

В.В. ЗАХАРОВ, Ю.Е. БАЛЫКИНА
**БАЛАНСОВАЯ МОДЕЛЬ ЭПИДЕМИИ COVID-19 НА ОСНОВЕ
ПРОЦЕНТНОГО ПРИРОСТА**

Захаров В.В., Балыкина Ю.Е. **Балансовая модель эпидемии COVID-19 на основе процентного прироста.**

Аннотация. В статье изучается возможность использования альтернативного подхода к прогнозированию статистических показателей эпидемии вируса нового типа. Представлен систематический обзор моделей прогнозирования эпидемий новых инфекций в зарубежной и российской научной литературе. Анализируется точность модели SIR при прогнозировании весенней волны эпидемии COVID-19 в России. В качестве альтернативного подхода к моделированию эпидемии предлагается использование вместо традиционной модели SIR новой дискретной стохастической модели распространения эпидемии CIR, основанной на балансе показателей эпидемии в текущий и прошлые моменты времени. Новая модель описывает динамику общего количества заболевших (С), общего количества выздоровевших и умерших (R) и числа активных случаев (I). Параметрами системы являются процентный прирост величины $S(t)$ и характеристика динамического баланса эпидемиологического процесса, впервые введенная в этой статье. Сформулирован принцип динамического баланса эпидемиологического процесса, предполагающий наличие у любого процесса свойства близости значений общего количества заболевших в прошлые периоды и значений общего количества выздоровевших и умерших в текущий момент времени. Для вычисления значений характеристики динамического баланса используется задача целочисленного программирования. Продемонстрировано, что в общем случае динамическая характеристика эпидемиологического процесса не является постоянной величиной. Эпидемиологический процесс, динамическая характеристика которого не является постоянной величиной, называется нестационарным. Для построения среднесрочных прогнозов показателей эпидемиологического процесса на промежутках стационарности эпидемиологического процесса разработан специальный алгоритм. Исследован вопрос об использовании этого алгоритма на промежутках стационарности и нестационарности. Приведены примеры применения модели CIR для построения прогнозов рассматриваемых показателей эпидемии в России в мае-июне 2020 года.

Ключевые слова: COVID-19, модели распространения, моделирование эпидемии новых вирусов, SIR-модели, прогнозирование.

1. Введение. Вспышка коронавирусной инфекции COVID-19, вызванная новым вирусом SARS-CoV-2, быстро распространилась по миру в конце 2019 года, затронув уже более 200 стран. 31 января 2020 года Всемирная организация здравоохранения (ВОЗ) объявила эту вспышку чрезвычайной ситуацией, имеющей международное значение, а 11 марта 2020 года классифицировала ее как пандемию. По состоянию на середину мая 2021 года в мире зарегистрировано более 159 млн. подтвержденных случаев COVID-19. По признанию ВОЗ, наличие своевременных математических моделей играет ключевую роль в принятии решений, основанных на фактах, лицах, определяющими политику здравоохранения. Рассмотрение инфекционных заболеваний

с математической точки зрения может выявить не только важные закономерности пандемии, но и возможные варианты эпидемического контроля. Математические модели передачи болезней помогают получить представление о динамике распространения инфекционных заболеваний и потенциальной роли различных видов стратегий вмешательства со стороны общественного здравоохранения [1]. Важно в условиях максимальной неопределенности своевременно выявлять параметры распространения новых инфекционных заболеваний, чтобы оценить потенциальные опасности и масштабы эпидемии.

Особенность пандемии COVID-19 состоит в отсутствии достоверной статистической информации о распространении вируса SARS-CoV-2 за прошлые годы. Это влечёт за собой проблему корректного использования имеющихся данных о параметрах развивающейся эпидемии. Научно-исследовательские группы в разных странах мира ведут работы по созданию методов краткосрочного и долгосрочного прогнозирования динамики распространения нового вируса.

Во втором разделе проанализировано современное состояние исследований по прогнозированию распространения эпидемий и описаны основные подходы к моделированию. В третьем разделе проводится анализ точности прогнозирования весенней волны эпидемии в России при использовании дискретной модели SIR. В четвертом разделе представлена новая балансовая модель распространения эпидемии на основе процентного прироста (CIR), вводится понятие характеристики динамического баланса эпидемии. В пятом разделе обсуждаются результаты применения модели CIR и метода прецедентов (CBR – case based reasoning) для построения краткосрочных прогнозов динамики эпидемии COVID-19 в России в мае-июне 2020 года в реальном времени.

2. Обзор научной литературы. Мы провели систематический обзор моделей прогнозирования эпидемий новых инфекций. Поиск производился в наукометрических базах данных Web of Science, Scopus, RCI. Дополнительно были проанализированы результаты поиска в базе Elsevier – одного из крупнейших европейских издательств. В качестве ключевых слов использовались “forecasting”, “prediction”, “model”, “emerging infection”, “coronavirus”, “covid”. Большая часть исследований, удовлетворяющих критериям поиска, касалась распространения эпидемии нового коронавируса SARS-CoV-2 (COVID-19). Модели эпидемии различаются по типу модели, способу получения исходных данных, гипотезе и распределению ключевых входных параметров. Все исследования старались ответить на следующие вопросы: (1) заразность рассматриваемого вируса, и связанная с этим оценка

значения базового репродуктивного числа; (2) оценка ключевых временных периодов распространения эпидемии (достижение пика, плато, затухание); (3) прогнозирование распространения инфекции (включая краткосрочное и долгосрочное прогнозирование); (4) оценка воздействия различных мер, направленных на сдерживание эпидемии. Также востребованным вопросом является влияние неизвестной ранее эпидемии на макро- и микроэкономические показатели регионов и стран.

В целом можно выделить четыре основных подхода к моделированию распространения инфекционных заболеваний [2]: камерные модели, модели на основе временных рядов, агентно-ориентированные модели и модели, построенные с использованием методов машинного обучения и эвристических подходов (рис. 1).



Рис. 1. Основные подходы к моделированию распространения эпидемий инфекционных заболеваний

Модели на основе временных рядов. Говоря о моделях на основе временных зависимостей, следует отметить, что регрессионный анализ и анализ временных рядов являются одними из самых известных методов прогнозирования распространения заболеваемости. Одной из наиболее часто используемых моделей является модель $ARIMA(p,d,q)$ – интегрированная модель авторегрессии – скользящего среднего. Аббревиатура $ARIMA$ носит описательный характер и отражает ключевые аспекты самой модели. AR (AutoRegressive) отвечает за авторегрессионную часть, т.е. за порядок запаздывания (p); I (Integrated) отвечает за степень разности (d); MA (Moving Average) отвечает за размер окна скользящего среднего, также называемого порядком скользящего среднего (q). Каждый из этих компонентов явно указывается в модели в качестве параметра.

Некоторые исследователи использовали $ARIMA$ для прогнозирования распространения пандемии новой коронавирусной инфекции. Так, в [3] авторы оценивали количество ежедневных новых случаев

заболевания в Иране. Результаты прогнозирования показывали в районе 9 558 новых случаев в день на конец апреля 2020 года. Официально на данную дату было зафиксировано 1 073 новых случая заражения. В работе [4] использовался метод простого скользящего среднего для прогнозирования подтвержденных случаев COVID-19 в Пакистане. Авторы прогнозировали более 35 000 случаев заболевания к концу мая 2020 года. (Фактически на конец мая было зафиксировано в 2 раза больше, 72 460 случаев новой коронавирусной инфекции). Тандон с соавторами [5] сравнивает результаты прогнозирования распространения эпидемии COVID-19 в Индии с помощью 7 различных моделей временных рядов. Горизонт моделирования составлял 20 дней. Авторы заключают, что лучшие результаты среди сравниваемых показала модель ARIMA. По данным авторов, к 3 мая 2020 года число случаев заражения должно было быть около 30 990 случаев. Фактически, на данную дату было зафиксировано 42 505 случаев заражения.

Очевидно, что хотя модели временных рядов и являются популярным инструментом прогнозирования, применение данного подхода для оценки распространения новых инфекций имеет свои ограничения. В частности, отсутствие статистики за предыдущие периоды и, как следствие, неизвестные значения параметров не позволяют построить модели достаточной степени точности.

Камерные модели. В базовой камерной модели рассматривается популяция из N человек. В каждый момент времени каждый человек относится к одной из трех групп (камер): в группу восприимчивых S (Susceptible) входят люди, которые еще не сталкивались с инфекцией, затем, по мере того как вирус распространяется среди населения, они переходят в группу I инфицированных (Infected), а затем – в группу R (Removed) выбывших (выздоровевших или умерших), при этом возможность повторного заражения в данной модели не предусматривается. Также считается, что размер популяции остается неизменным: $N = S(t) + I(t) + R(t)$. Впервые так называемая модель эпидемии SIR, описываемая системой из трех обыкновенных дифференциальных уравнений для переменных S , I , R , была представлена в статье У. Кермака и А. Маккендрика в 1927 году [6, 7].

Модель SIR определяется всего двумя параметрами: эффективная частота контактов (β), которая влияет на переход из группы восприимчивых в группу инфицированных, и скорость выздоровления (γ), которая влияет на переход из группы инфицированных во множество выздоровевших или умерших (рис. 2). Соответ-

ствующее соотношение $R_0 = \frac{\beta}{\gamma}$ носит название базового репродуктивного числа. Оно определяется как среднее количество новых случаев, вызванных одним случаем, в течение инфекционного периода среди полностью восприимчивого населения. Эпидемия возникает тогда и только тогда, когда $R_0 > 1$.



