

ОЦЕНКА РАССЕЛЕНИЯ НА ОСНОВЕ ДЕМОГРАФИЧЕСКИХ МОДЕЛЕЙ И ГЕОИНФОРМАЦИОННОГО АНАЛИЗА

А.И. Рюмкин, Е.С. Тябаев

1. Введение

При решении многочисленных задач управления крупным городом или регионом важнейшее значение имеют сведения о его населении, которые необходимы при формировании прогнозов социально-экономического развития, градостроительного проектирования, формировании отраслевых схем размещения объектов и решении многочисленных текущих вопросов социального обеспечения, образования, здравоохранения [1-8]. При этом для эффективного управления необходима подобная информация в пространственном аспекте, нужны сведения о расселении населения по территории, например по планировочным элементам города (кварталам, микрорайонам, слободам, поселкам). Вычисление прогнозной оценки такого расселения распадается на два этапа. На первом этапе рассчитывается демографический прогноз населения, на втором – производится расчет его распределения по территории.

Большинство известных моделей воспроизводства населения использует описание лишь чисто демографических событий и ситуаций, не учитывая явно влияние социально-экономического окружения. Таковы модели передвижки возрастов и модели потенциальной демографии [9], модели, анализирующие брачное состояние [10,11], экстраполяционные модели миграции [12]. Для систем перспективного планирования необходимо явное описание влияния социально-демографических факторов на демографические процессы, в особенности на межрайонную миграцию. Оценка влияния факторов на демографические показатели проводилась в [14]. Факторные модели миграции изучены в [15] и более полно – в [12], но почти все они статические, не учитывающие изменение во времени. Новые приложения предъявляют ряд специфических требований к моделям демографических процессов и, прежде всего, ориентацию на существующую статистическую отчетность. Между тем многие из существующих моделей информационно не обеспечены, например [11,16].

В настоящей работе описаны прикладные модели демографических процессов, учитывающие отмеченные выше особенности приложений. Основное внимание уделяется конструктивному решению вопросов идентификации моделей по доступным статистическим данным, оценке их точности и реализации. Данные модели впервые введены авторами в процессе создания территориальной подсистемы моделирования демографических процессов региона [14,17-19].

2. Основная модель движения населения

При описании демографических процессов обычно выделяют естественное движение населения (рождаемость, дожитие, брачность, смертность) и миграцию [9]. Каждый из этих видов движения характеризуется набором показателей. В соответствии с конечной целью моделирования демографических процессов в качестве выходных показателей модели естественного движения населения (ЕДН) выбраны половозрастные структуры [17]. Годичные половозрастные структуры (ПВС) дают достаточно подробную количественную характеристику населения города, и в то же время – это максимально доступная информация соответствующей статистики населения Госкомстата, включая данные переписей.

Прикладные модели демографических процессов должны давать текущие и прогнозные количественные оценки населения, учитывать влияние социально-экономической среды на динамику основных демографических параметров. Из моделей такого класса наиболее полно информационно обеспечена модель движения населения О.В. Староверова [10], включающая динамику возрастных структур, рождаемость и брачность. Поскольку полные данные о брачной и внебрачной рождаемости отсутствуют в статистике населения местных органов Госкомстата, а статистические оценки брачной рождаемости для небольших популяций непредставительны, мы исключаем описание брачности из моделей.

Кроме того, по сравнению с [10] вводится детализация описания внешней миграции на уровне ПВС и функциональное представление параметров модели, зависящих от социально-экономических факторов.

Построенная с учетом этих замечаний модель демографических процессов представляет собой систему разностных уравнений, отражающих взаимодействие основных демографических процессов рождаемости, смертности и миграции.

2.1. Моделирование естественного движения населения

При формулировке модели ЕДН предполагается выполнение обычных для подобных моделей предположений взаимной независимости демографического поведения отдельных индивидов, ординарности и т.п. [10].

Воспроизводство населения. Пусть $V(t, y)$ – функция численности женщин возраста y в момент времени t , а $f(t, y)$ – аналогичная функция рождаемости этих женщин, γ_t – доля рожденных мальчиков. Тогда численность рожденных за год t мальчиков $U(t, 0)$ и девочек $V(t, 0)$ определится следующим образом:

$$U(t,0) = \gamma_t \sum_{y=y_n}^{y_g} f(t,y)V(t,y), \quad (1)$$

$$(t,0) = (1 - \gamma_t) \sum_{y=y_n}^{y_g} f(t,y)V(t,y), \quad (2)$$

где $[y_n, y_g]$ – детородный возраст.

Смертность и миграция. Численность населения уменьшается за счет умерших и выбывших и увеличивается за счет прибывших. Введем демографические показатели:

$\mu(t, x)$ – показатель смертности мужчин возраста x в течение года t ;

$\zeta(t, x)$ – функция распределения по возрасту мужчин, прибывших в году t ;

$\xi(t, x)$ – функция распределения по возрасту мужчин, выбывших в году t ;

$\mu(t, y)$, $\zeta(t, y)$, $\xi(t, y)$ – соответствующие функции смертности, прибытия и выбытия для женщин.

Пусть $I_m(t)$, $Q_m(t)$, $I_f(t)$, $Q_f(t)$ – общая численность прибывших и выбывших в регион за год t мужчин и женщин соответственно. С учетом гипотезы ординарности уравнения динамики возрастных структур запишутся для мужчин:

$$U(t+1, x+1) = (U(t, x) - Q_m(t)\xi(t, x))(1 - \mu(t, x)) + I_m(t)\zeta(t, x) \quad (3)$$

и для женщин:

$$(t+1, y+1) = (V(t, y) - Q_f(t)\xi(t, y))(1 - \mu(t, y)) + I_f(t)\zeta(t, y). \quad (4)$$

В качестве начальных условий уравнений (3)-(4) обычно выбираются данные переписей населения.

Функциональное представление параметров модели. Демографические параметры модели движения населения (9)-(11) $f(t, y)$, $\mu(t, \cdot)$, $\zeta(t, \cdot)$, $\xi(t, \cdot)$ представляют собой векторы большой размерности, компоненты которых сильно зависимы между собой. Целесообразно сократить размерность описания, используя какие-либо функциональные приближения этих параметров. Подобная аппроксимация упрощает также решение задачи восстановления зависимости параметров от социально-экономических факторов и прогноза этих параметров.

