

УДК 51–76

ПРОГНОЗИРОВАНИЕ ПРОЦЕССА РАСПРОСТРАНЕНИЯ КОРОНАВИРУСНОЙ ИНФЕКЦИИ С ПОМОЩЬЮ УРАВНЕНИЙ С ПОСЛЕДЕЙСТВИЕМ

А. Н. Авлас, А. К. Деменчук, С. В. Лемешевский, Е. К. Макаров

Институт математики НАН Беларуси

e-mail: artolomiay@tut.by, demenchuk@im.bas-net.by, svl@im.bas-net.by, jcm@im.bas-net.by

Поступила 01.12.2023

Приводятся результаты прогнозирования первой волны распространения коронавирусной инфекции COVID-19 на основе упрощенной модели Барояна–Рвачева.

Введение. Пандемия COVID-19 и ее информационное сопровождение создали небывалую ранее ситуацию широкой доступности статистических данных о заболеваемости и смертности, вызванных однотипной инфекцией, одновременно распространявшейся во всех странах мира. Несмотря на неоднородность и неполноту этих данных, а также и не всегда достаточную достоверность, их доступность и сопровождавшие рост заболеваемости панические настроения не могли не стимулировать значительное расширение исследований в области математической эпидемиологии, посвященных изучению COVID-19. Последовавшее в 2022 г. постепенное снижение интенсивности эпидемических процессов и уменьшение накала информационной кампании вокруг COVID-19 привели к снижению актуальности данной тематики и возвращению исследований в этой области в относительно нормальное русло.

Основным результатом проведенных в мире работ стало установление того факта, что COVID-19 является обычной инфекцией, подчиняющейся общим законам эпидемиологии и допускающей моделирование на основе уже существующих моделей. Еще одним важным результатом стала разработка множества новых подходов и моделей инфекционных процессов и их проверка на натуральных данных, ставшая возможной благодаря их обилию и доступности. При этом, однако, следует отметить, что находящиеся в открытом доступе данные о заболеваемости коронавирусной инфекцией не детализированы и отражают лишь суммарные величины, относящиеся к отдельным странам. Вследствие этого большинство из огромного количества работ, опубликованных за время после начала пандемии COVID-19 и посвященных различным аспектам математической эпидемиологии, направлено на исследование и прогнозирование макроскопической динамики эпидемии, отражаемой общими количествами инфицированных, заболевших, выздоровевших и умерших.

Значительное количество работ ориентировано на рассмотрение случаев развития эпидемии с дополнительными обстоятельствами в виде противоэпидемических мер, вакцинации, завоза инфекции. Более усложненные модели применяются для детального рассмотрения хода процесса распространения заболеваний с учетом географического распределения населения, его возрастной структуры и т. п. Математически это выражается в разработке все новых и все более усложненных вариантов классической модели SIR (восприимчивые – инфицированные – выздоровевшие). Содержательную характеристику таких работ и многочисленные ссылки можно найти в [1]. Другим магистральным направлением работ последнего времени, как следует из обзора в [1], является переход к использованию в эпидемических моделях уравнений с последствием. Не остались в стороне от мировых тенденций и белорусские ученые. В частности, были опубликованы работы [2] и [3], посвященные прогнозированию распространения коронавирусной инфекции на материале ее

первой волны 2020 г., наиболее близкие по тематике к нашим исследованиям. Нами в июне 2020 г. в Отделение физики, математики и информатики НАН Беларуси был направлен отчет о выполнении внеплановых работ по прогнозированию течения эпидемии COVID-19 в Республике Беларусь с оправдавшимся впоследствии прогнозом снижения количества выявленных случаев инфекции COVID-19 до 100 в сутки в первой декаде августа. Часть материалов этого отчета использована при подготовке статьи [4], а также при проведении исследований, результаты которых представлены в настоящей статье.

1. Модель SIR. Модель SIR основана на следующей системе трех дифференциальных уравнений:

$$N \frac{dS}{dt} = -\beta SI, \quad N \frac{dI}{dt} = \beta SI - \gamma IN, \quad \frac{dR}{dt} = \gamma I.$$

Здесь S – число восприимчивых к заболеванию; I – число инфицированных; R – число выздоровевших; N – общая численность популяции. Коэффициенты β и γ определяют интенсивность инфицирования и скорость выздоровления соответственно.

