

УДК 33:51-77

С. С. Михайлова

ФГБОУ ВО «Финансовый университет при Правительстве Российской Федерации»,
Москва, e-mail: ssmihajlova@fa.ru

И. Р. Шулунова

ФГБОУ ВО «Восточно-Сибирский государственный университет технологий
и управления», Улан-Удэ, e-mail: rubinovna@mail.ru

М. В. Балданова

ФГБОУ ВО «Восточно-Сибирский государственный университет технологий
и управления», Улан-Удэ, e-mail: baldanova@mail.ru

РАЗРАБОТКА МОДЕЛЕЙ, АЛГОРИТМОВ И ПРОГРАММНЫХ СРЕДСТВ ПРОГНОЗИРОВАНИЯ ЧИСЛЕННОСТИ НАСЕЛЕНИЯ РЕГИОНА

Ключевые слова: прогнозирование, генетический алгоритм, численность популяции, математические модели.

Работа посвящена прогнозированию численности населения с помощью генетического алгоритма и построения математической модели Лотки-Вольтерры. Особенностью исследования является разработка методического подхода к региональному прогнозированию численности населения, который представляет собой альтернативное использование генетического алгоритма и математической модели Лотки-Вольтерры. Данный подход позволяет, во-первых, повысить точность прогноза, во-вторых расширить возможности корректировки информационных моделей региональных демографических данных. Реализованы варианты генетического алгоритма: с удалением лишних особей и без удаления, а также построены системы двух дифференциальных уравнений где, в качестве переменных попарно представлены показатели рождаемости, смертности и миграции. Разработан прогноз значений показателей рождаемости, смертности и миграции для Республики Бурятия и ее районов по трехлетним и пятилетним периодам. Для оценки точности алгоритмов и прогнозных значений предложено использовать величину разности интеграла от прогнозируемой и от реальной функции, что позволит определить наличие резких скачков в расчетных данных.

S. S. Mikhaylova

Financial University under the Government of the Russian Federation, Moscow,
e-mail: ssmihajlova@fa.ru

I. R. Shulunova

East Siberia State University of Technology and Management, Ulan-Ude,
e-mail: rubinovna@mail.ru

M. V. Baldanova

East Siberia State University of Technology and Management, Ulan-Ude,
e-mail: baldanova@mail.ru

DEVELOPMENT OF MODELS, ALGORITHMS, AND SOFTWARE FOR FORECASTING THE POPULATION OF THE REGION

Keywords: forecasting, genetic algorithm, population size, mathematical models.

The work is devoted to population forecasting using a genetic algorithm and the construction of a Lotka-Volterra mathematical model. A feature of the study is the development of a methodological approach to regional population forecasting, which is an alternative use of the genetic algorithm and the Lotka-Volterra mathematical model. This approach allows, firstly, to improve the accuracy of the forecast, and secondly, to expand the possibilities for correcting information models of regional demographic data. Variants of the genetic algorithm have been implemented: with the removal of extra individuals and without removal, and systems of two differential equations have been constructed where, as variables, fertility, mortality and migration indicators are presented in pairs. A forecast has been developed for the values of birth, death and migration indicators for the Republic of Buryatia and its regions for three and five year periods. To assess the accuracy of algorithms and predictive values, it is proposed to use the value of the difference between the integral from the predicted and from the real function, which will make it possible to determine the presence of sharp jumps in the calculated data.

Введение

Человечество представляет собой сложную систему, которая постоянно развивается и изменяется. Динамика численности населения каждого региона обладает своей спецификой в силу влияния различных факторов: уровень развития региона, благосостояние населения, болезни и эпидемии, стихийные бедствия. Различные виды математических моделей позволяют выявлять закономерности развития тех или иных демографических процессов, а также служат для анализа состава населения в целом. Развитие теории прогнозирования в последние десятилетия привело к разработке многих методов, процедур и приемов построения прогнозов, различных по своим параметрам [1-3, 5-8, 10, 11, 13-15, 17-19].

К основным моделям численности населения, используемым в демографии, относятся модели стабильного и стационарного населения, экспоненциального и гиперболического роста. Экспоненциальная модель не учитывает факторы окружающей среды и поэтому подходит для малого периода времени прогнозирования. Логистическая модель показывает, что рост численности населения уменьшается по мере того, как население достигает пропускной способности или предельной точки окружающей среды. Широкое распространение имеет логистическая модель с применением уравнения ошибки [2, 6]. Подобная модель также была открыта американским экологом Р.Перлем, в которой отображена зависимость темпов роста от численности населения в конкретный момент и от наличия «неиспользованных резервов жизнеобеспечения населения», например, в виде невозделанных земель. В 1945 г. Лесли были предложены матричные модели, в основе которых лежит понятие популяционной системы, описывающей основные взаимосвязи между представителями той или иной группы особей, проживающие на одной территории, имеющие единое происхождение и свободно скрещивающиеся. Матричный способ затрагивает возрастную структуру наблюдаемого объекта. Носовой М.Г. разработана и исследована математическая модель процесса изменения демографической ситуации в виде автономной немарковской системы массового обслуживания с неограниченным числом приборов и РН-распределением времени обслуживания заявок [11].

