

## Прогноз динамики пандемии COVID-19 по России на основе простых математических моделей эпидемий

Д. А. Томчин<sup>а</sup>, канд. техн. наук, [orcid.org/0000-0002-4223-373X](https://orcid.org/0000-0002-4223-373X), [dtomchin@yandex.ru](mailto:dtomchin@yandex.ru)

М. С. Ситчихина<sup>б</sup>, студент, [orcid.org/0000-0003-0798-7541](https://orcid.org/0000-0003-0798-7541)

М. С. Ананьевский<sup>б,а</sup>, канд. физ.-мат. наук, доцент, [orcid.org/0000-0003-2355-9184](https://orcid.org/0000-0003-2355-9184)

Т. А. Свенцицкая<sup>в</sup>, канд. физ.-мат. наук, доцент, [orcid.org/0000-0002-1931-5825](https://orcid.org/0000-0002-1931-5825)

А. Л. Фрадков<sup>а,б</sup>, доктор техн. наук, профессор, [orcid.org/0000-0002-5633-0944](https://orcid.org/0000-0002-5633-0944)

<sup>а</sup>Институт проблем машиноведения РАН, Большой пр. В. О., 61, Санкт-Петербург, 191778, РФ

<sup>б</sup>Санкт-Петербургский государственный университет, Университетская наб., 7–9, Санкт-Петербург, 199034, РФ

<sup>в</sup>Военно-космическая академия им. А. Ф. Можайского, Ждановская наб., 13, Санкт-Петербург, 197198, РФ

**Постановка проблемы:** пандемия нового коронавируса COVID-19, начавшаяся в 2020 г. и унесшая уже более пяти миллионов жизней, стала угрозой самому существованию человечества. Поэтому прогнозирование распространения COVID-19 в каждой отдельной стране является весьма актуальной задачей. Сложность ее решения обусловливается требованием быстрой обработки больших объемов данных и тем, что данные в большинстве своем неточны и не обладают статистическими свойствами, необходимыми для успешного применения статистических методов. Необходимо разработать простые методики прогнозирования на основе классических простых моделей эпидемиологии, слабо чувствительных к неточности данных. Важно также продемонстрировать работоспособность подхода применительно к данным о заболеваемости в России. **Цель:** получение прогнозных данных на основе классических простых моделей эпидемий – SIR и SEIR. **Методы:** для дискретных версий моделей SIR и SEIR предлагается оценивать параметры моделей с использованием редуцированной версии метода наименьших квадратов и применять сценарный подход к прогнозированию. Простота и небольшое число параметров являются преимуществами моделей SIR и SEIR, весьма существенными в условиях недостатка численных исходных данных и структурной неполноты моделей. **Результаты:** построен прогноз распространения COVID-19 в России на основе опубликованных данных по заболеваемости с 10 марта по 20 апреля 2020 г., а затем выборочно, по данным октября 2020 г. и данным октября 2021 г. Приводятся результаты сравнения прогноза по моделям SIR и SEIR. Тем же методом построены и представлены прогнозы на основе данных по заболеваемости осенью 2020 г. и осенью 2021 г. для России и для Санкт-Петербурга. Для задания параметров моделей, которые затруднительно определить из официальных данных, используется сценарный подход: анализируется динамика эпидемии при нескольких возможных значениях параметров. **Практическая значимость:** полученные результаты показывают, что предложенная методика хорошо предсказывает время наступления пика заболеваемости, несмотря на неточность исходных данных.

**Ключевые слова** – прогнозирование эпидемий, COVID-19, SIR, SEIR.

**Для цитирования:** Томчин Д. А., Ситчихина М. С., Ананьевский М. С., Свенцицкая Т. А., Фрадков А. Л. Прогноз динамики пандемии COVID-19 по России на основе простых математических моделей эпидемий. *Информационно-управляющие системы*, 2021, № 6, с. 31–41. doi:10.31799/1684-8853-2021-6-31-41

**For citation:** Tomchin D. A., Sitchikhina M. S., Ananievsky M. S., Svetsitskaya T. A., Fradkov A. L. Prediction of COVID-19 pandemic dynamics in Russia based on simple mathematical models of epidemics. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2021, no. 6, pp. 31–41 (In Russian). doi:10.31799/1684-8853-2021-6-31-41

### Введение

В 2020 г. человечество столкнулось с невиданной ранее угрозой: пандемией нового коронавируса COVID-19. Системы здравоохранения целого ряда развитых стран с трудом справляются с большим количеством тяжелобольных. Пандемия получила и продолжает собирать многочисленные жертвы. Ученые проводят обширные исследования, связанные с пандемией. На 1 ноября 2021 г. объем публикаций в базе данных «Скопус», у которых в заголовке есть термин «COVID-19», превысил 152 тысячи! Интерес специалистов в различных областях, равно как и

широкой публики, вызывает вопрос: как долго и насколько тяжело будет поражено человечество коронавирусной инфекцией. В сложной, труднопрогнозируемой ситуации может оказаться важной роль исследований с использованием математических моделей эпидемий. Основные модели в этой области известны уже более столетия и до сих пор привлекают внимание специалистов.

В настоящей работе для прогнозирования распространения COVID-19 в России применяются классические математические модели эпидемий SIR и SEIR. Несмотря на простоту, они часто используются в эпидемиологии, в том числе для анализа текущей эпидемии коронавируса

COVID-19 [1–4]. В некоторых работах говорится о преимуществах более простой модели SIR перед более сложной моделью SEIR [5]. Простота и небольшое число параметров являются преимуществами моделей SIR и SEIR, весьма существенными в условиях недостатка и недостоверности численных исходных данных и структурной неполноты моделей. Модели SIR и SEIR применялись для прогнозирования распространения COVID-19 в Китае, Франции, Италии, Германии, Португалии и ряде других стран [1–5]. Некоторые результаты по прогнозированию распространения COVID-19 в России в целом приводятся в работах [2, 6–10].

В предлагаемой работе представлен прогноз распространения COVID-19 в России по официальным статистическим данным о суточной заболеваемости и проведено сравнение результатов прогноза по моделям SIR и SEIR. Для задания параметров моделей, которые затруднительно определить из официальных данных, используется сценарный подход: анализируется динамика эпидемии при нескольких возможных значениях параметров. При выборе некоторых параметров модели SEIR используются результаты статьи [2]. Предварительные результаты были представлены авторами в работе [11], здесь эти результаты уточняются и строятся новые прогнозы по данным заболеваемости осенью 2020 г. и осенью 2021 г.

### Прогнозирование на основе модели SIR

В классической SIR-модели Кермака — Маккендрика [12] рассматривается три группы индивидов: восприимчивые к заболеванию (Susceptible), инфицированные (Infected) и переболевшие/умершие (Recovered/Removed). Передача инфекции осуществляется от инфицированных индивидов к восприимчивым. Считается, что переболевшие индивиды приобретают иммунитет и не могут быть заражены вторично. Модель описывается дифференциальными уравнениями

$$\begin{cases} \frac{dS}{dt} = \frac{-\beta}{N} S(t)I(t); \\ \frac{dI}{dt} = \frac{\beta}{N} S(t)I(t) - \gamma I(t); \\ \frac{dR}{dt} = \gamma I(t), \end{cases} \quad (1)$$

где  $S(t)$ ,  $I(t)$ ,  $R(t)$  — число восприимчивых, инфицированных и переболевших соответственно. Положительные величины  $\beta$ ,  $\gamma$  интерпретируются как параметры, значения которых определяются

скоростью инфицирования и выздоровления соответственно. Подробнее о свойствах SIR-модели см. [13, 14].

Для применения модели к конкретной ситуации необходимо провести ее калибровку — определение значений параметров и начальных условий из экспериментальных данных. Будем использовать официальные данные Всемирной организации здравоохранения (ВОЗ), представленные в удобной форме на сайте Worldometers [15]. Поскольку данные ВОЗ обновляются один раз в день, удобно перейти от дифференциальной модели (1) к дискретной:

$$\begin{cases} S(t+1) - S(t) = \frac{-\beta}{N} S(t)I(t); \\ R(t+1) - R(t) = \gamma I(t), \quad t = 1, 2, \dots \end{cases} \quad (2)$$

Используем для калибровки данные [15] за период с 10 марта по 20 апреля 2020 г. Начальные условия равны значениям переменных 10 марта 2020 г.