Эта модель предложена в 1927 г. А. Г. МакКендриком и У. О. Кермаком [5]. Она стандартно применяется для ситуации наибольшей неопределенности, когда нет информации о тонких деталях эпидемической обстановки. В своей основе модель SIR является развитием простейшей логистической модели

$$\frac{dN}{dt} = aN(K - N),$$

где N – численность популяции; K – ее максимально возможная численность; a – коэффициент размножения, известной еще с XIX в. и до сих пор применяемой для моделирования в популяционной динамике. Обе модели основаны на одном и том же вероятностном законе «действующих масс», утверждающем, что вероятность и, следовательно, интенсивность взаимодействия двух категорий людей, выделяемых в модели эпидемического процесса, пропорциональна их численности. Принципиальным недостатком логистической модели при ее применении к моделированию эпидемий является то, что в рамках такой модели в явном противоречии с наблюдаемыми фактами переболевают все восприимчивые, а скорости разрастания и затухания эпидемии совпадают. Модификация логистической модели, включающая в себя итеративный запуск вторичных субэпидемий, была использована авторами работы [2] для построения прогноза первой волны эпидемии COVID-19 в Республике Беларусь.

Главным отличием модели SIR и ее обобщений от простейшей логистической модели состоит в том, что в эпидемический процесс втягиваются не все восприимчивые. Уравнения этих моделей не имеют явного аналитического решения, но могут быть представлены в параметризованном виде в квадратурах. Качественный и численный анализ показывают, что модель SIR и ее модификации обеспечивают правильное асимптотическое поведение решений: и начальный рост, и финальное затухание в этих моделях экспоненциальные.

Общеизвестным недостатком данного класса моделей является отсутствие разработанной методики определения значений входящих в них коэффициентов и начального количества восприимчивых [6; 7]. Стандартных методик для их определения не существует. Отыскание этих величин по имеющимся в открытом доступе данным о заболеваемости достаточно затруднительно, поскольку в них отсутствуют сведения о динамике числа восприимчивых, а данные по выздоровевшим могут давать очень большой разброс для значений коэффициента γ .

В связи с этим основной областью применения различных вариантов модели SIR являются теоретические исследования и ретропрогнозирование. При использовании их для прогнозирования реальной ситуации приходится учитывать опыт прошлых эпидемий,

из которого должны быть извлечены как сценарии развития событий, так и количественные характеристики эпидпроцесса, позволяющие задать значения параметров модели. С другой стороны, модель SIR и ее варианты хорошо зарекомендовали себя на практике и дают результаты, близкие к натурным данным. Преодоление указанных недостатков компартментальных моделей является основной целью многих работ последнего времени, посвященных данной тематике. Поэтому представляется целесообразным использовать основные идеи этой модели, но внести в нее такие изменения, которые позволят проводить обучение модели на имеющихся данных.

2. Модель Барояна–Рвачева. В конце 60-х годов XX в. в СССР была разработана и впоследствии успешно применялась математическая модель протекания эпидемий гриппа в масштабах всей страны, созданная О. В. Барояном и Л. А. Рвачевым [8–11]. В ее основу положена аналогия между гидродинамическими явлениями и распространением эпидемии по территории страны.

Модель Барояна–Рвачева состоит из двух подмоделей: локальной модели распространения инфекции в отдельно взятом населенном пункте, рассматриваемом как однородная область активного перемешивания и тесных контактов, и сетевой модели переноса инфекции по транспортным коммуникациям между населенными пунктами. Локальная подмодель в ее реализации для случая непрерывного времени описана в [8–10; 11, с. 12–16] и имеет (после упрощений, описанных в [11, с. 14–15]) следующий вид:

$$\begin{aligned} \frac{dx}{dt} &= -\frac{\lambda x}{p} \int_0^T y(\tau, t) g(\tau) d\tau, \\ \frac{\partial y}{\partial t} + \frac{\partial y}{\partial \tau} &= 0, \\ y(0, t) &= \frac{\lambda x}{p} \int_0^T y(\tau, t) g(\tau) d\tau, \\ x(0) &= \alpha p, \quad y(\tau, 0) = a(\tau), \quad 0 \leq \tau \leq T, \end{aligned}$$

где p – общая численность населения рассматриваемого города; $x(t)$ – количество здоровых (и при этом восприимчивых к данной инфекции) людей в этом городе в момент времени t ; $y(\tau, t)$ – скорость прироста числа инфицированных в момент $t - \tau$; $g(\tau)$ – коэффициент, характеризующий исходящую от инфицированного опасность заражения окружающих по истечении времени τ после инфицирования; λ – коэффициент пропорциональности, который интерпретируется как средняя частота передачи инфекции; α – начальная доля восприимчивых в населении города; $a(\tau)$ – начальная заболеваемость; T – максимальная длительность заболевания.