В работе «A Mathematical and Statistical Approach for Predicting the Population

Growth» для прогнозирования численности населения применены модель Мальтуса (простой экспоненциальный рост), логистическая модель и уравнение Ферхюльста, проведено сравнение результатов расчетов трех моделей [1]. Tsoularis A. показал, что наиболее успешные прогностические модели основаны на уравнении логистического роста Ферхюльста, модель простого экспоненциального роста может быть использована для начального периода [13].

В XX веке работы Маккендрика внесли важный вклад в современную экологию и демографию, так с помощью модели эпидемии возможно предсказать количество и распределение случаев инфекционного заболевания [7]. Модели Malthusian, Logistic, Logistic controlled by resources, Biotic interactions, Lotka-Volterra, Leslie, Leslie controlled by resources, Ricker, McKendrick проанализированы Rui Dilao с учетом возможности описания вымирания и взрыва популяций, ненулевых состояний равновесия, колебаний, хаоса и квазипериодичности [14].

В работе Иванюка В.А. и Андрейчикова А.В. для полноценного прогнозирования и анализа предложено использовать комбинированную систему нейронная сеть – статистические методы – генетический алгоритм. Генетический алгоритм в данной системе настроен на уборку мусора в данных и выявление наиболее жизнеспособных особей среди данных. Оптимизация и ускорение вычислений в системе способствуют получению наиболее достоверных и точных прогнозов [5].

Основными показателями демографической статистики являются показатели численности и состава населения, количество родившихся, умерших, браков, разводов, миграционного движения. Перечисленные показатели позволяют формировать производные и расчетные показатели, которые способствуют в полном объеме отразить структуру и движение населения, социально-демографические процессы. При анализе социально-экономической ситуации, сложившейся в рассматриваемой территории, численность населения выступает исходным базисным показателем.

Актуальность работы обусловлена необходимостью формирования целостной структуры социально-демографических групп, которые в свою очередь служат основой для формирования социально-экономической политики региона и государства. Для развития национальной экономики, экономического

планирования территории и анализа перспектив развития необходима разработка инструментария, позволяющего прогнозировать численность населения региона.

В данном исследовании разработаны модели прогноза численности населения на основе модели Лотки-Вольерры и генетического алгоритма, программно реализованы алгоритмы расчета построенных моделей, проведена оценка точности прогнозирования и получен краткосрочный прогноз численности населения Республики Бурятия.

В качестве исходных данных для прогнозирования численности населения используются показатели рождаемости, смертности и миграции по региону за длительный период. Исследование состоит из двух направлений (рис. 1).

Одним из известных примеров описания динамики взаимодействующих популяций являются уравнения Лотки-Вольтерры, модель взаимодействия двух видов типа «хищник-жертва» (1):

$$\begin{cases} \frac{dx}{dt} = (\alpha - \beta y)x \\ \frac{dy}{dt} = (\delta x - \gamma)y \end{cases} \quad (1)$$

где x – число жертв; y – число хищников; t – время; $\alpha, \beta, \gamma, \delta$ – коэффициенты взаимодействия между видами.

Система уравнений Лотки – Вольтерры находит применение не только в природе,

но также часто используется при моделировании экономических систем [17]. Она применима для описания различных процессов в биологии, экологии, медицине, в социальных исследованиях и других науках. Варианты этой модели и сходных с ней применительно к анализу ряда природных и социальных явлений исследованы Д.И. Трубецковым. Так, например, в математической модели взаимодействия загрязнений с окружающей средой в качестве жертвы выступает природа, а в качестве хищника загрязнение. В модели классовой борьбы хищник – земельная рента, жертва – площадь землепользования [18].

В данной работе предложено рассматривать в качестве видов «хищник-жертва» показатели рождаемости, смертности и миграции, так эти показатели попарно являются переменными системы типа (1). Таким образом, построены модели (рис. 2): (differential equations system, DES) Рождаемость-Смертность DES Birth rate-Mortality (DES Br-Mo) (2), Рождаемость-Миграция DES Birth rate-Migration (DES Br-Mi) (3) и Смертность-Миграция DES Mortality-Migration (DES Mo-Mi) (4). При построении моделей учтена взаимосвязь переменных в каждой паре показателей. Например, в модели Рождаемость-Миграция в первом уравнении коэффициенты при переменных x (рождаемость) и переменной y (миграция) положительные, так как положительный миграционный прирост увеличивает рождаемость.

