

ПРИКЛАДНАЯ МАТЕМАТИКА

УДК 004.94, 616-036.22
MSC 92D30, 93A30, 68T05

Интегральная модель притока и оттока и ее приложения*

Ю. Е. Балькина^{1,2}, В. В. Захаров¹

¹ Санкт-Петербургский государственный университет,

Российская Федерация, 199034, Санкт-Петербург, Университетская наб., 7–9

² Государственный университет морского и речного флота имени адмирала С. О. Макарова,
Российская Федерация, 198035, Санкт-Петербург, ул. Двинская, 5/7

Для цитирования: Балькина Ю. Е., Захаров В. В. Интегральная модель притока и оттока и ее приложения // Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления. 2024. Т. 20. Вып. 2. С. 121–135.
<https://doi.org/10.21638/spbu10.2024.201>

Описана общая интегральная модель притока и оттока динамической системы, параметры которой имеют стохастический характер. Для такого типа динамических систем формулируется общий принцип динамического баланса, а также вводятся понятия интервальной динамической сбалансированности интегральных объемов притока и оттока и характеристики динамического баланса. Класс стохастических динамических процессов и систем притока и оттока, удовлетворяющих принципу динамического баланса, достаточно широк (распространение эпидемий вирусов и динамика заболеваемости в медицине, процессы изменения численности и структуры населения в демографии, динамика спроса-предложения в экономике и т. д.). Возможности применения предлагаемой модели для построения кратко- и долгосрочных прогнозов демонстрируются на примерах распространения эпидемии COVID-19 в Москве и Санкт-Петербурге, а также прогнозирования роста населения Земли и отдельных стран. Приводятся результаты вычислительных экспериментов по построению ретроспективных прогнозов состояния динамических систем с использованием метода динамических трендов стохастических параметров интегральной модели и классического метода ARIMA. Проводится сравнительный анализ точности прогнозирования.

Ключевые слова: динамические системы притока и оттока, принцип динамического баланса, характеристика динамического баланса, математическое моделирование, прогнозирование.

1. Введение. Класс стохастических динамических процессов и систем притока и оттока достаточно широк. К нему, например, относятся многочисленные процес-

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

© Санкт-Петербургский государственный университет, 2024

сы в области здравоохранения, такие как эволюция вирусных эпидемий и заболеваний невирусного характера, в области демографии — процессы роста численности и изменения структуры населения стран и Земли в целом, в экономике — процессы движения капиталов.

Пусть динамическая система X в момент времени $t = t_0 \geq 0$ состоит из $X(t_0) \geq 0$ элементов определенного типа. Предположим, что в каждый момент времени $t = t_0 + 1, t_0 + 2, \dots, t_0 + T$ ($1 \leq T < \infty$) задано число $x_{\text{inf}}(t) > 0$ новых элементов того же типа, поступающих в систему, и число $x_{\text{of}}(t) > 0$ элементов, покидающих систему, $x_{\text{inf}}(t_0) = x_{\text{of}}(t_0) = 0$. Временные ряды $x_{\text{inf}}(t)$ и $x_{\text{of}}(t)$ будем соответственно называть притоком и оттоком. Рассмотрим дискретную модель динамики системы $X(t)$ при наличии притока $x_{\text{inf}}(t)$ и оттока $x_{\text{of}}(t)$. Очевидно, что состояние системы можно описать дискретным уравнением

$$X(t) = X(t - 1) + x_{\text{inf}}(t) - x_{\text{of}}(t). \quad (1)$$

Если приток и отток представляют собой недетерминированные динамические процессы, то эволюция системы $X(t)$ есть случайный процесс, определяемый динамикой временных рядов притока и оттока, а сама система (1) в общем случае имеет стохастический характер.

Прогнозирование динамики такой системы не представляет труда, если значения притока и оттока $x_{\text{inf}}(t)$ и $x_{\text{of}}(t)$ являются детерминированными функциями времени. Для анализа и прогнозирования систем, для которых указанные временные ряды стационарные, существуют много статистических методов, позволяющих генерировать прогнозы достаточно высокой точности. Однако если условие стационарности не выполняется и/или функции распределения неизвестны, то принятие решений о построении прогнозов динамики системы будет связано со значительной неопределенностью.

К наиболее известным и часто используемым моделям прогнозирования временных рядов относится модель $\text{ARIMA}(p, d, q)$. Аббревиатура ARIMA носит описательный характер и отражает ключевые аспекты самой модели ($\text{AutoRegressive Integrated Moving Average}$, интегрированная модель авторегрессии — скользящего среднего). AR отвечает за авторегрессионную часть, т. е. за порядок запаздывания (p), I — за степень разности (d), MA — за размер окна скользящего среднего, также называемого порядком скользящего среднего (q). Получается, что разности временного ряда порядка d подчиняются модели $\text{ARMA}(p, q)$.

В общем виде модель $\text{ARIMA}(p, d, q)$ можно представить следующим образом:

$$\Delta^d X_t = c + \sum_{i=1}^p a_i \Delta^d X_{t-i} + \sum_{j=1}^q b_j \varepsilon_{t-j} + \varepsilon_t,$$

где a_i, b_j, c — параметры модели; ε_t — стационарный временной ряд; Δ^d — оператор разности временного ряда порядка d (последовательное взятие d раз разностей первого порядка); X_t соответствует значению $X(t)$.

Автокорреляционные методы регрессионного анализа достаточно широко применялись для прогнозирования, например, динамики временного ряда притока новых случаев заболевания во время пандемии COVID-19 . В статьях [1–4] рассматривается использование модели ARIMA при построении прогнозов распространения пандемии в ряде стран мира. Следует отметить, 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.

Модели анализа временных рядов широко внедрены для прогнозирования тенденций, структурных сдвигов, циклов и ненаблюдаемых значений и доказали свою полезность в области медицины [5, 6]. Было показано, что модель авторегрессионно-скользящего среднего обладает многообещающей точностью для прогнозирования динамики различных инфекционных заболеваний [7, 8]. При этом исследователи отмечают, что ARIMA способна давать качественные результаты на краткосрочном горизонте прогнозирования [9].

