Информатика и системы управления, 2021, №3(69) Моделирование систем



УДК 51-76

© 2021 г. **А.А. Яковлев**,

А.И. Абакумов, д-р физ.-мат. наук (Институт автоматики и процессов управления ДВО РАН, Владивосток)

МАТРИЧНАЯ МОДЕЛЬ ДЕМОГРАФО-ЭПИДЕМИЧЕСКОЙ ДИНАМИКИ НАСЕЛЕНИЯ РОССИИ

Анализируются данные и основные демографические характеристики динамики численности населения России по данным Росстата за 1995 - 2018 гг. Построена матрица G перехода вектора возрастных групп от года к следующему году, выяснены ее характеристики. Получены модельные оценки вероятности заболеть туберкулезом и вероятности выздороветь.

Ключевые слова: матричная модель, демография, возрастная структура, туберкулез.

DOI: 10.22250/isu.2021.69.53-66

Введение

Модели динамики популяций с дискретной возрастной структурой и дискретным временем исторически связаны с именем П. Лесли, изучавшего простейшие варианты подобных моделей, и Л. Лефковича, изучавшего стадийные структуры [1 – 4]. Классические матричные модели динамики популяций одного биологического вида опирались на весьма жесткие постулаты о «расписании» основных событий в жизненном цикле организмов (рождение, созревание, воспроизводство, смертность) и о постоянстве соответствующих демографических параметров во времени [5]. Математика классических моделей Лесли изучена досконально. В частности, определено понятие репродуктивного потенциала популяции, т.е. такой величины, которая легко вычисляется по демографическим параметрам и диагносцирует рост или убывание популяции в целом. Известны формулы для предельного распределения численностей возрастных классов, а также критерии, устанавливающие наличие или отсутствие циклов в асимптотическом поведении тра-

екторий модели [6]. Однако на практике приходится сталкиваться с более сложными жизненными циклами организмов – с ускорениями в развитии и возвратами к пройденным стадиям – и даже с двойной структурой популяции [3, 7].

В настоящей работе рассматривается матричная модель динамики населения России, которая разработана на основании матриц Лесли. Нами была предпринята попытка проанализировать демографическую ситуацию с учетом такой социально значимой болезни, как туберкулез. Динамика заболеваемости и течения болезни зависят от условий жизни [8 – 10], но может быть исследована «внешним» описанием с применением матричной модели во взаимосвязи с демографической ситуацией. Именно такой подход применен в этой статье.

Динамика половозрастной структуры населения

Все данные, используемые в работе, взяты с официального сайта Федеральной службы государственной статистики [11]. Динамика численности населения представлена с 1995 г. по 2018 г. Все население мужского и женского пола разделено на пятнадцать пятилетних групп. Исключение составляет последняя группа. В ней собраны люди от 70 лет и старше. Численность каждой группы считается в тысячах человек.

Возраст деторождения у женщин начинается в 15 и заканчивается в 49 лет. Женщины фертильного возраста также разбиты на пятилетние группы. Нам известен коэффициент рождаемости каждой группы по годам наблюдений, который рассчитывается как количество родившихся младенцев в среднем на 1000 женщин. Для наших расчетов мы посчитали среднее коэффициента рождаемости по каждой группе и полученное значение использовали в дальнейшем. График коэффициента рождаемости представлен на рис. 1.

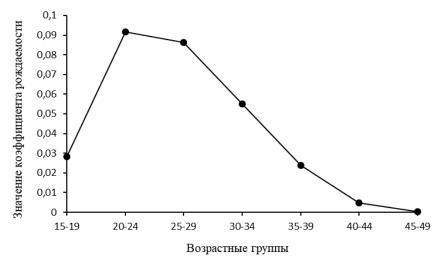


Рис. 1. График коэффициента рождаемости для женщин фертильного возраста.

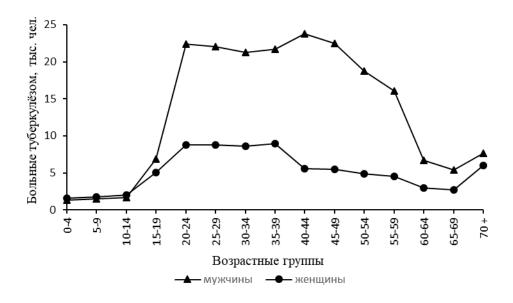
Как видно из графика, пик детородности приходится на возраст от 20 до 24 лет.