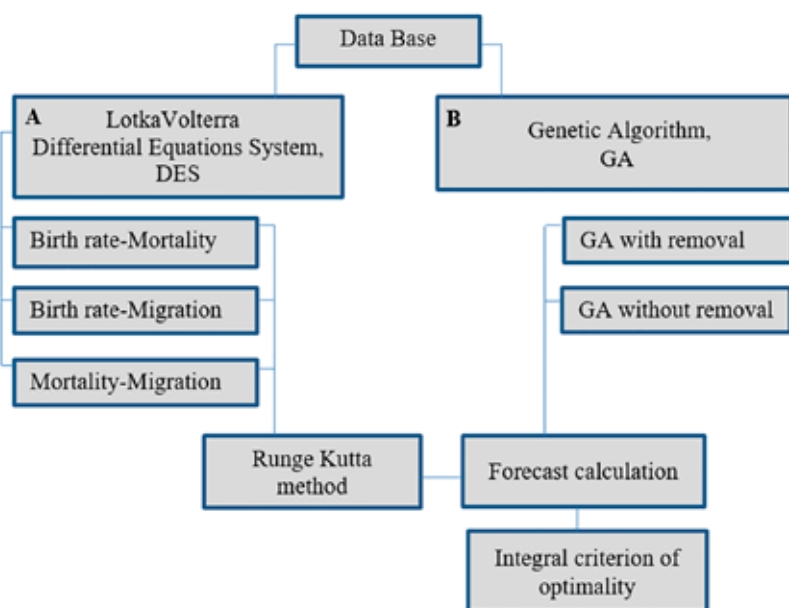


Рис. 1. Схема исследования

$$\begin{cases} \frac{dx}{dt} = \alpha x - \beta y \\ \frac{dy}{dt} = \delta x + \gamma y \end{cases} \quad (2)$$

$$\begin{cases} \frac{dx}{dt} = \alpha x + \beta y \\ \frac{dy}{dt} = \delta x + \gamma y \end{cases} \quad (3)$$

$$\begin{cases} \frac{dx}{dt} = \alpha x - \beta y \\ \frac{dy}{dt} = \delta x - \gamma y \end{cases} \quad (4)$$

Рис. 2. Модели DES рождаемости, смертности и коэффициента миграции

Для решения систем (2-4) использован метод Рунге-Кутты четвертого порядка [12].

В настоящее время для решения задач прогнозирования применяют методы машинного обучения, и одним из его методов, является нейросетевое моделирование. Для обучения и совершенствования нейронных сетей используется генетический алгоритм [3].

Генетический алгоритм (GA) основывается на идее эволюции с помощью естественного отбора и работает с совокупностью особей – популяцией, где каждая особь представляет возможное решение данной проблемы. GA включает в себя следующие этапы: начальная популяция, селекция (отбор), скрещивание (кроссинговер), мутация, формирование нового поколения. Алгоритм повторяется до тех пор, пока будет получен удовлетворительный результат или не произойдет одно из условий: количество поколений (циклов) достигнет заранее выбранного максимума или исчерпано время на мутацию. Основным блоком в GA является блок генетических операторов, который выполняет преобразование популяции альтернативных решений задач оптимизации и искусственного интеллекта [4, 9].

В качестве оператора кроссинговера выбран оператор на основе деления значения пополам и обмена полученными значениями между родителями. На этом этапе у двух родителей будет один потомок, за счет повтора результативных значений.

Оператор мутации реализован на основе изменения хромосомы случайным выбором числа из интервала [0,1] для каждой особи и отбором для выполнения этой операции тех особей, для которых случайное число оказывается меньшим значению вероятности мутации.

Следует отметить, что вероятность мутации устанавливается весьма малой (обычно $0 \leq p \leq 0.1$) и именно от нее зависит, будет данная особь мутировать или нет. Это следует из аналогии с миром живых организмов, где мутации происходят чрезвычайно редко.

Для оценки погрешностей прогнозируемых показателей рассчитаны прогнозы для временных периодов, на которых значения показателей известны. Для прогнозирования показателей демографических процессов исследован GA алгоритм с применением метода удаления лишних особей и без удаления. Входными данными выступали данные за предшествующие 3 или 5 лет прогноза.