Рис. 2. Общая схема камерной модели SIR

Некоторые исследователи предпочитают использовать SIR из-за небольшого количества требуемых на входе параметров. Однако, это преимущество возникает из-за порой чрезмерного упрощения модели за счет относительно нереалистичных предположений. Например, модель предполагает однородное смешивание популяции, а это означает, что все особи в популяции имеют равную вероятность вступить в контакт друг с другом. Это не отражает человеческие социальные структуры, в которых большинство контактов происходит в ограниченных сообществах. Модель SIR также предполагает замкнутую популяцию без миграции, рождений или смертей по причинам, отличным от эпидемии.

В качестве примера можно привести исследование Дил и Макена [8], в котором использовали модель SIR для прогнозирования подтвержденных случаев COVID-19 в регионе Восточного Средиземноморья (Иран, Ирак, Саудовская Аравия, ОАЭ, Ливан, Египет и Пакистан). По оценкам авторов, к 20 июня 2020 года предполагалось достичь 2,12 миллиона случаев в Иране, 0,58 миллиона в Саудовской Аравии и 0,51 миллиона в Пакистане. В действительности, число зафиксированных случаев заражения на 20.06.2020 официально составляло 202 584 случаев заражения в Иране, 176 617 случаев – в Пакистане и 154 233 случаев – в Саудовской Аравии (по данным Центра CSSE Университета Джонса Хопкинса [9]).

Кроме того, параметры в традиционной модели SIR не позволяют количественно оценить неопределенность параметров модели. Распространенной стратегией прогнозирования развития эпидемии является расчет модели SIR по нескольким возможным значениям для каждого параметра, результатом при этом является диапазон будущих траекторий. Более сложные модели используют распределения для каждого параметра вместо точечной оценки для характеристики веро-

ятности различных будущих траекторий. Также, в более сложных адаптациях камерной структуры SIR, можно формально включить наблюдаемые данные, чтобы значения параметров калибровались на основе поступающих данных [10].

Модель SEIR [11] похожа на модель SIR, за исключением того, что в неё добавляется камера E (Exposed) для людей, которые уже были инфицированы, но сами еще не являются заразными (за переход в данную группу отвечает параметр α). Далее, по истечении инкубационного периода, люди из группы E переходят уже в группу распространителей вируса I (рис. 3). При этом модель допускает возможность того, что вероятность передачи при контакте с зараженным человеком может отличаться от вероятности передачи при контакте с инфицированным человеком.

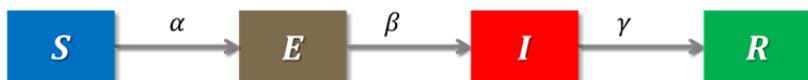


Рис. 3. Общая схема камерной модели SEIR

Помимо SEIR, часто используется модель SIRD, в которой предусмотрены отдельные камеры для выздоровевших людей (тех, кто пережил болезнь и теперь имеет иммунитет) и умерших людей. В качестве примера можно привести исследование [12].

Основная проблема всех камерных моделей состоит в том, что полученные с их помощью прогнозы достаточно чувствительны к значениям параметров на входе модели, таким как R_0 и другим. При этом, данные характеристики возможно получить, лишь набрав достаточный объем статистических сведений о динамике заболевания. В случае рассмотрения новых инфекций это представляет собой большую проблему. Например, в случае коронавирусной инфекции COVID-19 в первую очередь рассматривали данные Китая и Италии как стран, где эпидемия развивалась раньше всех. Но при этом использовать напрямую полученные значения параметров при моделировании в других странах нельзя, в силу демографических и географических особенностей рассматриваемых регионов (например, средний уровень иммунитета жителей, возрастной состав, вероятность мутаций вируса, климат и т.д.) [13].

В качестве примера использования модели SEIR, можно привести исследование Ченга и Бурку [14], в котором представлена платформа `icmonitoring.ch` для прогнозирования загруженности отделений интенсивной терапии (ОИТ) на уровне больниц на основе модели

SEIR. Данные обновляются каждые 3-4 дня, и предполагается, что такая платформа может помочь менеджерам ICU оценить потребность в дополнительных ресурсах. Простая модель SIRD была предложена в [12] для прогнозирования вспышек COVID-19 в Китае, Италии и Франции. В данном исследовании также была проведена оценка потребностей медицинских учреждений (например, оценивалось необходимое количество аппаратов ИВЛ). На сегодняшний день опубликовано несколько десятков моделей SIR и её разновидностей только по COVID-19. В частности, модели SIR [15-17], SIR-X [18], SEIR [19-25], SIRD [26], SEIQR [27], SEIHR [28].

В целом, камерные модели хорошо себя зарекомендовали и являются своего рода «золотым стандартом» при моделировании динамики распространения эпидемии. В то же время, в случае рассмотрения новых инфекций данный тип моделей показывает не лучшие результаты, особенно на ранних стадиях эпидемии. Этому есть несколько причин, основная из которых – недостаток достоверной информации о значениях параметров. Фактически, недостоверные исходные данные могут дать нереалистичные значения параметров и, следовательно, ненадежные прогнозы модели [29-30].

Сетевые модели. В современной математической эпидемиологии сетевые модели представляют собой один из новейших методов анализа и моделирования сложных эпидемиологических систем. По сравнению с камерными моделями они являются более детальными и позволяют рассматривать каждого участника в отдельности, при этом взаимодействие между людьми представляется в виде сложного графа социальных связей. В такие модели можно включать различную степень детализации: разбивать популяцию на возрастные группы, учитывать количество социальных связей, наличие мест с большей плотностью населения и эффективность ограничительных мер. В то время как камерные модели позволяют оценить общую динамику распространения эпидемии, сетевые модели дают возможность симитировать эффективность тех или иных мер по сдерживанию распространения инфекции [31].

Модели с использованием методов машинного обучения. Искусственный интеллект и машинное обучение давно используются в эпидемиологии. Это мощный инструмент для поиска взаимосвязи между входными и выходными данными в случаях, когда аналитическое исследование затруднительно. Применение таких эвристических подходов для раннего обнаружения эпидемиологических рисков в некоторых случаях позволяет улучшить качество прогнозирования. За последние годы в методах машинного обучения было достигнуто много

успехов. Важным представителем моделей на основе машинного обучения являются искусственные нейронные сети (ИНС). Они представляют собой модель искусственного интеллекта, «вдохновенную» структурой биологических нейронов человека. В случае прогнозирования распространения эпидемии обучение ИНС заключается в вычислении коэффициентов связей между входными сигналами, определяющих их силу, и выполняется на основе эмпирических данных, включающих статистику заболеваемости и значения различных факторов, способных оказывать влияние на распространение заболеваемости. В исследовании [32] рассматриваются возможности применения нейронных сетей для прогнозирования распространения COVID-19. Результаты работы сети разработанной архитектуры для некоторых регионов достигали 87%. При этом, авторы акцентируют внимание, что для корректного обучения ИНС необходим большой объем исторических данных. Также следует упомянуть и подход к моделированию на основе прецедентов (англ. case-based reasoning). Данный подход основывается на идее поиска возможных решений задачи на основе уже существующих решений для аналогичных ситуаций [33]. В работах [34, 35] описана новая модель CBRR (case-based rate reasoning) на базе данного подхода для прогнозирования будущих значений основных параметров эпидемии коронавируса в России, позволяющая строить краткосрочные прогнозы на основе аналогов динамики процентного прироста в других странах. Также описан новый эвристический метод оценивания продолжительности переходного процесса процентного прироста между заданными уровнями, учитывающий информацию о динамике эпидемиологических процессов в странах цепочки распространения. Подробный обзор возможностей использования методов машинного обучения для прогнозирования распространения COVID-19 можно найти в [36].

В целом, каждый подход имеет свои достоинства и недостатки. При этом, в случае попыток прогнозирования вновь возникающих эпидемий, основной проблемой является отсутствие исторических данных, на основании которых можно было бы оценить значения входных параметров моделей. Другой общей проблемой являются «грязные» исходные данные. Информация берется из разных источников, а разные регионы и страны используют отличные друг от друга методики учета новых случаев заражения, смерти, тестирования и т.д. Учитывая тот факт, что для вновь возникающих инфекционных заболеваний отсутствует накопленная за многие годы статистика, данное ограничение является критичным при определении ключевых параметров моделей. В качестве примера можно привести различные оцен-

ки базового репродуктивного числа для нового коронавируса ин-
фекции [37-39].

3. Модель SIR и её представление. Рассмотрим детерминиро-
ванную систему дифференциальных уравнений, традиционно исполь-
зуемую для моделирования эпидемий, известную как модель SIR [6]:

$$\frac{dS(t)}{dt} = -\beta \frac{S(t)I(t)}{N}, \quad (1)$$

$$\frac{dI(t)}{dt} = \beta \frac{S(t)I(t)}{N} - \gamma I(t), \quad (2)$$

$$\frac{dR(t)}{dt} = \gamma I(t), \quad (3)$$

$$N = I(t) + S(t) + R(t). \quad (4)$$

В этой модели переменная S принимает значения, равные чис-
ленности восприимчивого к вирусу населения популяции, I – количе-
ства инфицированных людей, R – общего количества выздоровевших
и умерших людей. При этом численность популяции N считается
неизменной в течение всего эпидемиологического процесса. Коэффи-
циенты β и γ в модели, как правило, являются неотрицательными
величинами, не изменяющимися во времени. Значения переменных в
начальный момент времени t_0 заданы.

