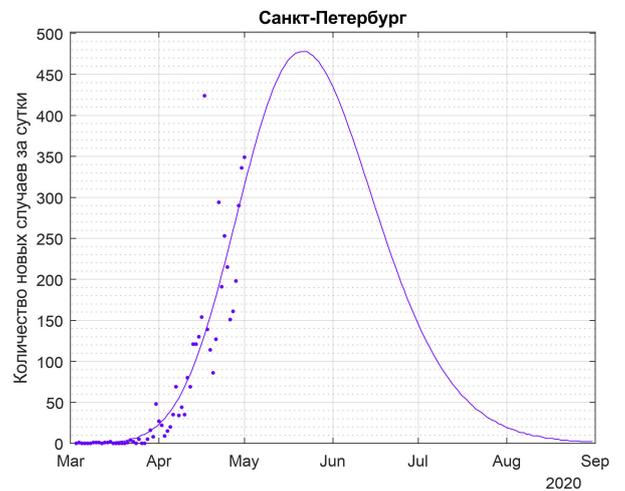


Рис. 3. Количество новых случаев в сутки (пример 1)

Согласно полученным результатам, перевес выздоровевших над болеющими произойдет в первой половине июня, а общее количество зафиксированных случаев заражения коронавирусом стабилизируется (перестанет возрастать) к началу сентября и приблизится к значению в 30 тысяч человек.

Для прогнозирования эпидемиологической ситуации важным показателем является также ежедневный прирост количества заболевших. Согласно результатам моделирования (рис. 3) пик эпидемии ожидается во второй половине мая, после чего эпидемия пойдет на спад.

Для оценки качества построенной модели построен график отклонений модельных значений от фактических (рис. 4) для данных за период с 20 апреля по 18 мая и включает в себя модельные значения как на ретроспективном, так и на прогнозном периоде. Как следует из графика, ошибки модели не превышают 10%, при этом средняя ошибка модели за весь рассматриваемый период составляет 4,9%.

Пример 2.

Построим еще одну модель для Санкт-Петербурга с автоматическим подбором параметров. Теперь рассмотрим данные за более длинный период: с 3 марта по 10 мая 2020 года. Новый ряд данных содержит 69 значений.

Применив автоматический расчет параметров, получим следующие значения:

$$\begin{aligned}\alpha &= 0,0344; & \beta &= 0,9518; \\ \gamma &= 0,1858; & \delta &= 0,0812; \\ \lambda_0 &= 0,0396; & \lambda_1 &= 0,0159; \\ \kappa_0 &= 1,0000; & \kappa_1 &= 0,1682.\end{aligned}$$

Также как и в предыдущем примере, проведем анализ полученных

параметров. Значения α стали ненамного больше, т.е. можно сказать, что скорость выздоровления с 1 мая выросла. Скорость заражения β снизилась со значения 1,1453 до 0,9518. На основе параметра γ рассчитаем среднее время инкубационного периода: $\gamma^{-1} = 5,38$ дней. Полученная длительность латентного периода близко к среднему значению, равному 5-7 дней. Рассчитаем среднее время болезни на основе параметра δ : $\delta^{-1} = 12$ дней – тоже входит в промежуток средних значений. Параметры $\lambda(t)$ и $\kappa(t)$ остались близкими к значениям предыдущего примера.

Используя полученные параметры, построим обобщенную SEIR-модель на следующие 3 месяца (до 31 августа). В результате построения модели получаем модельные значения каждой из рассматриваемых в работе групп: заболевшие, болеющие, выздоровевшие и умершие. Модельные и фактические значения представлены на рисунке 5, фактические значения показаны на графике точками.