Для анализа точности алгоритмов рассчитана площадь под кривой прогнозной функции, вычисленная по методу трапеций для определенного интеграла по каждому показателю [12]. В качестве критерия оптимальности принято отклонение площади прогнозной функции от площади функции реальных значений.

Материалы и методы исследования

Генетический алгоритм и решение модели Лотки Вольтерры методом Рунге-Кутты реализованы на языке C++ в среде разработки Visual Studio 2019. Проведены тестовые расчеты показателей рождаемости, смертности и миграции по данным Республики Бурятия за период 2000-2018 гг. [16]. Расчеты проведены с применением нескольких моделей: Генетический алгоритм без удаления лишних особей GA 1, Генетический алгоритм с удалением лишних особей GA 2, Лотки Вольтерры differential equations system (DES) Рождаемость-Смертность DES Birth rate-Mortality (DES Br-Mo), Рождаемость-Миграция DES Birth rate-Migration (DES Br-Mi) и Смертность-Миграция DES Mortality-Migration (DES Mo-Mi). Проведена оценка погрешностей прогнозируемых показателей относительно реальных данных.

Прогнозирование на основе модели Лотки Вольтера (DES) в качестве начальных условий взяты значения показателей за предшествующий год. Результаты расчета

модели DES Birth rate-Mortality за период 2003-2017 гг. Максимальная погрешность составляет 10,95% для показателя рождаемости и 9,17% для показателя смертности (рис. 3).

Результаты расчета модели DES рождаемости, миграции за период 2003-2017 гг. Для показателя рождаемости максимальное значение погрешности составляет 11,35% и 21,75% для показателя миграции (рис. 4).

Результаты расчета модели DES Mortality-Migration за период 2003-2017 гг. Максимальная погрешность для показателя смертности равна 8,18% и 20,72% для показателя миграции (рис. 5).

Так же проведены расчеты Генетическим алгоритмом без удаления лишних особей GA 1. Показатели рождаемости и смертности варьируются в диапазоне от 0,18% до 15,2%. Показатель миграции имеет значительные отклонения от 0,48% до 28,7% (рис. 6).

Результаты расчета Генетическим алгоритмом с удалением лишних особей GA 2. Погрешности показателей рождаемости и смертности варьируются в диапазоне от 1,12% до 17,36%. Погрешность прогнозирования показателя миграции достаточно велика от 16,82% до 43,27% (рис.7).