Выбор значений коэффициентов β и γ при использовании
этой модели для прогнозирования динамики эпидемии конкретного
типа вируса осуществляется на основании имеющейся статистики
прошлых периодов. Построенные для выбранных значений коэффици-
ентов траектории такой системы являются, как правило, детерминиро-
ванными функциями времени, которые, как показывают многочислен-
ные примеры из научной литературы, часто являются достаточно хоро-
шими приближениями фактических траекторий, развивающихся в
реальном времени. Вместе с тем, на точность оценки влияют много-
численные факторы, такие, например, как качество имеющейся стати-
стики и эволюционная изменчивость самого вируса и его различных
штаммов. В случае появления нового вируса подходящей статистики
нет. Неопределенность параметров модели его распространения стано-
вится существенной преградой для построения прогнозов динамики
эпидемиологической ситуации. Коэффициенты β и γ фактически
являются случайными величинами. Функции распределения этих слу-
чайных величин не известны заранее.

Естественно предположить, что количество восприимчивых к новому вирусу людей в начале эпидемии близко к значению численности популяции N , то есть $S \approx N$. Тогда, с учетом близости значений S и N , получим систему дифференциальных уравнений типа SIR вида:

$$\frac{dI(t)}{dt} = (\beta - \gamma)I(t), \quad (2')$$

$$\frac{dR(t)}{dt} = \gamma I(t). \quad (3')$$

Данная система уравнений описывает динамику количества инфицированных $I(t)$ и общего количества $R(t)$ выздоровевших и умерших людей на начальном этапе эпидемии, когда количество невосприимчивых к вирусу людей относительно мало. Разность $\beta - \gamma$ является показателем экспоненциального роста функции $I(t)$, а отношение $R_0 = \frac{\beta}{\gamma}$ называется базовым репродуктивным числом (индексом) вируса. В начале эпидемии значение $R_0 > 1$.

Решение системы (2')-(3') имеет простой аналитический вид

$$I(t) = I(t_0)e^{(\beta - \gamma)(t - t_0)},$$

$$R(t) = \frac{\gamma I(t_0)}{\beta - \gamma} \left(e^{(\beta - \gamma)(t - t_0)} - 1 \right) + R(t_0),$$

в том числе и для дискретных значений времени $t = 0, 1, 2, \dots$.

Рассмотрим вопрос об оценке точности приближений фактической динамики эпидемии с помощью дискретных решений системы (2')-(3). Предположим, что информация мониторинговой системы здравоохранения обновляется ежедневно (то есть $t = 0, 1, 2, \dots$) как, например, это происходит на странице стопкоронавирус.рф/information/ оперативных данных сайта стопкоронавирус.рф оперативного штаба Правительства РФ. После начала эпидемии мониторинговыми системами здравоохранения ежедневно регистрируются общее количество выявленных случаев инфицирования, количество новых случаев заболевания Δ , количество инфицированных людей

(число активных случаев) I , количество умерших и выздоровевших людей R .

Рассмотрим данные о числе активных случаев и общем количестве выздоровевших и умерших в России в период с 20 апреля по 19 мая 2020 года и внесем их в таблицу 1 (составлена по данным ежедневных Отчетов Коммуникационного центра Правительства РФ по ситуации с коронавирусом, размещенным в открытом доступе на портале стопкоронавирус.рф.). Выберем такие постоянные значения коэффициентов β и γ , при которых дискретная траектория системы (2')-(3') наиболее точно соответствует фактическим данным, имеющимся в этой таблице.

Таблица 1. Данные оперативного штаба и отклонение от расчетных данных модели SIR (период с 20 апреля по 19 мая 2020 года)*

t	Дата	Данные $I(t)$	Данные $I(t)$ модели SIR	Отклонение	Данные $R(t)$	Данные $R(t)$ модели SIR	Отклонение
1	20 апр	43270	43270	0	3851		
2	21 апр	48434	46332	-2102	4329	2545	-1784
3	22 апр	53066	49612	-3454	4933	5271	338
4	23 апр	57327	53122	-4205	5446	8189	2743
5	24 апр	62421	56881	-5540	6201	11314	5113
6	25 апр	67657	60906	-6751	6931	14660	7729
7	26 апр	73435	65215	-8220	7514	18243	10729
8	27 апр	79007	69828	-9179	8140	22079	13939
9	28 апр	84235	74767	-9468	9323	26186	16863
10	29 апр	88138	80055	-8083	11261	30584	19323
11	30 апр	93806	85716	-8090	12692	35293	22601
12	01 май	100042	91777	-8265	14389	40336	25947
13	02 мая	107819	98265	-9554	16235	45734	29499
14	03 мая	116768	105211	-11557	17919	51515	33596
15	04 мая	125817	112647	-13170	19451	57703	38252
16	05 мая	134054	120607	-13447	21316	64330	43014
17	06 мая	143065	129128	-13937	22864	71424	48560
18	07 мая	151732	138249	-13483	25428	79020	53592
19	08 мая	159528	148012	-11516	28331	87152	58821
20	09 мая	164933	158462	-6471	33743	95859	62116
21	10 мая	173467	169647	-3820	36221	105180	68959
22	11 мая	179534	181619	2085	41810	115159	73349
23	12 мая	186615	194432	7817	45628	125843	80215
24	13 мая	192056	208144	16088	50215	137280	87065
25	14 мая	196410	222819	26409	55835	149524	93689

26	15 мая	202199	238524	36325	60644	162631	101987
27	16 мая	206304	255329	49025	65739	176662	110923
28	17 мая	211748	273311	61563	70004	191681	121677
29	18 мая	217747	292551	74804	72931	207758	134827
30	19 мая	220974	313136	92162	78967	224967	146000

При выборе значений коэффициентов β и γ будем исходить из следующих общих оценок:

- в качестве оценки разности будем использовать линейную аппроксимацию для $\ln I = (\beta - \gamma)t + \ln I_0$;
- в качестве оценки параметра, отвечающего за скорость выздоровления, можно взять $\frac{1}{T}$, где T – среднее время перехода в состояние R . Данная оценка получается, исходя из предположения о неизменности значений коэффициентов.

В итоге для построения модели SIR были взяты следующие параметры: $T = 17$, $\gamma = 0.06$, $\beta = 0.13$. Данные значения согласуются со значениями, используемыми другими исследователями при моделировании динамики распространения COVID-19 [40]. Результаты расчётов модели SIR представлены на рисунке 4 и в таблице 1. Можно видеть, что в течение 20 дней, с 20 апреля по 9 мая, модель достаточно хорошо отражает динамику распространения эпидемии. В последние 10 дней наблюдается нарастающее расхождение между расчетными данными и фактическими. Средняя абсолютная ошибка в процентах (mean absolute percentage error, MAPE) расчетной траектории числа активных случаев заболевания на интервале с 20 апреля по 19 мая 2020 года составила 11%.

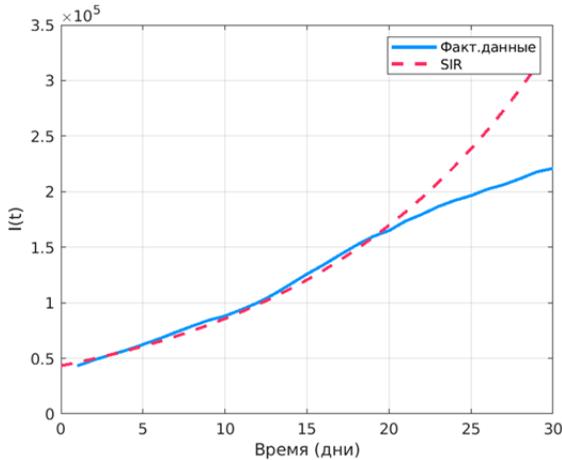


Рис. 4. Фактическая траектория числа активных случаев заболевания и расчетная траектория по модели SIR

4. Балансовая модель эпидемии на основе процентного прироста. Обозначим через $C(t)$ общее число подтвержденных случаев инфицирования (Cumulative Cases, Confirmed Cases) с момента начала эпидемии до дня t включительно, а $\Delta(t)$ – количество новых случаев, зарегистрированных в день t . С учетом введенных обозначений для $t = 1, 2, 3, \dots$ имеем

$$C(t) = C(t-1) + \Delta(t).$$

Запишем $C(t)$ в виде

$$C(t) = \alpha(t)C(t-1),$$

где

$$\begin{aligned} \alpha(t) &= 1 + \frac{r(t)}{100}, \\ r(t) &= 100 \frac{\Delta(t)}{C(t-1)}, \\ \Delta(t) &= C(t) - C(t-1). \end{aligned} \tag{5}$$

То есть окончательно имеем следующую формулу

$$C(t) = \left(1 + \frac{r(t)}{100}\right) \times C(t-1). \quad (6)$$

Здесь параметр $r(t)$ интерпретируется как отношение в процентах абсолютного прироста общего количества выявленных случаев инфицирования до дня t включительно к общему количеству выявленных случаев инфицирования до предыдущего дня включительно.