Коэффициент смертности для мужчин и женщин по каждой возрастной группе рассчитывается как количество умерших на 1000 человек соответствующей возрастной группы. В нашем исследовании мы использовали средние значения по возрастным группам. Графики коэффициента смертности для мужчин и женщин представлены на рис. 2.



Рис. 2. Графики коэффициента смертности для мужчин и женщин.

Количество мужчин и женщин, больных туберкулезом, разбитых на пятилетние группы в тысячах человек, представлены на рис. 3. Пик заболеваемости приходится на возраст от 20 до 44 лет.



Puc. 3. Графики больных туберкулезом мужчин и женщин по возрастным группам.

Гендерные различия в заболеваемости у мужчин и женщин обусловливаются уровнем иммунной защиты, поведением и социальными ролями [12].

Демографическая модель

В матричных моделях популяции с возрастной структурой единицы измерения времени и возраста совпадают с временным шагом Δt . Поэтому $\Delta t = 1$ и в следующий момент времени (t+1) все люди, выжившие за этот промежуток времени, становятся на единицу старше [13]. Будем рассматривать дискретное время в годах. Тогда за индекс t обозначим текущий такт времени.

С другой стороны, все люди, достигшие максимального возраста n, умирают к следующему моменту времени. Тем самым можно определить размерность вектора возрастной структуры человеческой популяции. Как уже говорилось, по данным Федеральной службы государственной статистики [11] население классифицировано на n=15 пятилетних возрастных групп.

У мужчин и у женщин рассматривается n синхронизированных по числу лет возрастных групп. Вектор численности мужчин в каждый такт времени обозначим как $y_t \in R^n$, вектор численности женщин как $z_t \in R^n$. В векторах $y = \begin{bmatrix} y_1, y_2, ..., y_n \end{bmatrix}^T$ и $z = \begin{bmatrix} z_1, z_2, ..., z_n \end{bmatrix}^T$ компоненты y_i и z_i показывают номер возрастной группы в тысячах человек. Для описания численности мужчин и женщин в каждой возрастной группе образуем вектор $x_t = \begin{pmatrix} y_t \\ z_t \end{pmatrix} \in R^{2n}$.

В таком случае получаем, что $x_{t,i}$ — это численность особей из возрастной группы i(i=1,2,...,n) в момент t.

Обозначим \overline{d}_i среднее значение коэффициента смертности по каждой возрастной группе i в каждый такт времени t.

Далее опишем рождаемость. Пусть $b_1, b_2, ..., b_n$ — это возрастноспецифические коэффициенты рождаемости. По данным Росстата возраст фертильности женщин начинается в 15 и заканчивается в 49 лет. Соответственно получается 7 пятилетних возрастных периода. Мы примем за $b_1, b_2, ..., b_n$ средние значения коэффициента рождаемости по каждой возрастной группе. Обозначим буквой a (adult) порядковый номер самой младшей возрастной фертильной группы, а буквой f (final) — самой старшей. В таком случае формально получаем, что коэффициенты b_i положительны только в диапазоне

возрастных групп от a до f, а вне этого диапазона нулевые. Таким образом получаем, что суммарная численность пополнения от родителей всех репродуктивных возрастов есть

$$x_{t+1} = b_a x_{t,a} + b_{a+1} x_{t,a+1} + \dots + b_f x_{t,f}.$$
(1)

Коэффициенты рождаемости удобно организовать в отдельную квадратную матрицу \boldsymbol{F} (fertility) размерностью $(n \times n)$, у которой только первая строчка имеет ненулевые элементы среди $b_1, b_2, ..., b_n$, а на всех остальных местах стоят нули. Тогда матрица \boldsymbol{F} принимает вид:

$$F = \begin{vmatrix} b_1 & b_2 & \cdots & b_{n-1} & b_n \\ 0 & 0 & \cdots & \cdots & 0 \\ \vdots & 0 & \ddots & 0 & \vdots \\ \vdots & \cdots & 0 & \ddots & 0 \\ 0 & \cdots & \cdots & 0 & 0 \end{vmatrix}.$$
 (2)

Как уже говорилось, возраст деторождения женщины начинается с 15 и заканчивается в 49 лет, т.е. получаем семь пятилетних периодов. Принимая это во внимание, получаем, что коэффициенты $b_5 - b_{11}$ будут положительные, остальные будут равны нулю.