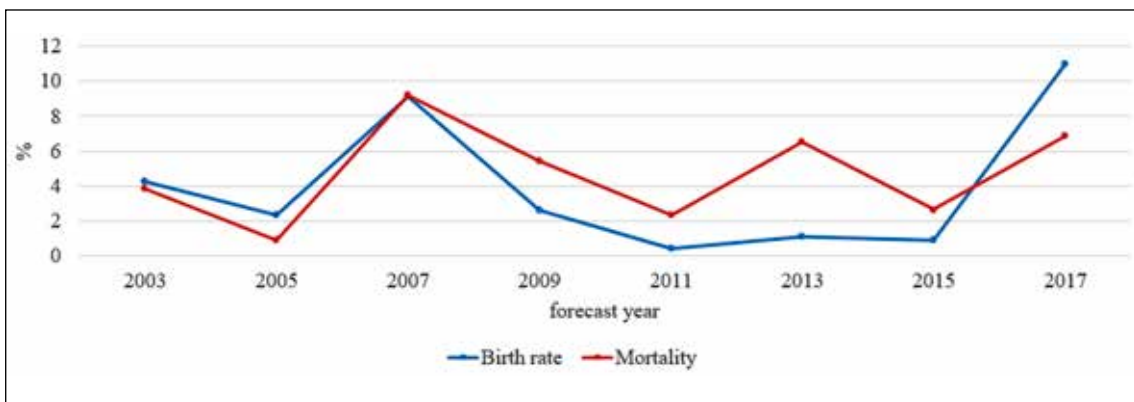


Рис. 3. Значения погрешности DES модели рождаемости, смертности

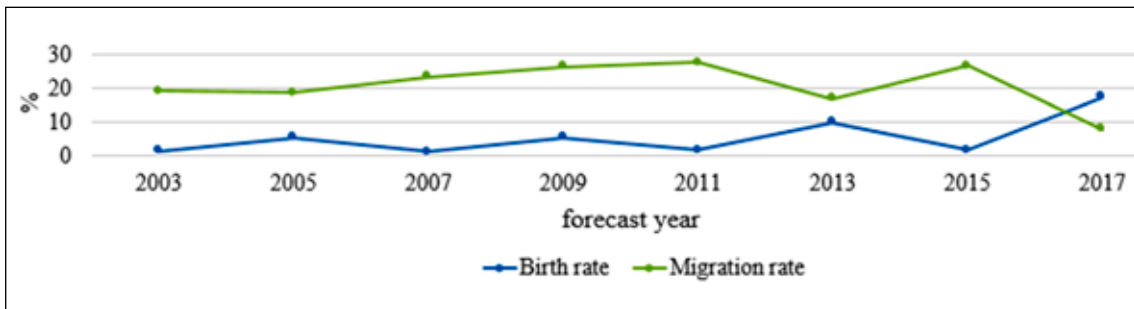


Рис. 4. Значения погрешности DES модели рождаемости, миграции

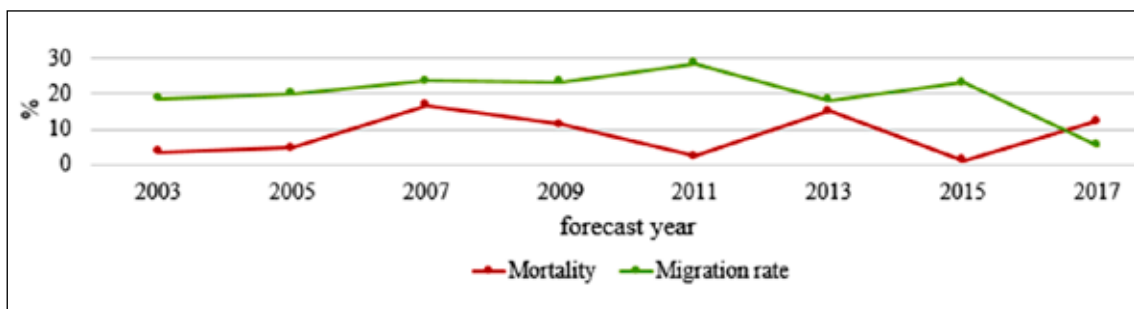


Рис. 5. Значения погрешности DES модели смертности, миграции

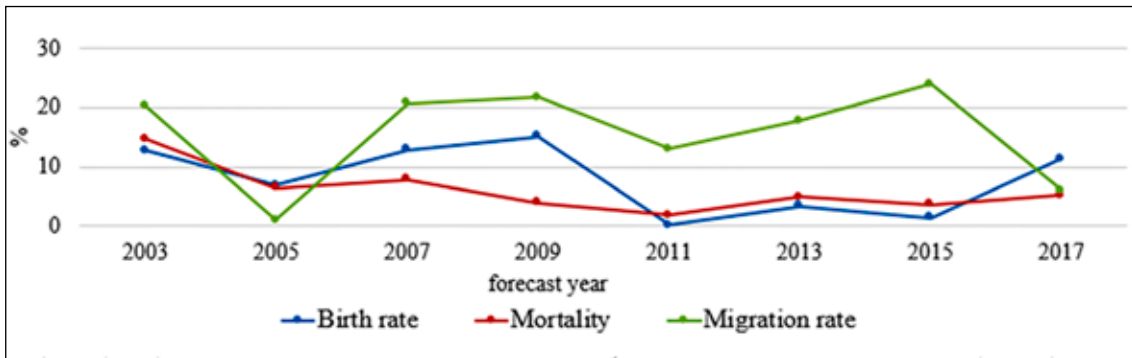


Рис. 6. Значения погрешности GA 1

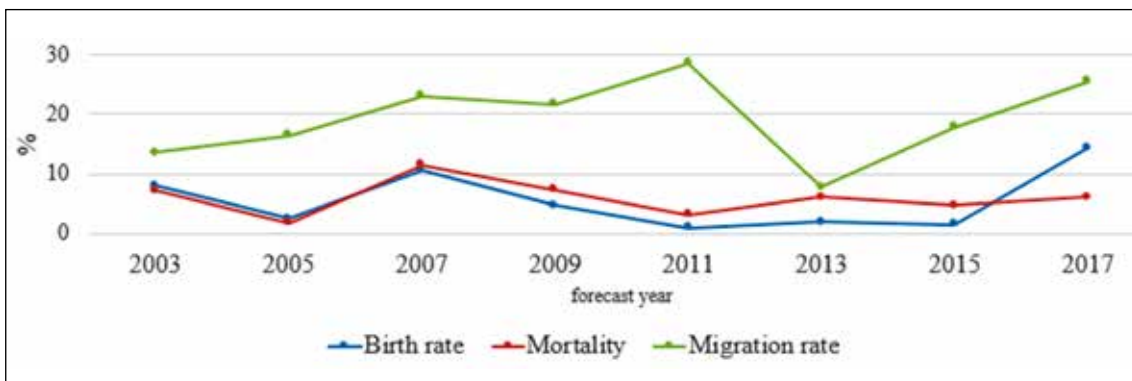


Рис. 7. Значения погрешности GA 2

Результаты прогнозирования позволяют сделать вывод, что применение метода GA 2 дает более точный результат.