Для оценки параметров используем метод наименьших квадратов (МНК). Удобно применить его в два этапа: сначала оценить  $\gamma$  по формуле

$$\hat{\gamma} = \frac{\sum_{t=1}^T (R(t+1) - R(t))I(t)}{\sum_{t=1}^T I(t)^2},$$

а затем оценить  $\beta$  по формуле

$$\hat{\beta} = N \frac{\sum_{t=1}^T (I(t+1) - I(t) + \hat{\gamma}I(t))S(t)I(t)}{\sum_{t=1}^T (S(t)I(t))^2}.$$

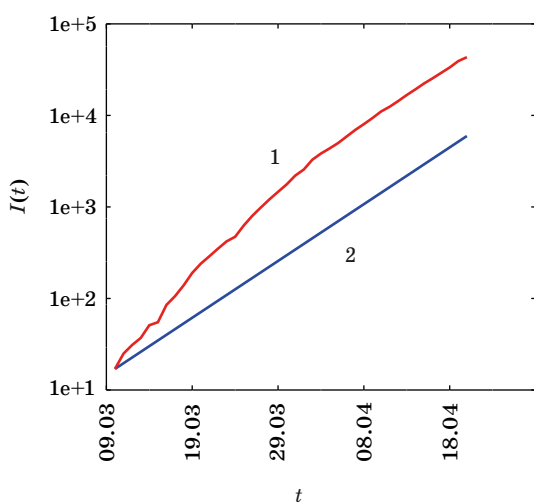
Через  $N$  обозначим начальное число восприимчивых, совпадающее с численностью населения Российской Федерации:  $N = 146,745$  млн. Расчет дает следующие значения параметров:  $\beta = 0,175$ ,  $\hat{\gamma} = 0,0164$ .

Результаты калибровки модели приведены на рис. 1. На первый взгляд ошибка кажется большой, но надо принимать во внимание, что модель должна в первую очередь учитывать не общее число инфицированных, а скорость роста. Тем не менее можно попытаться уточнить оценки. Из графиков видно, что точность аппроксимации снижается после 27 марта, поскольку снижается реальная скорость роста заболеваемости. Очевидно, это связано с введением в 20-х числах марта в РФ частичной самоизоляции. Для более точного прогнозирования с учетом самоизоляции временной отрезок наблюдений был разбит на две части: до 28 марта и с 28 марта по 20 апреля, — и параметры модели на каждом множестве оценивались отдельно. Если оценивать параметры модели отдельно по данным с 11 по 20 марта,

получим  $\hat{\beta} = 0,26$ , тогда как по данным с 21 марта по 15 апреля получим  $\hat{\beta} = 0,175$ , т. е. ту же величину, что и в первом случае. Оценки параметра  $\gamma$  также в обоих случаях совпадают:  $\hat{\beta} = 0,0164$ . Результаты калибровки модели по усеченному набору данных представлены на рис. 2.

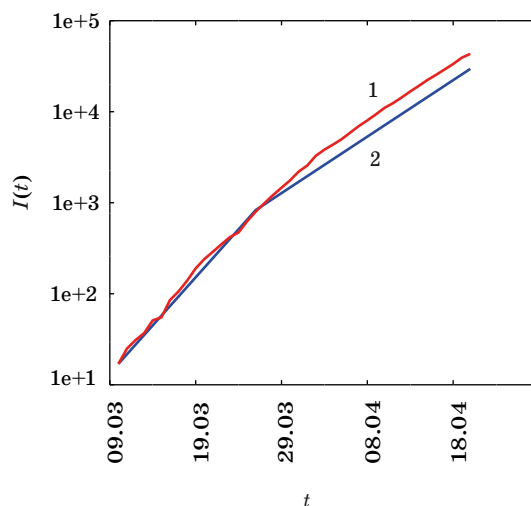
Результаты прогноза развития эпидемии на 120 дней вперед представлены на рис. 3, а и б. На рис. 3, а прогноз строится по данным с 10 марта по 20 апреля на период после 20 апреля. Из гра-

фика видно, что пик заболеваемости (выход на плато) прогнозируется примерно на 70-й день, т. е. 30 июня 2020 года. Для сравнения на рис. 3, б для прогноза на период после 28 марта используются значения параметров модели, рассчитанные по данным 10–28 марта, т. е. до введения режима самоизоляции. Из графика видно, что если бы режим самоизоляции не вводился, то пик наступил



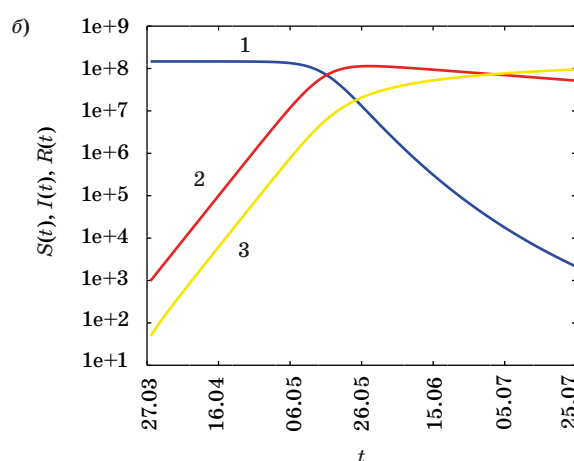
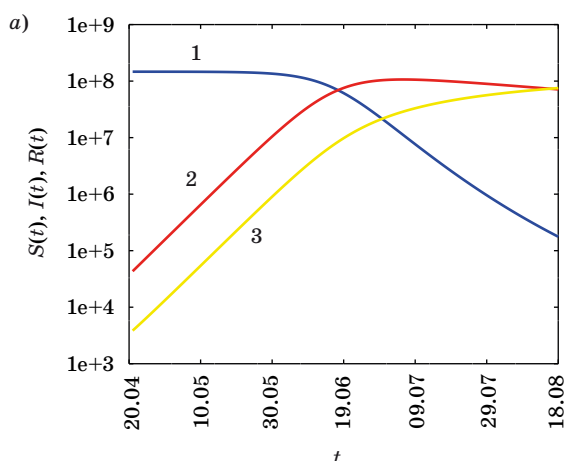
■ **Рис. 1.** Калибровка SIR-модели по данным 10 марта–20 апреля 2020 г.: 1 — реальные данные; 2 — график переменной  $I(t)$  по модели

■ **Fig. 1.** Calibration of the SIR model according to the data from 10 Mar. till 20 Apr. 2020: 1 — corresponds to the real data; 2 — the graph of the variable  $I(t)$  according to the model



■ **Рис. 2.** Калибровка SIR-модели отдельно по данным 10–27 марта 2020 г. и 28 марта–16 апреля 2020 г.: 1 — реальные данные; 2 — график переменной  $I(t)$  по скорректированной модели

■ **Fig. 2.** Calibration of the SIR model separately according to data from 10 Mar. till 27 Mar. 2020 and from 28 Mar. till 16 Apr. 2020: 1 — corresponds to the real data; 2 — the graph of the variable  $I(t)$  according to the updated model



■ **Рис. 3.** Прогноз по SIR-модели-1 на 120 дней: а — с 20.04.2020 по 18.08.2020 по данным 10 марта–20 апреля; б — с 27.03.2020 по 25.07.2020 по данным 11–20 марта: 1 — число восприимчивых  $S(t)$ ; 2 — число инфицированных  $I(t)$ ; 3 — число выведенных  $R(t)$  (сумма числа выздоровевших и числа умерших)

■ **Fig. 3.** Forecast from SIR model-1 for 120 days: а — from 20.04.2020 till 18.08.2020 based on the data 10 Mar.–20 Apr.; б — from 27.03.2020 till 25.07.2020 based on the data 11–20 Mar.: 1 — the number of susceptible  $S(t)$ ; 2 — the number of infected  $I(t)$ ; 3 — the number of removed  $R(t)$  (sum of the number of recovered and the number of the deceased)

бы раньше, на 60-й день, т. е. 26 мая. Полученные результаты можно использовать для оценки интенсивности карантинных мер, требуемых для заданного замедления развития вспышек эпидемий и пандемий.