При изучении COVID-19 с помощью модели ARIMA проводилось прогнозирование количества новых случаев заболевания, смертей и выздоровлений на основе ежедневных данных из разных стран для оценки будущей динамики эпидемии [10–13]. В течение первой волны эпидемии COVID-19 было несколько попыток спрогнозировать дальнейшее развитие эпидемической ситуации благодаря моделям ARIMA. Например, в работе [14] модель ARIMA применялась для краткосрочного прогноза общего количества новых случаев заболевания на основе ежедневных данных ООН с 21 января по 16 марта 2020 г. Для оценки построенных моделей использовались информационный критерий Акаике (AIC) и тест Лjung — Бокса. Чтобы оценить достоверность предложенной модели, были рассчитаны средняя абсолютная процентная ошибка (MAPE) и среднеквадратическая ошибка (RMSE) между наблюдаемыми и спрогнозированными значениями числа новых случаев COVID-19. Авторы [14] предложили использовать модель с параметрами ARIMA(1, 2, 1), при этом горизонт прогнозирования составлял 5 дней. В [15] была проанализирована динамика развития эпидемии COVID-19 в Южной Африке. С помощью модели ARIMA(11, 1, 9) был осуществлен 15-дневный прогноз новых случаев заболевания COVID-19, обучающими историческими данными служили сведения за период с 7 марта 2020 г. по 3 августа 2021 г.

2. Интегральная модель притока и оттока. Рассмотрим суммы членов временных рядов притока и оттока до момента времени $t \leq T$.

Определение 1. Будем называть $X_{\text{inf}}(t)$ интегральным значением притока в систему X с момента времени $t_0 + 1$ до момента времени $t \leq T$ сумму членов временного ряда притока:

$$X_{\text{inf}}(t) = \sum_{\tau=t_0}^t x_{\text{inf}}(\tau).$$

Определение 2. Будем называть $X_{\text{of}}(t)$ интегральным значением оттока из системы X с момента времени $t_0 + 1$ до момента времени $t \leq T$ сумму членов временного ряда оттока:

$$X_{\text{of}}(t) = \sum_{\tau=t_0}^t x_{\text{of}}(\tau).$$

Следует заметить, что полученные таким образом временные ряды $X_{\text{inf}}(t)$ и $X_{\text{of}}(t)$ зависят от всех членов временных рядов $x_{\text{inf}}(\tau)$ и $x_{\text{of}}(\tau)$, $t_0 + 1 \leq \tau \leq t$ соответственно.

Уравнение (1) с учетом введенных ограничений можно преобразовать к виду

$$X(t) = X(t_0) + X_{\text{inf}}(t) - X_{\text{of}}(t).$$

Введем также для любого $t > t_0 + 1$ следующие обозначения:

$$r_{\text{inf}}(t) = 100 \cdot \frac{X_{\text{inf}}(t) - X_{\text{inf}}(t-1)}{X_{\text{inf}}(t-1)}, \quad (2)$$

$$r_{\text{of}}(t) = 100 \cdot \frac{X_{\text{of}}(t) - X_{\text{of}}(t-1)}{X_{\text{of}}(t-1)}. \quad (3)$$

Величина $r_{\text{inf}}(t)$ называется процентным приростом интегрального значения притока в момент времени t , а $r_{\text{of}}(t)$ — процентным приростом интегрального значения оттока.

Рассмотрим систему дискретных уравнений переменных $X(t)$, $X_{\text{inf}}(t)$ и $X_{\text{of}}(t)$ с недетерминированными параметрами $r_{\text{inf}}(t)$ и $r_{\text{of}}(t)$ (см. (2) и (3)):

$$X(t) = X(t_0) + X_{\text{inf}}(t) - X_{\text{of}}(t), \quad (4)$$

$$X_{\text{inf}}(t) = \left(1 + \frac{r_{\text{inf}}(t)}{100}\right) X_{\text{inf}}(t-1), \quad (5)$$

$$X_{\text{of}}(t) = \left(1 + \frac{r_{\text{of}}(t)}{100}\right) X_{\text{of}}(t-1) \quad (6)$$

при $X_{\text{inf}}(t_0) = X_{\text{of}}(t_0) = 0$, $t > t_0 + 1$. Систему (4)–(6) будем называть интегральной моделью притока и оттока системы X .

Тогда справедливо следующее утверждение.

Утверждение. Пусть $X(t_0) = 0$, $x_{\text{inf}}(t_0) = x_{\text{of}}(t_0) = 0$, $x_{\text{inf}}(t) > 0$ и $x_{\text{of}}(t) > 0$ для всех $t > t_0$. Для того чтобы множество динамической системы X , состоящие которой удовлетворяет системе (4)–(6), было не пусто при $t > t_0$, необходимо и достаточно, чтобы для любого $t > t_0$ было выполнено неравенство

$$X_{\text{inf}}(t) > X_{\text{of}}(t). \quad (7)$$

Доказательство. Необходимость. Пусть множество системы X не является пустым при $t > t_0$, это означает, что $X(t) > 0$. Тогда, поскольку $X(t_0) = 0$, с учетом системы (4)–(6) получаем справедливость неравенства (7) для $t > t_0$.

Достаточность. Пусть для всех $t > t_0$ неравенство (7) выполняется. Тогда $X(t) > 0$ при любом значении $X(t_0) \geq 0$. То есть множество элементов системы X не является пустым при любом при $t > t_0$. \square

Рассмотрим для всех значений t , таких, что $t_0 + 1 \leq t \leq t_0 + T$, задачу целочисленного линейного программирования

$$\min_{t_0+1 \leq \tau \leq t} \tau \quad (8)$$

при условии, что

$$X_{\text{inf}}(\tau) \geq X_{\text{of}}(t). \quad (9)$$

Обозначим через $\tau(t)$ решение задачи (8), (9).

Справедлива следующая теорема.

Теорема (принцип динамического баланса систем притока и оттока). Пусть для любых целых значений t , таких, что $t_0 + 1 \leq t \leq t_0 + T$, выполнено неравенство $X_{\text{inf}}(t) > X_{\text{of}}(t)$, и $\tau(t)$ есть решение задачи целочисленного программирования (8), (9). Тогда для любых значений t , таких, что $t_0 + 1 \leq t \leq t_0 + T$, имеет место неравенство

$$X_{\text{inf}}(\tau(t)) \geq X_{\text{of}}(t) \geq X_{\text{inf}}(\tau(t) - 1). \quad (10)$$

Доказательство этой теоремы аналогично доказательству теоремы 1 в [16].

Замечание. В соответствии с принципом динамического баланса систем притока и оттока временные ряды $X_{\text{inf}}(t)$ и $X_{\text{of}}(t)$ удовлетворяют условию (10), которое содержательно означает, что величины интегральных объемов оттока в момент времени t находятся между последовательными значениями интегральных объемов притока в моменты времени $\tau(t) - 1$ и $\tau(t)$. Такое свойство будем называть интервальной динамической сбалансированностью притока и оттока.