Результаты исследования и их обсуждение

Анализ точности алгоритмов

Для анализа точности генетического алгоритма и модели Лотки Вольтерра каждого прогнозируемого показателя методом трапеций вычисляется определенный интеграл [12]. Определяется минимальное отклонение площади прогнозной функции от площади функции реальных значений показателей. Также вычисляется погрешность прогнозирования на каждом шаге расчетов и итоговая средняя погрешность. Результаты оценки точности прогнозных расчетов для двух диапазонов (трехгодичный и пятигодичный) на рис. 8.

Прогнозные расчеты показателя миграции модели GA 2 имеют наименьшее значение отклонения площади для трехлетнего периода 0,58% (средняя погрешность 12,61%) и для пятилетнего периода 5,46%

(средняя погрешность 15,97%). Однако, наименьшую среднюю погрешность в пятилетнем периоде имеет модель GA 1 (9,63%). Этот факт позволяет предположить, что процесс пятилетнего прогнозирования показателя миграции с помощью GA 1 содержит точечные пики больших значений погрешности (рис. 9), которые влияют на увеличение площади под расчетной кривой. Таким образом, использование значения отклонения площади позволяет определить резкие скачки расчетных данных.

Демографические данные обладают рядом особенностей: предсказуемость пределов изменений, отсутствие значительных перепадов между ближайшими значениями, временная упорядоченность, зависимость от социально-экономических и политических процессов. Нейросетевое моделирование и машинное обучение позволяют на основе большого числа накопленных данных выявлять закономерности процессов. Но прогнозирования с применением этих технологий необходима предварительная обработка и подготовка данных [6].

Method	triennium		five-year	
	area, %	inaccuracy, %	area, %	inaccuracy, %
birth rate				
GA 1	5,20	8,01	7,70	12,01
GA 2	1,85	5,61	0,70	6,41
DES Br-Mo	1,54	3,97	0,18	3,92
DES Br-Mi	0,98	5,59	0,70	4,25
mortality				
GA 1	1,17	6,06	2,07	6,74
GA 2	4,23	6,01	9,31	9,06
DES Br-Mo	3,69	4,72	4,75	4,93
DES Mo-Mi	2,48	8,40	3,52	9,64
migration				
GA 1	8,60	14,13	12,04	9,63
GA 2	0,58	12,61	5,46	15,97
DES Br-Mi	0,69	16,8	11,94	17,1
DES Mo-Mi	1,44	15,97	11,83	17,45

Рис. 8. Результаты точности оценки

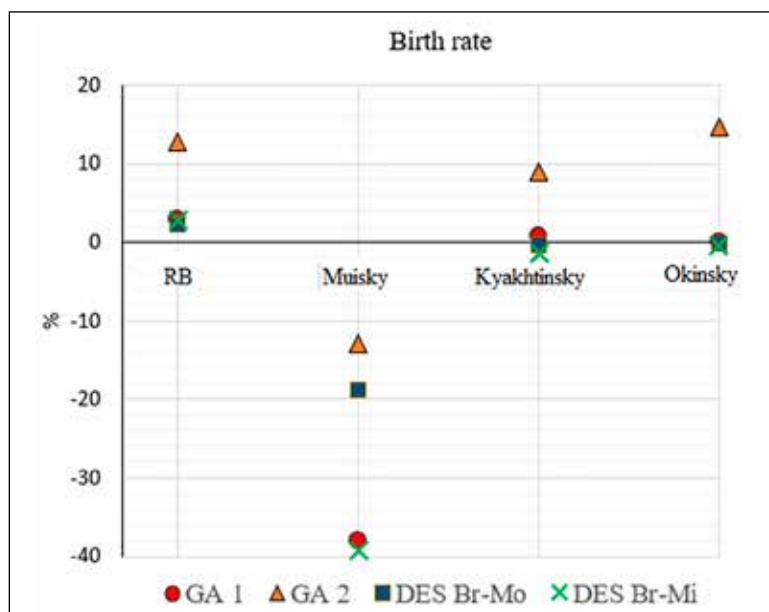


Рис. 9. Ошибка прогноза скорости миграции GA 1 и GA 2

Прогнозные расчеты

По трехлетнему периоду исходных данных (за 2015-2017 гг.) представлены прогнозные расчеты на 2018 год для трех районов Республики Бурятия имеющих различное географическое расположение: Муйский – северный район, Кяхтинский – южный район, Окинский – западный район. Прогноз рассчитан для показателей рождаемости, смертности и миграции по не-

скольким моделям (рис. 10-12). Показатели рождаемости и смертности для Республики Бурятия и Кяхтинского района прогнозируются практически всеми моделями с достаточной точностью (уровень значимости 0,05). Точность результатов прогнозирования показателя миграции для всех рассматриваемых территорий оказалась более 10%, можно отметить достаточно кучное расположение погрешностей всех моделей.