Таким образом, даже простейшая SIR-модель показывает эффективность введения режима самоизоляции с точки зрения растяжения во времени фазы роста заболеваемости и позволяет дать, пусть и грубые, но полезные оценки времени роста заболеваемости. Что касается числа инфицированных, его прогноз на первый взгляд выглядит неутешительно: для того чтобы начать снижаться, это число должно приблизиться к общей численности населения страны. Однако надо учитывать, что в популяции имеется большое число бессимптомных инфицированных, которые легко переносят болезнь, зачастую даже не догадываясь о ней. Возможность количественного учета бессимптомных больных будет показана в следующем разделе с использованием модели SEIR.

### Прогнозирование на основе модели SEIR

Модель SEIR (Susceptible — Exposed — Infected — Recovered) учитывает инкубационный период заболевания [13, 14]. Эта модель применялась для прогнозирования распространения COVID-19 в Китае [1] и в европейских странах [2]. Мы будем использовать модифицированную SEIR-модель с дискретным временем [2], описываемую уравнениями

$$\begin{cases} S(t+1) - S(t) = \frac{-\beta}{N}(p_c I(t) + r_t E(t))S(t); \\ E(t+1) - E(t) = -\sigma E(t) + \frac{\beta}{N}(p_c I(t) + r_t E(t))S(t); \\ I(t+1) - I(t) = \sigma E(t) - \gamma I(t); \\ R(t+1) - R(t) = \gamma I(t), \quad t = 1, 2, \dots \end{cases} \quad (3)$$

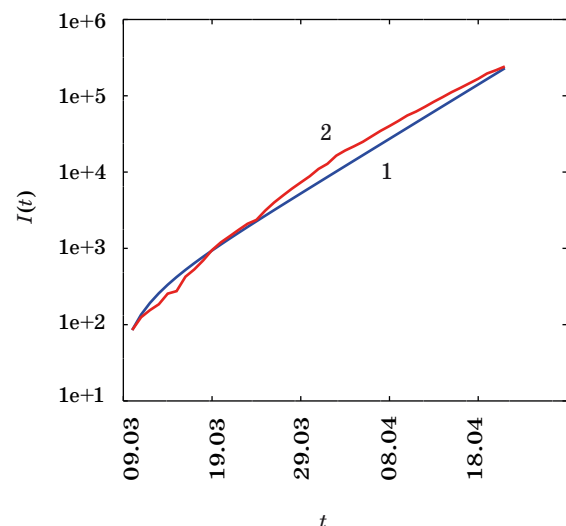
Здесь  $S(t)$ ,  $I(t)$ ,  $R(t)$  имеют тот же смысл, что и в SIR-модели (2), а  $E(t)$  — численность инфицированных индивидов, находящихся в стадии инкубационного (латентного) периода. Общая численность населения  $N = S(t) + E(t) + I(t) + R(t)$ , как и ранее, считается постоянной.

Модель SEIR (3) имеет два параметра, значения которых можно оценить на основе реальных данных о заболеваемости:  $\gamma$ ,  $\beta$ . Параметр  $\gamma > 0$  представляет интенсивность смертности и выздоровления, параметр  $\beta > 0$  соответствует скорости заражения вирусом восприимчивых людей во время контакта с инфицированными или латентными. Параметр  $\sigma > 0$  определяет скорость инкубации, с которой появляются симптомы у индивидов, находящихся

в латентном периоде. Величину  $\sigma$  можно выбрать обратно пропорциональной средней длительности инкубационного периода COVID-19:  $\sigma = 1/7$ .

Число  $p_c > 0$  соответствует числу контактов на человека в день для инфицированных  $I(t)$  (предполагается, что если инфицированные люди с симптомами находятся в карантине (самоизоляции), то количество контактов уменьшается);  $r_t > p_c$  — число контактов на человека в день для находящихся в латентном периоде  $E(t)$ . Параметры  $r_t$ ,  $p_c$  можно изменять путем применения мер, регулирующих социальную дистанцию (карантин, самоизоляция). Они зависят также от плотности населения и социальных традиций.

Важной особенностью пандемии COVID-19, как и ряда предыдущих эпидемий, является существенное отличие реального числа инфицированных от документированного. Это связано с наличием значительного числа бессимптомных инфицированных, с невозможностью осуществить полное тестирование, а также с неточностью и недостаточной чувствительностью тестов. Статистические данные по Европе и США говорят о том, что доля недокументированных случаев может меняться от 40 до 90 % [16]. Обозначим через  $\alpha$  отношение общего числа инфицированных к числу документированных инфицированных. Учитывая значительную неопределенность и трудность определения  $\alpha$ , проведем расчеты для  $\alpha = 5$  (близкие значения принимались и в исследованиях развития пандемии COVID-19 во Франции [2]) и  $\alpha = 10$ .



■ **Рис. 4.** Калибровка SEIR-модели при  $\alpha = 5$  по данным 10 марта–20 апреля 2020 г.: 1 — число инфицированных по модели; 2 — реальные данные  $I(t)$

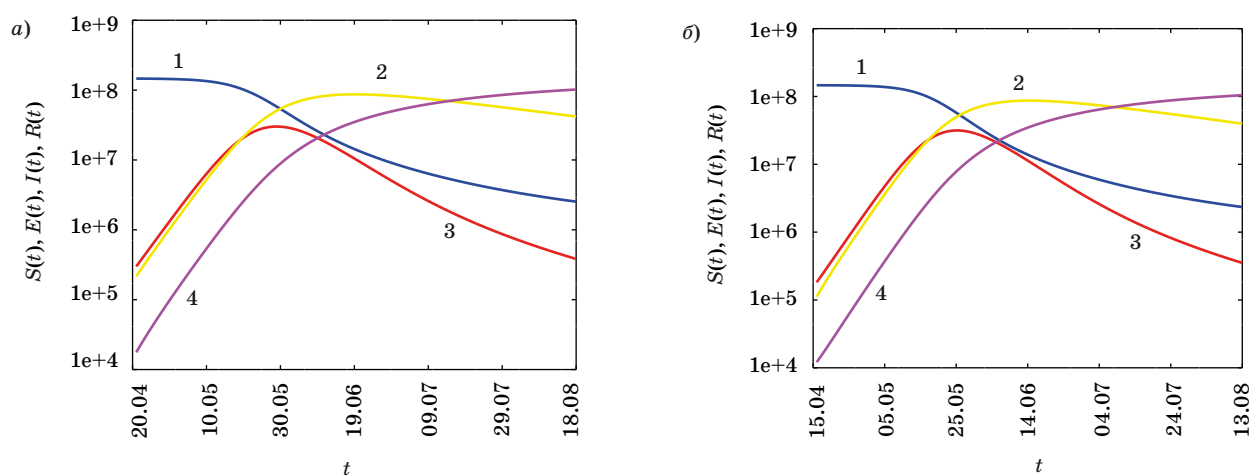
■ **Fig. 4.** Calibration of the SEIR model at  $\alpha = 5$  according to data from 10 Mar. till 20 Apr. 2020: 1 — the number of infected evaluated from the model; 2 — the real data  $I(t)$

Рассмотрим сначала случай  $\alpha = 5$ . В результате оценки параметров модели SEIR по МНК по данным с 10 марта по 20 апреля получим:  $\beta = 0,027$ ,  $\gamma = 0,017$ . Следуя [2] и исходя из близости социальных традиций во Франции и в России, примем следующие значения остальных параметров в качестве базовых:  $\sigma = 1/7$ ,  $p_c = 2$ ,  $r = 10$ . Эти значения соответствуют текущему карантинному режиму (модель-1). Начальные условия  $S(t)$ ,  $I(t)$ ,  $E(t)$ ,  $R(t)$  при  $t = 0$  также задаются аналогично [2].

