



# Ретроспективный анализ и прогнозирование распространения вирусов в реальном времени: на примере COVID-19 в Санкт-Петербурге и в Москве в 2020–2021 гг.

Захаров В.В., Балыкина Ю.Е.✉

ФГБОУ ВО «Санкт-Петербургский государственный университет», 199034, г. Санкт-Петербург, Россия

## Резюме

**Цель** исследования – применение математических методов для построения прогнозов динамики случайных значений процентного прироста общего количества заболевших и процентного прироста общего количества выздоровевших и умерших пациентов с проверкой методов при ретроспективном прогнозировании динамики эпидемического процесса COVID-19 в Санкт-Петербурге и в Москве.

**Материалы и методы.** При ретроспективном прогнозировании динамики общего количества заболевших COVID-19 и динамики общего количества пациентов, завершивших болезнь, использованы прогнозные значения процентных приростов этих показателей. Ретроспективный анализ и вычислительные эксперименты по прогнозированию динамики эпидемического процесса COVID-19 проводили на промежутках длиной 14 сут, начиная с 25 марта 2020 г. до 20 января 2021 г., с использованием метода прогнозирования временных рядов, предложенного авторами данной статьи.

**Результаты и обсуждение.** Представленные в работе ретроспективные 2-недельные прогнозы общего количества заболевших и количества активных случаев COVID-19 продемонстрировали достаточно высокую точность как в Москве, так и в Санкт-Петербурге. Ошибка MAPE (mean absolute percentage error) общего количества заболевших на пиках заболеваемости, как правило, не превышала 1%. Показано, что точность полученных ретроспективных прогнозов общего количества заболевших в Санкт-Петербурге, построенных начиная с мая 2020 г., значительно возросла по сравнению с апрельскими прогнозами. Аналогичное заключение можно сделать и относительно прогнозов общего количества заболевших в Москве в апреле и мае 2020 г.

**Ключевые слова:** SARS-CoV-2; COVID-19; эпидемический процесс; заболеваемость; активные случаи болезни; ретроспективный анализ; прогнозирование; моделирование

**Для цитирования:** Захаров В.В., Балыкина Ю.Е. Ретроспективный анализ и прогнозирование распространения вирусов в реальном времени: COVID-19 в Санкт-Петербурге и в Москве в 2020–2021 гг. *Вопросы вирусологии*. 2024; 69(6): 500–508. DOI: <https://doi.org/10.36233/0507-4088-265> EDN: <https://elibrary.ru/faiorq>

**Финансирование.** Исследование выполнено за счет гранта Российского научного фонда № 23-21-10049 (<https://rscf.ru/project/23-21-10049/>) и гранта Санкт-Петербургского научного фонда.

**Конфликт интересов.** Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

## ORIGINAL STUDY ARTICLE

DOI: <https://doi.org/10.36233/0507-4088-265>

# Retrospective analysis and forecasting of the spread of viruses in real time: the case of COVID-19 in St. Petersburg and Moscow in 2020–2021

Victor V. Zakharov, Yulia E. Balykina✉

St. Petersburg State University, 99034, St. Petersburg, Russia

## Abstract

The **aim** of the study is to apply mathematical methods to generate forecasts of the dynamics of random values of the percentage increase in the total number of infected people and the percentage increase in the total number of recovered and deceased patients. The obtained forecasts are used for retrospective forecasting of COVID-19 epidemic process dynamics in St. Petersburg and in Moscow.

**Materials and methods.** When conducting a retrospective analysis and forecasting the dynamics of the total number of cases and the dynamics of the total number of patients who have either died or recovered, the values of

percentage increases in these indicators were used. Retrospective analysis and forecasting of the dynamics of the COVID-19 epidemic process were carried out over 14-day time intervals, starting from March 25, 2020 to January 20, 2021, using the time series forecasting method proposed by the authors.

**Results and discussion.** The retrospective two-week forecasts of the total number of cases and the number of active cases presented in the paper demonstrated a high accuracy performance, both in Moscow and St. Petersburg. The MAPE (mean absolute percentage error) for the total number of cases at the peaks of incidence, generally, did not exceed 1%. It is shown that the accuracy of the obtained retrospective forecasts of the total number of cases in St. Petersburg, built starting from May 2020, has increased significantly compared to the April forecasts. A similar conclusion can be made regarding the forecasts of the total number of cases in Moscow in April and May 2020.

**Keywords:** SARS-CoV-2; COVID-19; epidemic process; incidence; active cases of the disease; retrospective analysis; forecasting

**For citation:** Zakharov V.V., Balykina Yu.E. Retrospective analysis and forecasting of the spread of viruses in real time: COVID-19 in St. Petersburg and Moscow in 2020–2021. *Problems of Virology (Voprosy Virusologii)*. 2024; 69(6): 500–508. DOI: <https://doi.org/10.36233/0507-4088-265> EDN: <https://elibrary.ru/faiopq>

**Funding.** The study was supported by the Russian Science Foundation grant No. 23-21-10049 (<https://rscf.ru/project/23-21-10049/>) and the St. Petersburg Science Foundation grant.

**Conflict of interest.** The authors declare no apparent or potential conflicts of interest related to the publication of this article.

## Введение

Эпидемия новой коронавирусной инфекции (COVID-19) в Российской Федерации, которая началась в марте 2020 г., вовлекла в эпидемический процесс население всех регионов страны. К концу декабря 2020 г. в стране было зарегистрировано более 3 млн случаев заболевания COVID-19. Такие темпы распространения эпидемии привели к значительному увеличению нагрузки на систему здравоохранения. Вместе с этим во весь рост встала и проблема отсутствия на тот период эффективных методов прогнозирования динамики распространения нового вируса, параметры которого еще не были изучены. По признанию Всемирной организации здравоохранения (ВОЗ), наличие адекватных математических моделей распространения вирусов и разработка методов прогнозирования динамики эпидемий может сыграть ключевую роль при подготовке принятия решений лицами, определяющими ресурсное обеспечение и политику здравоохранения. Такие модели помогают получить представление о будущей динамике распространения инфекционных заболеваний и потенциальной роли разных видов стратегий вмешательства со стороны общественного здравоохранения [1, 2]. Для описания процессов распространения инфекционных заболеваний многие исследователи используют трехкамерную модель SIR. Население страны, в которой распространяется инфекция, разделяется на 3 группы (камеры): восприимчивые к вирусу (Susceptible), инфицированные (Infected) и выздоровевшие или умершие (Removed). Предполагается, что численность населения страны равна сумме численностей этих 3 групп. Впервые модель была в общем виде описана в работе W. Kermack и A. McKendrick в 1927 г. [3], однако наибольший интерес исследователей был в дальнейшем сосредоточен на частном случае этой модели в виде системы из 3 дифференциальных уравнений с постоянными коэффициентами [4]. Некоторые специалисты предпочитают модель SIR