Обозначим за p_i (i=1,2,...,n) коэффициент остающихся особей в i-й возрастной группе. В нашем случае все группы по 5 лет. В следующую группу переходит 1/5 от количества группы и остается 4/5. Следовательно, во всех группах, кроме последней, $p_i=0.8$. В последней группе люди никуда не переходят, поэтому в ней коэффициент $p_n=1-d_n$.

Составим квадратную матрицу $P(n \times n)$, в которой элементы p_i поставим по главной диагонали, а на всех остальных местах стоят нули.

$$P = \begin{vmatrix} p_1 & 0 & \cdots & \cdots & 0 \\ 0 & p_2 & 0 & \cdots & \vdots \\ \vdots & 0 & \ddots & 0 & \vdots \\ \vdots & \cdots & 0 & p_{n-1} & 0 \\ 0 & \cdots & \cdots & 0 & p_n \end{vmatrix}.$$
 (3)

Обозначим как \mathbf{s}_i (коэффициент выживания) долю особей из возрастной группы i, выживших за один такт времени и перешедших в следующий возрастной класс. Коэффициент выживаемости s_i рассчитывается следующим образом:

$$s_i = 1 - d_i - p_i, \tag{4}$$

где s_i – доля выживших в i-й возрастной группе; d_i – смертность в i-й воз-

растной группе.

Составим матрицу переходов $S(n \times n)$. Она имеет специальное строение. Коэффициенты выживаемости s_i располагаются под главной диагональю, а остальные элементы равны нулю.

$$S = \begin{vmatrix} 0 & \cdots & \cdots & 0 \\ s_1 & 0 & \cdots & \cdots & \vdots \\ 0 & s_2 & \ddots & \cdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & s_{n-1} & 0 \end{vmatrix}.$$
 (5)

Теперь перейдем к динамической модели населения с учетом половой и возрастной структуры. Сначала составим модифицированные матрицы, описывающие динамику возрастной структуры мужчин, женщин и матрицу рождаемости.

Модифицированная матрица возрастной структуры мужчин $M\left(n\times n\right)$ представляется в виде

$$M = P + S, (6)$$

где матрицы P и S описаны соответственно в (3) и (5).

Матрица возрастной структуры женщин $W(n \times n)$ составляется следующим образом:

$$W = P + S + \delta F, \tag{7}$$

где матрицы P, S, F описаны соответственно в (3), (5), (2), а δ – это доля девочек среди новорожденных.

Матрица рождаемости для мужской части населения имеет вид

$$B = (1 - \delta)F,\tag{8}$$

где $(1-\delta)$ – это доля мальчиков среди новорожденных.

Сконструируем матрицу перехода к новому такту времени G следующим образом:

$$G = \begin{vmatrix} M & B \\ 0 & W \end{vmatrix}, \tag{9}$$

где матрицы M, B, W описаны соответственно в (6), (8) и (7).

Скажем несколько слов о матрице G. Эта матрица является неотрицательной и разложимой, так как у нее левая нижняя четверть является нулевой. Здесь уместно будет упомянуть о разложимых и неразложимых неотрицательных матрицах.

Определение и свойства неразложимых неотрицательных матриц даны

в теореме Перрона — Фробениуса, описанной в фундаментальной монографии по теории матриц Ф.Р. Гантмахера «Теория матриц». В ней говорится, что неразложимая неотрицательная матрица всегда имеет положительное характеристическое число r, которое является простым корнем характеристического уравнения. Модули других характеристических чисел не превосходят это число r. Максимальному характеристическому числу r соответствует собственный вектор z с положительными координатами [14].