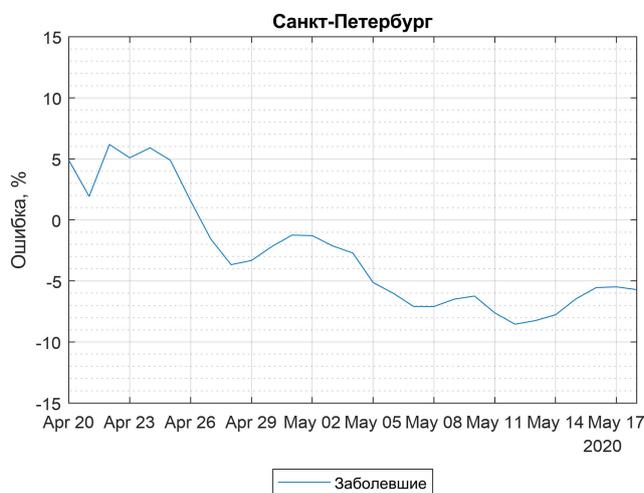


Рис. 4. Отклонения модельных значений от фактических (пример 1)

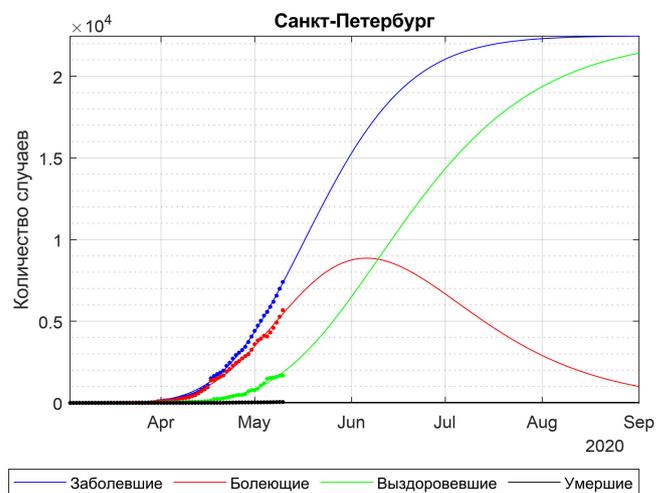


Рис. 5. Моделирование развития коронавирусной инфекции (пример 2)

Согласно полученным результатам, количество выздоровевших превысит количество болеющих в середине июня. Общее число заболевших коронавирусной инфекцией, также как и в предыдущем примере, стабилизируется к сентябрю, однако составит немного меньшее число случаев – 23 тысячи человек.

Число новых выявленных случаев заражения коронавирусом, спрогнозированное с помощью обобщенной SEIR-модели, вместе с фактическими значениями представлено на рисунке 6. Согласно построенной модели, пик эпидемии ожидается в середине мая. Далее можно ожидать постепенное снижение числа новых заболевших. Начиная с июля модельные значения не будут превышать 100 новых случаев в день.

Оценим качество построенной модели, построив график отклонения модельных значений от фактических (рис. 7) для данных с 20 апреля по 18 мая 2020 года. По данным, представленным на графике, можно сделать вывод, что ошибки модели не превышают значения в 5%. Средняя ошибка модели на рассматриваемом периоде при этом составляет 1,98%.

Примеры 1 и 2 демонстрируют развитие эпидемиологической ситуации в Санкт-Петербурге, описанное с помощью обобщенной SEIR-модели с автоматическим подбором параметров. Анализ параметров и характеристик двух построенных моделей показывает, что модель из примера 2, содержащая более актуальные данные, точнее описывает переход заболевших коронавирусной инфекцией людей между тремя классами (болеющих, выздоровевший или умерший). Параметры моделей изменились, однако эти изменения совпадают с выявленными на примерах других стран закономерностями. Например, с течением времени эпидемические процессы имеют свойство затухать, что связано в том числе с повышением скорости

выздоровления и снижением скорости заражения.

Количество заболевших коронавирусной инфекцией в самом начале развития эпидемии имеет довольно продолжительный по времени «хвост» с небольшим числом зафиксированных случаев. При моделировании с помощью обобщенной SEIR-модели избавление от части данных может оказывать положительное влияние на характеристики эпидемиологического процесса. В следующем примере будет рассмотрен случай, фактические значения в котором начинают учитываться с более поздней даты.

Пример 3.

Построим обобщенную SEIR-модель с автоматическим подбором параметров. В качестве фактических значений будем использовать ограниченный особым образом ряд: актуальные на 10 мая данные, ограниченные слева значением, равным 4% от актуального количества заболевших.