Дискретная версия, предназначенная для компьютерной реализации, представлена в [11]. Она получена из непрерывного варианта после дискретизации и некоторых упрощений и включает в себя уравнения

$$y(0, t+1) = \frac{\lambda x(t)}{p} \sum_{\tau=0}^T y(\tau, t) g(\tau), \quad (1)$$

$$x(t+1) = x(t) - y(0, t+1), \quad (2)$$

$$y(\tau, t+1) = y(\tau-1, t), \quad \tau = 1, \dots, T, \quad (3)$$

с начальными условиями

$$x(0) = \alpha p, \quad y(\tau, 0) = a(\tau), \quad 0 \leq \tau \leq T.$$

Переменные и коэффициенты, входящие в эти уравнения, являются аналогами соответствующих им величин в непрерывном варианте модели и имеют тот же или близкий смысл: t и τ – дискретное время, измеряемое в днях; $x(t)$ – количество здоровых восприимчивых в этом городе к исходу дня с номером t ; $y(\tau, t)$ – количество людей, которые заразились в течение дня с номером $t - \tau$; $g(\tau)$ – коэффициент, характеризующий исходящую от инфицированного опасность заражения окружающих по истечении времени τ после инфицирования; величины $p, \lambda, \alpha, a(\tau), T$ в дискретном случае имеют прежнее определение.

Как и для модели SIR основной и трудноразрешимой проблемой применения модели Барояна–Рвачева является оценка параметров модели: начального числа восприимчивых к новому варианту вируса гриппа, выражаемого коэффициентом α , и скорости передачи инфекции, выражаемой коэффициентом λ . Для решения этой проблемы разработчиками модели создан полуэвристический метод, позволяющий в каждом конкретном случае подобрать пару соответствующих друг другу значений λ и α , обеспечивающую необходимое качество прогноза [11, с. 32–35]. Основной вывод, который можно сделать из примеров применения этого метода, представленных в [11, с. 36–42], состоит в том, что в определенных пределах ошибки в выборе коэффициента α могут быть компенсированы подходящим выбором коэффициента λ и наоборот.

3. Модель, предложенная в [3], и наши исследования. Автором работы [3] предложена оригинальная модель распространения инфекции, проведено ее исследование и верификация на натуральных данных. В упомянутом во введении отчете нами независимо предложена аналогичная модель. В настоящей статье мы представляем результаты, полученные нами и дополняющие результаты [3].

В [3] рассматриваемая модель представлена как в дифференциальной форме с использованием дифференциальных уравнений с запаздыванием, так и в дискретной. В отличие от подхода работы [3] наша модель исходно разрабатывалась как дискретная с ориентацией на имеющиеся данные статистики о ежедневных приростах заболеваемости. В ее основу положено предположение, что источником заражения являются не все инфицированные, как в логистической модели, и не все инфицированные, но не выздоровевшие, как в классической модели SIR, а лишь те, кто заразился недавно, исходя из того, что заразившиеся в предшествующий период либо выздоровели, либо госпитализированы, и тем самым не могут выступать источником заражения. Обоснованность такого предположения в условиях, имевших место в ходе развития первой волны эпидемии COVID-19 в 2020 г. не вызывает сомнения и вполне очевидна.

Сформулированное предположение приводит к следующему выражению для численности категории источников инфекции E_k :

$$E_k = X_k - X_{k-l},$$

где X_k – суммарное количество инфицированных, выявленных к исходу k -х суток эпидемии; l – параметр модели, имеющий смысл длительности периода времени, в течение которого инфицированный представляет опасность для окружающих. При этом суммарная сила источника заражения принимается в виде aE_k , где a – коэффициент передачи инфекции. Выражение для числа восприимчивых S_k было принято в виде

$$S_k = K - X_k,$$

где K – общее число людей, восприимчивых к инфекции. Наличие явных выражений для численности этих категорий через суммарную численность инфицированных и дает возможность применить предлагаемое уравнение к решению задачи прогнозирования течения эпидемического процесса на основе имеющихся данных.