Определение 3. Пусть $\tau(t)$ есть решение задачи (8), (9). Величина $\theta(t) = t - \tau(t)$ называется характеристикой динамического баланса интегральных объемов притока и оттока в системе X .

3. Применение интегральной модели притока и оттока.

3.1. Прогнозирование динамики эпидемий новых вирусов. Наиболее популярной моделью для описания процессов распространения инфекционных заболеваний в некоторой популяции является трехкамерная модель SIR (Susceptible — Infected — Removed). В текущий момент времени каждый человек из этой популяции принадлежит одной из трех групп: восприимчивые к вирусу (Susceptible), инфицированные (Infected) и группа, для которых болезнь завершилась (выздоровевшие или умершие (Removed)). В группу восприимчивых входят все люди, которые еще не подверглись заражению инфекцией. По мере распространения вируса часть из них переходят в группу инфицированных, а затем в группу выбывших (выздоровевших или умерших). Предполагается, что полное число людей в популяции равно сумме численностей этих трех групп. Впервые модель была в общем виде описана в работе У. Кермака и А. Мак-Кендрика в 1927 г. [17], однако наибольший интерес исследователей был в дальнейшем сосредоточен на частном случае этой модели в виде системы из трех дифференциальных уравнений с постоянными коэффициентами [18].

Следует заметить, что публикуемые статистические данные о заболеваемости COVID-19 содержат ежедневное число новых случаев заболевания, ежедневное количество выздоровевших и умерших, общее количество заболевших и общее количество выздоровевших и умерших с начала эпидемии. На основании этих данных можно вычислить текущее число болеющих людей (активных случаев болезни). Фактически статистические данные содержат значения временных рядов притока (новых случаев заболевания) $x_{\text{inf}}(t)$ и оттока (общее число выздоровевших и умерших) $x_{\text{of}}(t)$ в системе, текущее число элементов в которой $X(t)$ равно численности группы инфицированных (болеющих людей) в момент времени t .

Вместо трехкамерной модели SIR рассмотрим модель, описываемую при $t > t_0$ системой дискретных уравнений следующего вида:

$$I(t) = C(t) - R(t), \quad (11)$$

$$C(t) = \left(1 + \frac{r_{\text{inf}}(t)}{100}\right) C(t-1), \quad (12)$$

$$R(t) = \left(1 + \frac{r_{\text{of}}(t)}{100}\right) R(t-1), \quad (13)$$

$$I(t_0) = C(t_0) = R(t_0) = 0.$$

В этой модели переменная $C(t)$ характеризует интегральный приток (общее количество) новых случаев заболевания (Confirmed Cases), а переменная $R(t)$ — интегральный отток суммарного количества выздоровевших и умерших (Removed), начиная с момента времени $t_0 + 1$ до момента времени t (включительно):

$$C(t) = \sum_{t=t_0+1}^t x_{\text{inf}}(t),$$

$$R(t) = \sum_{t=t_0+1}^t x_{\text{of}}(t).$$

Параметрами модели являются процентный прирост интегрального объема притока (общего количества выявленных случаев заболевания) $r_{\text{inf}}(t)$ и процентный прирост интегрального объема оттока (общего количества выздоровевших и умерших пациентов) $r_{\text{of}}(t)$, меняющиеся во времени.

Будем называть модель динамики заболевания (11)–(13) интегральной моделью притока и оттока болеющих людей.

3.2. Результаты вычислительных экспериментов. Рассмотрим результаты вычислительных экспериментов по построению ретроспективных прогнозов состояния динамических систем с использованием прогнозов линейных по времени трендов стохастических параметров интегральной модели и классического метода ARIMA.

Приведем сначала пример применения интегральной модели притока и оттока болеющих людей для прогнозирования динамики заболевания COVID-19 в Москве и Санкт-Петербурге в период первой волны эпидемии при прогнозировании динамики процентных приростов интегральных объемов притока с помощью линейных трендов [16] этих параметров.

Таблица 1. Ретроспективные прогнозы количества болеющих людей в Москве в период первой волны эпидемии COVID-19 в 2020 г. при использовании линейных трендов процентных приростов интегральных объемов притока и оттока

Дата	ИП, чел.	Прирост ИП, %	ИО, чел.	Прирост ИО, %	$I(t)$, чел.	$I(t)$, чел. (прогноз)	Точность прогноза, %
1	2	3	4	5	6	7	8
14 апреля	13 002	12.93	1111	20.89	11 891		
15 апреля	14 776	13.64	1311	18.00	13 465		
16 апреля	16 146	9.27	1507	14.95	14 639		
17 апреля	18 105	12.13	1644	9.09	16 461		
18 апреля	20 754	14.63	1827	11.13	18 927		
19 апреля	24 324	17.20	1939	6.13	22 385		
20 апреля	26 350	8.33	2042	5.31	24 308		
21 апреля	29 433	11.70	2290	12.14	27 143		
22 апреля	31 981	8.66	2528	10.39	29 453		
23 апреля	33 940	6.13	2736	8.23	31 204		
24 апреля	36 897	8.71	3060	11.84	33 837		
25 апреля	39 509	7.08	3413	11.54	36 096		
26 апреля	42 480	7.52	3579	4.86	38 901		
27 апреля	45 351	6.76	3959	10.62	41 392		
28 апреля	48 426	6.78	4609	16.42	43 817		
29 апреля	50 646	4.58	5156	11.87	45 490		
30 апреля	53 739	6.11	5746	11.44	47 993		
1 мая	57 300	6.63	6424	11.80	50 876	50 876	100.00
2 мая	62 658	9.35	7069	10.04	55 589	53 813	96.81
3 мая	68 606	9.49	7758	9.75	60 848	56 793	93.34
4 мая	74 401	8.45	8337	7.46	66 064	59 802	90.52
5 мая	80 115	7.68	8686	4.19	71 429	62 827	87.96