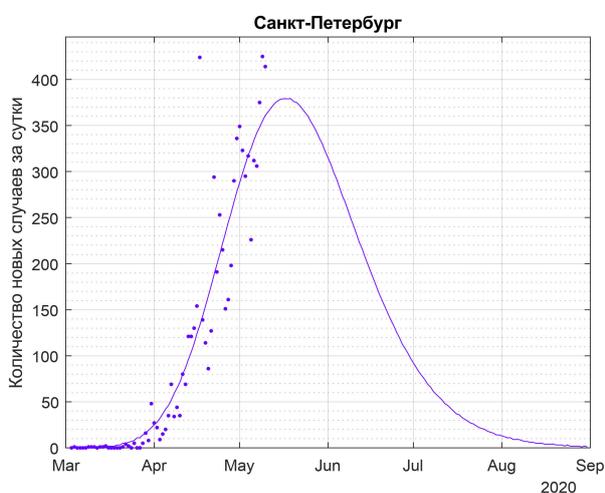


Рис. 6. Количество новых случаев в сутки (пример 2)

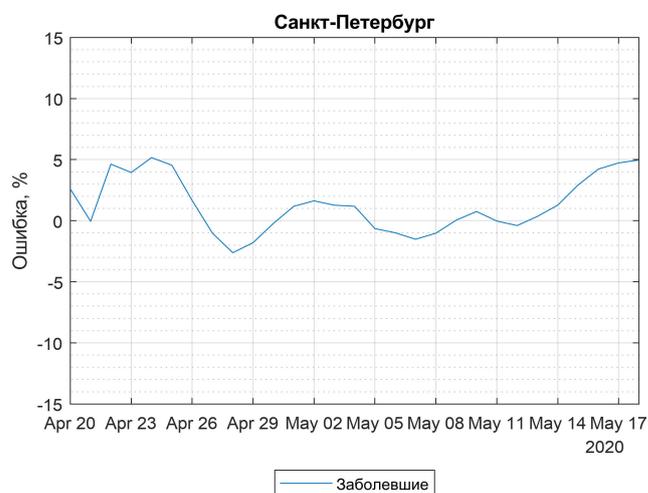


Рис. 7. Отклонения модельных значений от фактических (пример 2)

Общее число выявленных случаев заражения коронавирусом на 10 мая – 7404 человека. 4% от этого количества составляет 296 человек. Значит для построения модели используем данные с 8 апреля (в этот день 329 заболевших, что больше порогового значения 296) по 10 мая. Ограниченный таким образом ряд содержит 33 значения.

В результате применения программы с автоматическим расчетом параметров получим следующие значения:

$$\begin{aligned}\alpha &= 0,0064; & \beta &= 0,2850; \\ \gamma &= 0,6278; & \delta &= 0,1962; \\ \lambda_0 &= 0,0250; & \lambda_1 &= 0,1734; \\ \kappa_0 &= 0,1883; & \kappa_1 &= 1,9901.\end{aligned}$$

Рассмотрим подробнее полученные параметры. По сравнению с примерами 1 и 2 коэффициенты α и β значительно уменьшились, что объясняется введенным ограничением на точку начала построения модели. Рассчитаем на основе параметров γ и λ средние значения инкубационного периода и периода карантина. Среднее значение длительности пребывания вируса в латентном состоянии составляет 1,6 день. Среднее время болезни 5 дней. Полученные значения снова несравнимы с полученными для другого ряда данных значениями, однако значительно меньшие значения позволяют сделать вывод о том, что даже при небольшом ограничении временного ряда пороговым значением происходит следующее: модель начинает учитывать людей, заболевших ранее введенного порога, а выздоровевшие после него. За счет такого сдвига и получается, что средний инкубационный период и период карантина значительно меньше реальных значений. Из-за того же введенного порогового значения поменялись и параметры λ и κ .

Используя полученные автоматическим подбором параметры, постро-

им обобщенную SEIR-модель на следующие несколько месяцев. Для того, чтобы видеть приближающийся конец эпидемии, построим модель до 30 сентября. Полученные модельные значения рассматриваемых групп вместе с фактическими значениями (на графике показаны точками) представлены на рисунке 8.

Согласно графику полученных результатов, количество болеющих достигнет пикового значения в конце июня - начале июля. Тогда же количество выздоровевших превысит количество болеющих. Общее число заболевших стабилизируется к началу октября на значении, приближенном к 43 тысячам человек.