Функциональное представление рождаемости женщин. Функцию рождаемости $f(t, y)$ (время t в дальнейшем будет опускаться) можно представить состоящей из трех компонент:

$$f(y) = \begin{cases} 0, & \text{при } 0 < y < y_n, \\ > 0, & \text{при } y_n \leq y \leq y_e, \\ 0, & \text{при } y > y_e. \end{cases}$$

На интервале $[y_n, y_e]$ функцию рождаемости можно аппроксимировать кривыми Пирсона [20], гамма-функцией или кривой Хадвигера [21]. При этом статистические данные рассматриваются как эмпирические функции распределения. Коэффициенты указанных приближений находятся по первым четырем моментам этих распределений – математическому ожиданию m , среднеквадратическому отклонению σ и моментам третьего и четвертого порядка r_3, r_4 . Предпринятое нами исследование показало, что для аппроксимации подходят кривые Пирсона I рода:

$$\hat{f}(y_1) = \tilde{n} \left(1 + \frac{y_1}{l_1}\right)^{q_1} \left(1 - \frac{y_1}{l_2}\right)^{q_2}, \quad (5)$$

где

$$y_1 = y - m + \frac{\sigma r_3}{2} \left(\frac{s+2}{s-2}\right),$$

$$q_1, q_2 = \frac{1}{2} (s - 2 \mp s(s+2)r_3/t),$$

$$t = \sqrt{r_3^2 (s+2)^2 + 16(s+1)},$$

$$s = 6(r_4 - r_3^2 - 1)/(3r_3^2 - 2r_4 + 6),$$

$$l_1 = q_1 l / (s-2), \quad l_2 = q_2 l / (s-2), \quad l = \sigma t / 2,$$

$$\tilde{n} = c l_1^{q_1} l_2^{q_2} \Gamma(q_1 + q_2 + 2) / (l^{q_1+q_2+1} \Gamma(q_1+1) \Gamma(q_2+1)),$$

$\Gamma(\cdot)$ – гамма функция.

Пусть $F(y_i), i = \overline{1, N}$ – дискретные отсчеты параметра $f(y)$. Тогда необходимые для расчета моменты m, σ, r_3, r_4 определяются из соотношений:

$$m = \sum_{i=1}^N F(y_i) y_i,$$

$$\sigma = \sqrt{\sum_{i=1}^N (y_i - m)^2 F(y_i)},$$

$$r_3 = \frac{1}{-\rho^3} \sum_{i=1}^N (y_i - m)^3 F(y_i),$$

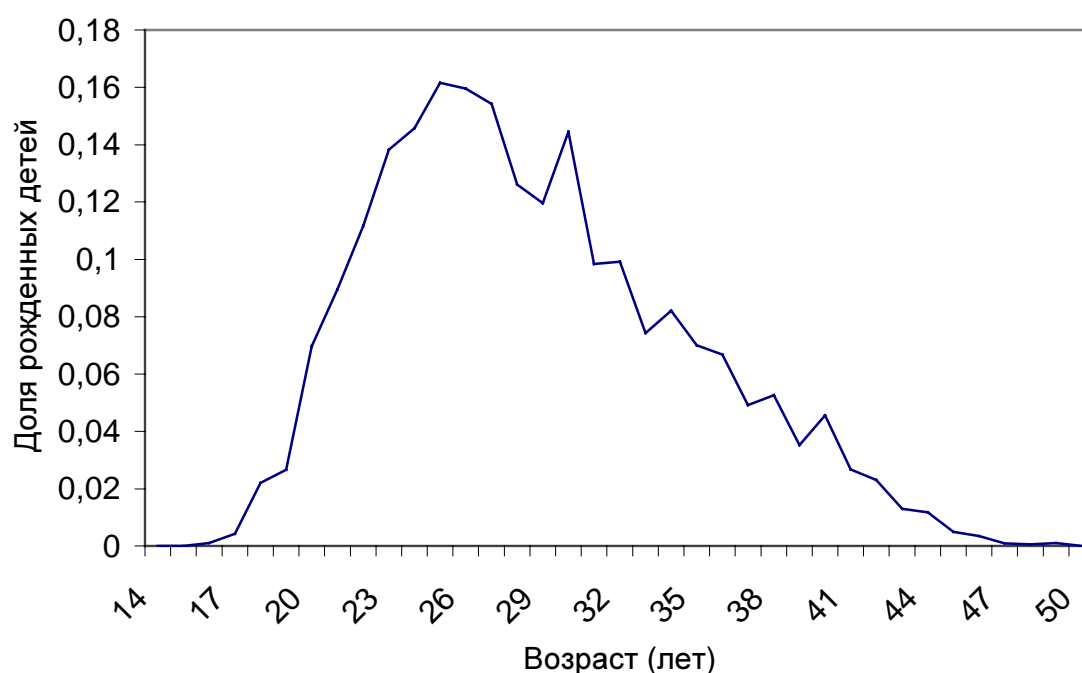


Рис. 1. Одна из реальных функций рождаемости.

$$r_4 = \frac{1}{-\sigma^4} \sum_{i=1}^N (y_i - m)^4 F(y_i),$$

$$c = \sum_{i=1}^N F(y_i). \quad (6)$$

На рис. 1 приведена для иллюстрации типичная функция рождаемости женщин.

Функциональное представление смертности. Нужно выделить три компоненты в этом параметре модели: смертность в возрасте 0-5 лет, смертность в возрасте 6-20 лет, смертность людей старше 20 лет. Первая компонента характеризуется уменьшением смертности, на второй компоненте смертность мало меняется, третья компонента характеризуется увеличением смертности. На рис. 2 приведена исходная и сглаженная функция смертности.

Для функционального представления параметра $\mu(x)$ удобнее перейти от вероятностей к числам умирающих:

$$d(x) = \prod_{i=0}^{x-1} (1 - \mu(i)) \mu(x).$$

Обратный переход от чисел $d(x)$ к значениям параметра $\mu(x)$ осуществляется таким образом:

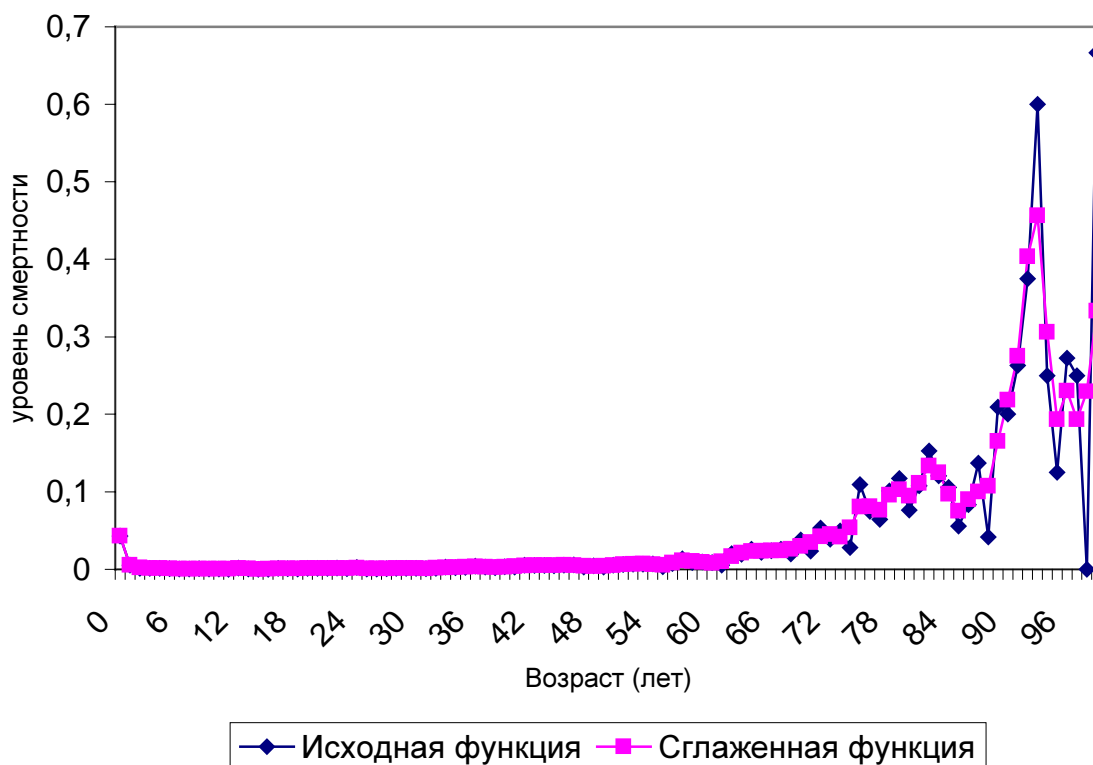


Рис. 2. Реальная и сглаженная функция смертности (сглаживание данных осуществлялось методом скользящего взвешенного среднего).