из-за небольшого количества требуемых на входе параметров и широких возможностей для применения традиционных классических подходов к моделированию и использованию вычислительных алгоритмов. Однако это преимущество одновременно становится и недостатком, поскольку упрощение модели порождает неточность получаемых на их основе прогнозов по сравнению с фактическими значениями основных переменных эпидемического процесса [5–7]. В ряде статей [8–11] приводятся примеры применения авторегрессионной модели ARIMA при построении прогнозов распространения пандемии в ряде стран мира. Отмечается, что эта модель по сравнению с другими инструментами дает более точные прогнозы, однако, хотя модели временных рядов и являются популярным инструментом прогнозирования, их применение для оценки распространения новых инфекций не всегда позволяет построить прогнозы высокой степени точности.

Помимо различных разновидностей моделей семейства ARIMA, для прогнозирования динамики COVID-19 могут также использоваться разные методы машинного обучения, включая деревья решений и ансамблевые модели, такие как случайный лес и градиентный бустинг [12, 13]. Эти техники позволяют учитывать множество факторов, включая уровень заражения, мобильность населения и применяемые меры ограничения социальной дистанции. Искусственные нейронные сети также находят широкое применение, благодаря своей способности обрабатывать временные ряды и делать прогнозы на основе последовательных данных [14–16].

Изучению закономерностей эпидемического распространения SARS-CoV-2 на начальном этапе пандемии COVID-19 в Российской Федерации, в Москве и Санкт-Петербурге в 2020 г. посвящены несколько статей [17–20], в которых выделен ряд важных периодов в динамике выявления новых случаев COVID-19: рост заболеваемости, период эпидемического затишья, период осеннего подъема,

период устойчиво высокого уровня заболеваемости COVID-19. Следует отметить, что важной мотивацией для проведения настоящего исследования явились в том числе публикации о вкладе академика В.Д. Белякова [21, 22] в развитие представлений о единой («общемедицинской» в его определении) эпидемиологии, включающей неинфекционные заболевания. В более ранних статьях [23–25] авторами настоящей статьи был предложен подход, основная идея которого заключается в прогнозировании сначала будущих прогнозных трендов стохастических параметров эпидемического процесса, а затем в вычислении будущих траекторий основных переменных этого процесса.

**Цель** настоящей работы заключается в построении прогнозов динамики временных рядов случайных значений процентного прироста общего количества заболевших и процентного прироста общего количества выздоровевших и умерших пациентов и описании результатов применения разработанных авторами методов для ретроспективного прогнозирования динамики эпидемических процессов в Санкт-Петербурге и в Москве.

### Материалы и методы

Работа выполнена научным коллективом Центра аналитики динамических процессов и систем Санкт-Петербургского государственного университета. При проведении вычислительных экспериментов по прогнозированию динамики эпидемического процесса COVID-19 в Санкт-Петербурге в период с марта 2020 г. по июнь 2020 г. и в Москве с сентября 2020 г. по январь 2021 г. использовали временные ряды данных официальной регистрации ежедневного количества новых случаев заболевания и количества случаев завершения болезни (выздоровления или смерти пациентов), представленные на сайте стопкоронавирус.рф и статистические данные Центра CSSE Университета Джонса Хопкинса<sup>1</sup>.

С точки зрения теории математической статистики, данные о новых случаях заболевания образуют временные ряды значений  $x_{inf}(t)$  (количество новых случаев в день  $t$ ), а данные о количестве случаев завершения болезни (выздоровления или смерти пациентов) образуют временные ряды  $x_{of}(t)$ , где  $t$  – это конкретный день эпидемии. Обозначим через  $X_{inf}(t)$  общее количество зарегистрированных случаев заболевания с начала эпидемии до дня  $t$  включительно, а через  $X_{of}(t)$  – общее количество пациентов, завершивших болезнь к этому дню. Полученные временные ряды  $X_{inf}(t)$  и  $X_{of}(t)$  называются интегральной формой временных рядов  $x_{inf}(t)$  и  $x_{of}(t)$  соответственно. Выраженность тенденции динамики новых случаев заболевания (рост/снижение) принято описывать по методике, предложенной В.Д. Беляковым [26].

Следует отметить, что поскольку члены временного ряда  $x_{inf}(t)$  принимают случайные значения, которые могут то возрастать, то убывать, то процентный прирост количества новых случаев, так же как и тенденция роста/снижения, может принимать как отрицательные значения, так и положительные (или равные 0). Процентный прирост числа новых случаев в день  $t$  по отношению к числу новых случаев в день  $t - 1$  мы обозначаем  $r(t)$  и вычисляем по формуле:

$$r(t) = \frac{x_{inf}(t) - x_{inf}(t - 1)}{x_{inf}(t - 1)} \times 100\%.$$

Тенденция роста/снижения динамики количества новых случаев заболевания описывается как среднее значение процентного прироста на заданных промежутках времени и используется в качестве важной характеристики при осуществлении мониторинга эпидемического процесса [17]. Однако ее использование в процессе прогнозирования будущей динамики эпидемического процесса в реальном времени представляется проблематичным, в основном ввиду того, что ее значения случайным образом меняют знак с плюса на минус. Можно заметить, что временные ряды  $X_{inf}(t)$  и  $X_{of}(t)$ , в отличие от временных рядов  $x_{inf}(t)$  и  $x_{of}(t)$ , являются неотрицательными и возрастают с течением времени. В настоящем исследовании в качестве основных параметров динамики эпидемического процесса рассматривался процентный прирост общего количества заболевших  $r_{inf}(t)$  и процентный прирост  $r_{of}(t)$  общего количества пациентов, для которых болезнь завершилась [23–25]. Значения этих параметров для любого дня  $t$  эпидемии можно вычислить по следующим формулам:

$$r_{inf}(t) = \frac{X_{inf}(t) - X_{inf}(t - 1)}{X_{inf}(t - 1)} \times 100\% = \frac{x_{inf}(t)}{X_{inf}(t - 1)} \times 100\%$$

$$r_{of}(t) = \frac{X_{of}(t) - X_{of}(t - 1)}{X_{of}(t - 1)} \times 100\% = \frac{x_{of}(t)}{X_{of}(t - 1)} \times 100\%$$