На рисунке 9 представлен график модельных значений количества новых случаев заражения коронавирусом вместе с отмеченными точками фактическими значениями. Согласно построенной модели, пик эпидемии ожидается в первой трети июня, после чего постепенно эпидемия будет

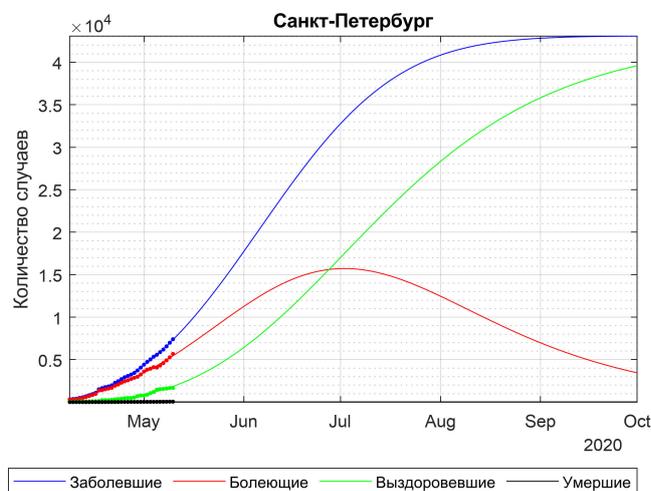


Рис. 8. Моделирование развития коронавирусной инфекции (пример 3)

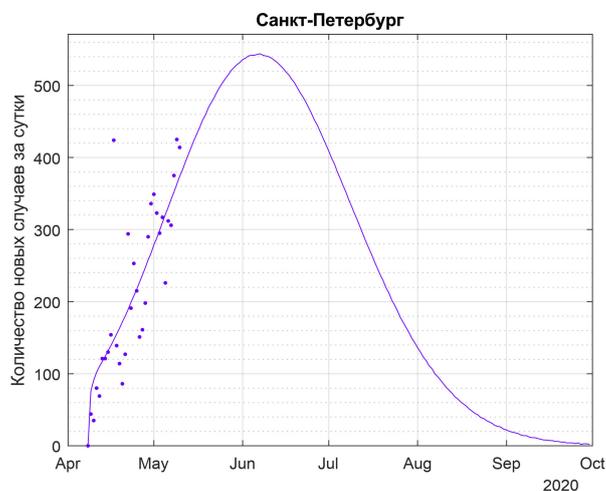


Рис. 9. Количество новых случаев в сутки (пример 3)

затухать.

Построим график отклонения построенных модельных значений от фактических (рис. 10) для данных с 20 апреля до 18 мая. По данным, представленным на графике, можно сказать, что ошибки модели на рассматриваемом периоде не превышают 5%. При этом рассчитанная средняя ошибка модели на данном периоде составляет 1,38%.

Рассмотренный в примере 3 способ построения модели позволяет получить прогнозные значения, более близкие к реальным значениям за этот период, однако имеет и недостатки. Одним из основных недостатков можно считать недостаточную интерпретируемость автоматически рассчитанных параметров. Несмотря на этот недостаток, построение таких моделей применяется с целью получения более точного прогноза.

Следует отметить, что от выбранного порогового значения также будет зависеть и сам прогноз. При выборе слишком высоких пороговых значений, модель становится неадекватной в части продолжительного роста

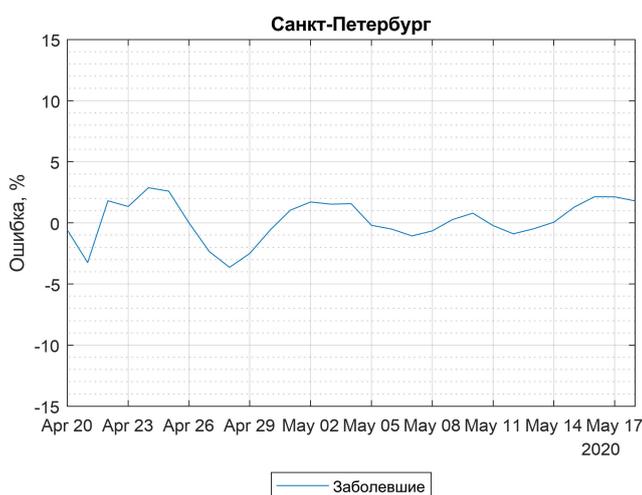


Рис. 10. Отклонения модельных значений от фактических (пример 3)

количества заболевших вплоть до достижения значения численности населения Санкт-Петербурга. Поскольку реальные условия отличаются от модельных, с учетом введения ограничительных мер, работы ученых над разработкой вакцины и многих других причин, охват эпидемией всей численности населения города практически невозможен.