В итоге, прогноз динамики заражения предлагается осуществлять с помощью дискретного уравнения

$$X_{k+1} - X_k = aE_k S_k$$

или, в развернутой форме,

$$X_{k+1} - X_k = a(X_k - X_{k-l})(K - X_k) \quad (4)$$

с начальным условием $X_k = \varphi_k, k = 0, \dots, l$, где φ_k – начальная заболеваемость. В дальнейшем мы будем для удобства считать, что день с номером 0 предшествует началу эпидемии, т. е. $X_0 = 0$, а в первый день эпидемии уже выполнено условие $X_1 > 0$.

Это уравнение является аналогом второго уравнения модели SIR

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I = \frac{\beta SI}{N} - \gamma I,$$

а также и простейшего логистического уравнения, отличаясь от них принятием сформулированных выше предположений и их следствий. Несложные выкладки и численные эксперименты показывают, что, как и в модели SIR, в рассматриваемой модели переболевают не все восприимчивые, а скорости разрастания и затухания эпидемии являются экспоненциальными и не совпадают в общем случае, что соответствует данным медицинской статистики.

В процессе исследования построенной модели было выяснено, что она может быть получена из модели Барояна–Рвачева путем некоторых упрощений и преобразований. Некоторые из них были осуществлены нами еще до выбора окончательного вида уравнения (4). В частности, в модели Барояна–Рвачева суммарная сила источника заражения aE_k модулируется с помощью коэффициентов $g(\tau)$, определяющих относительную степень исходящей от группы одновременно инфицированных людей опасности заражения для окружающих в i -й день после инфицирования. Проведенные нами численные эксперименты на реальных данных подтвердили содержащееся в [11, с. 22] указание на малую чувствительность результатов моделирования к значениям коэффициентов $g(\tau)$, а также, в определенной мере, и к значению длины промежутка l , в течение которого инфицированный является источником заражения окружающих. Исходя из этого нами было принято соглашение о равенстве всех коэффициентов $g(\tau)$, $\tau = 0, \dots, l$, единице, что позволило заменить их и коэффициент λ/p модели Барояна–Рвачева на единственный коэффициент a . Параметр l при использовании реальных данных выбирался в пределах 7–14 дней. Его изменение приводило к необходимости изменения коэффициента a , но не оказывало значительного влияния на прогноз.

Из определения величины $y(\tau, t)$ для дискретного времени в модели Барояна–Рвачева и принятого нами определения величины X_k , которое эквивалентно уравнению (2), вытекает равенство $X_{t+1} - X_t = y(0, t + 1)$. Отсюда и из уравнения (3) при каждом $t \geq 0$ и $\tau = 0, \dots, T$ имеем $X_{t-\tau} - X_{t-1-\tau} = y(\tau, t)$. Принимая предположение $g(\tau) = 1, \tau = 0, \dots, T$, получаем равенство

$$\sum_{\tau=0}^T y(\tau, t)g(\tau) = \sum_{\tau=0}^T (X_{t-\tau} - X_{t-1-\tau}) = X_t - X_{t-T-1}.$$

Тогда уравнение (1) может быть записано в виде

$$X_{t+1} - X_t = \frac{\lambda x(t)}{p} (X_t - X_{t-T-1}).$$

В силу уравнения (2) имеем равенства $x(t+1) - x(t) = -y(0, t + 1) = -X_{t+1} + X_t$, суммируя которые по t в пределах от $t = 0$ до $t = s - 1$ с произвольным натуральным s , получаем

соотношение $x(s) - x(0) = -X_s + X_0$ и, возвращаясь к переменной t , а также учитывая, что $X_0 = 0$ и $x(0) = \alpha p = K$, представление $x(t) = K - X_t$. Отсюда окончательно имеем уравнение (4) с временной переменной t вместо k и параметрами $l = T + 1$ и $a = \lambda/p$, что и устанавливает связь рассматриваемой модели с моделью Барояна–Рвачева.