Данные о числе инфицированных  $I(t)$ , числе выздоровевших  $H(t)$  и числе умерших  $D(t)$  берут-

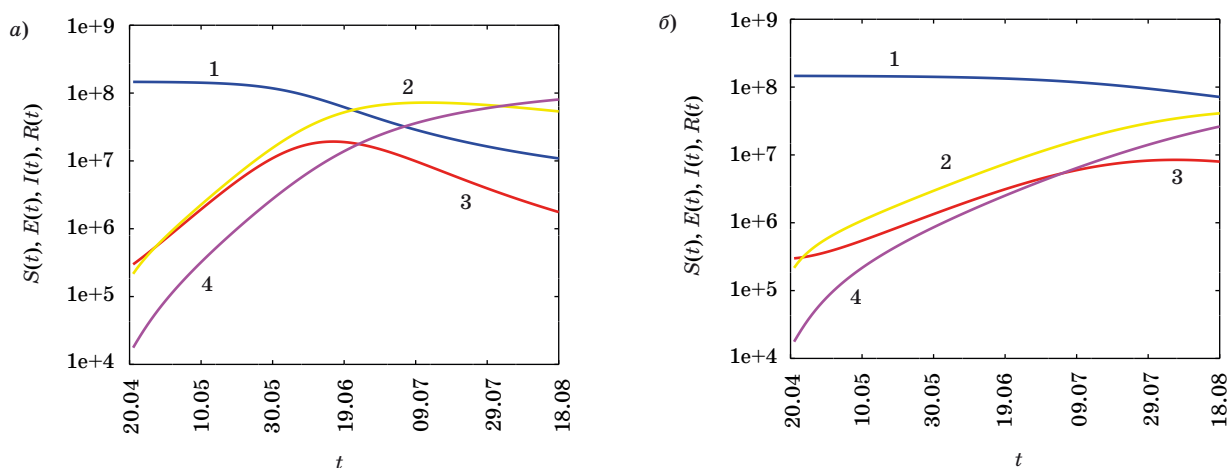
ся из официальных источников [15]. Результаты калибровки представлены на рис. 4. Они показывают хорошую точность аппроксимации реальных данных моделью. Результаты прогноза на 120 дней по модели представлены на рис. 5, а и б. Видно, что по данным с 10 марта по 20 апреля пик эпидемии прогнозируется на 40-й день — 30 мая 2020 г. (см. рис. 5, а).

Для сравнения на рис. 5, б показаны результаты прогноза по модели, построенной по данным с 10 марта по 15 апреля. Такую модель можно было построить, естественно, с использованием



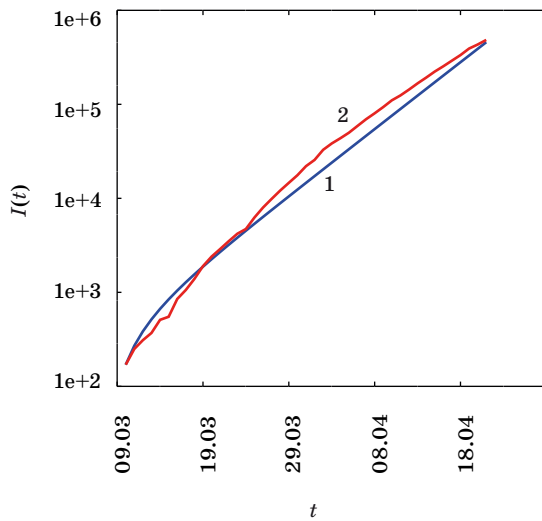
■ **Рис. 5.** Прогноз по SEIR-модели-1 на 120 дней при  $\alpha = 5$ : а — с 20.04.2020 по 18.08.2020 по данным 10 марта–20 апреля; б — с 15.04.2020 по 13.08.2020 по данным 10 марта–15 апреля: 1 — число восприимчивых  $S(t)$ ; 2 — число инфицированных в инкубационном периоде  $E(t)$ ; 3 — число инфицированных  $I(t)$ ; 4 — число выведенных  $R(t)$

■ **Fig. 5.** Forecast based on SEIR model-1 for 120 days at  $\alpha = 5$ : а — from 20.04.2020 till 18.08.2020 according to data from 10 Mar. till 20 Apr.; б — from 15.04.2020 till 13.08.2020 according to data from 10 Mar. till 15 Apr.: 1 — the number of susceptible  $S(t)$ ; 2 — the number of infected in the incubation period  $E(t)$ ; 3 — the number of infected  $I(t)$ ; 4 — the number of removed  $R(t)$



■ **Рис. 6.** Прогноз по SEIR-модели-2 (строгий карантин) (а) и по SEIR-модели-3 (очень строгий карантин) (б) при  $\alpha = 5$  по данным 10 марта–20 апреля 2020 г.: 1 — число восприимчивых  $S(t)$ ; 2 — число инфицированных в инкубационном периоде  $E(t)$ ; 3 — число инфицированных  $I(t)$ ; 4 — число выведенных  $R(t)$

■ **Fig. 6.** Forecast based on SEIR model-2 (strict quarantine) (а) and on SEIR model-3 (very strict quarantine) (б) for  $\alpha = 5$  according to data from 10 Mar. till 20 Apr. 2020: 1 — the number of susceptible  $S(t)$ ; 2 — the number of infected in the incubation period  $E(t)$ ; 3 — the number of infected  $I(t)$ ; 4 — the number of removed  $R(t)$



■ **Рис. 7.** Калибровка SEIR-модели при  $\alpha = 10$  по данным 10 марта–20 апреля 2020 г.: 1 — число инфицированных по модели; 2 — реальные данные  $I(t)$

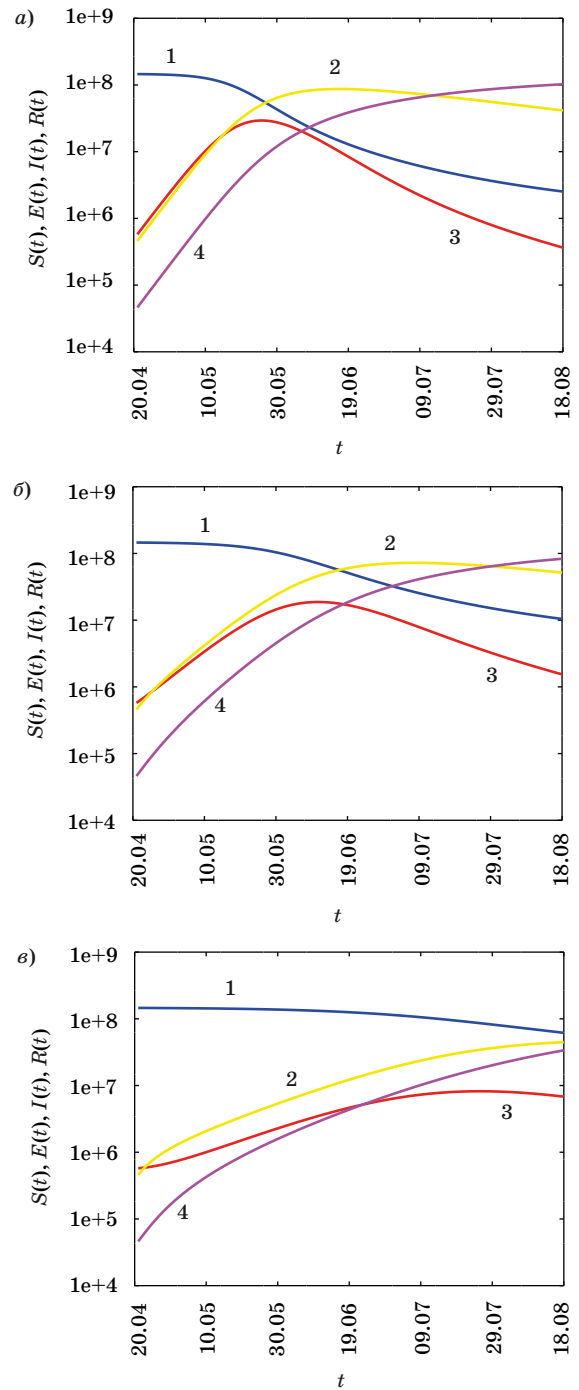
■ **Fig. 7.** Calibration of the SEIR model at  $\alpha = 10$  according to data from 10 Mar. till 20 Apr. 2020: 1 — the number of infected evaluated from the model; 2 — the real data  $I(t)$

меньшего количества данных. Казалось бы, она должна иметь меньшую предсказательную силу. Однако на графике видно, что даже такая «устаревшая» модель прогнозирует пик заболеваемости не менее точно, на 45-й день, или также на 30 мая.