1	2	3	4	5	6	7	8
6 мая	85 973	7.31	9324	7.35	76 649	65 852	85.91
7 мая	92 676	7.80	10 132	8.67	82 544	68 861	83.42
8 мая	98 522	6.31	11 215	10.69	87 307	71 835	82.28
9 мая	104 189	5.75	13 789	22.95	90 400	74 756	82.69
10 мая	109 740	5.33	14 858	7.75	94 882	77 604	81.79
11 мая	115 909	5.62	18 946	27.51	96 963	80 358	82.87
12 мая	121 301	4.65	20 821	9.90	100 480	MAPE = 13.24 %	
13 мая	126 004	3.88	22 738	9.21	103 266		
14 мая	130 716	3.74	24 617	8.26	106 099		
15 мая	135 464	3.63	25 920	5.29	109 544	109 544	100.00
16 мая	138 969	2.59	27 464	5.96	111 505	112 873	98.77
17 мая	142 824	2.77	28 993	5.57	113 831	116 069	98.03
18 мая	146 062	2.27	30 493	5.17	115 569	119 114	96.93
19 мая	149 607	2.43	33 147	8.70	116 460	121 991	95.25
20 мая	152 306	1.80	38 662	16.64	113 644	124 683	90.29
21 мая	155 219	1.91	42 476	9.86	112 743	127 175	87.20
22 мая	158 207	1.93	45 449	7.00	112 758	129 450	85.20
23 мая	161 397	2.02	49 347	8.58	112 050	131 495	82.65
24 мая	163 913	1.56	51 833	5.04	112 080	133 298	81.07
25 мая	166 473	1.56	55 620	7.31	110 853	134 845	78.36
26 мая	169 303	1.70	63 729	14.58	105 574	MAPE = 10.63 %	
27 мая	171 443	1.26	69 641	9.28	101 802		
28 мая	173 497	1.20	73 505	5.55	99 992		
29 мая	175 829	1.34	77 055	4.83	98 774		
30 мая	178 196	1.35	80 732	4.77	97 464		
31 мая	180 791	1.46	82 656	2.38	98 135		
1 июня	183 088	1.27	84 792	2.58	98 296		
2 июня	185 374	1.25	90 893	7.20	94 481		
3 июня	187 216	0.99	94 339	3.79	92 877		
4 июня	189 214	1.07	97 464	3.31	91 750	91 750	100.00
5 июня	191 069	0.98	100 164	2.77	90 905	90 582	99.64
6 июня	193 061	1.04	103 295	3.13	89 766	89 379	99.57
7 июня	195 017	1.01	105 633	2.26	89 384	88 147	98.62
8 июня	197 018	1.03	107 317	1.59	89 701	86 892	96.87
9 июня	198 590	0.80	112 766	5.08	85 824	85 620	99.76
10 июня	199 785	0.60	116 618	3.42	83 167	84 338	98.59
11 июня	201 221	0.72	118 907	1.96	82 314	83 053	99.10
12 июня	202 935	0.85	121 211	1.94	81 724	81 771	99.94
13 июня	204 428	0.74	122 789	1.30	81 639	80 500	98.60
14 июня	205 905	0.72	124 972	1.78	80 933	79 244	97.91
15 июня	207 264	0.66	126 574	1.28	80 690	78 013	96.68
16 июня	208 680	0.68	129 500	2.31	79 180	76 811	97.01
17 июня	209 745	0.51	131 819	1.79	77 926	75 645	97.07
18 июня	210 785	0.50	133 914	1.59	76 871	74 523	96.94
19 июня	211 921	0.54	135 965	1.53	75 956	73 449	96.70
20 июня	212 978	0.50	138 114	1.58	74 864	72 429	96.75
21 июня	213 946	0.45	139 153	0.75	74 793	71 470	95.56
22 июня	215 014	0.50	140 287	0.81	74 727	70 576	94.44
23 июня	216 095	0.50	142 891	1.86	73 204	69 752	95.28
24 июня	216 906	0.38	144 925	1.42	71 981	69 003	95.86
25 июня	217 791	0.41	145 863	0.65	71 928	68 333	95.00
26 июня	218 604	0.37	147 885	1.39	70 719	67 746	95.80
27 июня	219 354	0.34	149 757	1.27	69 597	67 244	96.62
28 июня	220 071	0.33	150 833	0.72	69 238	66 831	96.52
29 июня	220 853	0.36	151 863	0.68	68 990	66 508	96.40
30 июня	221 598	0.34	154 193	1.53	67 405	66 278	98.33
							MAPE = 2.71 %

В табл. 1 в столбцах 2–5 находятся фактические данные об интегральных объемах притока (ИП) и оттока (ИО), значениях их процентных приростов и количестве активных случаев в Москве, полученные из ежедневных отчетов Коммуникационного центра Правительства РФ по ситуации с коронавирусом, размещенных в открытом доступе на портале стопкоронавирус.рф. В столбце 7 представлены ретроспективные прогнозы количества болеющих людей, построенные при использовании линейных трендов процентных приростов интегральных объемов притока и оттока на основании статистики, предшествующей датам построения прогнозов 1 мая, 15 мая и 4 июня 2020 г., в столбце 8 — ежедневные абсолютные значения точности полученных прогнозов, а также MAPE — средняя абсолютная ошибка в процентах для соответствующего горизонта прогнозирования.

В табл. 2 приведены аналогичные данные для Санкт-Петербурга. Однако прогнозы количества активных случаев болезни (см. столбец 7) получены при использовании для прогнозирования процентных приростов притока и оттока модели ARIMA. Для этого были построены модели ARIMA отдельно для прогнозирования динамики изменения интегральных притока и оттока. Гиперпараметры моделей обновлялись при рассмотрении различных интервалов для моделирования. Так, для прогнозирования прироста интегрального притока в начале мая 2020 г. применялась модель ARIMA(7, 1, 7), для интервала прогнозирования с 1 по 30 июня 2020 г. — модель ARIMA(2, 1, 1). Для каждого горизонта прогнозирования также приведены точность прогноза и показатель MAPE.

Таблица 2. Ретроспективные прогнозы количества болеющих людей в Санкт-Петербурге в период первой волны эпидемии COVID-19 в 2020 г. при использовании ARIMA

Дата	ИП, чел.	Прирост ИП, %	ИО, чел.	Прирост ИО, %	$I(t)$, чел.	$I(t)$, чел. (прогноз)	Точность прогноза, %
1	2	3	4	5	6	7	8
14 апреля	799	17.85	82	0.00	717		
15 апреля	929	16.27	100	21.95	829		
16 апреля	1083	16.58	117	17.00	966		
17 апреля	1507	39.15	138	17.95	1369		
18 апреля	1646	9.22	240	73.91	1406		
19 апреля	1760	6.93	247	2.92	1513		
20 апреля	1846	4.89	249	0.81	1597		
21 апреля	1973	6.88	291	16.87	1682		
22 апреля	2267	14.90	341	17.18	1926		
23 апреля	2458	8.43	385	12.90	2073		
24 апреля	2711	10.29	431	11.95	2280		
25 апреля	2926	7.93	483	12.06	2443		
26 апреля	3077	5.16	516	6.83	2561		
27 апреля	3238	5.23	516	0.00	2722		
28 апреля	3436	6.11	571	10.66	2865		
29 апреля	3726	8.44	737	29.07	2989		
30 апреля	4062	9.02	808	9.63	3254		
1 мая	4411	8.59	812	0.50	3599	3599	100.00
2 мая	4734	7.32	910	12.07	3824	3693	96.59
3 мая	5029	6.23	1095	20.33	3934	3858	98.07
4 мая	5346	6.30	1234	12.69	4112	4004	97.37
5 мая	5572	4.23	1505	21.96	4067	4138	98.25
6 мая	5 884	5.60	1572	4.45	4312	4265	98.91
7 мая	6190	5.20	1596	1.53	4594	4387	95.50
8 мая	6565	6.06	1645	3.07	4920	4506	91.59
9 мая	6990	6.47	1715	4.26	5275	4623	87.64