Будем далее называть параметр $r(t)$ *процентным приростом* общего количества выявленных случаев инфицирования в день t . Учитывая, что $\Delta(t)$ является случайной величиной, принимающей неотрицательные значения, процентный прирост также является неотрицательной случайной величиной.

Значение величины $C(t)$ в любой день T вычисляется по следующей формуле

$$C(T) = \prod_{t=1}^T \left(1 + \frac{r(t)}{100}\right). \quad (7)$$

Зафиксируем некоторые значения $t_0 \geq 0$ и $T > t_0$, такие, что $R(T) > C(t_0) > 1$. Учитывая не убывание функций $C(t)$ и $R(t)$, а также то, что $C(t) \geq R(t)$ при любом $t \geq t_0$, такое значение T существует. Фактически существование таких значений t_0 и T означает, что пациенты, заболевшие к моменту времени t_0 , выздоравливают или умирают за конечное время.

Рассмотрим следующую задачу:

$$\min_{t_0 \leq t \leq T} t, \quad (8)$$

$$C(t) \geq R(T). \quad (9)$$

Учитывая свойства функций $C(t)$ и $R(t)$ множество допустимых решений такой задачи не пусто. Обозначим решение задачи (8)-(9) через $\tau(T)$.

Обратим внимание на то, что для $t = \tau(T) - 1$ условие (9) не выполняется. Тогда с учетом не убывания функции $C(t)$, очевидно, выполняется неравенство $R(T) \geq C(\tau(T) - 1)$. Таким образом, справедлива следующая теорема.

Теорема 1 (Принцип динамического баланса). Пусть заданы значения $t_0 \geq 0$ и $T > t_0$, такие, что $R(T) > C(t_0) > 1$. Тогда для решения $\tau(T)$ задачи (8)-(9) значение $R(T)$ в день T удовлетворяет неравенствам

$$C(\tau(T)) \geq R(T) \geq C(\tau(T) - 1). \quad (10)$$

Следствие. В условиях теоремы 1 функция $R(T)$ представима в виде

$$R(T) = \lambda_r C(\tau(T) - 1) + (1 - \lambda_r) C(\tau(T)), \quad \lambda_r \in [0, 1]. \quad (11)$$

Его справедливость следует из известной в аналитической геометрии формулы представления точек, лежащих на отрезке.

Условие (10) означает, что количество выздоровевших и умерших людей в определенный день t зависит от общего количества зафиксированных заболевших в прошлом, а именно $t - \tau(t)$ дней назад. Таким образом, с помощью условия (10) можно установить динамический баланс между значениями функций $R(t)$, $C(\tau(t) - 1)$ и $C(\tau(t))$.

Заметим, что для любого значения t справедливо равенство

$$C(t) = I(t) + R(t). \quad (12)$$

Это балансовое соотношение означает, что группу выявленных заболевших в любой день можно разделить на тех, кто еще болеет, и тех, кто к этому дню выздоровел или, к сожалению, умер. Воспользуемся формулами (6), (11) и (12) и запишем систему дискретных уравнений

$$C(t) = \left(1 + \frac{r(t)}{100}\right) \times C(t-1), \quad (13)$$

$$I(t) = \left(1 + \frac{r(t)}{100}\right) \times C(t-1) - R(t), \quad (14)$$

$$R(t) = \lambda_i C(\tau(t)-1) + (1 - \lambda_i) C(\tau(t)). \quad (15)$$

Определение 1. Модель, динамика которой описывается системой дискретных уравнений (13)-(15), будем в дальнейшем сокращенно называть *CIR-модель*.

CIR-модель использует понятие процентного прироста общего количества заболевших $C(t)$ и учитывает балансовое соотношение между общим количеством заболевших в прошлые моменты времени $\tau(t)$ и $\tau(t)-1$ и общим количеством выздоровевших и умерших $R(t)$ в момент t . В связи с этим мы можем также называть эту модель *балансовой моделью эпидемии на основе процентного прироста*.

Определение 2. Функцию $\mathcal{G}(t) = t - \tau(t)$ будем называть *характеристикой динамического баланса эпидемии*.

Исследуя статистику пандемии коронавирусной инфекции COVID-19 в России и других странах можно отметить, что на отдельных промежутках времени характеристика динамического баланса $\mathcal{G}(t)$ является константой, однако, в общем случае она меняется во времени, хотя ее волатильность ограничена.

Определение 3. Будем называть эпидемиологический процесс *стационарным* на промежутке времени $[t_1, t_2]$, если $\mathcal{G}(t) = const$ для всех $t \in [t_1, t_2]$.

Рассмотрим данные эпидемиологического процесса из табл. 1. Оценим, каковы на промежутке $[23, 24, \dots, 30]$ значения функций $\tau(t)$, $\mathcal{G}(t)$, интервалы возможных значений функции $R(t)$, а также отклонения $R(t)$ от этих интервалов. На промежутке $[1, 2, \dots, 22]$ по данным таблицы 1 определить эти значения не представляется возможным.

По имеющимся в столбцах 3 и 4 таблицы 2 данным можно определить моменты времени, соответствующие допустимым решениям задачи (8)-(9) в период с 12 по 19 мая, то есть для $T = 23, 24, \dots, 30$ соответственно (для значений, меньших 23, множество допустимых решений является пустым). Эти решения представлены в столбце 5 таблицы 2, в столбце 6 – вычисленные значения характеристики динамического баланса $\mathcal{G}(t)$, в столбце 7 – оценка интервалов будущих возможных значений $R(T)$ при постоянной характеристике динамического баланса, в столбце 8 – величина отклонения фактического значения $R(T)$ от интервала оценки. Заметим, что в период с 12 по 19

мая функция $\mathcal{G}(t)$ не равна константе, то есть рассматриваемый эпидемиологический процесс в этот период времени не является стационарным.

Заметим, что по имеющимся в таблице данным интервальная оценка будущих значений общего количества выздоровевших и умерших в дни с 12 по 19 мая может быть получена при известных заранее значениях характеристики динамического баланса $\mathcal{G}(t)$ за 22 или 23 дня до этих дат. Увеличение значения функции $\mathcal{G}(t) = t - \tau(t)$ 18 и 19 мая на один день по отношению к периоду с 12 по 17 мая, когда она была постоянна (22 дня), привело к отклонению фактических значений $R(18)$ и $R(19)$ на 1697 и 1982 соответственно, или примерно на 30% от прироста новых случаев заболевания 20 апреля.

Таким образом, условие динамического баланса можно использовать для прогнозирования количества активных случаев на промежутке, продолжительность которого зависит от значения характеристики динамического баланса.

Таблица 2. Значения характеристики динамического баланса и оценки (период с 20 апреля по 19 мая 2020 года)

t	Дата	Общее число заболевших $C(t)$	Всего выздоровели или умерли $R(t)$	$\tau(t)$	$\mathcal{G}(t)$	Интервалы возможных значений функции $R(t)$	Отклонение $R(t)$ от интервала
1	20 апр	47121	3851				
2	21 апр	52763	4329				
3	22 апр	57999	4933				
4	23 апр	62773	5446				
5	24 апр	68622	6201				
6	25 апр	74588	6931				
7	26 апр	80949	7514				
8	27 апр	87147	8140				
9	28 апр	93558	9323				
10	29 апр	99399	11261				
11	30 апр	106498	12692				
12	01 мая	114431	14389				
13	02 мая	124054	16235				
14	03 мая	134687	17919				
15	04 мая	145268	19451				
16	05 мая	155370	21316				
17	06 мая	165929	22864				

18	07 мая	177160	25428				
19	08 мая	187859	28331				
20	09 мая	198676	33743				
21	10 мая	209688	36221				
22	11 мая	221344	41810				
23	12 мая	232243	45628	1	22	≤ 47121	0
24	13 мая	242271	50215	2	22	[47121;52763]	0
25	14 мая	252245	55835	3	22	[52763;57999]	0
26	15 мая	262843	60644	4	22	[57999;62773]	0
27	16 мая	272043	65739	5	22	[62773;68622]	0
28	17 мая	281752	70004	6	22	[68622; 74588]	0
29	18 мая	290678	72931	6	23	[74588;80949]	1657
30	19 мая	299941	78967	7	23	[80949;87147]	1982

В нашем примере, при имеющихся данных, 20 апреля можно сделать относительно точный прогноз на период с 12 мая по 19 мая. Для получения прогноза для периода с 21 апреля до 19 мая требуется дополнительно учитывать более ранние данные. Можно также отметить, что если бы эпидемиологический процесс на промежутке с 12 по 19 мая был стационарным, точность прогноза была бы выше.

5. CIR-модель и метод прецедентов для прогнозирования процентного прироста. В модели CBRR (Case-Based Rate Reasoning), подробно рассмотренной в статьях [34, 35] мы предложили новый подход для прогнозирования динамики показателя $C(t)$, основанный на методе прецедентов. Он предусматривает использование процедуры эвристического выбора будущих значений процентного прироста общего количества выявленных случаев инфицирования в России на основе информации о динамических процессах развития пандемии в других странах. В соответствии с этим алгоритмом была сформирована цепочка распространения эпидемии, в которую вошли Россия (страна-последователь), и страны-предшественники Италия, Испания, Великобритания и Франция. На рисунке 5 представлены графики процентного прироста заболевших в этих странах после выхода на уровень общего количества заболевших, равный пять тысяч.