Аналогичный прогноз для усиленного карантинного режима (модель-2,  $p_c = 1,5$ ,  $r = 7,5$ ) представлен на рис. 6, а, а на рис. 6, б — для еще более строгого карантинного режима (модель-3,  $p_c = 1$ ,  $r = 5$ ).

Таким образом, как видно из графиков рис. 5 и 6, при стандартном карантинном режиме пик заболеваемости наступает на 40-й день — 30 мая, при строгом режиме на 60-й день — 10 июня, а при очень строгом режиме на 100-й день — 20 июля 2020 года. Видно также, что более строгий режим приводит к снижению пика заболеваемости: при стандартном режиме число заболевших может достигать 50 млн, при строгом карантинном режиме величина пика падает до 20 млн, а при очень строгом режиме становится менее 10 млн. При этом доля документированных случаев еще в  $\alpha$  раз меньше.

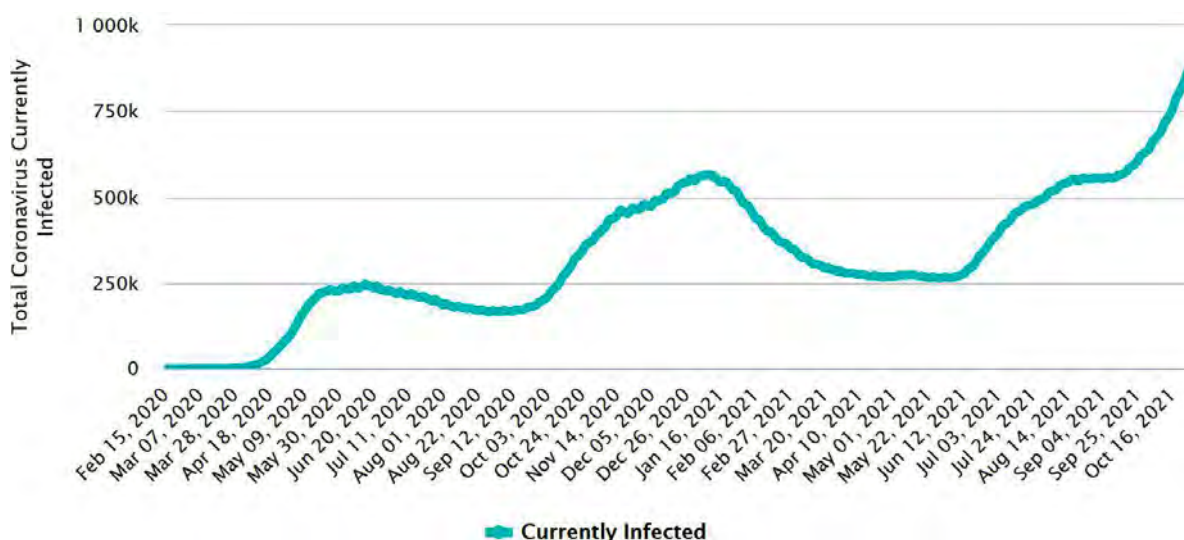
Перейдем к рассмотрению случая  $\alpha = 10$ . Калибровка модели проводится аналогично случаю  $\alpha = 5$ . Результаты калибровки представлены на рис. 7. Они также показывают хорошую точность аппроксимации реальных данных моделью везде, кроме начального участка. Результаты прогноза на 120 дней на период с 20 апреля по 18 августа 2020 г. по модели представлены на рис. 8, а. Из графиков видно, что прогнозируемые даты достижения максимального числа инфици-



■ **Рис. 8.** Прогноз по SIR-модели-1 (а), по SEIR-модели-2 (строгий карантин) (б) и по SEIR-модели-3 (очень строгий карантин) (в) при  $\alpha = 10$  на 120 дней с 20.04.2020 по 18.08.2020 по данным 10 марта–20 апреля 2020 г.: 1 — число восприимчивых  $S(t)$ ; 2 — число инфицированных в инкубационном периоде  $E(t)$ ; 3 — число инфицированных  $I(t)$ ; 4 — число введенных  $R(t)$

■ **Fig. 8.** Forecast based on SEIR model-1 (a), on SEIR model-2 (strict quarantine) (б) and on SEIR model-3 (very strict quarantine) (в) for  $\alpha = 10$  from 20.04.2020 till 18.08.2020 based on the data 10 Mar.–20 Apr. 2020: 1 — the number of susceptible  $S(t)$ ; 2 — the number of infected in the incubation period  $E(t)$ ; 3 — the number of infected  $I(t)$ ; 4 — the number of removed  $R(t)$





■ **Рис. 9.** Число инфицированных в России с 15.02.2020 по 28.10.2021 (по данным ВОЗ [15])  
 ■ **Fig. 9.** Number of infected in Russia from 15.02.2020 till 28.10.2021 (according to WHO data [15])

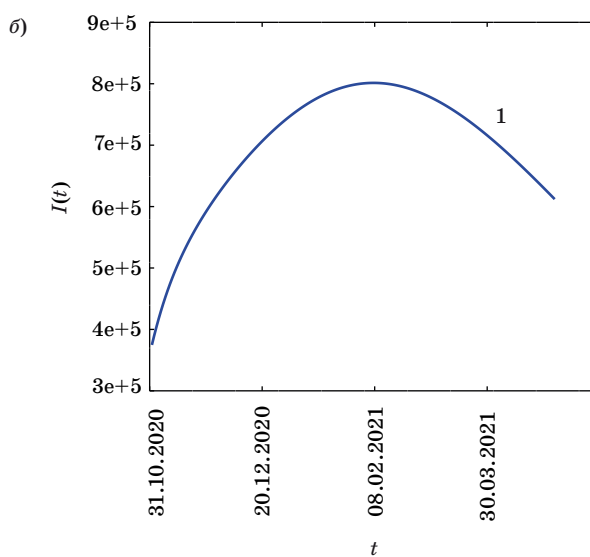
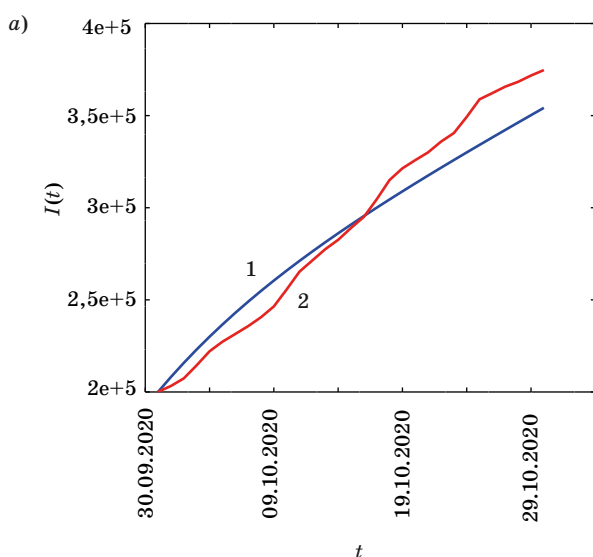
рованных приходится на 23–24 мая. На рис. 8, б представлен аналогичный прогноз для усиленного карантинного режима (модель-2,  $p_c = 1,5$ ,  $r = 7,5$ ), а на рис. 8, в — прогноз для очень строгого карантинного режима (модель-3,  $p_c = 1$ ,  $r = 5$ ). Видно, что изменение характера процессов при изменении величины  $\alpha$  от 5 до 10 невелико.