Учитывая, что в числителе и знаменателе представленных формул стоят члены временных рядов, указанные процентные приросты тоже образуют временные ряды случайных величин. В ходе эпидемии общее количество случаев заболевания и общее количество завершивших болезнь пациентов, как правило, строго возрастают, поэтому значения этих параметров можно считать положительными для любого дня, за исключением дня, в котором число новых случаев равно 0. Кроме того, если процентный прирост  $r_{inf}(t)$  больше процентного прироста  $r_{inf}(t - 1)$ , то имеет место неравенство:

$$\frac{x_{inf}(t)}{X_{inf}(t - 1)} > \frac{x_{inf}(t - 1)}{X_{inf}(t - 2)}$$

или, что то же самое:

$$\frac{x_{inf}(t)}{x_{inf}(t - 1)} > \frac{X_{inf}(t - 1)}{X_{inf}(t - 2)} \geq 1.$$

Из последнего неравенства можно сделать вывод, что при возрастании процентного прироста общего количества заболевших количество новых случаев заболевания тоже, как правило, возрастает. Однако утверждение, что при убывании процентного прироста

<sup>1</sup>COVID-19 Data Visualization Center [Internet]. Johns Hopkins Coronavirus Resource Center. Доступно по: <https://coronavirus.jhu.edu/data>

ста общего количества заболевших количество новых случаев убывает, справедливо не всегда.

То же верно и для процентного прироста общего количества выздоровевших и умерших пациентов. Еще одна тенденция, которую можно обнаружить при анализе статистических данных заболеваемости в Санкт-Петербурге и в Москве в мае 2020 г., состоит в том, что снижение процентного прироста общего количества заболевших началось примерно за неделю до наступления пика по числу новых случаев. Эта же ситуация повторилась и в декабре 2020 г. Поэтому мы полагаем, что мониторинг динамики изменения процентного прироста общего количества заболевших имеет существенное значение при прогнозировании динамики как общего количества заболевших, так и новых случаев заболевания.

В рамках настоящего исследования для статистических данных 2020 г. в Санкт-Петербурге и в Москве были построены 20 последовательных 2-недельных ретроспективных прогнозов процентного прироста общего количества заболевших и процентного прироста общего количества пациентов, для которых болезнь завершилась.

Если обратиться к **таблице**, то можно заметить, что процентный прирост общего количества заболевших в Санкт-Петербурге после 22 апреля 2020 г. имеет явную тенденцию к убыванию. Та же ситуация наблюдалась и в Москве. Аналогичную тенденцию, хотя и менее выраженного характера и с запаздыванием на 2–3 нед, имеет процентный прирост общего количества пациентов, для которых болезнь завершилась.

Наш метод построения прогноза в конкретный заданный день предполагает на 1-м этапе выбор подходящей функции (линейной, степенной или экспоненциальной) для аппроксимации значений процентных приростов  $r_{inf}(t)$  и  $r_{of}(t)$  за предшествующие 14 сут до дня построения прогноза. После завершения 1-го этапа с помощью выбранной аппроксимирующей функции вычисляются прогнозные значения процентных приростов  $\tilde{r}_{inf}(t)$  и  $\tilde{r}_{of}(t)$  в течение следующих после дня построения прогноза 14 сут. Полученные значения процентных приростов использовались для вычисления в течение горизонта прогнозирования прогнозных значений  $\tilde{X}_{inf}(t)$  и  $\tilde{X}_{of}(t)$  по формулам:

$$\tilde{X}_{inf}(t) = \tilde{X}_{inf}(t-1) \times \left(1 + \frac{\tilde{r}_{inf}(t)}{100}\right)$$

$$\tilde{X}_{of}(t) = \tilde{X}_{of}(t-1) \times \left(1 + \frac{\tilde{r}_{of}(t)}{100}\right).$$

Для каждого дня  $t$  промежутка прогнозирования можно вычислить отклонение  $e(t)$  прогнозного значения от фактического, например:

$$e(t) = \tilde{X}_{inf}(t) - X_{inf}(t),$$

и абсолютную ошибку прогноза:

$$\Delta(t) = \left(\frac{|X_{inf}(t) - \tilde{X}_{inf}(t)|}{X_{inf}(t)}\right) \times 100\%.$$

В качестве средней ошибки прогноза для каждого промежутка прогнозирования выбрали часто используемую в статистике ошибку MAPE (mean absolute

percentage error). Значения MAPE вычисляли для промежутка прогнозирования длиной 14 сут по формуле:

$$MAPE = \frac{1}{14} \sum_{t=1}^{14} \Delta(t).$$

## Результаты

Результаты проведенных вычислительных экспериментов представлены в **табл. S1–S8**, помещенных в **Приложение**. В табл. S1–S3 приведены прогнозы динамики процентных приростов общего количества заболевших в Санкт-Петербурге с начала эпидемии, построенные 8 и 22 апреля, 6 и 20 мая, 3 и 17 июня 2020 г.

Средняя ошибка MAPE для прогноза общего количества заболевших, построенного 8 апреля 2020 г., равна 5,15%, 22 апреля – 19,66%. Таким образом, точность прогнозов общего количества заболевших в апреле оказалась не очень высокой. Средняя ошибка MAPE для прогноза, построенного 6 мая 2020 г., равна 4,52% (**рис. 1**). Следует отметить, что июньские прогнозы, представленные в табл. S3, имеют уже более высокую точность, чем майские. Ошибка MAPE для прогноза, построенного 3 июня 2020 г., равна 3,17%, а для прогноза 17 июня – 0,77%.

Данные о некоторых прогнозах в осенне-зимний период в Москве приведены в табл. S4 и S5. Ошибка MAPE для прогноза в Москве, построенного 25 ноября 2020 г., равна 0,13%, построенного 9 декабря – 0,13%. Ошибка MAPE для прогноза, построенного 23 декабря 2020 г., равна 0,17%, а построенного 6 января 2021 г. – 0,30% (**рис. 2**).