1	2	3	4	5	6	7	8	
10 мая	7404	5.92	1727	0.70	5677	4739	83.47	
11 мая	7711	4.15	1737	0.58	5974	4854	81.25	
12 мая	8050	4.40	1842	6.04	6208	MAPE = 7 %		
13 мая	8485	5.40	1911	3.75	6574			
14 мая	8945	5.42	1968	2.98	6977			
15 мая	9486	6.05	2079	5.64	7407	7407	100.00	
16 мая	10 011	5.53	2189	5.29	7822	7710	98.57	
17 мая	10 462	4.51	2260	3.24	8202	8075	98.45	
18 мая	10 887	4.06	2299	1.73	8588	8437	98.24	
19 мая	11 340	4.16	2367	2.96	8973	8796	98.03	
20 мая	11 795	4.01	2579	8.96	9216	9152	99.31	
21 мая	12 203	3.46	2765	7.21	9438	9506	99.28	
22 мая	12 592	3.19	2966	7.27	9626	9856	97.61	
23 мая	12 955	2.88	3237	9.14	9718	10 203	95.01	
24 мая	13 339	2.96	3585	10.75	9754	10 548	91.86	
25 мая	13 713	2.80	3728	3.99	9985	10 889	90.95	
26 мая	14 076	2.65	3894	4.45	10 182	MAPE = 5 %		
27 мая	14 463	2.75	4616	18.54	9847			
28 мая	14 846	2.65	4950	7.24	9896			
29 мая	15 215	2.49	5291	6.89	9924			
30 мая	15 580	2.40	5564	5.16	10 016			
31 мая	15 949	2.37	5884	5.75	10 065			
1 июня	16 313	2.28	5961	1.31	10 352			
2 июня	16 689	2.30	6231	4.53	10 458			
3 июня	17 069	2.28	6571	5.46	10 498			
4 июня	17 444	2.20	6959	5.90	10 485	10 485	100.00	
5 июня	17 822	2.17	7367	5.86	10 455	10 709	97.57	
6 июня	18 169	1.95	7808	5.99	10 361	10 816	95.61	
7 июня	18 509	1.87	8136	4.20	10 373	10 924	94.68	
8 июня	18 835	1.76	8307	2.10	10 528	11 035	95.18	
9 июня	19 153	1.69	8534	2.73	10 619	11 148	95.02	
10 июня	19 466	1.63	9003	5.50	10 463	11 262	92.36	
11 июня	19 769	1.56	9384	4.23	10 385	11 379	90.43	
12 июня	20 043	1.39	9837	4.83	10 206	11 497	87.35	
13 июня	20 305	1.31	10 130	2.98	10 175	11 616	85.83	
14 июня	20 561	1.26	10 274	1.42	10 287	11 738	85.90	
15 июня	20 813	1.23	10 374	0.97	10 439	11 861	86.38	
16 июня	21 047	1.12	10 660	2.76	10 387	11 986	84.61	
17 июня	21 275	1.08	11 414	7.07	9861	12 112	77.18	
18 июня	21 506	1.09	12 080	5.83	9426	12 239	70.15	
19 июня	21 734	1.06	12 964	7.32	8770	12 369	58.97	
20 июня	21 966	1.07	13 779	6.29	8187	12 499	47.33	
21 июня	22 195	1.04	14 303	3.80	7892	12 631	39.95	
22 июня	22 412	0.98	14 516	1.49	7896	12 765	38.34	
23 июня	22 632	0.98	14 772	1.76	7860	12 899	35.89	
24 июня	22 850	0.96	15 453	4.61	7397	13 035	23.78	
25 июня	23 071	0.97	16 061	3.93	7010	13 172	12.09	
26 июня	23 294	0.97	16 503	2.75	6791	13 311	4.00	
27 июня	23 518	0.96	17 276	4.68	6242	13 450	0.00	
28 июня	23 735	0.92	17 895	3.58	5840	13 591	0.00	
29 июня	23 954	0.92	18 161	1.49	5793	13 732	0.00	
30 июня	24 207	1.06	18 446	1.57	5761	13 875	0.00	
							MAPE = 45 %	

На рисунке представлены результаты моделирования для Москвы и Санкт-Петербурга.

При использовании краткосрочных прогнозов линейных трендов процентных приростов ошибка MAPE при оценке будущего количества болеющих людей для