Интересно сравнить результаты прогнозирования с фактическими данными заболеваемости. На рис. 9 показано число инфицированных по официальным данным ВОЗ [15]. Пик первой (весенней) волны пандемии 2020 г. был достигнут

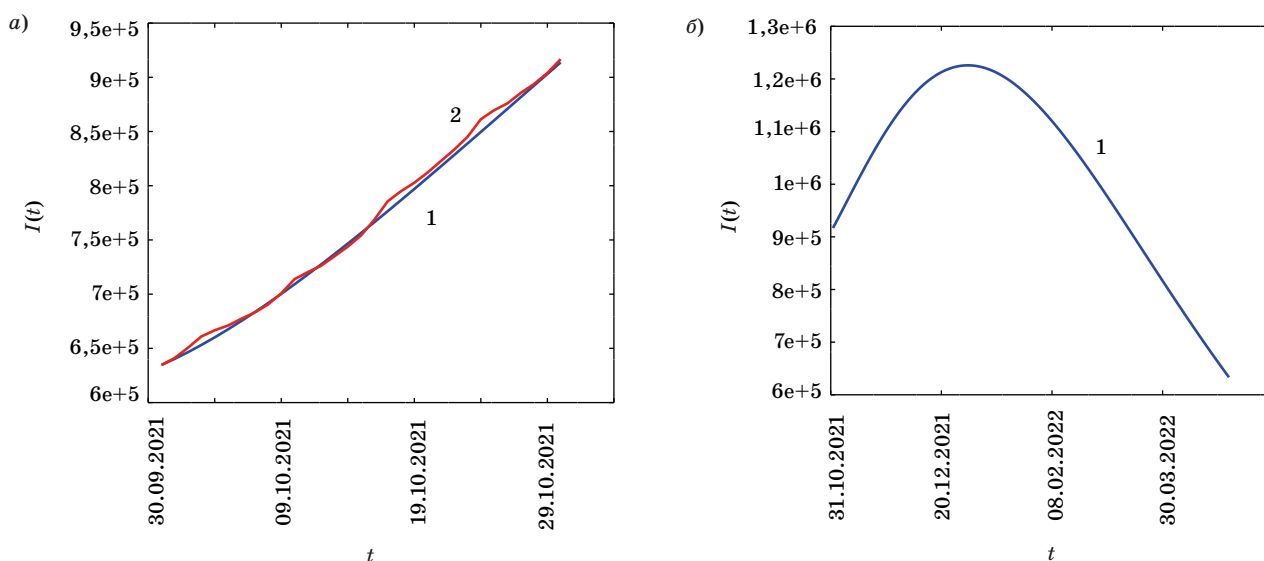
в конце июня, т. е. предложенный в работе подход дает более точный прогноз, чем прогнозы, появлявшиеся в апреле-мае 2020 г. в СМИ.

### Прогноз по осенним данным 2020 и 2021 года

Представленный подход применяется далее к прогнозированию числа инфицированных COVID-19 в России по модели SEIR по осенним данным 2020 и 2021 г. Построим прогноз по дан-



■ **Рис. 10.** Калибровка (а) и прогноз (б) SEIR-модели на 240 дней по данным 1–31 октября 2020 г.: 1 — число инфицированных по модели; 2 — реальные данные  $I(t)$   
 ■ **Fig. 10.** Calibration (a) and forecast (б) of the SEIR model for 240 days according to data from 1 till 31 Oct. 2020: 1 — the number of infected evaluated from the model; 2 — the real data  $I(t)$



■ **Рис. 11.** Калибровка (а) и прогноз (б) SEIR-модели на 240 дней по данным 1–31 октября 2021 г.: 1 — число инфицированных по модели; 2 — реальные данные  $I(t)$

■ **Fig. 11.** Calibration (a) and forecast (б) of the SEIR model for 240 days according to data from 1 till 31 Oct. 2021: 1 — the number of infected evaluated from the model; 2 — the real data  $I(t)$

ным за октябрь 2020 г. с параметрами модели SEIR по МНК:  $\beta = 0,01$ ,  $\gamma = 0,03$ . Результаты калибровки представлены на рис. 10, а. На рис. 10, б видим заметный рост числа инфицированных, что соответствует наблюдаемой в 2020 г. ситуации. При этом пик заболеваемости по модели приходится на начало февраля 2021 г., в то время как реальный пик был отмечен в начале января 2021 г. (см. рис. 9).

Проведем аналогичные исследования на данных за октябрь 2021 г. из [15]. Оценивая параметры модели SEIR по МНК, получим  $\beta = 0,018$ ,  $\gamma = 0,02$ . На рис. 11, а представлена калибровка модели, видно, что точность аппроксимации весьма высока.

Отметим, что октябрь 2020-го и октябрь 2021 г. похожи тем, что заболеваемость COVID-19 в эти периоды возрастает. Возможно, именно это обуславливает правдоподобие прогнозов.

По рис. 11, б можно сделать вывод, что число заболевших в конце 2021 г. растет.

## Заключение

В статье приведены результаты расчета параметров простых математических моделей динамики пандемии коронавируса COVID-19 — классических моделей эпидемиологии SIR и SEIR. В качестве исходных данных для расчета взяты официальные данные ВОЗ по заболеваемости в России с 10 марта по 20 апреля 2020 г., а также за октябрь 2020 и 2021 гг. Получены вари-

анты прогноза числа инфицированных на 120 и 240 дней при изменении параметра социальной дистанции, характеризующего различные режимы строгости карантинных ограничительных мер, и параметра  $\alpha$ , определяющего долю недокументированных случаев инфицирования.

Сравнение результатов прогноза с реальными данными показывает, что простые модели SIR и SEIR работоспособны в условиях ограниченного объема и недостаточной достоверности данных и дают возможность прогнозировать процессы распространения вируса не только качественно, но и количественно. Результаты прогноза могут использоваться для оценки эффективности ограничительных мер, для выбора оптимальной стратегии, например путем периодического усиления и ослабления карантинных мер [17], и т. д.

Расчет по данным за октябрь 2021 г. показал, что пик числа инфицированных в России зимой 2021/2022 года следует ожидать не ранее января 2022 года. На момент написания статьи судить о достоверности данного прогноза невозможно.

Следует заметить, что исследование проводилось при ряде дополнительных предположений: считалось, что все население страны восприимчиво к вирусу, что вирус не обладает сезонностью и т. п. При таких предпосылках итоговым сценарием является то, что в конечном итоге переболеет значительная часть населения, а от принимаемых мер зависит только темп распространения болезни. В реальности же распространение инфекции может быть ограничено какими-то факторами, которые в настоящий момент неизвест-



ны и здесь не учитывались. С другой стороны, на динамику могут влиять и другие разнонаправленные факторы, такие как повторное заражение, мутации вируса и т. д. При этом наиболее правдоподобным оказывается пессимистический сценарий, чему способствует недостаточная информированность не только населения и властей, но и специалистов о ходе развития пандемии. При этом пессимизм прогнозов специалистов [18] усугубляется объективными данными о низкой культуре соблюдения профилактических мер населением.

За прошедшее с начала пандемии время арсенал математических моделей, предлагаемых специалистами, значительно расширился. Используются регрессионные модели [6], модели с применением машинного обучения [9, 19], модели более высоких порядков [20] и др. Казалось

бы, учет дополнительных факторов, в том числе региональной специфики, должен приводить к повышению точности прогнозов. Однако, по мнению авторов, увеличение сложности модели оправдано лишь при повышении достоверности исходных данных для расчета.

Тем не менее в дальнейших исследованиях целесообразно учесть дополнительные факторы: пространственные неоднородности (сетевую структуру) моделей, неполную восприимчивость населения к инфекции и эффекты повторного заражения, а в исследованиях на основе данных 2021 г. и более поздних следует учесть влияние степени вакцинации населения.