$$\mu(0) = d(0),$$

$$\mu(x) = d(x) / \prod_{i=0}^{x-1} (1 - \mu(i)), \quad x = \overline{1, 100}. \quad (7)$$

Представление первой компоненты. По нашим оценкам [17] на интервале 0-5 лет лучшие результаты дает следующая модель:

$$s_1 = a_1 + a_2 \ln(x + 1), \quad \text{где } s_1(x) = \sum_{i=0}^x d(i),$$

а коэффициенты a_1, a_2 разыскиваются методом наименьших квадратов. Переход от коэффициентов a_1, a_2 к числам $\hat{d}(x)$ осуществляется по следующим формулам:

$$\begin{aligned} \hat{d}(0) &= a_1, \\ \hat{d}(x) &= a_2 (\ln(x + 1) - \ln(x)), \quad x = \overline{1, 5}. \end{aligned} \quad (8)$$

Представление второй компоненты. Ее предлагается описать квадратичной зависимостью:

$$\hat{d}(x) = a_3 + a_4(x - 5) + a_5(x - 5)^2, \quad (9)$$

в которой коэффициенты a_3, a_4, a_5 определяются методом наименьших квадратов. Условия сшивания локальных компонент накладывают такие ограничения на значения коэффициентов:

а) сшивание первой и второй компонент:

$$a_3 = a_1 + a_2 \ln(6),$$

б) сшивание второй и третьей компонент:

$$a_3 + 15a_4 + 225a_5 = \hat{d}_3(20), \tag{10}$$

где $\hat{d}_3(20)$ – значение $d_3(20)$, вычисленное с использованием функционального представления третьей компоненты.

Чтобы отыскать коэффициенты a_3, a_4, a_5 в формуле (9), нужно сначала идентифицировать функциональные представления первой и третьей локальных компонент, найти значение первой компоненты в точке $x = 5$, а третьей – в точке $x = 20$, затем отыскать a_3 по формуле (10).

Коэффициенты a_4, a_5 определяются при минимизации квадратичного функционала вида

$$J = \frac{1}{16} \sum_{i=5}^{20} (d(i) - a_3 - a_4(i - 5) - a_5(i - 5)^2)^2 + \lambda(a_3 + 15a_4 + 225a_5),$$

где λ – множитель Лагранжа. Второе слагаемое функционала появляется в связи с учетом ограничения (10).

Представление третьей компоненты. Здесь стоит воспользоваться методом, описанным в [22]. В этой работе процесс старения населения предполагается зависящим от случайной величины η . Распределение этой величины нормально для каждой когорты населения. Математическое ожидание величины η уменьшается с возрастом по закону $m_\eta = e^{-a_6x}$, а среднеквадратическое отклонение постоянно: $\sigma_\eta = a_7$. Смерть члена когорты наступает тогда, когда его значение η достигает определенного минимума a_8 , превышение которого величиной η необходимо для выживания. Накопленное относительное число смертей до возраста x выразится так:

$$\sum_{i=0}^x d(i) = \int_{-\infty}^{a_8} z(x, \eta) d\eta, \tag{11}$$

$$z(x, \eta) = \exp\left(\frac{(\eta - e^{-a_6x})^2}{2a_7^2 e^{-2a_6x}}\right) / (\sqrt{2\pi} a_7 e^{-a_6x}). \tag{12}$$

Отсюда распределение времени дожития будет равно

$$\frac{d}{dx} \int_{-\infty}^{a_8} z(x, \eta) d\eta. \quad (13)$$

Применяя разложение в ряд Тейлора в точке $\eta = a_8$ к выражению (12), подставляя это выражение в (13), можно получить представление для третьей компоненты в виде $\hat{d}(x) = a_6 a_8 z(x, a_7)$.

2.2. Моделирование миграции

При моделировании миграции населения региона целесообразно выделить два уровня его описания:

1. Потокное (валовое) описание межрайонных миграционных взаимодействий.
2. Детализированное описание структуры суммарного потока внешней миграции.

При этом в процессе моделирования на первом уровне учитывается структура причинно-следственных связей и их специфика, определяющая направление и интенсивность миграционного обмена данного региона с другими регионами системы. Мера описания – численность прибывающих, выбывающих, сальдо.

На втором уровне описания осуществляется агрегирование попарных миграционных потоков регионов окружения в общий поток внешней миграции. Характеристикой этого агрегированного потока являются ПВС мигрантов. Выбор такого показателя определяется необходимостью стыковки моделей ЕДН и миграции по данным. Различие показателей по уровням описания определяется действующей статистикой: половозрастная структура мигрантов рассчитывается только для региона в целом (по прибытию и выбытию), а данные о межрайонных потоках представлены лишь численностью мигрантов. Соответственно различаются описания миграции на каждом уровне.

2.2.1. Описание межрайонной миграции

Перед решением этой задачи предварительно должна быть проведена структуризация внешнего миграционного потока на составляющие. Для каждой пары взаимодействующих регионов определяются специфические для них причины миграции и совокупность характеризующих их факторов. Соответствующие вопросы подробно рассматриваются при описании результатов идентификации моделей миграции.

Ориентируясь на имеющуюся статистику межрайонной миграции, мы фактически принимаем предположения о следующих свойствах миграционных процессов [18].

1. Регионы территориально однородны в отношении условий жизни населения.
2. Различия в миграционном поведении отдельных социальных групп региона по отношению к выбранному множеству факторов пренебрежимо малы.
3. Поведение каждого мигранта независимо от поведения других жителей региона.

Кроме того, сделаем ряд предположений, необходимых для упрощения моделей и процедур идентификации.

4. Показатели каждой пары межрайонных миграционных связей зависят лишь от значений факторов взаимодействующих регионов и не зависят от влияния других факторов.
5. Показатели межрайонной миграции не изменяются при изменении значений факторов в регионах обмена на одну и ту же величину.

Наиболее полно отмеченным требованиям удовлетворяет описание миграции с помощью разностного уравнения

$$y(t + \tau) = \sum_{l=0}^n a_l y(t - l) + \sum_{i=1}^m \sum_{j=0}^{dl_i} b_{ij} x_i(t - j). \quad (13)$$

Здесь $y(t - l)$ – значение показателя миграции в момент $t - l$, $x_i(t - j)$ – значение i -го фактора в момент $t - j$, τ – интервал прогнозирования, m – количество факторов, dl_i – количество учитываемых точек предыстории по i -му фактору, a_l, b_{ij} – коэффициенты модели. Уравнение (13) обобщает описание значительного числа моделей миграции. Например, при $m = 0$ получаем экстраполяционные модели, при $n = 0, dl_i = 0$ – статические факторные модели.