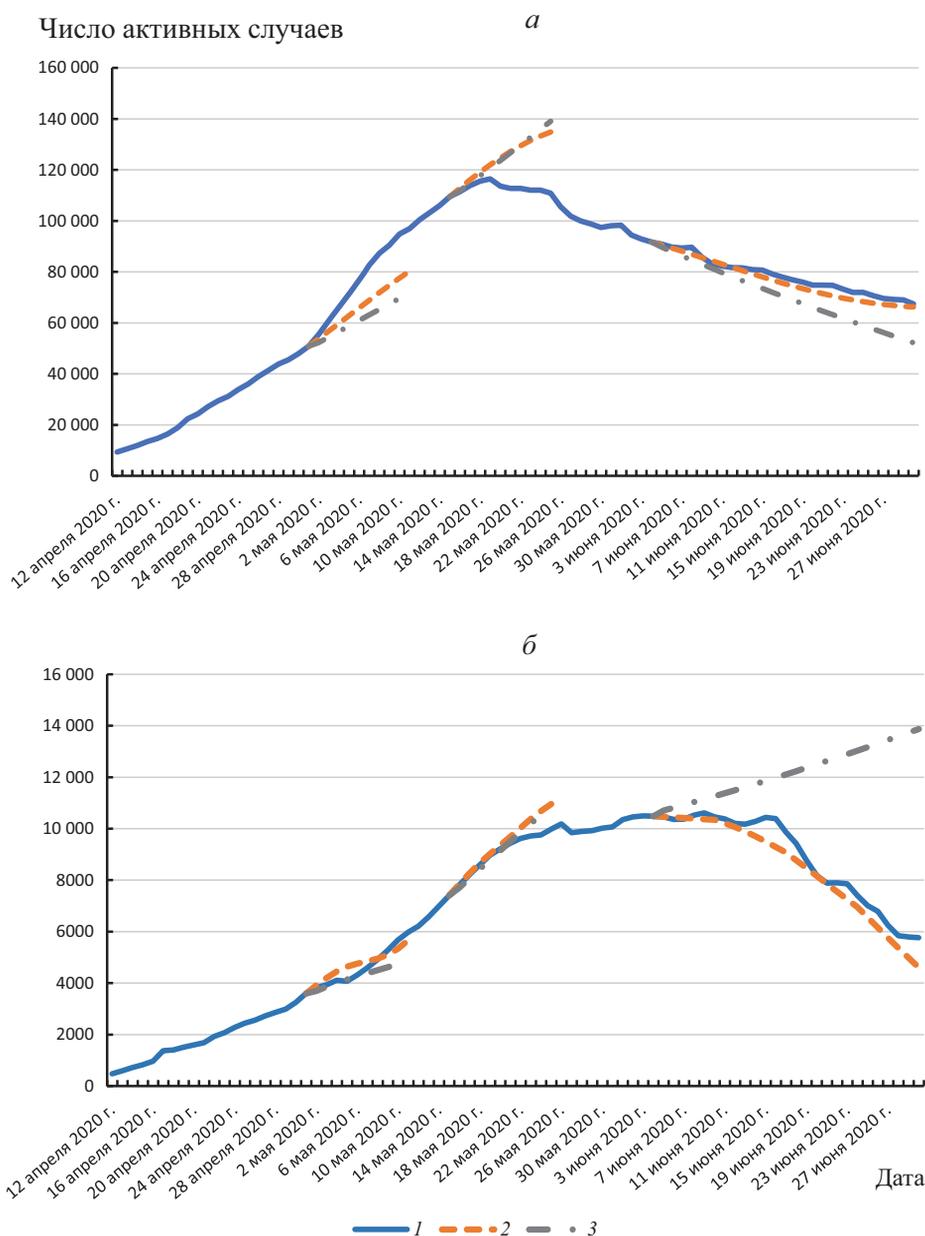


Рис. 1. Фактические и расчетные траектории изменения числа активных случаев заболевания COVID-19 в Москве (а) и Санкт-Петербурге (б) на рассматриваемых интервалах прогнозирования (2020 г.)

Количество активных случаев: 1 — фактическое, 2 — прогноз с помощью линейных трендов, 3 — прогноз, модель ARIMA.

Москвы в периоды 1–11 и 15–20 мая 2020 г. первой волны составила 13.24 и 10.63 % соответственно. Показатели для модели ARIMA были равны 18 и 10 % соответственно. При увеличении горизонта прогнозирования до 27 дней (период с 4 по 30 июня 2020 г.)

прогнозы линейных трендов показали более высокие результаты — ошибка MAPE составила 2.7 % против 10 % при применении модели ARIMA. Схожие результаты были получены и при рассмотрении ситуации в Санкт-Петербурге. Ошибка краткосрочных прогнозов при оценке с помощью линейных трендов не превышала 5.6 %, а модели ARIMA — не более 7 %. В случае более далекого горизонта прогнозирования (27 дней) ошибка модели ARIMA возрастала до 45 %, в то время как при использовании модели оценки линейных трендов процентных приростов MAPE = 3.6 %.

Таким образом, следует отметить, что и для Москвы, и для Санкт-Петербурга при коротких горизонтах прогнозирования как модель притока и оттока, так и модель ARIMA показывали схожие результаты. При этом при увеличении интервала прогнозирования ошибка модели ARIMA начинала нарастать, а модель оценки линейных трендов стохастических параметров интегральной модели притока и оттока демонстрировала значительно более высокую точность.

3.3. Прогнозирование динамики численности населения. Для прогнозирования численности населения Земли и отдельных стран построим интегральную модель динамики численности населения. Временные ряды ежегодных значений количества родившихся детей $B(t)$ и умерших людей $D(t)$, формирующих приток и отток в систему народонаселения Земли, содержатся в базе данных ООН [19]. При построении модели следует дополнительно учесть отток населения за счет чистой миграции $NM(t)$. Систему народонаселения страны обозначим N , численность ее элементов — $N(t)$. Приток в систему N обеспечивает годовая рождаемость $B(t)$, отток из системы $DNM(t)$ равен $D(t) + NM(t)$.

Тогда система дискретных уравнений, описывающих динамику численности населения $N(t)$, будет иметь следующий вид:

$$N(t) = N(t_0) + B_{\text{int}}(t) - DNM_{\text{int}}(t), \quad (14)$$

$$B_{\text{int}}(t) = \left(1 + \frac{r_{\text{inf}}(t)}{100}\right) B_{\text{int}}(t-1), \quad (15)$$

$$DNM_{\text{int}}(t) = \left(1 + \frac{r_{\text{of}}(t)}{100}\right) DNM_{\text{int}}(t-1), \quad (16)$$

где

$$B_{\text{int}}(t) = \sum_{t_0+1}^t B(t),$$

$$DNM_{\text{int}}(t) = \sum_{t_0+1}^t DNM(t).$$

Результаты моделирования и прогнозирования динамики численности населения Земли в целом представлены в статье [16], а численности населения двух стран Западной Африки — в [20]. Важными особенностями полученных результатов применения модели (14)–(16) при прогнозировании численности населения Земли и отдельно взятых стран являются достаточно высокая точность ретроспективного прогнозирования и достаточно хорошо предсказуемая будущая динамика характеристики динамического баланса рассмотренных демографических процессов. Следует отметить, что в процессе вычислительных экспериментов по ретроспективному прогнозированию интегральных объемов притока и оттока значения характеристики динамического баланса, рассчитанные на основе полученных прогнозов, практически полностью совпадают с фактическими.

4. Заключение. Описанная в работе интегральная модель притока и оттока динамической системы может быть использована при прогнозировании динамики достаточно широкого класса процессов со стохастическими параметрами. Сформулированный общий принцип динамического баланса систем притока и оттока, а также наличие у системы свойства интервальной динамической сбалансированности интегральных объемов притока и оттока вместе со свойством монотонного убывания значений процентных приростов этих объемов позволяют обеспечить высокую точность прогнозирования динамики таких систем на достаточно продолжительных временных горизонтах.