2.2. Модель развития эпидемии в Санкт-Петербурге

В данном разделе будет построена модель развития коронавирусной инфекции в Санкт-Петербурге за период до 18 мая 2020 года, а также спрогнозированы значения общего числа заболевших, болеющих, выздоровевших и умерших до 31 августа. Для приближения модели к реальным данным воспользуемся описанным в предыдущем разделе автоматическим подбором параметров для обобщенной SEIR-модели. Будем использовать данные без ограничения на временную ось, т.е. с 1 зафиксированного случая.

В результате автоматического подбора параметров получим следующие значения:

$$\begin{aligned}\alpha &= 0,0100; & \beta &= 1,1167; \\ \gamma &= 0,3086; & \delta &= 0,4977; \\ \lambda_0 &= 0,0191; & \lambda_1 &= 1,6154; \\ \kappa_0 &= 0,9973; & \kappa_1 &= 0,1963.\end{aligned}$$

С помощью рассчитанных автоматически параметров построим модель и выведем на графики модельные и фактические значения количества человек в каждой из рассматриваемых групп (рис. 11) и ежедневный прирост заболевших (рис. 12).

Построим также график отклонения модельных значений от фактических на промежутке с 1 по 18 мая (рис. 13). Из полученных значений видно, что отклонения количества умерших растет. Это происходит из-за выбранных автоматически параметров: в данном случае подобранная комбинация параметров κ_0 и κ_1 приводит к резкому росту смертности в начале развития эпидемии и останавливается уже к середине периода построения модели на значении 27. Для устранения этой проблемы настроим пара-

метры модели вручную, основываясь на подобранных автоматически. В результате настройки получаем следующие параметры:

$$\alpha = 0,0089; \quad \beta = 1,0694;$$

$$\gamma = 0,3544; \quad \delta = 0,5158;$$

$$\lambda_0 = 0,0195; \quad \lambda_1 = 1,9997;$$

$$\kappa_0 = 0,0053; \quad \kappa_1 = 0,0375.$$

Построенная с новыми параметрами модель представлена на рисунке 14. Из сравнения графиков двух построенных моделей с разными параметрами следует, что общее число заболевших в новой модели увеличится до значения в 29 тысяч человек к началу сентября. Из графиков прироста заболевших (рис. 12 и 15 – для новой модели) можно сделать вывод, что Санкт-Петербург в конце мая находится на пике эпидемии, после чего последует постепенное снижение новых заболевших.