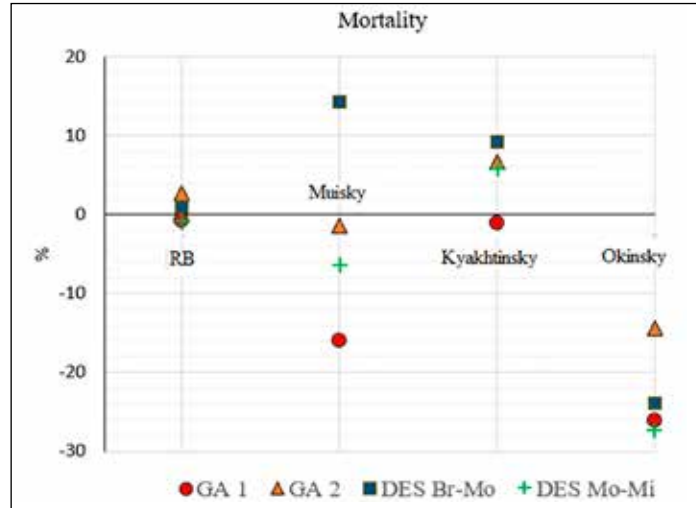


Рис. 10. Ошибка прогноза рождаемости

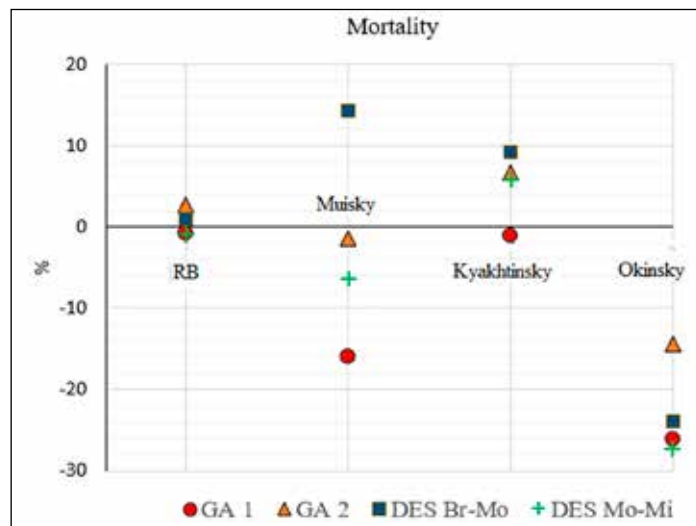


Рис. 11. Ошибка прогноза смертности

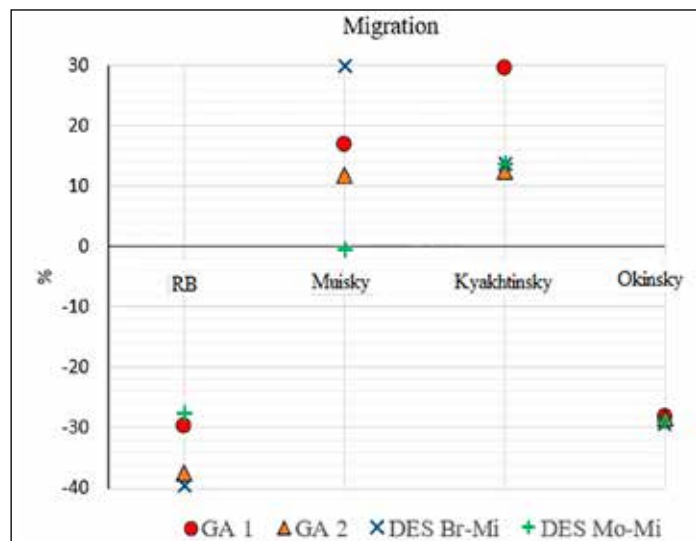


Рис. 12. Ошибка прогноза миграции

Также по результатам расчетов можно сделать вывод об уникальности каждой территории, так для Муйского района погрешности всех моделей имеют отрицательные значения и очевиден разброс значений.