Как известно, количество активных случаев болезни в любой день можно вычислить, вычитая из общего числа заболевших общее число выздоровевших и умерших пациентов. Прогнозируемое количество активных случаев болезни вычисляли как разность между спрогнозированным общим числом заболевших и спрогнозированным общим числом пациентов, завершивших болезнь. На **рис. 3** и в табл. S6 представлены прогнозные и фактические значения количества активных случаев болезни в Санкт-Петербурге в период прохождения пика эпидемии в мае 2020 г. В табл. S7 содержатся данные о прогнозах количества новых случаев заболевания и количества активных случаев в Москве с 24 декабря 2020 г. по 6 января 2021 г. в период прохождения пиков эпидемии (**рис. 4, 5**). Жирным шрифтом в таблицах выделены максимальные значения фактических и прогнозируемых величин активных случаев болезни в окрестностях пиков эпидемии первой и второй волны.

Для того чтобы оценить возможность предсказания резкого роста заболеваемости при распространении штамма Delta в Москве в июне 2021 г., мы обратились к статистическим данным о динамике COVID-19 за этот период. Резкий рост заболеваемости наблюдался с 8 по 15 июня. Мониторинг динамики процентного прироста общего количества заболевших показал, что изменение тенденции устойчивого убывания

**Таблица.** Общее количество заболевших COVID-19 в Санкт-Петербурге в апреле–июне 2020 г.

**Table.** Total number of registered COVID-19 cases in St. Petersburg in April–June 2020

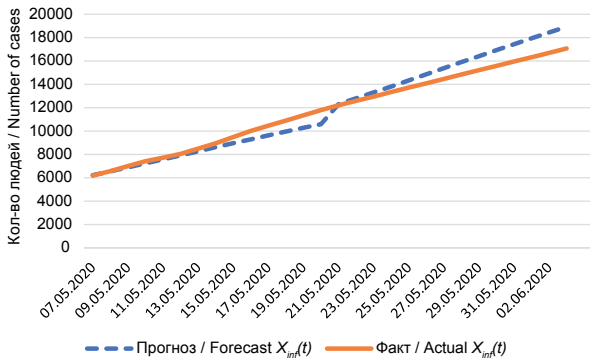
Дата Date	Общее число заболевших, абс. Total number of cases, abs.	$r_{ин}(t)$ , %	Дата Date	Общее число заболевших, абс. Total number of cases, abs.	$r_{ин}(t)$ , %
12.04.2020	557	14,14	21.05.2020	12 203	3,46
13.04.2020	678	21,72	22.05.2020	12 592	3,19
14.04.2020	799	17,85	23.05.2020	12 955	2,88
15.04.2020	929	16,27	24.05.2020	13 339	2,96
16.04.2020	1083	16,58	25.05.2020	13 713	2,80
17.04.2020	1507	39,15	26.05.2020	14 076	2,65
18.04.2020	1646	9,22	27.05.2020	14 463	2,75
19.04.2020	1760	6,93	28.05.2020	14 846	2,65
20.04.2020	1846	4,89	29.05.2020	15 215	2,49
21.04.2020	1973	6,88	30.05.2020	15 580	2,40
22.04.2020	2267	14,90	31.05.2020	15 949	2,37
23.04.2020	2458	8,43	01.06.2020	16 313	2,28
24.04.2020	2711	10,29	02.06.2020	16 689	2,30
25.04.2020	2926	7,93	03.06.2020	17 069	2,28
26.04.2020	3077	5,16	04.06.2020	17 444	2,20
27.04.2020	3238	5,23	05.06.2020	17 822	2,17
28.04.2020	3436	6,11	06.06.2020	18 169	1,95
29.04.2020	3726	8,44	07.06.2020	18 509	1,87
30.04.2020	4062	9,02	08.06.2020	18 835	1,76
01.05.2020	4411	8,59	09.06.2020	19 153	1,69
02.05.2020	4734	7,32	10.06.2020	19 466	1,63
03.05.2020	5029	6,23	11.06.2020	19 769	1,56
04.05.2020	5346	6,30	12.06.2020	20 043	1,39
05.05.2020	5572	4,23	13.06.2020	20 305	1,31
06.05.2020	5884	5,60	14.06.2020	20 561	1,26
07.05.2020	6190	5,20	15.06.2020	20 813	1,23
08.05.2020	6565	6,06	16.06.2020	21 047	1,12
09.05.2020	6990	6,47	17.06.2020	21 275	1,08
10.05.2020	7404	5,92	18.06.2020	21 506	1,09
11.05.2020	7711	4,15	19.06.2020	21 734	1,06
12.05.2020	8050	4,40	20.06.2020	21 966	1,07
13.05.2020	8485	5,40	21.06.2020	22 195	1,04
14.05.2020	8945	5,42	22.06.2020	22 412	0,98
15.05.2020	9486	6,05	23.06.2020	22 632	0,98
16.05.2020	10 011	5,53	24.06.2020	22 850	0,96
17.05.2020	10 462	4,51	25.06.2020	23 071	0,97
18.05.2020	10 887	4,06	26.06.2020	23 294	0,97
19.05.2020	11 340	4,16	27.06.2020	23 518	0,96
20.05.2020	11 795	4,01	28.06.2020	23 735	0,92

процентного прироста наблюдалось до 2 июня, затем его значение остановилось на 3 сут на отметке 0,24%, после чего процентный прирост начал возрастать. К 8 июня значение процентного прироста уже достигло уровня 0,32%. Такой быстрый подъем процентного прироста в течение 4 сут, по-нашему мнению, может свидетельствовать о возможном дальнейшем экспоненциальном росте процентного прироста и числа новых случаев заболевания. Ретроспективный прогноз,

построенный 8 июня на 6 сут с использованием предлагаемых нами методов, представлен в табл. S8 и на **рис. 6**. Он имеет ошибку MARE для общего количества заболевших 0,06%, для новых случаев – 10,41%.