4. Определение параметров модели и результаты расчетов. Уравнение (4) может быть переписано в виде

$$W_k = a(K - X_k),$$

где

$$W_k = \frac{X_{k+1} - X_k}{X_k - X_{k-l}},$$

т. е. величина W_k должна линейно зависеть от величины X_k . Это позволяет воспользоваться стандартными процедурами построения линейной регрессии для определения значения коэффициента a передачи инфекции по данным о заболеваемости. Устойчивая работа алгоритмов, реализующих это построение, требует предварительной подготовки исходных данных, состоящей в снижении содержащегося в них уровня шума. Как показывают численные эксперименты, рекомендуемая ВОЗ стандартная процедура недельного усреднения не обеспечивает требуемого качества данных. Поэтому нами разработан и использован специальный метод сглаживания для данной задачи. Сглаживание осуществляется на графике ежедневных приростов. Предполагается, что шум в этих данных в основном вызван несвоевременным обнаружением инфицированных, а также, что инфицированные организованы в относительно изолированные кластеры с более быстрым, чем в среднем распространением вируса. Таким образом эпидемия представляется в виде серии микровспышек, результаты которых обнаруживаются постфактум. Сглаживание представляет собой процесс, эквивалентный гомогенизации этих кластеров. Для этого график аппроксимируется короткими геометрическими прогрессиями, со знаменателями, изменяющимися во времени. Всплески, превышающие нормированный таким образом рост, опрокидываются в прошлое, путем переноса данных назад, в предположении, что люди, дающие всплеск, очевидно, инфицированы раньше, но обнаружены только сейчас.

Характер зависимости между W_k и X_k для реальных данных хорошо виден на следующем примере данных 2020 г. по Ирландии (рис. 1).

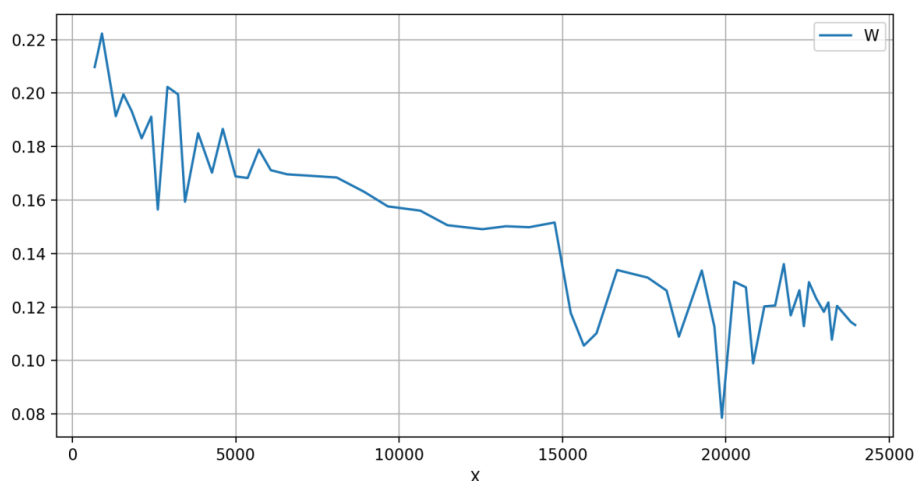


Рис. 1. Зависимость параметра W от переменной X

Некоторые результаты прогнозирования первой волны коронавирусной инфекции в Германии, Италии и Республике Беларусь по данным, доступным на различные даты, приведены на рис. 2–7.