2.2.2. Функциональное представление ПВС агрегированного миграционного потока

Параметры миграции $\zeta(x)$, $\xi(x)$, $\zeta(y)$ и $\xi(y)$ удобно рассматривать на трех интервалах: 0-15, 15-35, 35-100 и старше. Первая и третья компонента характеризуются малым уровнем миграционной активности, а вторая – большим (рис. 3). Для аппроксимации миграционных параметров используем полиномы Чебышева, ортогональные на системе равноотстоящих точек [20]:

$$\zeta(x) = \sum_{i=0}^{m_l} a_{li} P_{N_{li}}(x), x \in l\text{-му интервалу}, l = \overline{1,3}, \quad (14)$$

где m_l – максимальный порядок полинома на l -м интервале (оценивается эмпирически), a_{li} – неизвестные коэффициенты разложения, N_{li} – число равноот-

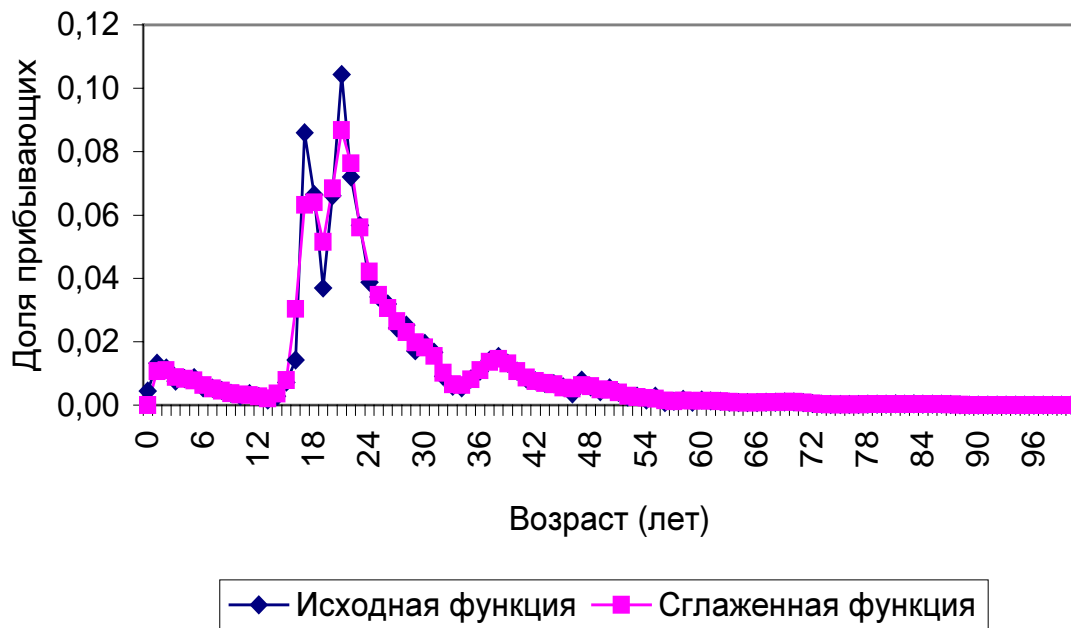


Рис. 3. Реальная и сглаженная функция миграции.

Сглаживание проводилось методом скользящего среднего.

стоящих точек на l -м интервале. Полином $P_{N_i}(x)$ можно оценить по рекуррентным формулам:

$$P_{N_{l0}}(x) = 1, P_{N_{l1}}(x) = x - (N_l + 1)/2,$$

$$P_{N_{li+1}}(x) = P_{N_{l1}}(x)P_{N_{li}}(x) - \frac{i^2(N_l^2 - i^2)P_{N_{li-1}}(x)}{4(2i-1)(2i+1)}. \quad (15)$$

Для определения коэффициентов $a_{li}, l = \overline{1,3}, i = \overline{0, m_l}$ предлагается следующий квазиоптимальный алгоритм. Сначала на первом и третьем интервалах определяются коэффициенты a_{ki} :

$$a_{ki} = \frac{\sum_{x=x_h^k}^{x_g^k} \zeta(x)P_{N_{ki}}(x)}{\sum_{x=x_h^k}^{x_g^k} P_{N_{ki}}^2(x)},$$

где $[x_h^k, x_g^k]$ — k -й интервал, $k = \overline{1,3}$, затем определяются коэффициенты a_{2i} , причем начало и конец второй компоненты должны совпадать с концом первой и началом второй компоненты. Оптимальные a_{2i} определяют при минимизации квадратичного функционала

$$J = M \left\{ \left(\zeta(x) - \sum_{i=0}^{m_2} a_{2i} P_{N_{2i}}(x) \right)^2 \right\} + 2\lambda_1 \sum_{i=0}^{m_2} a_{2i} P_{N_{2i}}(x_h^2) + 2\lambda_2 \sum_{i=0}^{m_2} a_{2i} P_{N_{2i}}(x_g^2). \quad (16)$$

Здесь $M(\cdot)$ – оператор усреднения, λ_1, λ_2 – множители Лагранжа. Дифференцируя функционал (18) по $a_{2i}, i = \overline{0, m_2}$ и приравнявая производные нулю, получим $m_2 + 1$ уравнение. Добавив к ним уравнения – условия стыковки компонент, получим систему $m_2 + 3$ линейных уравнений с $m_2 + 3$ неизвестными:

$$\begin{cases} a_{2i}M\{P_{N_{2i}}^2\} + \lambda_1 P_{N_{2i}}(x_{2н}) + \lambda_{21} P_{N_{2i}}(x_{2б}) = M\{\zeta(x)P_{N_{2i}}(x)\}, i = \overline{1, m_2}, \\ \sum_{i=0}^{m_2} a_{2i}P(x_{2н}) = C_1, \\ \sum_{i=0}^{m_2} a_{2i}P(x_{2б}) = C_2, \end{cases} \quad (17)$$

где C_1 – значение представления первой компоненты в точке $x_{2н}$, C_2 – значение представления третьей компоненты в точке $x_{2б}$.