Более того, на основе модели можно планировать режим вакцинации и использовать математические методы для создания щадящих колебательных режимов вакцинации [21, 22].

## Литература

1. Yang Z., Zeng Z., Wang K., Wong S.-S., Liang W., Zanin M., Liu P., Cao X., Gao Z., Mai Z., Liang J., Liu X., Li S., Li Y., Ye F., Guan W., Yang Y., Li F., Luo S., Xie Y., Liu B., Wang Z., Zhang S., Wang Y., Zhong N., He J. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *Journal of Thoracic Disease*, 2020, vol. 12, no. 3. doi:10.21037/jtd.2020.02.64
2. Efimov D., Ushirobira R. On an interval prediction of COVID-19 development based on a SEIR epidemic model. *Annual Reviews in Control*, 2021, vol. 51, pp. 477–487. doi:10.1016/j.arcontrol.2021.01.006
3. Zhong L., Mu L., Li J., Wang J., Yin Z., Liu D. Early prediction of the 2019 novel coronavirus outbreak in the mainland China based on simple mathematical model. *IEEE Access*, 2020, vol. 8, pp. 51761–51769. doi:10.1109/ACCESS.2020.2979599
4. Teles P. A time-dependent SEIR model to analyse the evolution of the SARS-CoV-2 epidemic outbreak in Portugal. *arXiv:2004.04735*. <https://arxiv.org/abs/2004.04735> (дата обращения: 02.08.2021).
5. Roda W. C., Varughese M. B., Han D., Li M. Y. Why is it difficult to accurately predict the COVID-19 epidemic? *Infectious Disease Modelling*, 2020, no. 5, pp. 271–281. doi:10.1016/j.idm.2020.03.001
6. Попков Ю. С., Дубнов Ю. А., Попков А. Ю. Прогнозирование развития эпидемии COVID-19 в странах Европейского союза с использованием энтропийно-рандомизированного подхода. *Информатика и автоматизация*, 2021, т. 20, № 5, с. 1010–1033. doi:10.15622/20.5.1
7. Захаров В. В., Балыкина Ю. Е. Балансовая модель эпидемии COVID-19 на основе процентного прироста. *Информатика и автоматизация*, 2021, т. 20, № 5, с. 1034–1064. doi:10.15622/20.5.2
8. Aronov I. Z., Maksimova O. V., Galkina N. M. COVID-19 highest incidence forecast in Russia based on regression model. *International Journal of Mathematical, Engineering and Management Sciences*, 2020, vol. 5, no. 5, pp. 812–819. doi:10.33889/IJMEMS.2020.5.5.063
9. Wang P., Zheng X., Ai G., Liu D., Zhu B. Time series prediction for the epidemic trends of COVID-19 using the improved LSTM deep learning method: Case studies in Russia, Peru and Iran. *Chaos, Solitons and Fractals*, 2020, vol. 140, no. 110214. doi:10.1016/j.chaos.2020.110214
10. Кольцова Э. М., Куркина Е. С., Васецкий А. М. Математическое моделирование распространения эпидемии коронавируса COVID-19 в ряде европейских, азиатских стран, Израиле и России. *Проблемы экономики и юридической практики*, 2020, т. 16, № 2, с. 154–165.
11. Tomchin D. A., Fradkov A. L. Prediction of the COVID-19 spread in Russia based on SIR and SEIR models of epidemics. *OnLine*, 2020, vol. 53, no. 5, pp. 833–838. doi:10.1016/j.ifacol.2021.04.209
12. Kermack W. O., McKendrick A. G. A contribution to the mathematical theory of epidemics. *Proc. Roy. Soc. Lond., Ser. A*, 1927, no. 115, pp. 700–721.
13. Братусь А. С., Новожилов А. С., Платонов А. П. *Динамические системы и модели биологии*. М., Физматлит, 2010. 400 р.
14. Wang Z., Bauch C. T., Bhattacharyya S., d'Onofrio A., Manfredi P., Perc M., Perra N., Salathé M., Zhao D. Statistical physics of vaccination. *Physics Reports*, 2016, vol. 664, pp. 1–113. doi:10.1016/j.physrep.2016.10.006
15. *Worldometers*. <https://www.worldometers.info/coronavirus/country/russia/> (дата обращения: 02.08.2021).
16. Li R., Pei S., Chen B., Song Y., Zhang T., Yang W., Shaman J. Substantial undocumented infection fa-

- ilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science*, 2020, vol. 368, с. 489–493. doi:10.1126/science.abb3221
17. Eubank S., Eckstrand I., Lewis D., Venkatraman S., Marathe M., Barrett C. L. Commentary on Ferguson, et al. “Impact of Nonpharmaceutical Interventions (NPIs) to reduce COVID-19 mortality and healthcare demand”. *Bulletin of Mathematical Biology*, 2020, vol. 82, Article 52. doi:10.1007/s11538-020-00726-x
18. Даниленко Д. М., Комиссаров А. Б., Стукова М. А., Ллюзнов Д. А. Быть или не быть: прогноз развития эпидемии COVID-19 в России. *Журнал инфектологии*, 2020, т. 12, № 3, с. 6–11. doi:10.22625/2072-6732-2020-12-3-6-11
19. Al-qaness M. A. A., Saba A. I., Elsheikh A. H., Elaziz M. A., Ibrahim R. A., Lu S., Hemedan A. A., Shanmugan S., Ewees A. A. Efficient artificial intelligence forecasting models for COVID-19 outbreak in Russia and Brazil. *Process Safety and Environmental Protection*, 2021, vol. 149, pp. 399–409. doi:10.1016/j.psep.2020.11.007
20. Giordano G., Blanchini F., Bruno R., Colaneri P., Di Filippo A., Di Matteo A., Colaneri M. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nature Medicine*, 2020, vol. 26, pp. 855–860. doi:10.1038/s41591-020-0883-7
21. d’Onofrio A., Manfredi P., Poletti P. The interplay of public intervention and private choices in determining the outcome of vaccination programmes. *PLoS ONE*, 2012, vol. 7, iss. 10: e45653. https://doi.org/10.1371/journal.pone.0045653
22. Efimov D. V., Fradkov A. L. Oscillatory of nonlinear systems with static feedback. *SIAM J Control Optim.*, 2009, vol. 48, pp. 618–640. https://doi.org/10.1137/070706963

UDC 004.942

doi:10.31799/1684-8853-2021-6-31-41

### Prediction of COVID-19 pandemic dynamics in Russia based on simple mathematical models of epidemics

D. A. Tomchin<sup>a</sup>, PhD, Tech., orcid.org/0000-0002-4223-373X, dtomchin@yandex.ru

M. S. Sitchikhina<sup>b</sup>, Student, orcid.org/0000-0003-0798-7541

M. S. Ananievsky<sup>b,a</sup>, PhD, Phys.-Math., Associate Professor, orcid.org/0000-0003-2355-9184

T. A. Svetsitskaya<sup>c</sup>, PhD, Phys.-Math., Associate Professor, orcid.org/0000-0002-1931-5825

A. L. Fradkov<sup>a,b</sup>, Dr. Sc., Tech., Professor, orcid.org/0000-0002-5633-0944

<sup>a</sup>Institute of Mechanical Engineering Problems of the RAS, 61, Bol’shoi Pr. V. O., 199178, Saint-Petersburg, Russian Federation

<sup>b</sup>Saint-Petersburg State University, 7–9, Universitetskaya Emb., 199034, Saint-Petersburg, Russian Federation

<sup>c</sup>A. F. Mozhaiskii Military Space Academy, 13, Zhdanovskaya Emb., 197198, Saint-Petersburg, Russian Federation