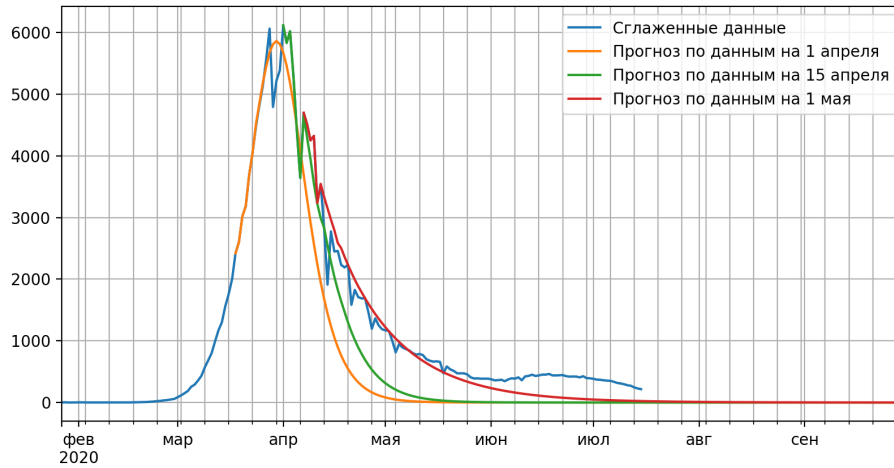


Рис. 2. Ежедневный прирост выявленных зараженных, Германия

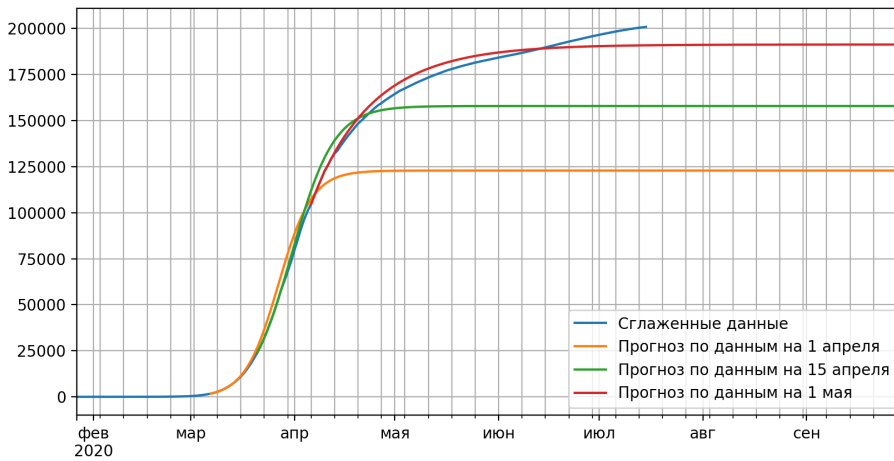


Рис. 3. Общее количество выявленных зараженных, Германия

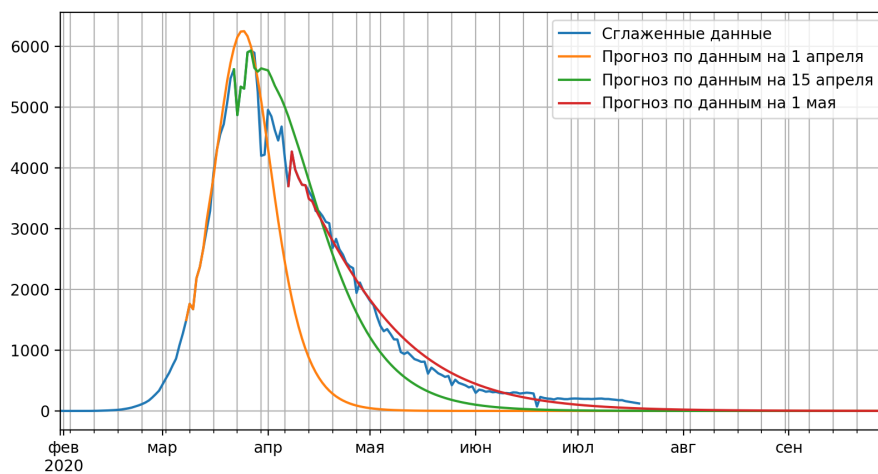


Рис. 4. Ежедневный прирост выявленных зараженных, Италия

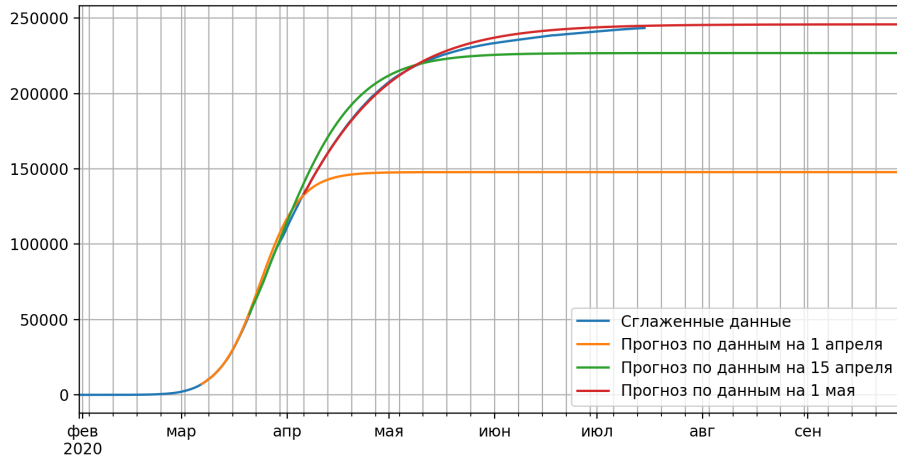


Рис. 5. Общее количество выявленных зараженных, Италия

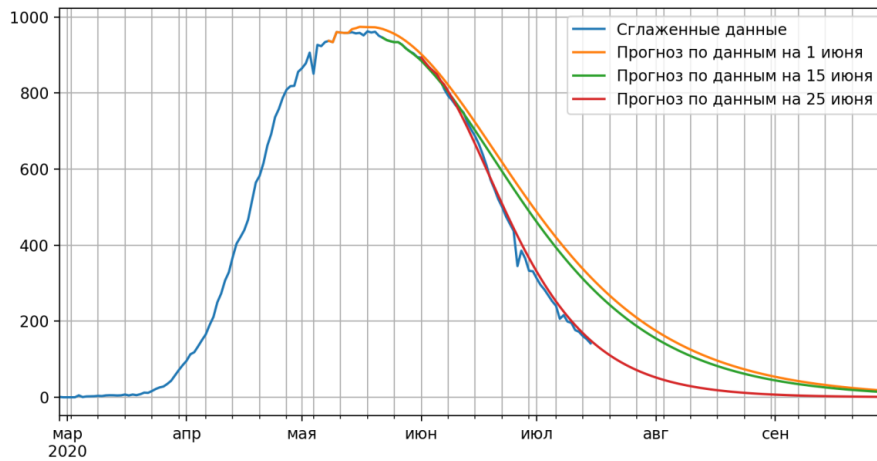


Рис. 6. Ежедневный прирост выявленных зараженных, Беларусь

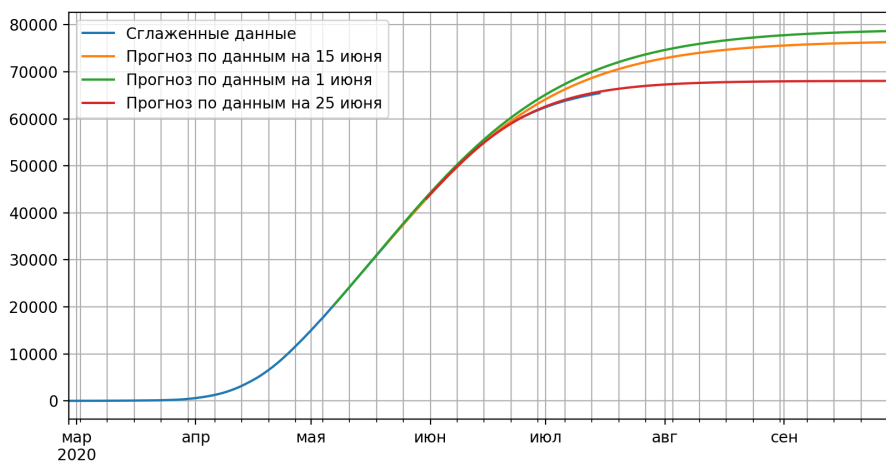


Рис. 7. Общее количество выявленных зараженных, Беларусь

Выбор для верификации метода именно первой волны 2020 г. обусловлен отсутствием в этот период как вакцинации населения, так и приобретенного естественного иммунитета, поскольку влияние этих факторов трудно выявить на основе имеющихся суммарных данных о заболеваемости. Это же обстоятельство позволяет считать параметр K , т. е. общее число людей, восприимчивых к данной инфекции, совпадающим с численностью населения рассматриваемой страны. Для последующих волн распространения эпидемии COVID-19 указанные условия, по-видимому, не выполняются и для них определение величины K представляет значительные сложности.

Отдельного обоснования, очевидно, требует применение локальной модели распространения инфекции для прогнозирования хода эпидемии в масштабах целой страны. Возможность успешного прогноза при таком, *a priori* не вполне корректном, использовании модели, по-видимому, определяется характером эпидемических процессов, протекавших в указанное время в рассматриваемых странах. Их наиболее заметной особенностью является протекание эпидемии в компактных регионах с высокой степенью транспортной связности, значительной концентрацией и активным перемешиванием населения, что и позволяет рассматривать эти регионы как аналог единого населенного пункта. В конечном же счете корректность того или иного применения любой модели определяется лишь степенью его успешности в каждом конкретном случае.