Но как уже было сказано выше, матрица G является неотрицательной разложимой. Для таких матриц максимальное характеристическое число r и соответствующий ему собственный вектор y могут быть неотрицательными [14, 15].

Задача, как выяснить, растет или убывает популяция, сводится к отысканию максимального собственного числа матрицы и сравнению его с единицей [16]. Назовем максимальное характеристическое число доминантным числом r и рассчитаем его для нашей матрицы G, написав скрипт в прикладной программе MATLAB R2021a. Если r > 1, то популяция экспоненциально растет, если r < 1, то экспоненциально убывает, если r = 1, то стремится к предельному распределению. После выполнения скрипта получаем, что доминантное число матрицы G равно 0,9938, что свидетельствует о некотором снижении общей численности населения.

Основное динамическое соотношение тогда представляется в виде

$$x_{t+1}^k = Gx_t^k, (10)$$

где надстрочный индекс k обозначает, что используются данные, взятые с сайта Росстата.

Равенство (10) есть матричная модель для наблюдаемой динамики численности населения.

Для наглядного представления различий наблюдаемой и модельной динамики вычислим поправочный коэффициент Θ_t из соотношения

$$x_{t+1}^k \approx \Theta_t G x_t^k \,, \tag{11}$$

при условии минимизации разности векторов в соотношении (11) для наблюдаемой численности населения

$$\left\| x_{t+1}^k - \Theta_t G x_t^k \right\|_e^2 \to \min, \tag{12}$$

где подстрочный индекс е обозначает евклидову норму.

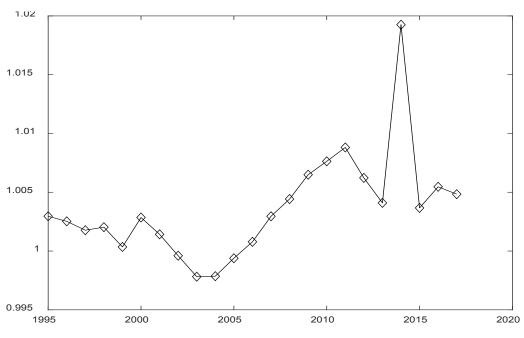
Продифференцировав (12) по Θ_t и приравняв полученное выражение к нулю, получаем:

$$-2\left(x_{t+1}^k - \Theta G x_t^k\right) G x_t^k = 0.$$

Отсюда имеем

$$\Theta_t = \frac{x_{t+1}^k G x_t^k}{\left\| G x_t^k \right\|_e^2}.$$
(13)

Рассчитывая Θ_t для (t-1) лет, получим зависимость, изображенную на рис. 4.



Puc. 4. Зависимость $\Theta(t)$.

На графике видно, что значение Θ меняется волнообразно. Присутствует плавный «провал», который приходится на 2003-2004 гг., и резкий выброс в 2014 г.

Получим модельные значения численности населения согласно уравнению (10). К каждому полученному вектору x применим «октаэдрическую» норму [14]:

$$||x|| = \sum_{i=1}^{n} |x_i|. \tag{14}$$

В результате получаем модельные значения численности населения РФ в период с 1995 г. по 2018 г. Графики наблюдаемой численности населения и модельной представлены на рис. 5.

Модельные графики считались по следующим формулам:

1) модельный тип 1

$$x_1 = x_{1995}^k, x_{t+1} = Gx_t,$$

где x_{1995}^k — вектор наблюдаемых значений численности мужчин и женщин в 1995 г.; x_t — вектор модельных значений предшествующего года; x_{t+1} — вектор модельных значений следующего года;

2) модельный тип 2

$$x_1 = x_{1995}^k, x_{t+1} = Gx_t^k,$$

где x_t^k — вектор наблюдаемых значений предшествующего года;

3) модельный, тип 3

$$x_1 = x_{1995}^k, x_{t+1} = \Theta_t G x_t^k,$$

где Θ_t – коэффициент Θ , соответствующий предыдущему году.