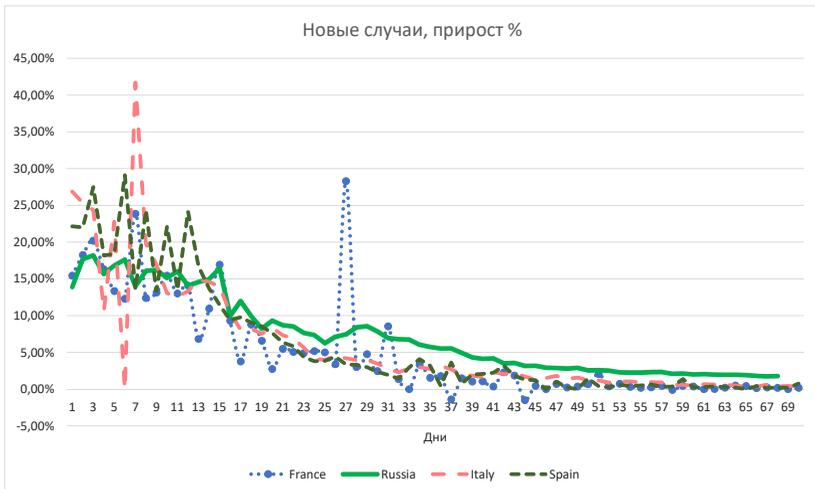


Рис. 5. Динамика процентного прироста новых случаев заболевания COVID-19 во Франции, Италии, Испании и России

Общая процедура формирования прогноза процентного прироста при использовании CBRR-модели заключается в следующем. Рассматриваемый временной горизонт развития эпидемии разбивается на M промежутков $(T_{m-1}, T_m]$, $0 \leq T_{m-1} < T_m \leq T$, $m = 1, 2, \dots, M$. Границы промежутков соответствуют прогнозируемым моментам времени прохождения процентным приростом заданных значений (например, 10, 5, 2, 1%). Количество дней в промежутке $(T_{m-1}, T_m]$ равно $T_m - T_{m-1}$. Динамику изменения прогнозируемых значений общего количества инфицированных внутри каждого промежутка можно описать, например, следующим образом:

$$\tilde{C}(T_{m-1} + k) = \left(1 + \frac{r(T_{m-1} + k)}{100} \right) \tilde{C}(T_{m-1} + k - 1), \quad k = 1, 2, \dots, (T_m - T_{m-1}),$$

а для вычисления общего количества инфицированных в конце промежутка $(T_{m-1}, T_m]$ воспользуемся формулой

$$\tilde{C}(T_m) = \left(1 + \frac{r(T_{m-1} + 1)}{100} \right) \left(1 + \frac{r(T_{m-1} + 2)}{100} \right) \dots \left(1 + \frac{r(T_m)}{100} \right) \tilde{C}(T_{m-1}).$$

На рисунках 6 и 7 представлены построенные на промежутке от 23.04.2020 до 30.06.2020 спрогнозированные траектории, которые сравниваются с фактическими траекториями. Среднее значение отклонения траектории прогноза общего количества заболевших от факти-

ческой составило 0,37%. 30 июня отклонение оказалось равным минус 1,17%.

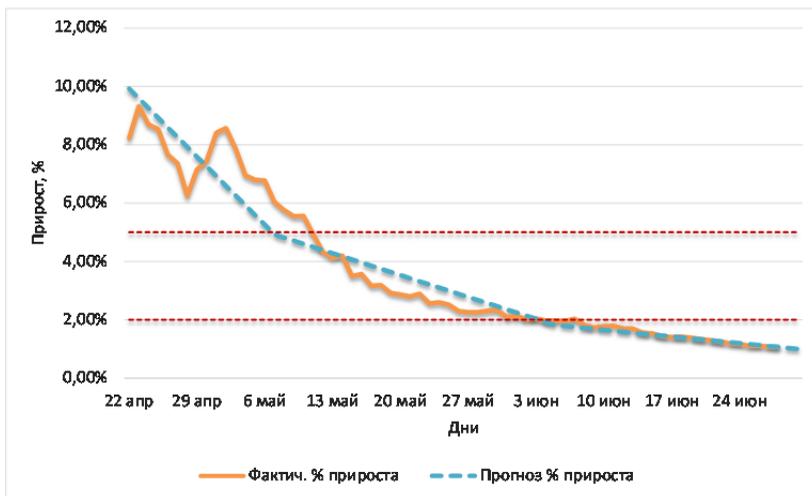


Рис. 6. Процентный прирост общего числа случаев заболевания

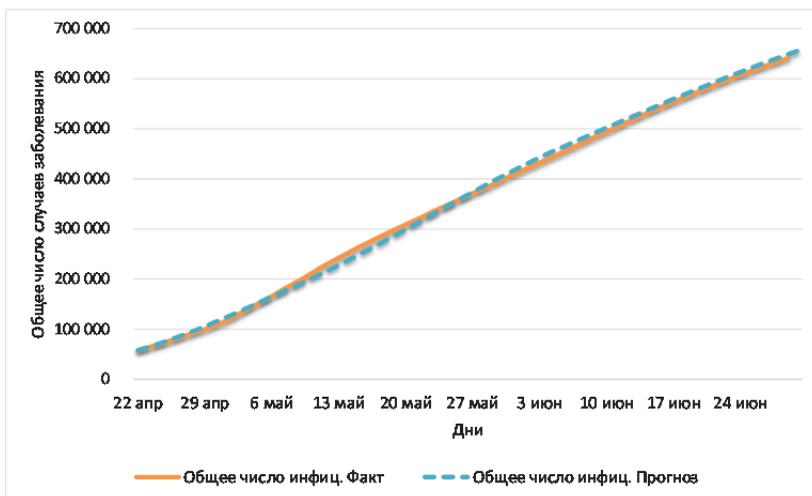


Рис. 7. Динамика общего числа зарегистрированных случаев заболевания – фактическая и расчётная траектории

Рассмотрим результаты применения разработанного подхода к построению прогноза показателей эпидемии в России на промежутке с

6 июня по 30 июня 2020 года на основании имеющейся 5 июня информации. При предположении о стационарности эпидемиологического процесса с 6 по 30 июня мы взяли для расчета постоянное значение характеристики динамического баланса, равное 25 дней. Спрогнозированные значения общего количества выздоровевших и умерших $R(t)$ и активных случаев $I(t)$ приведены, соответственно, во втором и в четвертом столбцах таблицы 3, а их фактически реализовавшиеся значения – в столбцах 3 и 5. Отклонение расчётной траектории от фактической траектории числа активных больных составило, в среднем, 3,63%. На пике эпидемии по количеству активных случаев (15 июня) отклонение достигало 6,38%. Значение MAPE для данного прогноза составило 4,14%.

Для динамики общего количества выздоровевших и умерших ошибка прогноза MAPE на рассматриваемом интервале составила 4,69%. При этом максимальное отклонение составляло 7,3%. Графики расчетных и фактических значений этих показателей приведены на рисунках 8 и 9.

Предположим, что значения характеристики динамического баланса $\vartheta(t)$ на промежутке прогнозирования нам известны заранее. Эти значения внесены в столбец 6 таблицы 3. График функции $\vartheta(t)$ представлен на рисунке 10.

Таблица 3. Прогноз от 5 июня 2020 года и фактические значения $R(t)$ и $I(t)$. Значения характеристики динамического баланса и оценки отклонения (период с 06 июня по 30 июня 2020 года)

Дата	Прогноз $R(t)$	Факт $R(t)$	Прогноз $I(t)$	Факт $I(t)$	$\vartheta(t)$	Отклонение $R(t)$	Отклонение $I(t)$
06 июн	226794	227113	241750	231576	25	-320	10174
07 июн	237257	232590	239645	235083	25	4667	4562
08 июн	247258	236659	237995	239999	26	10599	-2004
09 июн	257544	248539	236049	236714	26	9005	-665
10 июн	267443	259141	234471	234516	26	8302	-45
11 июн	276898	267682	233315	234754	26	9216	-1439
12 июн	286215	276085	232266	235338	26	10130	-3072
13 июн	295310	281470	231405	238659	27	13840	-7254
14 июн	304323	286998	230584	241966	27	17325	-11382
15 июн	313130	291630	229922	245580	27	21500	-15658
16 июн	322001	301590	229143	243868	27	20411	-14725
17 июн	331165	311820	228011	241481	27	19345	-13470

18 ИЮН	340182	321623	226962	239468	27	18559	-12506
19 ИЮН	348954	332333	226085	236730	27	16621	-10645
20 ИЮН	357885	342594	224972	234358	27	15291	-9386
21 ИЮН	366511	347822	224080	236858	27	18689	-12778
22 ИЮН	374866	352622	223369	239658	28	22244	-16289
23 ИЮН	383337	364788	222445	234917	27	18549	-12472
24 ИЮН	392099	377335	221127	229546	27	14764	-8419
25 ИЮН	401209	383769	219353	230225	27	17440	-10872
26 ИЮН	410361	392933	217423	227861	27	17428	-10438
27 ИЮН	419310	402321	215575	225325	27	16989	-9750
28 ИЮН	428009	408160	213849	226277	27	19849	-12428
29 ИЮН	436693	412596	212006	228560	28	24097	-16554
30 ИЮН	445471	421979	209929	225870	28	23492	-15941

Если использовать фактические значения $\vartheta(t)$ из таблицы 3, то, как видно на рисунке 11, графики фактических и спрогнозированных значений $R(t)$ практически совпадают. Аналогично на рисунке 12 практически совпадают графики общего количества активных больных. На интервале с 6 по 30 июня отклонение от фактической траектории числа активных больных составило в среднем 1,13%. На пике эпидемии (15 июня) отклонение составляет 1,01%.