Введем обозначения:

$$\begin{cases} M\{\zeta(x)P_{N_{2i}}(x)\} = \Theta_i, M\{P_{N_{2i}}^2\} = M_i, \Theta_i / M_i = a_i^*, a_{2i} = A_i, \\ P_{N_{2i}}(x_{2н}) = \rho_i, P_{N_{2i}}(x_{2б}) = (-1)^i \rho_i. \end{cases} \quad (18)$$

Последние два соотношения учитывают особенности значений полиномов Чебышева на концах интервала ортогональности. Решение системы (17) с учетом обозначений (18) имеет вид:

$$\begin{cases} \lambda_1 = \frac{\sum_{i=0}^{m_2} \rho_i^2 / M_i \left[\sum_{j=0}^{m_2} a_j^* \rho_i (1 - (-1)^{i+j}) - C_1 + (-1)^i C_2 \right]}{\sum_{i=0}^{m_2} (\rho_i^2 / M_i) \sum_{j=0}^{m_2} (\rho_j^2 / M_j) (1 - (-1)^{i+j})}, \\ \lambda_2 = \frac{\sum_{i=0}^{m_2} \rho_i^2 / M_i \left[\sum_{j=0}^{m_2} a_j^* \rho_i (-1)^i ((-1)^{i-j} - 1) - C_2 + (-1)^i C_1 \right]}{\sum_{i=0}^{m_2} (\rho_i^2 / M_i) \sum_{j=0}^{m_2} (\rho_j^2 / M_j) (1 - (-1)^{i+j})}. \end{cases} \quad (19)$$

Для A_{2i} и A_{2i+1} имеем:

$$\left\{ \begin{array}{l} A_{2i} = a_{2i}^* - \frac{\rho_{2i}}{M_i} \frac{\sum_{j=0}^{(m_2-1)/2} a_{2j}^* \rho_{2j} - (C_1 + C_2)/2}{\sum_{j=0}^{m_2/2} \rho_j^2 / M_j}, \\ A_{2i+1} = a_{2i+1}^* - \frac{\rho_{2i+1}}{M_{2i+1}} \frac{\sum_{j=0}^{(m_2-1)/2} a_{2j+1}^* \rho_{2j+1} - (C_1 + C_2)/2}{\sum_{j=0}^{m_2/2} \rho_{2j+1}^2 / M_{2j+1}}. \end{array} \right. \quad (20)$$

Выше $[E]$ – целая часть E . После определения коэффициентов A_l , $l = \overline{1, m_l}$ по формуле (14) можно восстановить параметр $\zeta(x)$. Для представления параметра ξ служат те же формулы (14).

2.2.3. Прогноз коэффициентов представления аппроксимирующего приближения параметров модели

Прогноз всего набора коэффициентов $\{a_i\}$, относящихся к конкретному представлению параметров модели (1)-(4), позволит определить оценки этих параметров в прогнозируемый момент времени.

Ранее отмечалось, что эти коэффициенты предполагаются зависящими от времени и некоторого набора влияющих факторов x_1, \dots, x_n . Ситуация осложняется тем, что в настоящее время нет какой-либо устоявшейся точки зрения о характере влияния социально-экономических факторов на демографические параметры f, μ, ζ, ξ , а следовательно, и вид функциональных зависимостей неопределен. При этом можно попытаться подобрать наиболее точную структуру в классе выбранных функций, например (13), или использовать непараметрические методы восстановления зависимостей, работоспособные при минимальной априорной информации о процессе [23]. Пусть имеется марковский процесс

$$y(t + \tau) = \phi(y(t), x_1(t), \dots, x_n(t), g(t)),$$

где t – текущее время, $t = 1, 2, \dots$, x_i -я переменная, влияющая на y , $g(t)$ – помеха с нулевым средним и независимыми значениями. При квадратичной функции потерь оптимальная оценка прогноза $\hat{y}(t + \tau)$ равна условному математическому ожиданию при известных $y(t)$ и $x_i(t)$:

$$y(t + \tau) = \int_{\Omega} y(t + \tau) P_t(y(t + \tau) | y(t), x_1(t), \dots, x_n(t)) d\Omega, \quad (21)$$

где Ω – область определения $y(t + \tau)$, $P_t(\cdot)$ – условная плотность распределения при известных $y(t)$ и $x_j(t)$. Непараметрическая оценка Розенблата–Парзена условной плотности распределения $P_t(y | x)$ имеет вид

$$P_t(y|x) = \frac{\sum_{i=1}^t [\Phi(y - y(i))/c_t^y] [\Phi(x - x(i))/c_t^x]}{\sum_{i=1}^t \Phi(x - x(i))/c_t^x}, \tag{22}$$

где Φ – колоколообразная функция (ядро). Подставляя (22) в (21), можно получить непараметрическую модель зависимости

$$y(t + \tau) = \frac{\sum_{i=1}^{t-\tau} y(t + \tau) \Phi\left(\frac{y(t) - y(i)}{c_t^y}\right) \prod_{j=1}^N \Phi\left(\frac{x_j(t) - x_j(i)}{c_t^x}\right)}{\sum_{i=1}^{t-\tau} \Phi\left(\frac{y(t) - y(i)}{c_t^y}\right) \prod_{j=1}^N \Phi\left(\frac{x_j(t) - x_j(i)}{c_t^x}\right)}.$$

Ядро $\Phi(\cdot)$ должно удовлетворять следующим требованиям:

$$c > 0, 0 < \Phi(y/c) < \infty, \text{ при } c \neq 0,$$

$$\int_{\Omega} \Phi(y/c) f(y) d\Omega \geq 0 \forall f(y) > 0, \int_{\Omega} \Phi(y/c) d\Omega = 1.$$

В.А. Епанечниковым [24] предложена оптимальная в некотором смысле колоколообразная функция, имеющая вид

$$\Phi(y/c) = \begin{cases} \frac{3}{4\sqrt{5}} - \frac{3}{20\sqrt{5}}(y/c), & \text{при } |y| < \sqrt{5}c, \\ 0, & \text{при } |y| > \sqrt{5}c. \end{cases}$$

Коэффициент размытости c определяется в эксперименте.

3. Идентификация моделей

3.1. Предварительная обработка данных

Необходимость этого этапа обусловлена прежде всего стандартизацией формата данных, с которыми работает модель (перегруппировка данных), с устранением случайных и систематических ошибок в данных (сглаживание, коррекция) и с восстановлением тех данных, которые отсутствуют в органах госстатистики (восстановление возрастных структур).

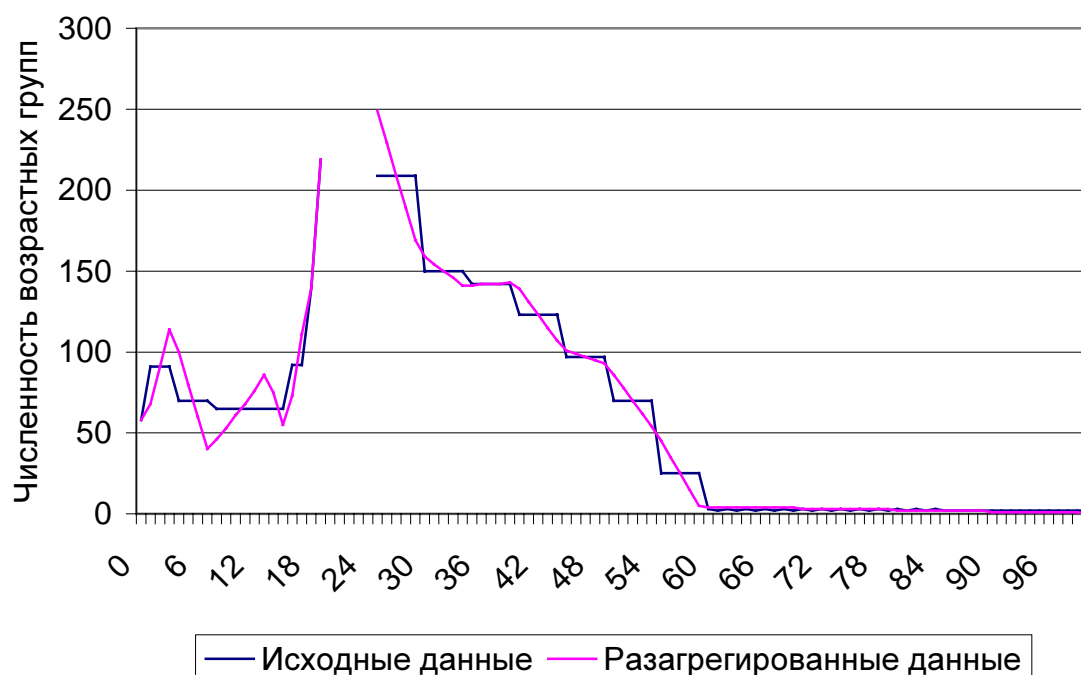


Рис. 4. Разагрегирование данных о миграции населения (приведены данные возрастных групп, которые разагрегируются).