Проанализировав график отклонений новой модели от фактических

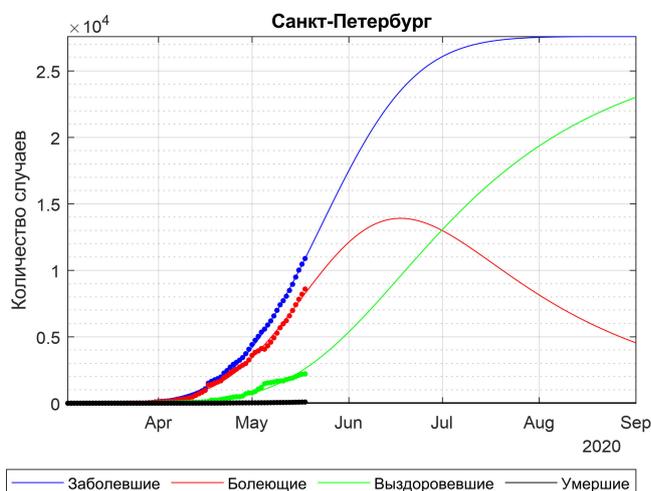


Рис. 11. Моделирование развития коронавирусной инфекции

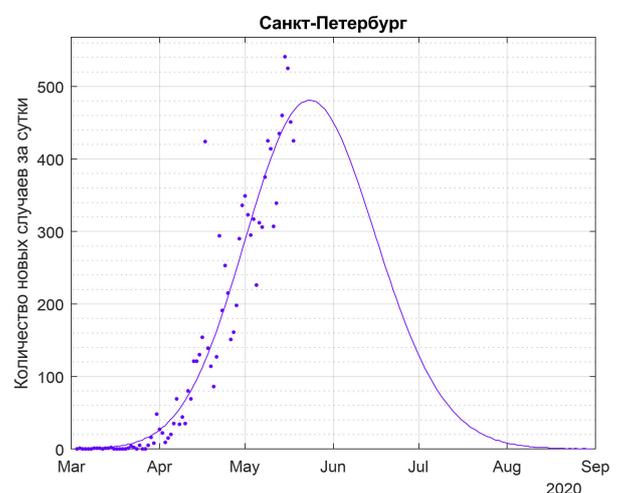


Рис. 12. Количество новых случаев в сутки

значений (рис. 16) можно сделать следующий вывод: модель с новыми параметрами лучше адаптирована к фактическим значениям количества умерших. Также можно заметить, что ряды данных по выздоровевшим и умершим имеют значительно большие ошибки, чем группы болеющих и выздоровевших. Такое влияние на поведение этих рядов оказывают выбранные функции (2) и (3) для параметров λ и κ . Измененный параметр κ новой модели обеспечивает запаздывание первых смертей от коронавирусной инфекции, после чего происходит более плавный рост модельных значений группы умерших. При необходимости, для еще большего приближения модельных значений к фактическим, функции (2) и (3) могут быть изменены.

Несмотря на существенное изменение модельных значений для количества умерших вследствие изменения параметров, вид оставшихся групп практически не изменился.

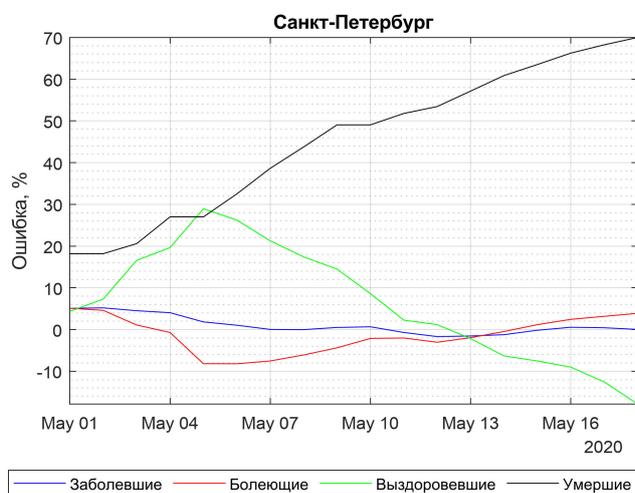


Рис. 13. Отклонения модельных значений от фактических

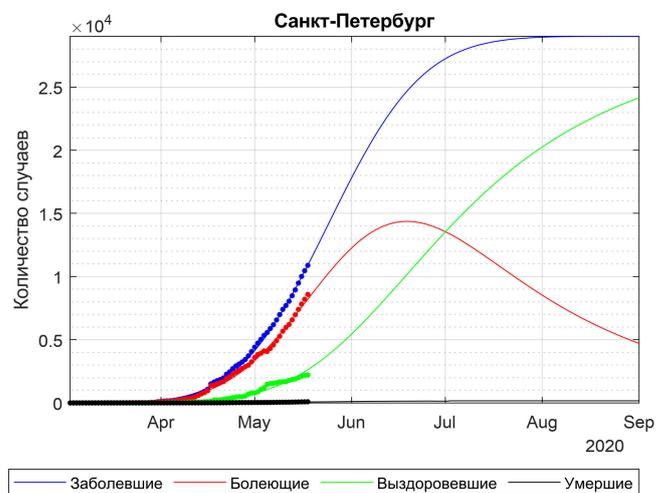


Рис. 14. Моделирование развития коронавирусной инфекции

Рассмотрим подробнее общее число заболевших. Как уже было выявлено ранее, изменение параметров модели не так сильно повлияло на этот ряд данных, по сравнению изменениями функции количества умерших: произошел небольшой рост итоговых значений числа заболевших в Санкт-Петербурге. Отклонения модельных значений от ретроспективных за последний месяц не превышает 10%, средняя ошибка аппроксимации за этот период составляет 3,9%, а за период с 1 по 18 марта 2,0%.