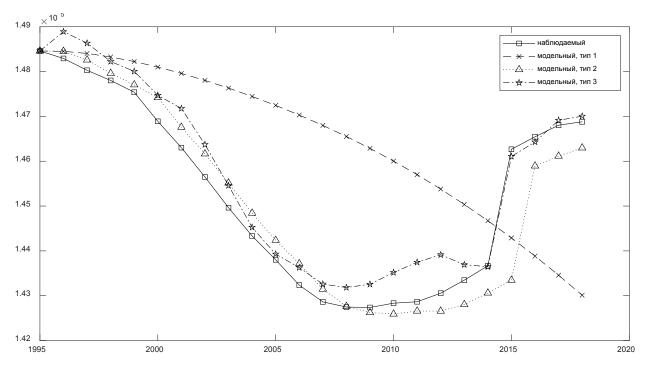


Рис. 5. Графики наблюдаемой и модельной численности населения РФ.

Из графика наблюдаемой численности населения видно, что с 1995 г. по 2007 г. численность населения РФ монотонно убывала, а в 2008-2009 гг. произошла стабилизация. С 2010 г. начинается незначительный рост численности населения. В 2014-2015 гг. произошло резкое увеличение населения России. Похожий скачок в 2014 г. мы наблюдали на рис. 4 и связываем этот резкий рост населения с вхождением Республики Крым в состав России.

Модельные графики тип 2 и тип 3 в целом повторяют наблюдаемый. Форма модельного графика тип 1 подтверждает вышесказанное утверждение, что при доминантном числе матрицы меньше 1 численность популяции убывает.

Матрица перехода G в уравнении (10) является примитивной [14]. Определим относительную возрастную структуру вектора x_t как вектора долей возрастных групп в общей численности населения (т.е. вектор нормируется по норме (14)). Для матричных моделей типа (10) известно асимптотическое свойство, состоящее в том, что при $t \to \infty$ относительная структура

вектора x_t стремится к относительной структуре собственного вектора для доминантного собственного числа матрицы G. Если матричная модель (10) соответствует реальной динамике населения России, то подобное свойство можно предположить выполняющимся для реальной динамики. На рис. 6, 7 такое сравнение сделано для укрупненных возрастных групп промежутков возраста 0-14 лет, 15-34 года, 35-59 лет и свыше 60 лет для мужчин и женщин.

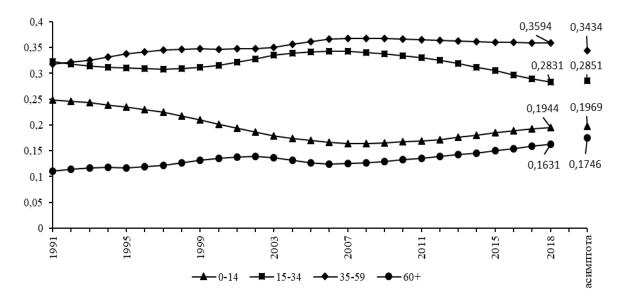


Рис. 6. Графики наблюдаемой динамики численности укрупненных возрастных групп мужчин по годам и их асимптотическое значение.

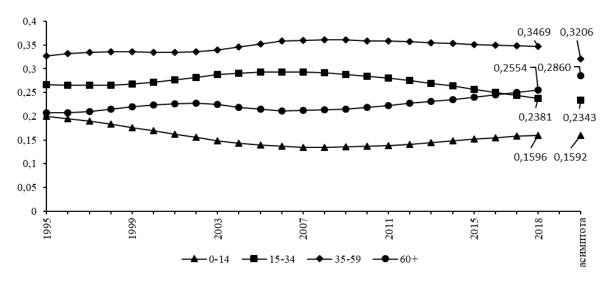


Рис. 7. Графики наблюдаемой динамики численности укрупненных возрастных групп женщин по годам и его асимптотическое значение.

В нашем случае асимптотическое свойство выполняется достаточно точно, несмотря на ограниченный промежуток лет наблюдений. Локально наблюдаемая динамика общей численности населения может значимо отли-

чаться от модельной (рис. 5), но относительная структура населения близка к предписываемой собственным вектором матрицы G (рис. 6, 7).