3.1.1. Перегруппировка данных

Часть данных по демографическому движению представлена в статистических формах в агрегированном (сгруппированном) виде. Для модели стандартным представлением данных являются одногодичные группировки. Возникшая задача по перегруппировке данных решена с помощью аппроксимирующего сплайна первого порядка. Для i -го ($i = \overline{1, N}$, N — число группировок длиной Δ_i) участка сплайн вида

$$\hat{f}_i(x) = \frac{F_i}{\Delta_i} + a_i(x - \Delta_i), x \in \Delta_i, \quad (23)$$

удовлетворяющего интегральному условию $\int_0^{\Delta_i} f_i(x) dx = F_i$ (здесь F_i — численность группировки), необходимо сопряжение его концов с $i-1$ и $i+1$ участками сплайна.

Коэффициенты a_i для сплайна (23) задаются выражением

$$a_i = \frac{1}{\Delta_i} \left(a_1 \Delta_1 (-1)^{i-1} + 2 \sum_{j=2}^i (F_j / \Delta_j - F_{j-1} / \Delta_{j-1}) (-1)^{i-j} \right), i = \overline{2, N}$$

или рекуррентно

$$a_{i+1} = F_{i+1} / \Delta_{i+1} - F_i / \Delta_i - a_i \Delta_i / 2.$$

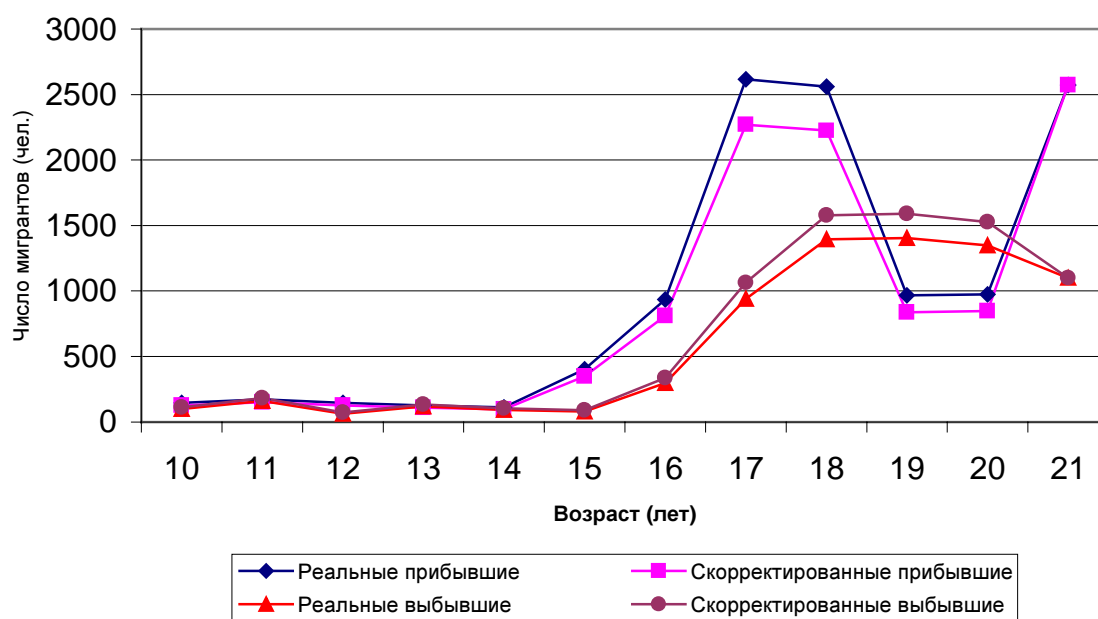


Рис. 5. Корректировка возрастных численностей мигрантов в период между переписями.

Коэффициент a_1 , чтобы сплайн не разбалтывался, лучше оценивать, руководствуясь экстремальным критерием. Например, критерием минимума сумм квадратов отрезков сплайна

$$J = \sum_{i=1}^N l_i^2 = \sum_{i=1}^N (\Delta_i^2 + a_i^2 \Delta_i^2).$$

В этом случае a_1 вычисляется по формуле

$$a_1 = \frac{2}{N\Delta_1} \sum_{i=2}^N (N-i+1)(-1)^i (F_i / \Delta_i - F_{i-1} / \Delta_{i-1}).$$

Для вычисления одногодичной группировки на i -ом интервале используется выражение

$$f_i(x) = F_i / \Delta_i + a_i(x - \Delta_i / 2 + 0,5).$$

Данный алгоритм используется для перегруппировки данных по возрастной миграции населения (рис. 4). На этом рисунке интервал с 18 по 25 лет представлен одногодичными возрастными группами.

3.1.2. Восстановление межпереписных возрастных структур

Данные о демографическом движении, хранящиеся в органах госстатистики, собраны в результате переписей и ежегодного статистического учета. Данные переписей – это возрастные структуры населения на начало 1959, 1970, 1979, 1989 года. Текущий ежегодный статистический учет по миграции, смертности и рождаемости населения в Томской области с 1945 года. Для изучения

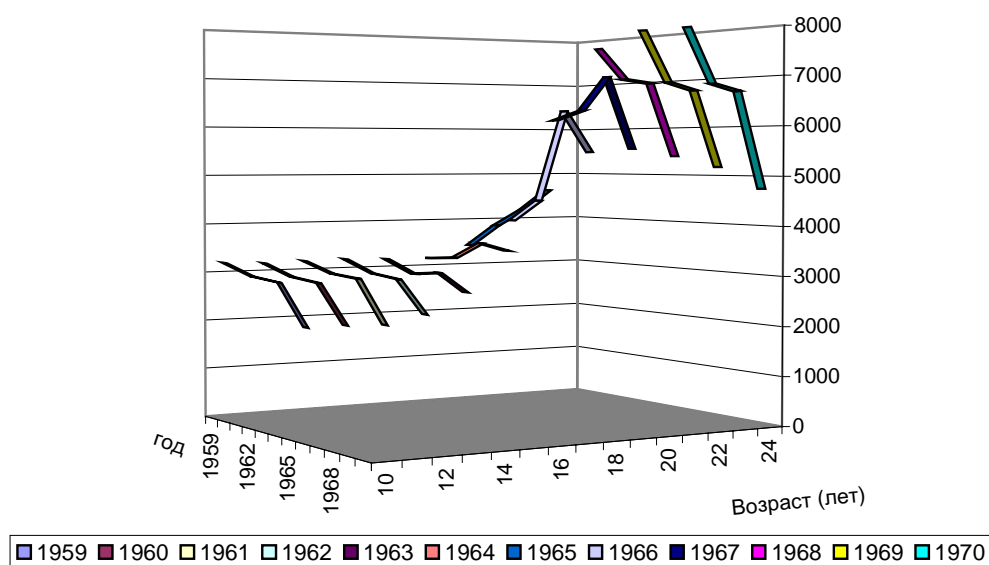


Рис. 6. Динамика некоторых возрастных групп населения.