Дополнительная настройка параметров модели после автоматического подбора позволила снизить отклонения модельных значений от фактических на ретроспективном периоде, тем самым улучшив качество построенной модели.

Близкие к фактическим полученные модельные значения позволяют использовать данную модель для построения прогноза. Наиболее качественными можно считать результаты прогнозирования на краткосрочном

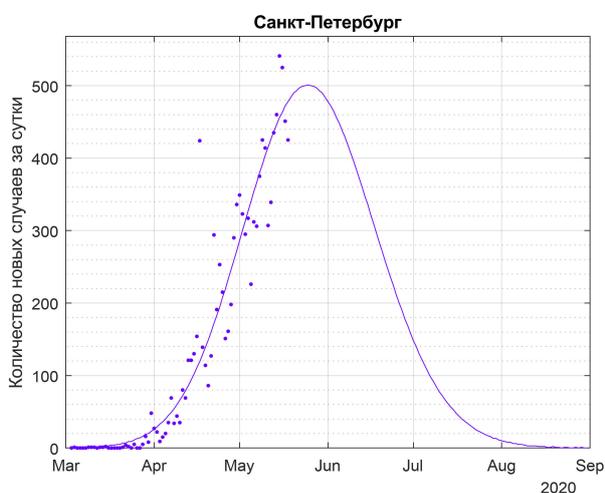


Рис. 15. Количество новых случаев в сутки

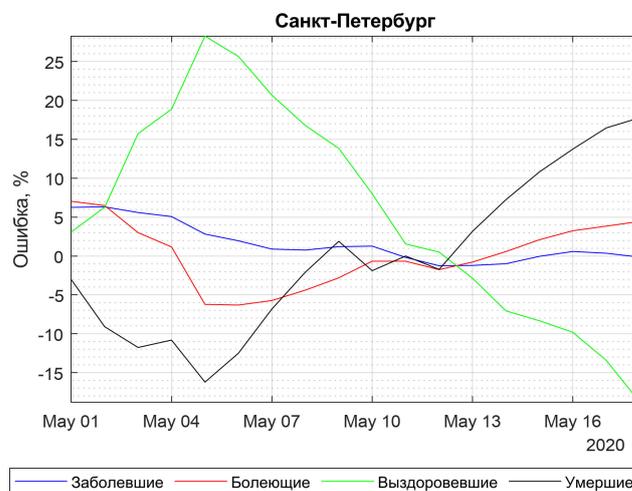


Рис. 16. Отклонения модельных значений от фактических

периоде, поскольку с течением времени могут измениться факторы, влияющие на распространение инфекции, например, ужесточение режима самоизоляции может снизить уровень заболеваемости, или же наоборот, несоблюдение или резкое снятие существующих ограничений приведет к новой волне роста количества заболевших.

Для каждой из четырех рассматриваемых групп (заболевшие, болеющие, выздоровевшие и умершие) в таблице приложения приведены фактические значения с 1 по 18 марта и численные значения построенной модели за период с 1 по 31 марта.

Выводы

В результате рассмотрения моделей SIR, SEIR и построения обобщенной SEIR модели для анализа и прогнозирования эпидемиологической ситуации в Санкт-Петербурге можно сделать следующие выводы:

- Модели SIR и SEIR в их классической постановке не учитывают многие параметры, необходимые для построения модели развития коронавирусной инфекции, а имеющиеся параметры являются фиксированными величинами.
- Обобщенная SEIR-модель за счет введения дополнительных групп населения и большего числа параметров, в том числе предусматривающих изменение с течением времени, позволяет учесть изменчивость условий распространения эпидемии.
- Применение программной реализации обобщенной SEIR-модели с автоматическим подбором параметров позволяет получить реалистичную модель изменения общей численности заболевших, количества болеющих, выздоровевших и умерших людей.
- Представление результатов работы программы в графическом виде дает возможность быстро оценивать качество построенной модели. Дополнительная настройка параметров модели позволяет скорректировать выявленные несоответствия модельных значений с реальными поступающими данными, что приводит к повышению качества модели.
- С учетом внесенных изменений построена качественная модель распространения коронавирусной инфекции для Санкт-Петербурга по данным с начала ее развития по 18 мая 2020 года. Построенная модель поз-

воляет спрогнозировать развитие эпидемии, однако с учетом возможности изменения влияющих на распространение коронавируса факторов полученный прогноз можно считать наиболее точным на краткосрочном периоде.