Динамика больных туберкулезом

Данные по заболеваемости туберкулезом (ТБ) среди мужчин и женщин распределены по возрастным группам аналогично демографии. Составим векторы численности здоровых и больных мужчин и женщин в каждый такт времени t аналогично действиям, описанным при анализе демографической ситуации.

Введем обозначения. Пусть $x_{t,1} = \begin{pmatrix} y_{t,1} \\ z_{t,1} \end{pmatrix} \in R^{2n}$ — вектор здоровых особей,

а
$$x_{t,2} = \begin{pmatrix} y_{t,2} \\ z_2 \end{pmatrix} \in R^{2n}$$
 — вектор больных особей в t -м году. Численность населе-

ния каждой возрастной группы состоит из здоровых и больных.

$$x_t = x_{t,1} + x_{t,2}$$
.

Для получения численности здорового населения вычтем вектор больных из вектора общей численности по каждой возрастной группе.

$$x_{t,1} = x_t - x_{t,2} \,. \tag{15}$$

Применяя полученную матрицу перехода к новому такту времени G, получим:

$$\begin{cases} x_{t+1,1} = Gx_{t+\frac{1}{2},1}, \\ x_{t+1,2} = \gamma Gx_{t+\frac{1}{2},2}. \end{cases}$$
 (16)

Промежуточные значения $x_{t+\frac{1}{2},1}$ и $x_{t+\frac{1}{2},2}$ получим из отношений:

$$\begin{cases} x_{t+\frac{1}{2},1} = (1-\alpha)x_{t,1} + \beta x_{t,2}, \\ x_{t+\frac{1}{2},2} = (1-\beta)x_{t,2} + \alpha x_{t,1}, \end{cases}$$
(17)

где α — доля заболевающих среди здоровых; β — доля выздоравливающих; γ — коэффициент жизнеспособности больных по сравнению со здоровыми.

Подставив (17) в (16), получим

$$\begin{cases} x_{t+1,1} = (1-\alpha)Gx_{t,1} + \beta Gx_{t,2}, \\ x_{t+1,2} = \alpha \gamma Gx_{t,1} + (1-\beta)\gamma Gx_{t,2}. \end{cases}$$
(18)

Для нахождения коэффициентов α , β , γ введем функцию $\Phi(\alpha,\beta,\gamma)$ и

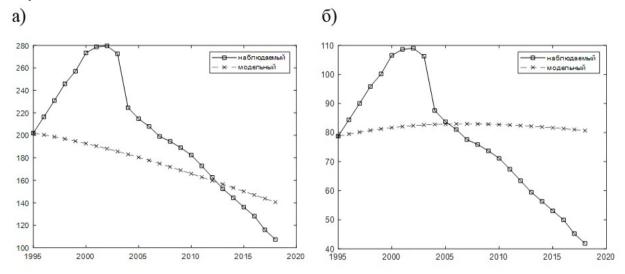
найдем ее экстремум:

$$\Phi(\alpha,\beta,\gamma) = \sum_{t} ||x_t^m - x_t^k||^2 \to \min,$$

где x_t^m — модельные значения численности населения; x_t^k — наблюдаемые значения численности населения.

После выполнения скрипта в программе MatLab получаем следующие значения: $\alpha = 0.0005$, $\beta = 0.03$, $\gamma = 0.9443$. Таким образом, усредненная доля заболевающих среди здоровых в течение года составляет 0.05%. По данным Росстата это значение равно 0.07%, что близко к модельному значению. Усредненная доля выздоравливающих среди больных в течение года составляет 3%. Жизнеспособность больных ТБ составляет 94% по сравнению со здоровыми.

На рис. 8 представлены графики наблюдаемых и модельных значений для мужчин и женщин.



Puc. 8. Графики наблюдаемых и модельных значений: а) мужчины с заболеванием ТБ; б) женщины с заболеванием ТБ.

Заключение

В работе приведена матричная модель динамики численности населения России. Анализируются данные и основные демографические характеристики динамики численности населения России по данным Росстата за 1995 – 2018 гг.