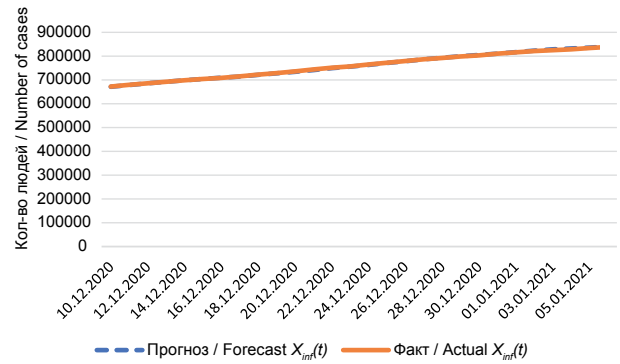
#### Обсуждение

Как показали вычислительные эксперименты, в апреле 2020 г. динамика процентного прироста общего количества заболевших в Санкт-Петербур-



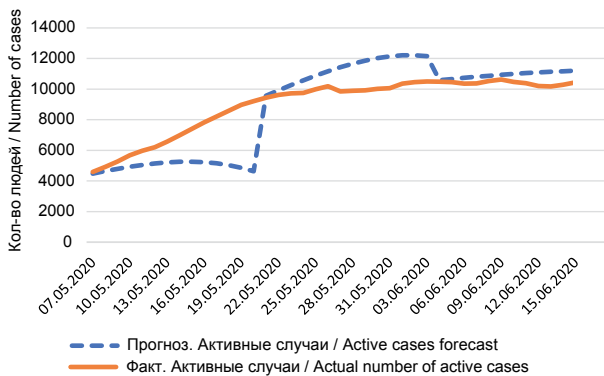
**Рис. 1.** Двухнедельные прогнозы общего количества заболевших в Санкт-Петербурге в мае 2020 г.

**Fig. 1.** Two-week forecasts of the total number of cases in St. Petersburg in May 2020.



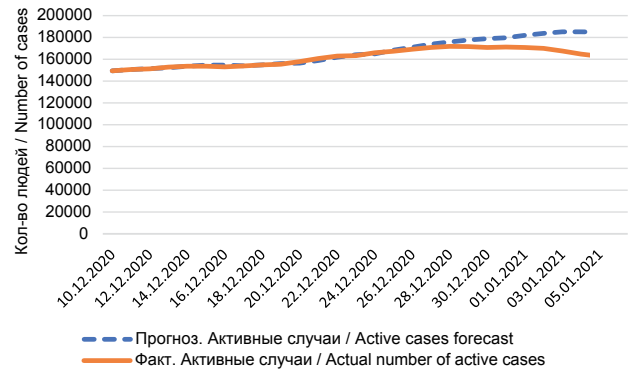
**Рис. 2.** Двухнедельные прогнозы общего количества заболевших в Москве в декабре 2020 г.

**Fig. 2.** Two-week forecasts of the total number of cases in Moscow in December 2020.



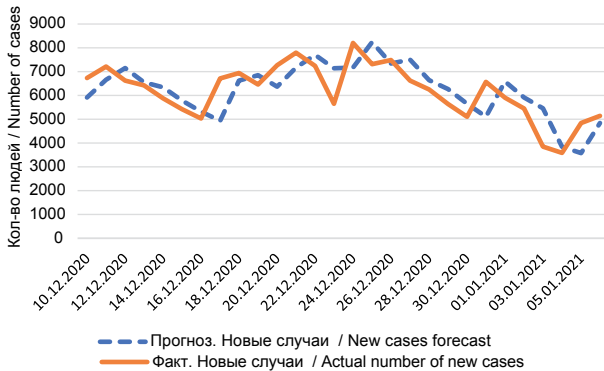
**Рис. 3.** Прогноз количества активных случаев болезни в Санкт-Петербурге в мае-июне 2020 г.

**Fig. 3.** Forecast of the number of active cases in St. Petersburg in May-June 2020.



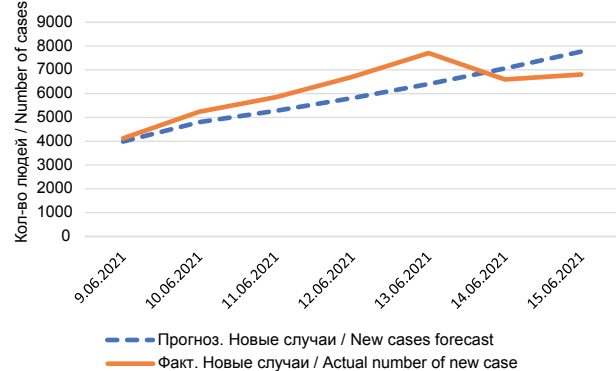
**Рис. 4.** Прогноз количества активных случаев болезни в Москве в декабре 2020 г.

**Fig. 4.** Forecast of the number of active cases in Moscow in December 2020.



**Рис. 5.** Прогноз новых случаев болезни в Москве в декабре 2020 г.

**Fig. 5.** Forecast of new cases in Moscow in December 2020.



**Рис. 6.** Прогноз новых случаев при распространении штамма дельта в Москве в июне 2021 г.

**Fig. 6.** Forecast of new cases in during the spread of delta strain in Moscow in June 2021

ге была трудно предсказуема, однако построенные в мае-июне прогнозы оказались уже достаточно точны. Точность полученных ретроспективных прогнозов общего количества заболевших, построенных в мае, значительно возросла по сравнению с апрелем (табл. S2 и рис. 1). Это, по-нашему мнению, объясняется тем, что выбранный метод прогнозирования прошел в апреле стадию обучения на текущих данных и после этого оказался адаптированным к ис-

пользованию для новых данных, полученных в мае. Из этого можно сделать вывод, что неопределенность в оценке будущей динамики развития эпидемического процесса к началу мая значительно уменьшилась. Эффективность предложенного метода прогнозирования подтвердилась и при построении ретроспективных 2-недельных прогнозов в Москве (табл. S4 и S5, рис. 2). Следует отметить, что в периоды начала распространения более вирулентных

штаммов вируса и при введении жестких мер противодействия распространению эпидемии, точность работы предложенного в статье метода может быть ограничена. Такие ситуации требуют дополнительного изучения. В рамках настоящего исследования проводили вычислительные эксперименты по построению 2-недельных ретроспективных прогнозов не только динамики общего количества заболевших, но также были построены прогнозы динамики числа ежедневных новых случаев заболевания, общего числа пациентов, завершивших болезнь, числа активных случаев болезни (текущего числа болеющих людей). Результаты прогнозирования продемонстрировали достаточно высокую точность полученных прогнозов. В качестве примеров в табл. S6 и S7 также приведены данные о прогнозах и фактических значениях количества активных случаев болезни в период прохождения пиков эпидемии в мае 2020 г. в Санкт-Петербурге (рис. 3) и в декабре 2020 г. в Москве (рис. 4). На этих рисунках видно, что спрогнозированные значения количества активных случаев при приближении к пику эпидемии отличаются незначительно от фактических значений. Ошибка MAPE прогнозирования активных случаев болезни в Москве на горизонте прогнозирования 25 декабря 2020 г. – 6 января 2021 г. равна 6,16%. В начальный период подъема заболеваемости фактический максимальный уровень (10 619) болеющих в Санкт-Петербурге был зафиксирован 6 июня, а по прогнозу он составил 12 210 активных случаев болезни и был спрогнозирован на 2 июня. Максимальное число новых случаев в период второго подъема заболеваемости в Москве было спрогнозировано на 25 декабря 2020 г. и составило 8237, а фактическое его значение (8203) было зафиксировано 24 декабря (рис. 5). Ошибка MAPE прогнозирования новых случаев в Москве на горизонте прогнозирования 25 декабря 2020 г. – 6 января 2021 г. равна 15,12%.