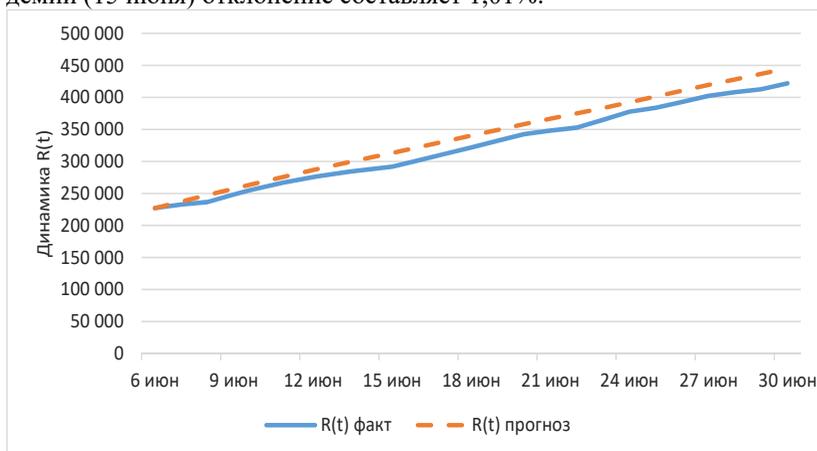


Рис. 8. Динамика $R(t)$ – фактическая и расчётная траектории на интервале с 6 по 30 июня 2020 года

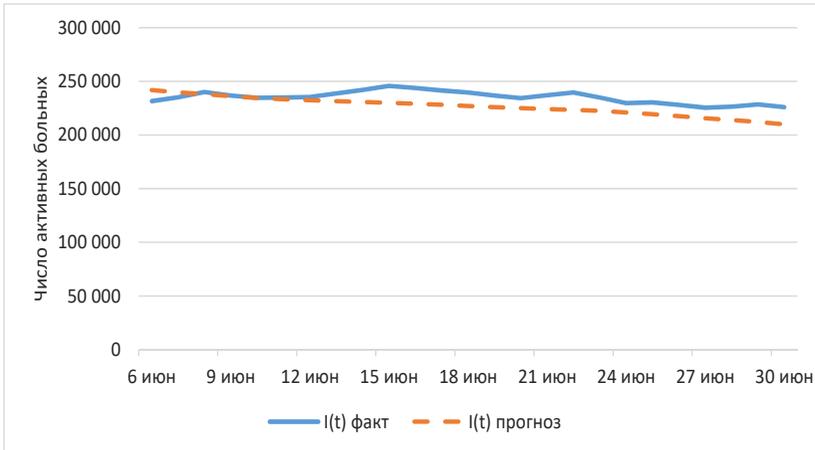


Рис. 9. Динамика $I(t)$ – фактическая и расчётная траектории на интервале с 6 по 30 июня 2020 года

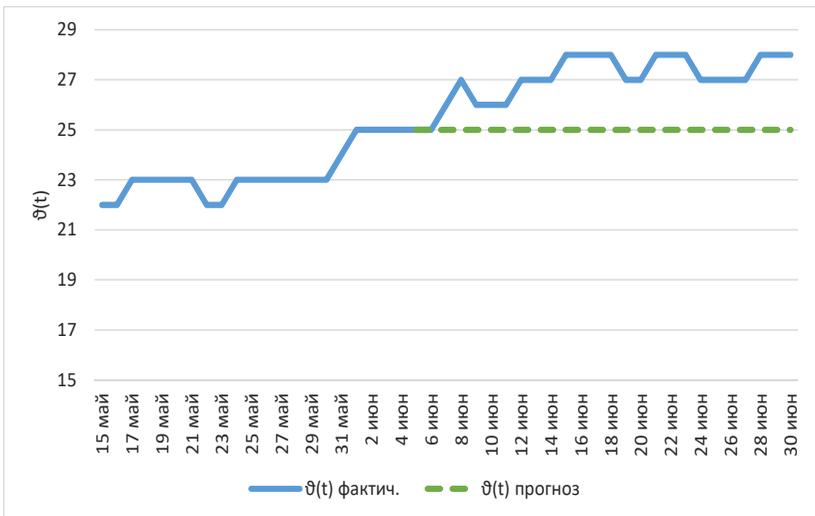


Рис. 10. Фактические и модельные значения характеристики динамического баланса $\vartheta(t)$

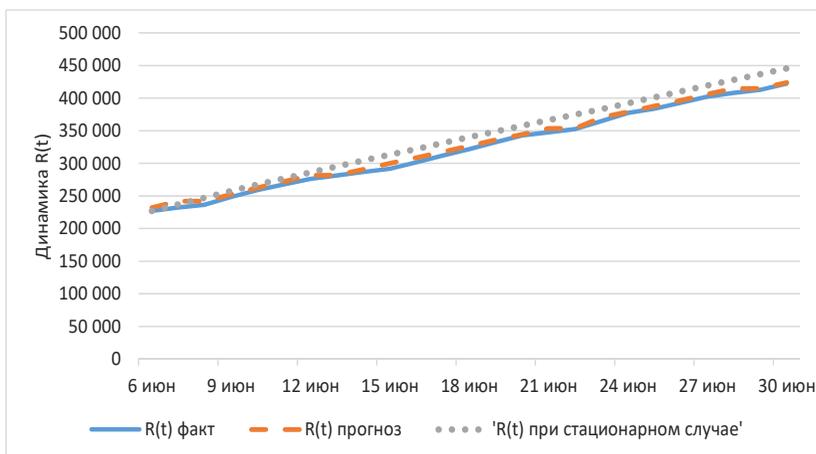


Рис. 11. Динамика $R(t)$ – фактическая и расчётная траектории на интервале с 6 по 30 июня 2020 года при известных заранее значениях $\mathcal{G}(t)$

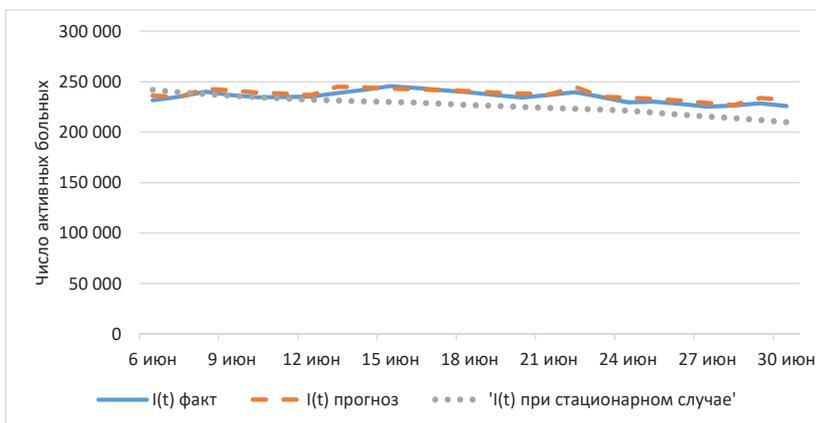


Рис. 12. Динамика $I(t)$ – фактическая и расчётная траектории на интервале с 6 по 30 июня 2020 года при известных заранее значениях $\mathcal{G}(t)$

6. Заключение. В статье представлена новая дискретная SIR-модель распространения эпидемии, описывающая с использованием случайных значений процентного прироста и характеристики динамического баланса динамику общего количества заболевших, общего количества выздоровевших и умерших и количества активных случаев болезни. Для прогнозирования значений общего количества заболе-

ших применяется CBRR-модель, ранее предложенная и апробированная авторами. Она сначала прогнозирует на среднесрочном промежутке времени значения процентного прироста общего количества заболевших с использованием метода прецедентов и интеллектуальных алгоритмов извлечения необходимых данных с использованием специального класса модели PDRM (predictive dynamic regression model) [34], а затем использует полученные значения для построения прогноза общего количества случаев заражения. Горизонт прогнозирования CIR-модели, как правило, ограничен величиной, равной значению характеристики динамического баланса эпидемии, вычисляемой на основании принципа динамического баланса эпидемиологического процесса, сформулированного в теореме 1. Поскольку характеристика динамического баланса является случайной величиной, зависящей от времени и принимающей целочисленные значения, возможности использования предложенной модели для прогнозирования динамики нестационарных эпидемиологических процессов пока представляются ограниченными. Однако, как показывают проведенные расчеты, точность прогнозирования при применении предложенной модели даже в случае нестационарности эпидемиологических процессов на промежутке прогнозирования достаточно высока. Так, отклонение расчётной траектории от фактической траектории числа активных больных составило, в среднем, 3,63%. На пике эпидемии по количеству активных случаев (15 июня) отклонение достигало 6,38%. Значение MAPE для данного прогноза составило 4,14%. В случае использования фактических значений $y(t)$ отклонение траектории числа активных больных сокращалось до 1,13%. Для динамики общего количества выздоровевших и умерших ошибка прогноза MAPE на рассматриваемом интервале с 6 по 30 июня 2020 г. составила 4,69%.