**Introduction.** The COVID-19 pandemic which began in 2020 and has taken more than five million lives has become a threat to the very existence of mankind. Therefore, predicting the spread of COVID-19 in each individual country is a very urgent task. The complexity of its solution is due to the requirement for fast processing of large amounts of data and the fact that the data are mostly inaccurate and do not have the statistical properties necessary for the successful application of statistical methods. Therefore, it seems important to develop simple forecasting methods based on classical simple models of epidemiology which are only weakly sensitive to data inaccuracies. It is also important to demonstrate the feasibility of the approach in relation to the incidence data in Russia. **Purpose:** Obtaining forecast data based on classical simple models of epidemics, namely SIR and SEIR. **Methods:** For discrete versions of SIR and SEIR models, it is proposed to estimate the parameters of the models using a reduced version of the least squares method, and apply a scenario approach to the forecasting. The simplicity and a small number of parameters are the advantages of SIR and SEIR models, which is very important in the context of a lack of numerical input data and structural incompleteness of the models. **Results:** A forecast of the spread of COVID-19 in Russia has been built based on published data on the incidence from March 10 to April 20, 2020, and then, selectively, according to October 2020 data and October 2021 data. The results of the comparison between SIR and SEIR forecasts are presented. The same method was used to construct and present forecasts based on morbidity data in the fall of 2020 and in the fall of 2021 for Russia and for St. Petersburg. To set the parameters of the models which are difficult to determine from the official data, a scenario approach is used: the dynamics of the epidemic is analyzed for several possible values of the parameters. **Practical relevance:** The results obtained show that the proposed method predicts well the time of the onset of the peak incidence, despite the inaccuracy of the initial data.

**Keywords** — forecasting epidemics, COVID-19, SIR, SEIR.

**For citation:** Tomchin D. A., Sitchikhina M. S., Ananievsky M. S., Svetsitskaya T. A., Fradkov A. L. Prediction of COVID-19 pandemic dynamics in Russia based on simple mathematical models of epidemics. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2021, no. 6, pp. 31–41 (In Russian). doi:10.31799/1684-8853-2021-6-31-41

### References

- Yang Z., Zeng Z., Wang K., Wong S.-S., Liang W., Zanin M., Liu P., Cao X., Gao Z., Mai Z., Liang J., Liu X., Li S., Li Y., Ye F., Guan W., Yang Y., Li F., Luo S., Xie Y., Liu B., Wang Z., Zhang S., Wang Y., Zhong N., He J. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *Journal of Thoracic Disease*, 2020, vol. 12, no. 3. doi:10.21037/jtd.2020.02.64
- Efimov D., Ushirobira R. On an interval prediction of COVID-19 development based on a SEIR epidemic model. *Annual Reviews in Control*, 2021, vol. 51, pp. 477–487. doi:10.1016/j.arcontrol.2021.01.006

3. Zhong L., Mu L., Li J., Wang J., Yin Z., Liu D. Early prediction of the 2019 novel coronavirus outbreak in the mainland China based on simple mathematical model. *IEEE Access*, 2020, vol. 8, pp. 51761–51769. doi:10.1109/ACCESS.2020.2979599
4. Teles P. A time-dependent SEIR model to analyse the evolution of the SARS-CoV-2 epidemic outbreak in Portugal. *arXiv:2004.04735*. Available at: <https://arxiv.org/abs/2004.04735> (accessed 2 August 2021).
5. Roda W. C., Varughese M. B., Han D., Li M. Y. Why is it difficult to accurately predict the COVID-19 epidemic? *Infectious Disease Modelling*, 2020, no. 5, pp. 271–281. doi:10.1016/j.idm.2020.03.001
6. Popkov Y., Dunnov Y., Popkov A. Forecasting development of COVID-19 epidemic in European Union using entropy-randomized approach. *Informatics and Automation*, 2021, vol. 20, no. 5, pp. 1010–1033 (In Russian). doi:10.15622/20.5.1
7. Zakharov V., Balykina Y. Balance model of COVID-19 epidemic based on percentage growth rate. *Informatics and Automation*, 2021, vol. 20, no. 5, pp. 1034–1064 (In Russian). doi:10.15622/20.5.2
8. Aronov I. Z., Maksimova O. V., Galkina N. M. COVID-19 highest incidence forecast in Russia based on regression model. *International Journal of Mathematical, Engineering and Management Sciences*, 2020, vol. 5, no. 5, pp. 812–819. doi:10.33889/IJMEMS.2020.5.5.063
9. Wang P., Zheng X., Ai G., Liu D., Zhu B. Time series prediction for the epidemic trends of COVID-19 using the improved LSTM deep learning method: Case studies in Russia, Peru and Iran. *Chaos, Solitons and Fractals*, 2020, vol. 140, no. 110214. doi:10.1016/j.chaos.2020.110214
10. Koltsova E. M., Kurkina E. S., Vasetsiy A. M. Mathematical modeling of the spread of COVID-19 coronavirus epidemic in a number of european, asian countries, Israel and Russia. *Problemy ekonomiki i yuridicheskoy praktiki*, 2020, vol. 16, no. 2, pp. 154–165 (In Russian).
11. Tomchin D. A., Fradkov A. L. Prediction of the COVID-19 spread in Russia based on SIR and SEIR models of epidemics. *OnLine*, 2020, vol. 53, no. 5, pp. 833–838. doi:10.1016/j.ifacol.2021.04.209
12. Kermack W. O., McKendrick A. G. A contribution to the mathematical theory of epidemics. *Proc. Roy. Soc. Lond., Ser. A*, 1927, no. 115, pp. 700–721.
13. Bratus A. S., Novozhilov A. S., Platonov A. P. *Dinamichesk- ie sistemy i modeli biologii* [Dynamical systems and models in biology]. Moscow, Fizmatlit Publ., 2010. 400 p.
14. Wang Z., Bauch C. T., Bhattacharyya S., d’Onofrio A., Manfredi P., Perc M., Perra N., Salathé M., Zhao D. Statistical physics of vaccination. *Physics Reports*, 2016, vol. 664, pp. 1–113. doi:10.1016/j.physrep.2016.10.006
15. *Worldometers*. Available at: <https://www.worldometers.info/coronavirus/country/russia/> (accessed 2 August 2021).
16. Li R., Pei S., Chen B., Song Y., Zhang T., Yang W., Shaman J. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science*, 2020, vol. 368, pp. 489–493. doi:10.1126/science.abb3221
17. Eubank S., Eckstrand L., Lewis D., Venkatramanan S., Marathe M., Barrett C. L. Commentary on Ferguson, et al. “Impact of Nonpharmaceutical Interventions (NPIs) to reduce COVID-19 mortality and healthcare demand”. *Bulletin of Mathematical Biology*, 2020, vol. 82, Article 52. doi:10.1007/s11538-020-00726-x
18. Danilenko D. M., Komissarov A. B., Stukova M. A., Lioznov D. A. To be or not to be: forecast of COVID-19 epidemic development in Russia. *Journal Infectology*, 2020, vol. 12, no. 3, pp. 6–11 (In Russian). doi:10.22625/2072-6732-2020-12-3-6-11
19. Al-qaness M. A. A., Saba A. I., Elsheikh A. H., Elaziz M. A., Ibrahim R. A., Lu S., Hemedan A. A., Shanmugan S., Ewees A. A. Efficient artificial intelligence forecasting models for COVID-19 outbreak in Russia and Brazil. *Process Safety and Environmental Protection*, 2021, vol. 149, pp. 399–409. doi:10.1016/j.psep.2020.11.007
20. Giordano G., Blanchini F., Bruno R., Colaneri P., Di Filippo A., Di Matteo A., Colaneri M. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nature Medicine*, 2020, vol. 26, pp. 855–860. doi:10.1038/s41591-020-0883-7
21. d’Onofrio A., Manfredi P., Poletti P. The interplay of public intervention and private choices in determining the outcome of vaccination programmes. *PLoS ONE*, 2012, vol. 7, iss. 10: e45653. <https://doi.org/10.1371/journal.pone.0045653>
22. Efimov D. V., Fradkov A. L. Oscillatority of nonlinear systems with static feedback. *SIAM J Control Optim.*, 2009, vol. 48, pp. 618–640. <https://doi.org/10.1137/070706963>