### Заключение

Анализ динамики ежедневного количества новых случаев заболевания COVID-19 показывает, что это изменение в ходе всей эпидемии имеет стохастический характер, является трудно предсказуемым и, как следствие, создает значительную неопределенность в оценке будущих значений этого показателя и эволюции эпидемического процесса. Рассмотренный в работе метод прогнозирования будущих значений процентных приростов общего количества заболевших и общего количества пациентов, завершивших болезнь, был проверен на статистических данных портала стопкоронавирус.рф в период с марта 2020 г. по июнь 2021 г. Построенные в ходе вычислительных экспериментов и описанные в работе ретроспективные 2-недельные прогнозы общего количества заболевших в течение двух пиковых периодов эпидемии, когда нагрузка на систему здравоохранения была максимальной, продемонстрировали достаточно высокую точность. Точность полученных ретроспективных прогнозов общего количества заболевших, построенных в мае, значительно возросла по сравнению

с точностью апрельских прогнозов. Это, по нашему мнению, свидетельствует о том, что выбранный метод прогнозирования динамики процентного прироста общего количества заболевших прошел в апреле стадию обучения на текущих данных и после этого оказался адаптированным к использованию для новых данных, полученных, начиная с мая 2020 г. Применение предлагаемого в работе метода прогнозирования также оказалось достаточно эффективным и при прогнозировании динамики активных случаев болезни и новых случаев в окрестности пиков эпидемии первой и второй волны.

### ЛИТЕРАТУРА

1. Foppa I.M. *A Historical Introduction to Mathematical Modeling of Infectious Diseases: Seminal Papers in Epidemiology*. London: Academic Press; 2016.
2. Shinde G.R., Kalamkar A.B., Mahalle P.N., Dey N., Chaki J., Hassanien A.E. Forecasting models for coronavirus disease (COVID-19): A survey of the state-of-the-art. *SN Comput. Sci.* 2020; 1(4): 197. <https://doi.org/10.1007/s42979-020-00209-9>
3. Kermack W.O., McKendrick A.G. A contribution to the mathematical theory of epidemics. *Proc. R. Soc. (London) A.* 1927; 115(772): 700–21. <https://doi.org/10.1098/rspa.1927.0118>
4. Anderson R.M., May R.M. *Infectious Diseases of Humans. Dynamics and Control*. Oxford: Oxford University Press; 1991.
5. Moein S., Nickaeen N., Roointan A., Borhani N., Heidary Z., Javanmard S.H., et al. Inefficiency of SIR models in forecasting COVID-19 epidemic: a case study of Isfahan. *Sci. Rep.* 2021; 11(1): 4725. <https://doi.org/10.1038/s41598-021-84055-6>
6. Melikechi O., Young A.L., Tang T., Bowman T., Dunson D., Johndrow J. Limits of epidemic prediction using SIR models. *J. Math Biol.* 2022; 85(4): 36. <https://doi.org/10.1007/s00285-022-01804-5>
7. 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; 12(6): e8582. <https://doi.org/10.7759/cureus.8582>
8. 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 J. Public Health.* 2020; 49(Suppl. 1): 92–100. <https://doi.org/10.18502/ijph.v49iS1.3675>
9. Ahmar A.S., Del Val E.B. SutteARIMA: Short-term forecasting method, a case: Covid-19 and stock market in Spain. *Sci. Total Environ.* 2020; 729: 138883. <https://doi.org/10.1016/j.scitotenv.2020.138883>
10. Chaudhry R.M., Hanif A., Chaudhary M., Minhas S. 2<sup>nd</sup>, Mirza K., Ashraf T., et al. Coronavirus disease 2019 (COVID-19): Forecast of an emerging urgency in Pakistan. *Cureus.* 2020; 12(5): e8346. <https://doi.org/10.7759/cureus.8346>
11. Tandon H., Ranjan P., Chakraborty T., Suhag V. Coronavirus (COVID-19): ARIMA-based time-series analysis to forecast near future and the effect of school reopening in India. *J. Health Manag.* 2022; 24(3): 373–88. <https://doi.org/10.1177/09720634221109087>
12. Özen F. Random forest regression for prediction of COVID-19 daily cases and deaths in Turkey. *Heliyon.* 2024; 10(4): e25746. <https://doi.org/10.1016/j.heliyon.2024.e25746>
13. Galasso J., Cao D.M., Hochberg R. A random forest model for forecasting regional COVID-19 cases utilizing reproduction number estimates and demographic data. *Chaos Solitons Fractals.* 2022; 156: 111779. <https://doi.org/10.1016/j.chaos.2021.111779>
14. Wiczorek M., Siłka J., Woźniak M. Neural network powered COVID-19 spread forecasting model. *Chaos Solitons Fractals.* 2020; 140: 110203. <https://doi.org/10.1016/j.chaos.2020.110203>
15. Dadyan E., Avetisyan P. Neural networks and forecasting COVID-19. *Opt. Mem. Neural Networks.* 2021; 30: 225–35. <https://doi.org/10.3103/S1060992X21030085>
16. Tamang S., Singh P., Datta B. Forecasting of COVID-19 cases based on prediction using artificial neural network curve fitting