Литература

1. Foppa I.M. A historical introduction to mathematical modeling of infectious diseases. Academic Press. 2017.
2. Shinde G.R., Kalamkar, A.B., Mahalle P.N., et al. Forecasting Models for Coronavirus (COVID-19): A Survey of the State-of-the-Art. SN Computer Science. 2020. vol. 1.
3. Mofakhar L., Seif M., Safe M.S. Exponentially increasing trend of infected patients with COVID-19 in Iran: a comparison of neural network and ARIMA forecasting models. Iran Journal of Public Health. vol. 49. pp. 92–100.
4. Chaudhry R.M. et al. Coronavirus disease 2019 (COVID-19): forecast of an emerging urgency in Pakistan. Cureus. 2020. vol. 12. no 5.
5. Tandon H., Ranjan P., Chakraborty T., Suhag V. Coronavirus (covid-19): Arima based time-series analysis to forecast near future. arXiv:2004.07859.

6. Kermack W.O., McKendrick A.G. A contribution to the mathematical theory of epidemics. Proceedings of the Royal Society of London. Series A. 1927. vol. 115. no. 772. pp. 700–721.
7. Anderson R.M., May R.M. Infectious diseases of humans: Dynamics and control. Oxford: Oxford University Press. 1991. P. 757
8. Dil S., Dil N., Maken Z.H. COVID-19 trends and forecast in the eastern mediterranean region with a particular focus on Pakistan. Cureus. 2020. vol. 12. no 6.
9. Johns Hopkins Coronavirus Resource Center. Available at: <https://coronavirus.jhu.edu/data> (accessed: 20.04.2021)
10. Liao Z., Lan P., Liao Z. et al. TW-SIR: time-window based SIR for COVID-19 forecasts. Sci Rep. 2020. vol. 10.
11. Rădulescu A., Williams C., Cavanagh K. Management strategies in a SEIR-type model of COVID 19 community spread. Sci Rep. 2020. vol. 10.
12. Fanelli D., Piazza F. Analysis and forecast of COVID-19 spreading in China, Italy and France. Chaos Solitons Fractals. 2020. vol. 134.
13. Miller A., et al. Correlation between universal BCG vaccination policy and reduced morbidity and mortality for COVID-19: an epidemiological study. medRxiv 2020.03.24.20042937.
14. Cheng Z., et al. Icumonitoring.ch: a platform for short-term forecasting of intensive care unit occupancy during the COVID-19 epidemic in Switzerland. Swiss Medical Weekly. 2020. vol. 150.
15. Rodrigues H.S. Application of SIR epidemiological model: new trends. International Journal of Applied Mathematics and Informatics. 2016. vol. 10. pp. 92-97.
16. Iwami S., Takeuchi Y., Liu X. Avian–human influenza epidemic model. Mathematical biosciences. 2007. vol. 207. no. 1. pp. 1–25.
17. Teles P. Predicting the evolution of SARS-COVID-2 in Portugal using an adapted SIR model previously used in South Korea for the MERS outbreak. medRxiv 2020.03.18.20038612.
18. Maier B.F., Brockmann D. Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. Science. 2020. vol. 368. no. 6492. pp. 742-746.
19. Chinazzi M., Davis J.T., et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. Science. 2020. vol. 368. no. 6489. pp. 395-400.
20. Tang B., Wang X., Li Q., Bragazzi N.L., Tang S., Xiao Y., et al. Estimation of the transmission risk of 2019-nCov and its implication for public health interventions. Journal of Clinical Medicine. 2020. vol. 9. no. 2.
21. Tian H., Liu Y., Li Y. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. Science. 2020. vol. 368. no. 6491. pp. 638-642.
22. López L., Rodó X. A modified SEIR model to predict the COVID-19 outbreak in Spain and Italy: Simulating control scenarios and multi-scale epidemics. Results in Physics. 2021. vol. 21.
23. Feng S., Feng Z., Ling C., Chang C., Feng Z. Prediction of the COVID-19 epidemic trends based on SEIR and AI models. PLoS ONE. 2021. vol. 16. no. 1.
24. Криворотько О.И., Кабаныхин С.И., Зятков Н.Ю., Приходько А.Ю., Прохошин Н.М., Шишленин М.А. Математическое моделирование и прогнозирование COVID-19 в Москве и Новосибирской области. URL: <https://arxiv.org/abs/2006.12619v1> (дата обращения: 30.04.2021)
25. Матвеев А.В. Математическое моделирование оценки эффективности мер против распространения эпидемии COVID-19. Национальная безопасность и стратегическое планирование. 2020. №. 1. С. 23-39.

26. Anastassopoulou C., Russo L., Tsakris A., Siettos C. Data-based analysis, modelling and forecasting of the COVID-19 outbreak. *PloS One*. 2020. vol. 15. no. 3.
27. Mandal S., Bhatnagar T., Arinaminpathy N. Prudent public health intervention strategies to control the coronavirus disease 2019 transmission in India A mathematical model-based approach. *Indian Council of Medical Research*. 2020. vol. 151. pp. 190-199.
28. Choi S., Ki M. Estimating the reproductive number and the outbreak size of COVID-19 in Korea. *Epidemiology and Health*. 2020. vol. 42.
29. Tolles J., Luong T. Modeling Epidemics with Compartmental Models. *JAMA*. 2020. vol. 323. no. 24. pp. 2515–2516.
30. Fudolig M., Howard R. The local stability of a modified multi-strain SIR model for emerging viral strains. *PLoS ONE*. 2020. vol. 15. no. 12.
31. Adam D. Special report: The simulations driving the world’s response to COVID-19. *Nature*. 2020. vol. 580. pp. 316-318.
32. Wiczorek M., Siłka J., Woźniak M. Neural network powered COVID-19 spread forecasting model. *Chaos, Solitons & Fractals*. 2020. vol. 140.
33. Кондратьев М.А. Методы прогнозирования и модели распространения заболеваний. Компьютерные исследования и моделирование. 2013. т. 5. № 5 с. 863-882.
34. Zakharov V., Balykina Y., Petrosian O., Gao H. CBRR Model for Predicting the Dynamics of the COVID-19 Epidemic in Real Time. *Mathematics*. 2020. vol. 8. no. 10.
35. Захаров В.В., Балыкина Ю.Е. Прогнозирование динамики эпидемии коронавируса (COVID-19) на основе метода прецедентов. Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления. 2020. т. 16. Вып. 3. С. 249-259.
36. Dairi A., Harrou F., Zeroual A., Hittawe M.M., Sun Y. Comparative study of machine learning methods for COVID-19 transmission forecasting. *Journal of Biomedical Informatics*. 2021. vol. 18.
37. Mizumoto K., Chowell G. Transmission potential of the novel coronavirus (COVID-19) onboard the diamond Princess Cruises Ship, 2020. *Infectious Disease Modelling*. 2020. vol. 5. pp. 264-270.
38. Zhang S., et al. Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: A data-driven analysis. *Int J Infect Dis*. 2020. vol. 93. pp. 201-204.
39. Jung S., et al. Real-Time Estimation of the Risk of Death from Novel Coronavirus (COVID-19) Infection: Inference Using Exported Cases. *J. Clin. Med*. 2020. vol. 9. no. 2.
40. Cooper I., Mondal A., Antonopoulos C. G. A SIR model assumption for the spread of COVID-19 in different communities. *Chaos, Solitons and Fractals*. 2020. vol. 139.

Захаров Виктор Васильевич — д-р физ.-мат. наук, профессор, заведующий кафедрой, кафедра математического моделирования энергетических систем, факультет прикладной математики – процессов управления, Санкт-Петербургский государственный университет. Область научных интересов: представление и обработка данных и знаний с неопределенностью, применение методов математики и информатики в социокультурных и эпидемиологических исследованиях. Число научных публикаций — 230. v.zaharov@spbu.ru; Университетский пр., 35, 198504, г. Санкт-Петербург, Россия; р.т.: +7 (812) 328-20-00.

Балыкина Юлия Ефимовна — канд. физ.-мат. наук, доцент, факультет прикладной математики – процессов управления, Санкт-Петербургский государственный

университет. Область научных интересов: математическое моделирование, анализ данных, применение методов математического моделирования в медицине. Число научных публикаций — 80. j.balykina@spbu.ru; Университетский пр., 35, 198504, г. Санкт-Петербург, Россия; р.т.: +7 (812) 328-20-00.

V. ZAKHAROV, Y. BALYKINA
**BALANCE MODEL OF COVID-19 EPIDEMIC BASED ON
PERCENTAGE GROWTH RATE**

Zakharov V., Balykina Y. Balance Model of COVID-19 Epidemic Based on Percentage Growth Rate.

Abstract. The paper examines the possibility of using an alternative approach to predicting statistical indicators of a new COVID-19 virus type epidemic. A systematic review of models for predicting epidemics of new infections in foreign and Russian literature is presented. The accuracy of the SIR model for the spring 2020 wave of COVID-19 epidemic forecast in Russia is analyzed. As an alternative to modeling the epidemic spread using the SIR model, a new CIR discrete stochastic model is proposed based on the balance of the epidemic indicators at the current and past time points. The new model describes the dynamics of the total number of cases (C), the total number of recoveries and deaths (R), and the number of active cases (I). The system parameters are the percentage increase in the C(t) value and the characteristic of the dynamic balance of the epidemiological process, first introduced in this paper. The principle of the dynamic balance of epidemiological process assumes that any process has the property of similarity between the value of the total number of cases in the past and the value of the total number of recoveries and deaths at present. To calculate the values of the dynamic balance characteristic, an integer linear programming problem is used. In general, the dynamic characteristic of the epidemiological process is not constant. An epidemiological process the dynamic characteristic of which is not constant is called non-stationary. To construct mid-term forecasts of indicators of the epidemiological process at intervals of stationarity of the epidemiological process, a special algorithm has been developed. The question of using this algorithm on the intervals of stationarity and non-stationarity is being examined. Examples of the CIR model application for making forecasts of the considered indicators for the epidemic in Russia in May-June 2020 are given.