5. Интерпретация полученных результатов и выводы. Скорость разрастания эпидемии на ее начальном этапе определяется достаточно надежно при условии наличия качественной медицинской статистики.

Динамика развития эпидемии вплоть до пика заболеваемости может быть спрогнозирована на ее начальных этапах.

Скорость затухания эпидемии и асимптотический порог заболеваемости (максимальное суммарное число инфицированных) не могут быть достоверно определены до наступления пика заболеваемости. Мы предполагаем, что это ограничение принципиально.

Динамика затухания эпидемии надежно прогнозируется лишь после прохождения пика заболеваемости.

Время наступления пика заболеваемости, его высота и длительность прогнозируются наименее точно.

Заметим, что эти выводы вполне согласуются с выводами о возможности прогнозирования хода эпидемии на основе метода аппроксимации, сделанными в [4].

Заключение. Таким образом, на основе общедоступных данных о количестве инфицированных может быть получен приемлемый по точности прогноз развития эпидемии COVID-19 в национальных рамках для достаточно больших отрезков времени. Тем не менее, существуют принципиальные препятствия для наращивания точности такого прогноза, зависящие от момента составления прогноза и доступных при этом данных. Представленные в настоящей работе методы прогнозирования эпидемии COVID-19 в национальных рамках для достаточно больших отрезков времени на основе общедоступных данных о количестве инфицированных могут быть использованы на практике после уточнения процедур выбора параметров.

Работа выполнена при финансовой поддержке Белорусского республиканского фонда фундаментальных исследований (проект № Ф21КОВИД-012).

Литература

1. Harendra Pal Singh, Sumit Kaur Bhatia, Yashika Bahri, Riya Jain. Optimal control strategies to combat COVID-19 transmission: A mathematical model with

incubation time delay // Results in Control and Optimization. 2022. Vol. 9. Art. 100176. <https://doi.org/10.1016/j.rico.2022.100176>.

2. Гринчук П. С., Фисенко С. П. Физическая кинетика и моделирование распространения эпидемии // Инженерно-физический журнал. 2021. Т. 94, № 1. С. 3–8.

3. Шнип А. И. Кинетическая модель динамики эпидемий и ее тестирование на данных распространения эпидемии COVID-19 // Инженерно-физ. журн. 2021. Т. 94, № 1. С. 9–21.

4. Авлас А. Н., Деменчук А. К., Лемешевский С. В., Макаров Е. К. Аппроксимация изолированной волны эпидемического процесса с помощью комбинации экспонент // Вес. Нац. акад. навук Беларусі. Сер. фіз.-мат. навук. 2021. Т. 57, № 4. С. 391–400. <https://doi.org/10.29235/1561-2430-2021-57-4-391-400>.

5. Kermack W. O., McKendrick A. G. A Contribution to the Mathematical Theory of Epidemics // Proc. Roy. Soc. Lond. Ser. A. 1927. Vol. 115, N 772. P. 700–721.

6. Десятков Б. М., Бородулин А. И., Котлярова С. С. и др. Математическое моделирование эпидемических процессов и оценка их статистических характеристик // Химическая и биол. безопасность. 2009. № 1–3 (43–45). С. 15–20.

7. Гришунина Ю. Б., Контаров Н. А., Архарова Г. В., Юминова Н. В. Моделирование эпидемической ситуации с учетом внешних рисков // Эпидемиология и вакцинопрофилактика. 2014. № 5 (78). С. 61–66.

8. Рвачев Л. А. Эксперимент по моделированию на УЦВМ эпидемии большого масштаба // Докл. АН СССР. 1968. Т. 180, № 2. С. 294–296.

9. Рвачев Л. А. Эксперимент по машинному прогнозированию эпидемии гриппа // Докл. АН СССР. 1971. Т. 198, № 1. С. 68–70.

10. Рвачев Л. А. Моделирование медико-биологических процессов в обществе как раздел динамики сплошных сред // Докл. АН СССР. 1972. Т. 203, № 3. С. 540–542.

11. Бароян О. В., Рвачев Л. А. Математика и эпидемиология. М.: Знание, 1977.

**A. N. Avlas, A. K. Demenchuk, S. V. Lemeshevskii, E. K. Makarov
Predicting the spread of coronavirus infection using equations with aftereffects**

Summary

The results of forecasting the first wave of the spread of COVID-19 coronavirus infection based on the simplified Baroyan–Rvachev model are presented.