Приведенные результаты ретроспективного прогнозирования распространения эпидемии COVID-19 на примере мегаполисов, к которым относятся Москва и Санкт-Петербург, могут быть положены в основу создания аналитической системы для обеспечения мероприятий органов здравоохранения по борьбе с эпидемиями как на общенациональном, так и на региональном уровне. Применение разработанной методики построения линейных трендов значений процентных приростов интегральных объемов притока и оттока вместе с моделью ARIMA дает возможность получения достаточно высокой точности будущей динамики основных переменных рассматриваемых процессов. Однако, как показали проведенные эксперименты, использование линейных трендов имеет существенные преимущества в точности прогнозирования на продолжительных горизонтах времени. Эта методика для построения прогнозов динамики численности населения Земли, ее стран и регионов также хорошо зарекомендовала себя в процессе проведения численных экспериментов. Перспективной задачей будущих исследований является расширение класса динамических процессов и систем притока и оттока, для которых с помощью предложенной модели можно будет обеспечить приемлемый уровень точности прогнозирования будущей динамики таких процессов и систем.

Литература

1. *Moftakhar L., Seif M., Safe M. S.* Exponentially increasing trend of infected patients with COVID-19 in Iran: a comparison of neural network and ARIMA forecasting models // *Iran Journal of Public Health*. 2020. Vol. 9. P. 92–100.
2. *Ahmar A. S., del Val E. B.* SutteARIMA: short-term forecasting method, a case: COVID-19 and stock market in Spain // *Science of the Total Environment*. 2020. Vol. 729. Art. N 138883.
3. *Chaudhry R. M., Hanif A., Chaudhary M., Minhas S. 2nd, Mirza K., Ashraf T., Gilani S. A., Kashif M.* Coronavirus disease 2019 (COVID-19): Forecast of an emerging urgency in Pakistan // *Cureus*. 2020. Vol. 12. N 5. Art. N e8346.
4. *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 // *Journal of Health Management*. 2022. Vol. 24. Iss. 3. P. 373–388.
5. *Earnest A., Chen M. I., Ng D., Leo Y. S.* Using autoregressive integrated moving average (ARIMA) models to predict and monitor the number of beds occupied during a SARS outbreak in a tertiary hospital in Singapore // *BMC Health Services Research*. 2005. Vol. 5. Art. N 36.
6. *Li X. J., Kang D. M., Cao J., Wang J. Z.* A time series model in incidence forecasting of hemorrhagic fever with renal syndrome // *Journal of Shandong University (Health Sciences)*. 2008. Vol. 46. N 5. P. 547–549.
7. *Heisterkamp S. H., Dekkers A. L., Heijne J. C.* Automated detection of infectious disease outbreaks: hierarchical time series models // *Statistics in Medicine*. 2003. Vol. 25. N 24. P. 4179–96.
8. *Zhang G. P.* Time series forecasting using a hybrid ARIMA and neural network model // *Neurocomputing*. 2003. Vol. 50. P. 159–175.
9. *De Beer J.* Projecting age-specific fertility rates by using time-series methods // *European Journal of Population*. 1990. Vol. 5. N 4. P. 315–346.

10. *Abonazel M., Darwish N.* Forecasting confirmed and recovered COVID-19 cases and deaths in Egypt after the genetic mutation of the virus: ARIMA Box-Jenkins approach // *Communications in Mathematical Biology and Neuroscience*. 2022. Vol. 2022. Art. N 17.

11. *Gecili E., Ziady A., Szczesniak R. D.* Forecasting COVID-19 confirmed cases, deaths and recoveries: revisiting established time series modeling through novel applications for the USA and Italy // *PLoS One*. 2021. Vol. 16. N 1. Art. N e0244173.

12. *Singh S., Parmar K. S., Makkhan S. J. S., Kaur J., Peshoria S., Kumar J.* Study of ARIMA and least square support vector machine (LS-SVM) models for the prediction of SARS-CoV-2 confirmed cases in the most affected countries // *Chaos, Solitons and Fractals*. 2020. Vol. 139. Art. N 110086.

13. *Aditya S. C. B., Darmawan W., Nadia B. U., Hanafiah N.* Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET // *Procedia Computer Science*. 2021. Vol. 179. P. 524–532.

14. *Duong N., Phuong Th. L., Nhu Q. D., Binh L., Ai L. C., Hong D. P.* Predicting the pandemic COVID-19 using ARIMA model // *VNU Journal of Science: Mathematics—Physics*. 2020. Vol. 36. N 4. Art. N 4492.

15. *Claris S., Peter N.* ARIMA model in predicting of COVID-19 epidemic for the Southern Africa region // *African Journal of Infectious Diseases*. 2022. Vol. 17. N 1. P. 1–9.

16. *Захаров В. В.* Принцип динамического баланса демографического процесса и пределы роста населения Земли // *Докл. РАН. Математика, информатика, процессы управления*. 2023. Т. 15. С. 108–114. <https://doi.org/10.31857/S2686954323600301>

17. *Kermack W. O., McKendrick A. G.* A contribution to the mathematical theory of epidemics // *Proceedings of the Royal Society A*. 1927. Vol. 115. P. 700–721.

18. *Anderson R. M., May R. M.* *Infectious diseases of humans: Dynamics and control*. Oxford: Oxford University Press, 1991. 757 p.

19. *United Nations, Department of Economic and Social Affairs, Population Division.* *World Population Prospects 2022: Methodology of the United Nations population estimates and projections*. New York: United Nations Publ., 2022. 64 p.

20. *Захаров В. В., Ндиайе С. М.* Прогнозирование численности населения и динамические игры против природы // *Математическая теория игр и ее приложения*. 2024. Т. 16. № 1. С. 17–37.

Статья поступила в редакцию 9 января 2024 г.

Статья принята к печати 12 марта 2024 г.