Keywords: COVID-19, epidemic modeling of new viruses, SIR models, forecast.

Zakharov Victor — Ph.D., Dr.Sci., Professor, Head of department, Mathematical modeling of energy systems department, Faculty of applied mathematics and control processes, Saint Petersburg State University. Research interests: uncertain knowledge and data representation and processing, application of mathematics and computer science in sociocultural and epidemiological studies. The number of publications — 230. v.zakharov@spbu.ru; 35, University pr., 198504, St. Petersburg, Russia; office phone: +7 (812) 328-20-00.

Balykina Yulia — Ph.D., Associate Professor, Faculty of applied mathematics and control processes, Saint Petersburg State University. Research interests: mathematical modeling, data mining, application of mathematical modeling in medicine. The number of publications — 80. j.balykina@spbu.ru; 35, University pr., 198504, St. Petersburg, Russia; office phone: +7 (812) 328-20-00.

References

1. Foppa I.M. A historical introduction to mathematical modeling of infectious diseases. Academic Press. 2017.
2. Shinde G.R., Kalamkar, A.B., Mahalle P.N., et al. Forecasting Models for Coronavirus (COVID-19): A Survey of the State-of-the-Art. SN Computer Science. 2020. vol. 1.

3. Moftakhar L., Seif M., Safe M.S. Exponentially increasing trend of infected patients with COVID-19 in Iran: a comparison of neural network and ARIMA forecasting models. *Iran Journal of Public Health*. vol. 49. pp. 92–100.
4. Chaudhry R.M. et al. Coronavirus disease 2019 (COVID-19): forecast of an emerging urgency in Pakistan. *Cureus*. 2020. vol. 12. no 5.
5. Tandon H., Ranjan P., Chakraborty T., Suhag V. Coronavirus (covid-19): Arima based time-series analysis to forecast near future. arXiv:2004.07859.
6. Kermack W.O., McKendrick A.G. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A*. 1927. vol. 115. no. 772. pp. 700–721.
7. Anderson R.M., May R.M. *Infectious diseases of humans: Dynamics and control*. Oxford: Oxford University Press. 1991. P. 757
8. Dil S., Dil N., Maken Z.H. COVID-19 trends and forecast in the eastern mediterranean region with a particular focus on Pakistan. *Cureus*. 2020. vol. 12. no 6.
9. Johns Hopkins Coronavirus Resource Center. Available at: <https://coronavirus.jhu.edu/data> (accessed: 20.04.2021)
10. Liao Z., Lan P., Liao Z. et al. TW-SIR: time-window based SIR for COVID-19 forecasts. *Sci Rep*. 2020. vol. 10.
11. Rădulescu A., Williams C., Cavanagh K. Management strategies in a SEIR-type model of COVID 19 community spread. *Sci Rep*. 2020. vol. 10.
12. Fanelli D., Piazza F. Analysis and forecast of COVID-19 spreading in China, Italy and France. *Chaos Solitons Fractals*. 2020. vol. 134.
13. Miller A., et al. Correlation between universal BCG vaccination policy and reduced morbidity and mortality for COVID-19: an epidemiological study. medRxiv 2020.03.24.20042937.
14. Cheng Z., et al. Icumonitoring.ch: a platform for short-term forecasting of intensive care unit occupancy during the COVID-19 epidemic in Switzerland. *Swiss Medical Weekly*. 2020. vol. 150.
15. Rodrigues H.S. Application of SIR epidemiological model: new trends. *International Journal of Applied Mathematics and Informatics*. 2016. vol. 10. pp. 92-97.
16. Iwami S., Takeuchi Y., Liu X. Avian–human influenza epidemic model. *Mathematical biosciences*. 2007. vol. 207. no. 1. pp. 1–25.
17. Teles P. Predicting the evolution of SARS-COVID-2 in Portugal using an adapted SIR model previously used in South Korea for the MERS outbreak. medRxiv 2020.03.18.20038612.
18. Maier B.F., Brockmann D. Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. *Science*. 2020. vol. 368. no. 6492. pp. 742-746.
19. Chinazzi M., Davis J.T., et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science*. 2020. vol. 368. no. 6489. pp. 395-400.
20. Tang B., Wang X., Li Q., Bragazzi N.L., Tang S., Xiao Y., et al. Estimation of the transmission risk of 2019-nCov and its implication for public health interventions. *Journal of Clinical Medicine*. 2020. vol. 9. no. 2.
21. Tian H., Liu Y., Li Y. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science*. 2020. vol. 368. no. 6491. pp. 638-642.
22. López L., Rodó X. A modified SEIR model to predict the COVID-19 outbreak in Spain and Italy: Simulating control scenarios and multi-scale epidemics. *Results in Physics*. 2021. vol. 21.
23. Feng S., Feng Z., Ling C., Chang C., Feng Z. Prediction of the COVID-19 epidemic trends based on SEIR and AI models. *PLoS ONE*. 2021. vol. 16. no. 1.

24. Krivorotko O.I., Kabanikhin S.I., Zyatkov N.Y., Prikhodko A.Y., Prokoshin N.M., Shishlenin M.A. Matematicheskoe modelirovanie i prognozirovanie COVID-19 v Moskve i Novosibirskoj oblasti. [Mathematical modeling and prediction of COVID-19 in Moscow city and Novosibirsk region]. Available at: <https://arxiv.org/abs/2006.12619v1> (accessed: 30.04.2021) (In Russ.).
25. Matveev A. V. Matematicheskoe modelirovanie ocenki jeffektivnosti mer protiv rasprostraneniya jepidemii COVID-19. Nacional'naja bezopasnost' i strategicheskoe planirovanie. [The mathematical modeling of the effective measures against the COVID-19 spread]. National Security and Strategic Planning. 2020. no. 1. pp. 23-39.
26. Anastassopoulou C., Russo L., Tsakris A., Siettos C. Data-based analysis, modelling and forecasting of the COVID-19 outbreak. PloS One. 2020. vol. 15. no. 3.
27. Mandal S., Bhatnagar T., Arinaminpathy N. Prudent public health intervention strategies to control the coronavirus disease 2019 transmission in India A mathematical model-based approach. Indian Council of Medical Research. 2020. vol. 151. pp. 190-199.
28. Choi S., Ki M. Estimating the reproductive number and the outbreak size of COVID-19 in Korea. Epidemiology and Health. 2020. vol. 42.
29. Tolles J., Luong T. Modeling Epidemics with Compartmental Models. JAMA. 2020. vol. 323. no. 24. pp. 2515-2516.
30. Fudolig M., Howard R. The local stability of a modified multi-strain SIR model for emerging viral strains. PLoS One. 2020. vol. 15. no. 12.
31. Adam D. Special report: The simulations driving the world's response to COVID-19. Nature. 2020. vol. 580. pp. 316-318.
32. Wiczorek M., Siłka J., Woźniak M. Neural network powered COVID-19 spread forecasting model. Chaos, Solitons and Fractals. 2020. vol. 140.
33. Kondratyev M.A. Metody prognozirovaniya i modeli rasprostraneniya zabojevanij. Komp'juternye issledovanija i modelirovanie. [Forecasting methods and models of disease spread]. Computer Research and Modeling. 2013. vol. 5. no. 5. pp. 863-882.
34. Zakharov V., Balykina Y., Petrosian O., Gao H. CBRR Model for Predicting the Dynamics of the COVID-19 Epidemic in Real Time. Mathematics. 2020. vol. 8. no. 10.
35. Zakharov V.V., Balykina Yu.E. Prognozirovanie dinamiki jepidemii koronavirusa (COVID-19) na osnove metoda precedentov. [Predicting the dynamics of the coronavirus (COVID-19) epidemic based on the case-based reasoning approach]. Vestnik of Saint Petersburg University. Applied Mathematics. Computer Science. Control Processes, 2020. vol. 16., no. 3. pp. 249–259. (In Russ.).
36. Dairi A., Harrou F., Zeroul A., Hittawe M.M., Sun Y. Comparative study of machine learning methods for COVID-19 transmission forecasting. Journal of Biomedical Informatics. 2021. vol. 18. Article ID 103791.
37. Mizumoto K., Chowell G. Transmission potential of the novel coronavirus (COVID-19) onboard the diamond Princess Cruises Ship, 2020. Infectious Disease Modelling. 2020. vol. 5. pp. 264-270.
38. Zhang S., et al. Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: A data-driven analysis. Int J Infect Dis. 2020. vol. 93. pp. 201-204.
39. Jung S., et al. Real-Time Estimation of the Risk of Death from Novel Coronavirus (COVID-19) Infection: Inference Using Exported Cases. J. Clin. Med. 2020. vol. 9. no. 2.
40. Cooper I., Mondal A., Antonopoulos C. G. A SIR model assumption for the spread of COVID-19 in different communities. Chaos, Solitons and Fractals. 2020. vol. 139.