эволюции демографических параметров модели, коррекции ошибок в статистических данных необходимо иметь возрастные структуры населения за каждый год этого учета. Эти структуры можно восстановить методом передвижки возрастов [9]. Можно получить три варианта восстановленных возрастных структур населения. При отсутствии ошибок в данных все варианты совпали бы между собой. Но практические расчеты показали существенные различия между вариантами, вызванные ошибками в данных. Для согласования вариантов разработана такая процедура: возрастные структуры 1945-1959 годов восстанавливаются от переписи 1959 года путем передвижки возрастов назад по шкале времени; возрастные структуры 1960-1970, 1971-1979, 1980-1989, 1990-1998 годов восстанавливались передвижкой возрастов вперед. Затем восстановленные возрастные структуры 1970, 1979, 1989 года сравнивались с соответствующими переписными структурами. Полученные в результате этого возрастные структуры невязок сводились к нулю путем пропорциональной коррекции элементов возрастных структур мигрантов. На рис. 5 и 6 показана эволюция четырех межпереписных возрастных структур, абсолютные значения невязок, значения корректирующих добавок для данных по миграции. Наблюдается существенная величина невязки. Значения корректирующих добавок в пределах 10-15% от общей численности возрастных групп мигрантов говорят о наличии существенного систематического смещения в этих данных.

3.1.3. Сглаживание данных

Традиционно используемые в демографии алгоритмы сглаживания типа скользящего среднего, скользящего взвешенного среднего предназначены для уменьшения случайных ошибок в данных по координате «возраст». Такое сглаживание параметров модели аппроксимируемыми функциональными кривыми, аналитические выражения которых приведены в (8)-(12), (14), (21), пока-

зано на рис. 4-6. Хотя качество сглаживания подобными алгоритмами достаточно хорошее, они не учитывают вариации параметров модели и случайных ошибок по координате «время». Поэтому более целесообразным следует считать применение двумерных алгоритмов сглаживания. Предлагается следующий девятиточечный алгоритм двумерного сглаживания:

$$\hat{f}(t, y) = \frac{\sum_{i=-1}^1 \sum_{j=-1}^1 f(t+i, y+j) a_{ij}}{\sum_{i=-1}^1 \sum_{j=-1}^1 a_{ij}}, \quad i, j = \overline{-1, 1}, \quad (24)$$

где $\{a_{ij}\} = \{b_i\} * \{c_j\}'$ – матрица весов, $\{b_i\}$, $\{c_j\}$ – векторы-строки сглаживающих коэффициентов. По умолчанию принято $\{b_i\} = \{c_j\} = \{0,5; 1; 0,5\}$.

При сглаживании краев таблицы двумерных данных в формуле (24) модифицируются границы изменения индексов, например, для верхнего края $j = \overline{-1, 0}$, для нижнего $j = \overline{0, 1}$.

Примеры сглаживания приведены на рис. 7-9.

3.2. Оценка параметров модели

Имея данные текущей демографической статистики и восстановленные возрастные структуры, динамические ряды параметров модели вычисляют таким образом:

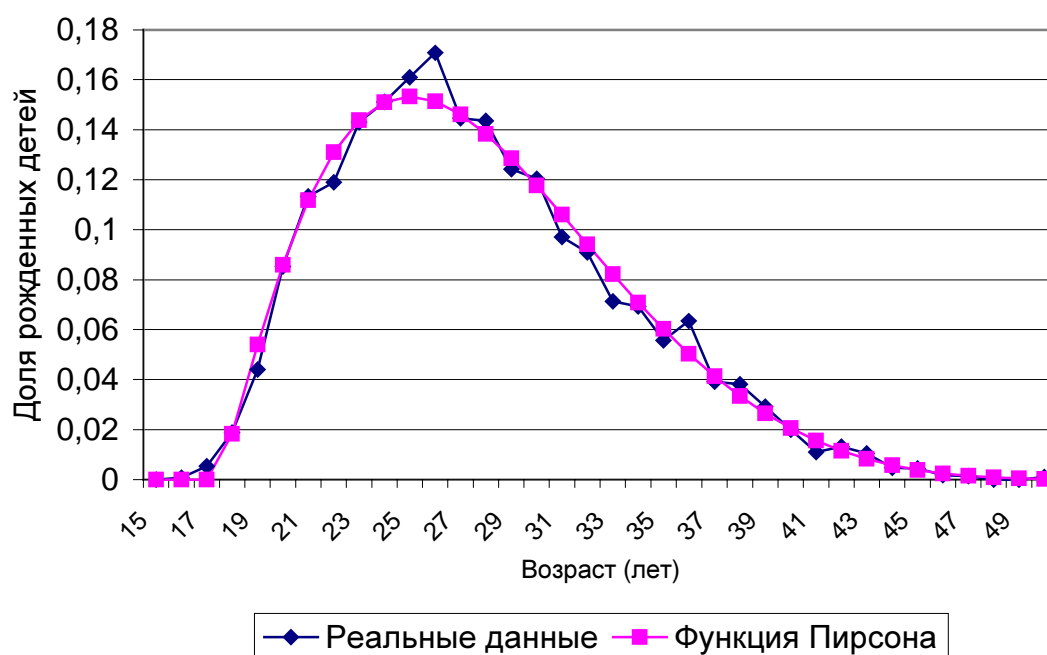


Рис. 7. Реальная функция рождаемости и аппроксимирующая ее функция.