Немонотонность наблюдаемой динамики приводит к локальным расхождениям ее с модельной динамикой. В то же время доминантное собственное число матрицы перехода в модели (10) близко к единице и равно 0.9938. Это на качественном уровне соответствует незначительному спаду численности населения за период наблюдений. Относительная возрастная

структура населения близка к предписанной собственным вектором доминантного собственного числа матрицы перехода. Этот результат говорит о хорошем соответствии построенной матричной модели наблюдаемой динамике численности населения России. Полученная конструкция применена к исследованию динамики заболевания туберкулезом. Коэффициент жизнеспособности больных по отношению к здоровым составляет 94%. Модельная оценка вероятности заболеть туберкулезом в течение года составляет 0.05%, а вероятности выздороветь — 3%. Полученные характеристики согласуются с другими данными Росстата.

ЛИТЕРАТУРА

- 1. *Leslie P.H.* On the use of matrices in certain population mathematics // Biometrika. 1945. 33/3. P. 183-212.
- 2. *Leslie P.H.* Some further notes on the use of matrices in population mathematics // Biometrika. $-1948. N_{\odot} 53/3-4. P. 213-245.$
- 3. *Логофет Д.О., Клочкова И.Н.* Математика модели Лефковича: репродуктивный потенциал и асимптотические циклы // Матем. моделирование. 2002. Т. 14, № 10. С. 116—126.
- 4. *Dmitrii O. Logofet*. Stronger-than-Lyapunov notions of matrix stability, or how "flowers" help solve problems in mathematical ecology // Linear Algebra and its Applications. 2005. № 398 P. 75–100.
- 5. *Белова И.Н.* Матричные модели динамики популяций с дискретными структурами: Автореф. дис... канд. физ.-мат. наук. М.: Институт физики атмосферы им. А.М. Обухова РАН, 2009.
- 6. *Абакумов А.И., Пидюра Т.А.* Магистральные характеристики матричных моделей экономической динамики // Вестник ТГЭУ. 2010. № 2. С. 96–103.
- 7. *Абакумов А.И., Худзик Т.А.* Асимптотика в матричных моделях динамических систем // Дальневосточный математический журнал. 2003. Т.4, №1. С. 44 51.
- 8. *Яковлев А.А.*, *Абакумов А.И*. Анализ данных о туберкулезе и влиянии на заболевание условий жизни в Приморском крае // Информатика и системы управления. 2019. № 3 (61). С. 42-53. DOI: 10.22250/isu.2019.61.42-53.
- 9. *Ермолицкая М.З.* Перспективы вылечиться от туберкулеза. Анализ данных средствами программы RSTUDIO // Информатика и системы управления. 2020. № 1 (63). C. 50-58. DOI: 10.22250/isu.2020.63.50-58.
- 10. *Ермолицкая М.З.* Использование искусственных нейронных сетей для прогнозирования положительной динамики при лечении больных туберкулезом // Информатика и системы управления. 2020. № 4 (66). С. 139-147. DOI: 10.22250/isu.2020.66.139-147.
- 11. Официальный сайт Федеральной службы государственной статистики РФ. URL: https://rosstat.gov.ru/ (дата обращения 25.03.2021).
- 12. Киселевская-Бабинина В.Я., Санникова Т.Е., Романюха А.А., Каркач А.С. Моделирование влияний гендерных различий на заболеваемость туберкулезом // Математиче-

- ская биология и биоинформатика. 2018. Т. 13, № 2. С. 308-321. DOI: /10.17537/2018.13.308
- 13. *Логофет Д.О., Уланова Н.Г.* Матричные модели в популяционной биологии: учебное пособие. М.: МАКС Пресс, 2017.
- 14. *Гантмахер Ф.Р.* Теория матриц. М.: Наука, 1966.
- 15. *Хитров Г.М.* Об определении разложимой матрицы и ее нормальной формы // Вестник СПбГУ. -2006. Сер. 10, вып. 3. С. 85-91.
- 16. *Клочкова И.Н.* Обобщение теоремы о репродуктивном потенциале для матриц Логофета // Вестник Московского университета. Серия 1 «Матем., мех.». 2004. № 3. С. 45-48.

E-mail:

Яковлев Анатолий Александрович – yakovlev-aa@iacp.dvo.ru; Абакумов Александр Иванович – abakumov@iacp.dvo.ru.