Заключение

В данной работе были рассмотрены несколько математических моделей распространения эпидемий. Произведен анализ этих моделей, после которого в качестве основы была выбрана обобщенная SEIR-модель.

На основании реальных данных распространения коронавирусной инфекции в Санкт-Петербурге модель была реализована в среде MATLAB. Программная реализация позволила также автоматически рассчитывать параметры модели. С помощью данной опции были проанализированы и визуализированы полученные на различных наборах данных результаты.

Также в работе построена модель распространения коронавируса в Санкт-Петербурге по состоянию на 18 мая 2020 года, учитывающая развитие эпидемии в четырех группах населения: заболевшие, болеющие, выздоровевшие и умершие. Автоматически рассчитанные, а затем скорректированные параметры позволили построить качественную модель и спрогнозировать развитие эпидемии.

Список литературы

1. Aalto A., Ameijeiras-Alonso J., Goncalves J., Husch A., Kemp F., Ley C., Magni S., Mombaerts L., Proverbio D., Skupin A. Assessing suppression strategies against epidemic outbreaks like COVID-19: the SPQEIR model // 2020.
2. Anderson R. M, May R. M. Infectious diseases of humans: Dynamics and control. Oxford University Press. 1991.
3. Borgonovo E., Lu X. Is Time to Intervention in the COVID-19 Outbreak Really Important? A Global Sensitivity Analysis Approach // 2020.
4. Chen L., Tan S. Real-Time Differential Epidemic Analysis and Prediction for COVID-19 Pandemic // 2020.
5. Edelstein-Keshet L. Mathematical Models in Biology // Society for Industrial and Applied Mathematics. 2005. P. 242-254.
6. Hethcote H. W. The Mathematics of Infectious Diseases // SIAM Review. 2000. V. 42. P. 599-653.
7. Hong L., Liangrong P., Yang W., Zhang D., Zhuge C. Epidemic analysis of COVID-19 in China by dynamical modeling // *medRxiv*. 2020.
8. Kermack W. O., McKendrick A. G. A contribution to the mathematical theory of epidemics // Proceedings of the royal society of London. 1927. P. 700–721.
9. Lopez L., Rodo X. A modified SEIR model to predict the COVID-19 outbreak in Spain: simulating control scenarios and multi-scale epidemics // 2020.

10. Weiss H. The sir model and the foundations of public health // MATerials MATematics. 2013. № 3. P. 1-17.
11. Петростат Социально-экономическое положение субъектов Российской Федерации, находящихся в пределах Северо-Западного федерального округа в 2019 году. Статистический бюллетень. 2020.
12. Сайт Роспотребнадзора РФ <https://www.rosпотребнадзор.ru/>
13. <https://стопкоронавирус.рф/>
14. <https://github.com/ECheynet/SEIR>

Приложение

Таблица фактических и модельных значений

Дата	Заболевшие		Выздоровевшие		Умершие		Болеющие	
	факт	модель	факт	модель	факт	модель	факт	модель
10.05	7404	7355	1674	1540	53	54	5677	5761
11.05	7711	7765	1681	1655	56	56	5974	6054
12.05	8050	8186	1784	1775	58	59	6208	6352
13.05	8485	8616	1848	1901	63	61	6574	6654
14.05	8945	9055	1899	2033	69	64	6977	6958
15.05	9486	9502	2005	2172	74	66	7407	7264
16.05	10011	9956	2109	2316	80	69	7822	7571
17.05	10462	10417	2175	2467	85	71	8202	7879
18.05	10887	10883	2209	2624	90	74	8588	8185
19.05		11354		2763		76		8515
20.05		11829		2930		79		8820
21.05		12307		3103		81		9123
22.05		12786		3281		84		9421
23.05		13267		3465		86		9716
24.05		13748		3655		89		10004
25.05		14228		3850		91		10287
26.05		14707		4051		94		10562
27.05		15183		4257		96		10830
28.05		15655		4467		99		11089
29.05		16124		4683		101		11340
30.05		16587		4904		103		11580
31.05		17044		5129		106		11809