- technique. *Glob. J. Environ. Sci. Manag.* 2020; 6(S): 53–64. <https://doi.org/10.22034/GJESM.2019.06.SI.06>
17. Акимкин В.Г., Кузин С.Н., Семенов Т.А., Шипулина О.Ю., Яцышина С.Б., Тиванова Е.В. и др. Закономерности эпидемиологического распространения SARS-CoV-2 в условиях мегаполиса. *Вопросы вирусологии.* 2020; 65(4): 203–11. <https://doi.org/10.36233/0507-4088-2020-65-4-203-211> <https://elibrary.ru/fxkaqf>
  18. Акимкин В.Г., Кузин С.Н., Семенов Т.А., Плоскирева А.А., Дубоделов Д.В., Тиванова Е.В. и др. Характеристика эпидемиологической ситуации по COVID-19 в Российской Федерации в 2020 г. *Вестник Российской академии медицинских наук.* 2021; 76(4): 412–22. <https://doi.org/10.15690/vramn1505> <https://elibrary.ru/zmowbe>
  19. Акимкин В.Г., Попова А.Ю., Плоскирева А.А., Углева С.В., Семенов Т.А., Пшеничная Н.Ю. и др. COVID-19: эволюция пандемии в России. Сообщение I: проявления эпидемиологического процесса COVID-19. *Журнал микробиологии, эпидемиологии и иммунологии.* 2022; 99(3): 269–86. <https://doi.org/10.36233/0372-9311-276> <https://elibrary.ru/zxgtfd>
  20. Акимкин В.Г., Кузин С.Н., Колосовская Е.Н., Кудрявцева Е.Н., Семенов Т.А., Плоскирева А.А. и др. Характеристика эпидемиологической ситуации по COVID-19 в Санкт-Петербурге. *Журнал микробиологии, эпидемиологии и иммунологии.* 2021; 98(5): 497–511. <https://doi.org/10.36233/0372-9311-154> <https://elibrary.ru/dtmnhz>
  21. Белов А.Б. Академик В.Д. Беляков – основоположник отечественной теории эпидемиологической науки XXI века. *Эпидемиология и вакцинопрофилактика.* 2016; 15(6): 9–15. <https://elibrary.ru/xemevf>
  22. Саркисов А.С. Академик В.Д. Беляков и его вклад в развитие эпидемиологии. *Бюллетень национального научно-исследовательского института общественного здоровья имени Н.А. Семашко.* 2020; (4): 68–72. <https://doi.org/10.25742/NRIPH.2020.04.010> <https://elibrary.ru/qeazcn>
  23. Захаров В.В., Балькина Ю.Е. Балансовая модель эпидемии COVID-19 на основе процентного прироста. *Информатика и автоматизация.* 2021; 20(5): 1034–64. <https://doi.org/10.15622/20.5.2> <https://elibrary.ru/zczxuw>
  24. Zakharov V., Balykina Y., Ilin I., Tick A. Forecasting a new type of virus spread: a case study of COVID-19 with stochastic parameters. *Mathematics.* 2022; 10(20): 3725. <https://doi.org/10.3390/math10203725>
  25. Балькина Ю.Е., Захаров В.В. Интегральная модель притока и оттока и ее приложения. *Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления.* 2024; 20(2): 121–35. <https://doi.org/10.21638/spbu10.2024.201>
  26. Беляков В.Д., Дегтярев А.А., Иванников Ю.Г. *Качество и эффективность противоэпидемических мероприятий.* Ленинград: Медицина; 1981. <https://elibrary.ru/zferwn>
  27. Foppa I.M. *A Historical Introduction to Mathematical Modeling of Infectious Diseases: Seminal Papers in Epidemiology.* London: Academic Press; 2016.
  28. Shinde G.R., Kalamkar A.B., Mahalle P.N., Dey N., Chaki J., Hassanien A.E. Forecasting models for coronavirus disease (COVID-19): A survey of the state-of-the-art. *SN Comput. Sci.* 2020; 1(4): 197. <https://doi.org/10.1007/s42979-020-00209-9>
  29. Kermack W.O., McKendrick A.G. A contribution to the mathematical theory of epidemics. *Proc. R. Soc. (London) A.* 1927; 115(772): 700–21. <https://doi.org/10.1098/rspa.1927.0118>
  30. Anderson R.M., May R.M. *Infectious Diseases of Humans. Dynamics and Control.* Oxford: Oxford University Press; 1991.
  31. Moein S., Nickaeen N., Roointan A., Borhani N., Heidary Z., Javanmard S.H., et al. Inefficiency of SIR models in forecasting COVID-19 epidemic: a case study of Isfahan. *Sci. Rep.* 2021; 11(1): 4725. <https://doi.org/10.1038/s41598-021-84055-6>
  32. Melikechi O., Young A.L., Tang T., Bowman T., Dunson D., Johndrow J. Limits of epidemic prediction using SIR models. *J. Math Biol.* 2022; 85(4): 36. <https://doi.org/10.1007/s00285-022-01804-5>
  33. 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; 12(6): e8582. <https://doi.org/10.7759/cureus.8582>
  34. 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 J. Public Health.* 2020; 49(Suppl. 1): 92–100. <https://doi.org/10.18502/ijph.v49iS1.3675>
  35. Ahmar A.S., Del Val E.B. SutteARIMA: Short-term forecasting method, a case: Covid-19 and stock market in Spain. *Sci. Total Environ.* 2020; 729: 138883. <https://doi.org/10.1016/j.scitotenv.2020.138883>
  36. Chaudhry R.M., Hanif A., Chaudhary M., Minhas S. 2<sup>nd</sup>, Mirza K., Ashraf T., et al. Coronavirus disease 2019 (COVID-19): Forecast of an emerging urgency in Pakistan. *Cureus.* 2020; 12(5): e8346. <https://doi.org/10.7759/cureus.8346>
  37. Tandon H., Ranjan P., Chakraborty T., Suhag V. Coronavirus (COVID-19): ARIMA-based time-series analysis to forecast near future and the effect of school reopening in India. *J. Health Manag.* 2022; 24(3): 373–88. <https://doi.org/10.1177/09720634221109087>
  38. Özen F. Random forest regression for prediction of COVID-19 daily cases and deaths in Turkey. *Heliyon.* 2024; 10(4): e25746. <https://doi.org/10.1016/j.heliyon.2024.e25746>
  39. Galasso J., Cao D.M., Hochberg R. A random forest model for forecasting regional COVID-19 cases utilizing reproduction number estimates and demographic data. *Chaos Solitons Fractals.* 2022; 156: 111779. <https://doi.org/10.1016/j.chaos.2021.111779>
  40. Wiczorek M., Silka J., Woźniak M. Neural network powered COVID-19 spread forecasting model. *Chaos Solitons Fractals.* 2020; 140: 110203. <https://doi.org/10.1016/j.chaos.2020.110203>
  41. Dadyan E., Avetisyan P. Neural networks and forecasting COVID-19. *Opt. Mem. Neural Networks.* 2021; 30: 225–35. <https://doi.org/10.3103/S1060992X21030085>
  42. Tamang S., Singh P., Datta B. Forecasting of COVID-19 cases based on prediction using artificial neural network curve fitting technique. *Glob. J. Environ. Sci. Manag.* 2020; 6(S): 53–64. <https://doi.org/10.22034/GJESM.2019.06.SI.06>
  43. Akimkin V.G., Kuzin S.N., Semenenko T.A., Shipulina O.Yu., Yatsyshina S.B., Tivanova E.V., et al. Patterns of the SARS-CoV-2 epidemic spread in a megacity. *Voprosy virusologii.* 2020; 65(4): 203–11. <https://doi.org/10.36233/0507-4088-2020-65-4-203-211> <https://elibrary.ru/fxkaqf> (in Russian)
  44. Akimkin V.G., Kuzin S.N., Semenenko T.A., Ploskireva A.A., Dubodolov D.V., Tivanova E.V., et al. Characteristics of the COVID-19 epidemiological situation in the Russian Federation in 2020. *Vestnik Rossiiskoi akademii meditsinskikh nauk.* 2021; 76(4): 412–22. <https://doi.org/10.15690/vramn1505> <https://elibrary.ru/zmowbe> (in Russian)
  45. Akimkin V.G., Popova A.Yu., Ploskireva A.A., Ugleva S.V., Semenenko T.A., Pshenichnaya N.Yu., et al. Covid-19: the evolution of the pandemic in Russia. Report I: manifestations of the COVID-19 epidemic process. *Zhurnal mikrobiologii, epidemiologii i immunobiologii.* 2022; 99(3): 269–86. <https://doi.org/10.36233/0372-9311-276> <https://elibrary.ru/zxgtfd> (in Russian)
  46. Akimkin V.G., Kuzin S.N., Kolosovskaya E.N., Kudryavtceva E.N., Semenenko T.A., Ploskireva A.A., et al. Assessment of the COVID-19 epidemiological situation in St. Petersburg. *Zhurnal mikrobiologii, epidemiologii i immunobiologii.* 2021; 98(5): 497–511. <https://doi.org/10.36233/0372-9311-154> <https://elibrary.ru/dtmnhz> (in Russian)
  47. Belov A.B. The academician V.D. Belyakov – the founder of the domestic theory of epidemiological science of the XXI century. *Epidemiologia i Vaksino profilaktika.* 2016; 15(6): 9–15. <https://elibrary.ru/xemevf> (in Russian)
  48. Sarkisov A.S. Academician V.D. Belyakov and its contribution to the development of epidemiology. *Byulleten' natsional'nogo nauchno-issledovatel'skogo instituta obshchestvennogo zdorov'ya imeni N.A. Semashko.* 2020; (4): 68–72. <https://doi.org/10.25742/NRIPH.2020.04.010> <https://elibrary.ru/qeazcn> (in Russian)
  49. Zakharov V., Balykina Y. Balance model of COVID-19 epidemic based on percentage growth rate. *Informatika i avtomatizatsiya.* 2021; 20(5): 1034–64. <https://doi.org/10.15622/20.5.2> <https://elibrary.ru/zczxuw> (in Russian)
  50. Zakharov V., Balykina Y., Ilin I., Tick A. Forecasting a new type of virus spread: a case study of COVID-19 with stochastic parameters. *Mathematics.* 2022; 10(20): 3725. <https://doi.org/10.3390/math10203725>