На основе данных за 2016-2018 годы получен прогноз численности населения Республики Бурятия на 2019, 2020 и 2021 годы (рис. 13).

algorithm/year	2019	2020	2021
GA 1	984,9	984,9	984,9
GA 2	984,7	984,8	984,8
DES	984,4	984,3	984,2

Рис. 13. Прогнозные значения численности населения Республики Бурятия, тыс. чел.

Заключение

Разработка математической модели является основной составляющей моделирования процессов демографического разви-

тия территории. В свою очередь, моделирование демографического прогноза является важной частью при формировании демографической политики определенной территории. Выявлена необходимость применения в рассматриваемой области информационных технологий, а именно разработка инструментария для анализа и прогнозирования показателей.

Разработаны и программно реализованы алгоритмы прогнозирования показателей рождаемости, смертности и миграции на основе генетического алгоритма и математической модели в виде системы дифференциальных уравнений. Также произведен анализ показателей и оценка точности алгоритмов. По результатам оценки точности прогнозных расчетов показателей рождаемости, смертности и миграции сделан вывод о том, что для прогнозирования показателя рождаемости лучше применять модель DES, для показателя смертности GA 1 и для миграции GA 2. Выполнен прогнозный расчет численности населения Республики Бурятия на 2021 год.

Библиографический список

1. Abdelrahim M. Zabadi, Ramiz Assaf, Mohammad Kanan. A Mathematical and Statistical Approach for Predicting the Population Growth. World Wide Journal of Multidisciplinary Research and Development. August 2017. P. 50-59.
2. Ajay J. Patel, Dr. M.B. Prajapati. Estimation for Future Population Growth of China by using Logistic model. International Journal of Scientific Development and Research (IJS DR). September 2016. IJS DR. Vol. 1. Iss. 9. P. 52-56.
3. Акмудашева З.К. Сравнительный анализ методов прогнозирования демографических процессов // Актуальные вопросы науки и практики. 2017. С. 41-46.
4. Goldberg David E. Genetic Algorithms in Search, Optimization and Machine Learning. USA: Addison-Wesley Publishing Company. Inc., 1989.
5. Иванюк В.А., Андрейчиков А.В. Интеллектуальная система прогнозирования на основе методов искусственного интеллекта и статистики // Программные продукты и системы. 2008. С. 90-91.
6. Kapur J.N., Khan Q.J.A. Some Mathematical Model for Population Growth. Indian Journal of Pure Applied Mathematics. 2015. № 10(3). P. 277-286.
7. Kermack W.O., McKendrick A.G. A Contribution to the Mathematical Theory of Epidemics. Communicated by Sir Gilbert Walker, F.R.S. Received May 13. 1927. P. 700-721.
8. Корнина А.Е. Машинное обучение и нейронные сети в бизнесе // Хроноэкономика. 2018. № 2 (10). С. 111-116.
9. Курейчик В.М. Модифицированные генетические операторы // Известия СФУ. Технические науки. 2009. Т. 12. С. 7-14.
10. Носова М.Г. Применение математической модели к исследованию процесса изменения демографической ситуации в Российской Федерации // Молодой ученый. 2017. № 1. С. 1-4.
11. Носова М.Г. Автономная немарковская система массового обслуживания и ее применение в задачах демографии: дис. канд. физ.-мат. наук. Томск, 2010. 204 с.
12. Приумов Ю.Г. Численный метод, сборник вопросов. М.: Дрофа, 2007. 144 с.
13. Цуларис А. Анализ моделей логистического роста. Рез. лат. мат. наука. Том. 2. 2001. С. 23-46.

14. Руи Дилао. Математические модели в популяционной динамике и экологии // Биоматематика. Моделирование и имитация. 2006. Глава 15. С. 399-449.
15. Соловьев А.И., Соловьев С.А. О подходах к анализу демографических данных с использованием машинного обучения // Современные информационные технологии и ИТ-образование. 2018. Т. 14. № 4. С. 947-959.
16. Территориальное управление Федеральной службы государственной статистики по Республике Бурятия. URL: <https://burstat.gks.ru/demo> (дата обращения: 10.01.2023).
17. Титов В.А., Вайнберг Р.Р. Анализ существующих динамических моделей на основе системы уравнений Лотки-Вольтерра «хищник-жертва» // Фундаментальные исследования. 2016. С. 409-413.
18. Трубецков Д.И. Феномен математической модели Лотки-Вольтерра и подобных моделей // Методические заметки. Труды ВУЗа. 2011. Т. 19, № 2. С. 69-88.
19. Затонский А.В., Сиротина Н.А. Прогнозирование экономических систем по модели, основанной на дифференциальном уравнении регрессии // Математический анализ экономических моделей. 2014. Т. 50, № 1. С. 91-99.