К о н т а к т н а я и н ф о р м а ц и я:

Балыкина Юлия Ефимовна — канд. физ.-мат. наук, доц.; j.balykina@sbpu.ru

Захаров Виктор Васильевич — д-р физ.-мат. наук, проф.; v.zaharov@sbpu.ru

Integral inflow and outflow model and its applications*

Yu. E. Balykina^{1,2}, *V. V. Zakharov*¹

¹ St. Petersburg State University, 7–9, Universitetskaya nab., St. Petersburg, 199034, Russian Federation

² Admiral Makarov State University of Maritime and Inland Shipping, 5/7, Dvinskaya ul., St. Petersburg, 198035, Russian Federation

For citation: Balykina Yu. E., Zakharov V. V. Integral inflow and outflow model and its applications. *Vestnik of Saint Petersburg University. Applied Mathematics. Computer Science. Control Processes*, 2024, vol. 20, iss. 2, pp. 121–135. <https://doi.org/10.21638/spbu10.2024.201> (In Russian)

The article describes a general integral model of the inflow and outflow of a dynamic system, the parameters of which are stochastic in nature. For this type of dynamic systems, the general principle of dynamic balance is formulated, and the concepts of interval dynamic balance of integral volumes of inflow and outflow as well as the concept of dynamic balance

* This research was supported by the Russian Science Foundation, project N 23-21-10049 (<https://rscf.ru/project/23-21-10049/>), and Saint Petersburg Science Foundation.

characteristic are introduced. The class of stochastic dynamic processes and systems of inflow and outflow that satisfy the principle of dynamic balance is quite wide (the spread of viral epidemics and the dynamics of morbidity in medicine, processes of changes in the size and structure of the population in demography, the dynamics of supply and demand in the economy, etc.). The possibilities of using the proposed model for constructing short-term and long-term forecasts are demonstrated using examples of the spread of the COVID-19 epidemic in Moscow and Saint Petersburg, as well as using the example of forecasting the growth of the Earth population and population of countries. The results of computational experiments on constructing retrospective forecasts of the state of dynamic systems using the method of dynamic trends of stochastic parameters of the integral model and using the classical ARIMA method are presented. A comparative analysis of forecasting accuracy is provided.

Keywords: dynamic systems of inflow and outflow, principle of dynamic balance, dynamic balance characteristic, mathematical modeling, forecasting.

References

1. Moftakhar L., Seif M., Safe M. S. Exponentially increasing trend of infected patients with COVID-19 in Iran: a comparison of neural network and ARIMA forecasting models. *Iran Journal of Public Health*, 2020, vol. 9, pp. 92–100.
2. Ahmar A. S., del Val E. B. SutteARIMA: short-term forecasting method, a case: COVID-19 and stock market in Spain. *Science of the Total Environment*, 2020, vol. 729, art. no. 138883.
3. Chaudhry R. M., Hanif A., Chaudhary M., Minhas S. 2nd, Mirza K., Ashraf T., Gilani S. A., Kashif M. Coronavirus Disease 2019 (COVID-19): Forecast of an Emerging urgency in Pakistan. *Cureus*, 2020, vol. 12, iss. 5, art. no. e8346.
4. 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. *Journal of Health Management*, 2022, vol. 24, iss. 3, pp. 373–388.
5. Earnest A., Chen M. I., Ng D., Leo Y. S. Using autoregressive integrated moving average (ARIMA) models to predict and monitor the number of beds occupied during a SARS outbreak in a tertiary hospital in Singapore. *BMC Health Services Research*, 2005, vol. 5, art. no. 36.
6. Li X. J., Kang D. M., Cao J., Wang J. Z. A time series model in incidence forecasting of hemorrhagic fever with renal syndrome. *Journal of Shandong University (Health Sciences)*, 2008, vol. 46, no. 5, pp. 547–549.
7. Heisterkamp S. H., Dekkers A. L., Heijne J. C. Automated detection of infectious disease outbreaks: hierarchical time series models. *Statistics in Medicine*, 2003, vol. 25, no. 24, pp. 4179–96.
8. Zhang G. P. Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 2003, vol. 50, pp. 159–175.
9. De Beer J. Projecting age-specific fertility rates by using time-series methods. *European Journal of Population*, 1990, vol. 5, no. 4, pp. 315–346.
10. Abonazel M., Darwish N. Forecasting confirmed and recovered COVID-19 cases and deaths in Egypt after the genetic mutation of the virus: ARIMA Box-Jenkins approach. *Communications in Mathematical Biology and Neuroscience*, 2022, vol. 2022, art. no. 17.
11. Gecili E., Ziady A., Szczesniak R. D. Forecasting COVID-19 confirmed cases, deaths and recoveries: revisiting established time series modeling through novel applications for the USA and Italy. *PLoS One*, 2021, vol. 16, no. 1, art. no. e0244173.
12. Singh S., Parmar K. S., Makkhan S. J. S., Kaur J., Peshoria S., Kumar J. Study of ARIMA and least square support vector machine (LS-SVM) models for the prediction of SARS-CoV-2 confirmed cases in the most affected countries. *Chaos, Solitons and Fractals*, 2020, vol. 139, art. no. 110086.
13. Aditya S. C. B., Darmawan W., Nadia B. U., Hanafiah N. Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET. *Procedia Computer Science*, 2021, vol. 179, pp. 524–532.
14. Duong N., Phuong Th. L., Nhu Q. D., Binh L., Ai L. C., Hong D. P. Predicting the pandemic COVID-19 using ARIMA model. *VNU Journal of Science: Mathematics – Physics*, 2020, vol. 36, no. 4, art. no. 4492.
15. Claris S., Peter N. ARIMA model in predicting of COVID-19 epidemic for the Southern Africa region. *African Journal of Infectious Diseases*, 2022, vol. 17, no. 1, pp. 1–9.

16. Zaharov V. V. Printcip dinamicheskogo balansa demograficheskogo processa i predely rosta zemli [Dynamic balance principle of the demographic process and the limits of earth population growth]. *Papers of Russian Academy of Sciences*, 2023, vol. 15, pp. 108–114. <https://doi.org/10.31857/S2686954323600301> (In Russian)

17. Kermack W. O., McKendrick A. G. A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society A*, 1927, vol. 115, pp. 700–721.

18. Anderson R. M., May R. M. *Infectious diseases of humans: Dynamics and control*. Oxford, Oxford University Press, 1991, 757 p.

19. United Nations, Department of Economic and Social Affairs, Population Division. *World Population Prospects 2022: Methodology of the United Nations population estimates and projections*. New York, United Nations Publ., 2022, 64 p.

20. Zaharov V. V., Ndiaye S. M. Prognozirovanie chislennosti naseleniya i dinamicheskie igry protiv prirody [Population growth forecasting and dynamic games against nature]. *Mathematical Game Theory and its Applications*, 2024, vol. 16, no. 1, pp. 17–37. (In Russian)

Received: January 9, 2024.

Accepted: March 12, 2024.

A u t h o r s ' i n f o r m a t i o n :

Yulia E. Balykina — PhD in Physics and Mathematics, Associate Professor; j.balykina@spbu.ru

Victor V. Zaharov — Dr. Sci. in Physics and Mathematics, Professor; v.zaharov@sbpu.ru