## REFERENCES



25. Balykina Yu.E., Zakharov V.V. Integral inflow and outflow model and its applications. *Prikladnaya matematika. Informatika. Protsesty upravleniya*. 2024; 20(2): 121–35. <https://doi.org/10.21638/spbu10.2024.201> (in Russian)
26. Belyakov V.D., Degtyarev A.A., Ivannikov Yu.G. *Quality and Efficiency of Anti-Epidemic Measures [Kachestvo i effektivnost' protivoepidemicheskikh meropriyatiy]*. Leningrad: Meditsina; 1981. <https://elibrary.ru/zfepwn> (in Russian)

**Информация об авторах:**

**Захаров Виктор Васильевич** – д-р физ.-мат. наук, профессор кафедры математического моделирования энергетических систем ФГБОУ ВО «Санкт-Петербургский государственный университет», Санкт-Петербург, Россия. E-mail: [v.zaharov@spbu.ru](mailto:v.zaharov@spbu.ru); <https://orcid.org/0000-0002-2743-3880>

**Балыкина Юлия Ефимовна**✉ – канд. физ.-мат. наук, доцент кафедры математического моделирования энергетических систем ФГБОУ ВО «Санкт-Петербургский государственный университет», Санкт-Петербург, Россия. E-mail: [j.balykina@spbu.ru](mailto:j.balykina@spbu.ru); <https://orcid.org/0000-0003-2143-0440>

**Участие авторов:** Захаров В.В. – концепция и дизайн исследования, проведение экспериментов, сбор, анализ и интерпретация данных, подготовка текста; Балыкина Ю.Е. – концепция и дизайн исследования, проведение экспериментов, сбор, анализ и интерпретация данных, подготовка текста.

Поступила 24.09.2024  
Принята в печать 14.11.2024  
Опубликована 26.12.2024

**Information about the authors:**

**Victor V. Zakharov** – PhD, Professor, Department of Mathematical Modeling of Energy Systems, St. Petersburg State University, St. Petersburg, Russia. E-mail: [v.zaharov@spbu.ru](mailto:v.zaharov@spbu.ru); <https://orcid.org/0000-0002-2743-3880>

**Yulia E. Balykina**✉ – PhD, Associate Professor, Department of Mathematical Modeling of Energy Systems, St. Petersburg State University, St. Petersburg, Russia. E-mail: [j.balykina@spbu.ru](mailto:j.balykina@spbu.ru); <https://orcid.org/0000-0003-2143-0440>

**Contribution:** Zakharov V.V. – study concept and design, experiments, data collection, analysis and interpretation, text preparation; Balykina Yu.E. – study concept and design, experiments, data collection, analysis and interpretation, text preparation.

Received 24 September 2024  
Accepted 14 November 2024  
Published 26 December 2024