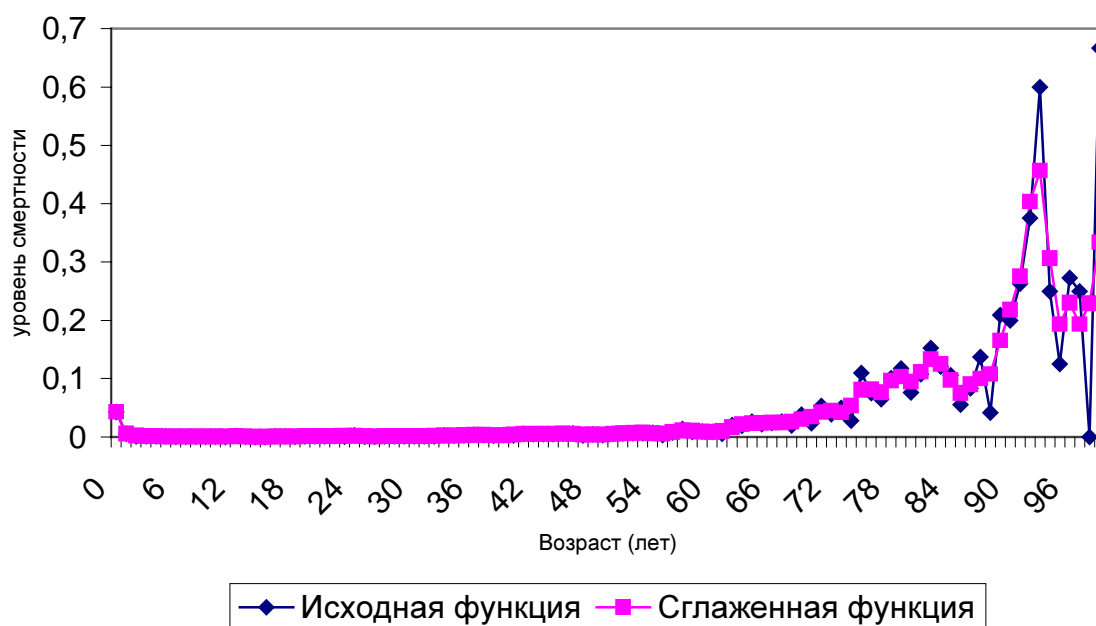


Рис. 8. Реальная и сглаженная функция смертности.

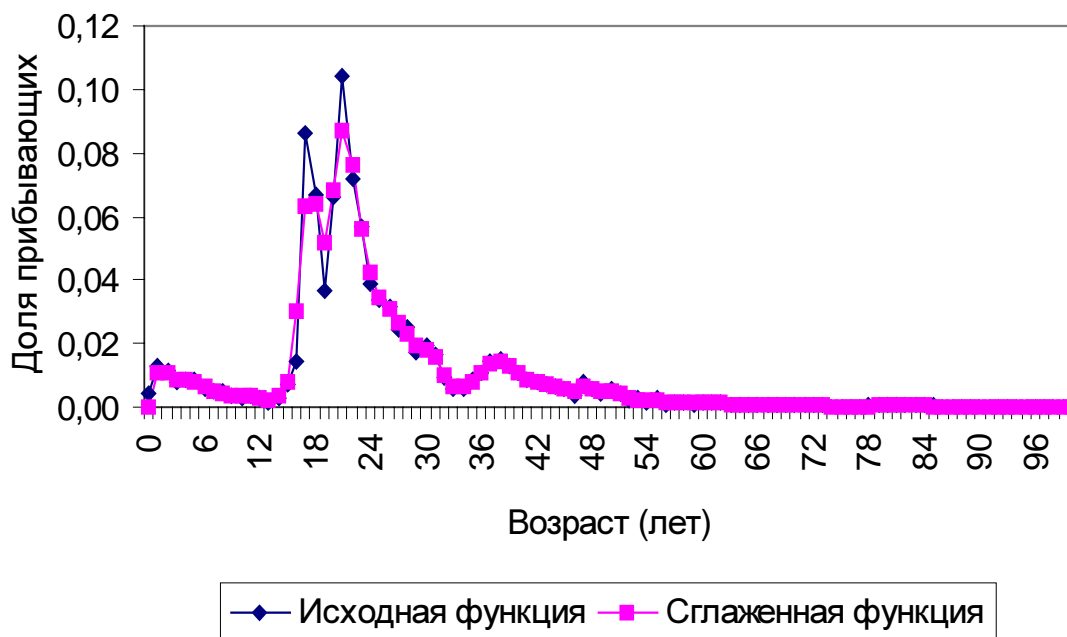


Рис. 9. Реальная и сглаженная функция миграции.

$$f(t, y) = \frac{R(t, y)}{V(t, y)}, \quad \mu(t, y) = \frac{y(t, y)}{V(t, y)}, \quad \mu(t, x) = \frac{y(t, x)}{U(t, x)}.$$

Здесь $y(t, y)$, $y(t, x)$ – возрастные структуры умерших; (t, y) , $U(t, x)$ – возрастные структуры живущих; $R(t, y)$ – возрастные структуры женщин, родивших детей.

3.2.1. Вычисление коэффициентов функциональных представлений демографических показателей

Коэффициенты представления рождаемости (КПР), как отмечалось в первой части, являются математическое ожидание, дисперсия, скос, эксцесс и сумма ординат функции $f(t, y)$ при фиксированном t . Динамические ряды КПР приведены на рис. 10. Видна существенная эволюция этих коэффициентов во времени.

За недостатком места коэффициенты функциональных представлений (КФП) смертности и миграции не приводятся.

3.2.2. Выбор влияющих на КФП факторов

В качестве доступных факторов, влияющих на КФП, были взяты следующие:

1. Численность населения региона (тыс. чел.).
2. Численность городского населения (тыс. чел.).
3. Плотность населения (чел./км²).
4. Общий объем капитальных вложений (млн. руб.).
5. Количество вводимого жилья (тыс. м²).
6. Объем товарооборота (млн. руб.).
7. Количество мест в детских садах и яслях.
8. Обеспеченность врачами (чел./1000 жителей).
9. Количество студентов вузов (чел.).
10. Количество учащихся техникумов (чел.).
11. Сальдо миграции (чел.).

Факторы были собраны для Томской области, начиная с 1963 года. Значения факторов 2, 4-8, 10, 11 пересчитаны на душу населения.

Выбор влияющих факторов на какой-либо КФП осуществляется путем направленного перебора всевозможных линейных регрессий. Фактор включается процедурой в модель, если вероятность подсчитанного для него значения F -статистики, отражающей повышение качества модели за счет включения этого фактора, превышает 0,9.

Был проведен отбор факторов, влияющих на КФП рождаемости ($R_1 - R_5$), смертности ($M_1 - M_8$), прибывших ($I_1 - I_{18}$) и выбывших ($Q_1 - Q_{18}$) по пятифакторным регрессионным моделям, дающих значение квадрата коэффициента множественной корреляции $r^2 > 0,81$.

Наиболее высокие значения r^2 получены для КФП миграции, затем следуют КФП рождаемости. Низкие значения r^2 , полученные для КФП смертности, требуют формирования своего набора факторов для этого параметра. По частоте встречаемости среди влияющих на КФП все факторы распределены, в общем-то, равномерно, поэтому, на первый взгляд, нет причин для выбрасывания того или иного фактора.

3.2.3. Оценка параметров модели межрайонной миграции

Для идентификации прогнозной модели миграции (13) введем составной вектор коэффициентов этой модели:

$$\theta^T = (-\alpha^T, \beta_1^T, \dots, \beta_m^T) = (-\alpha_1, -\alpha_2, \dots, -\alpha_n, \beta_{10}, \dots, \beta_{1d1}, \dots, \beta_{mdm}).$$

Пусть N – объем выборки,

$$y = (y_1, y_2, \dots, y_N)^T,$$

$$= \begin{bmatrix} y_0 & \dots & y_{0-n} \\ \vdots & \ddots & \vdots \\ y_{N-1} & \dots & y_{N-n} \end{bmatrix}, \quad X_i = \begin{bmatrix} x_1^i & \dots & x_1^i \\ \vdots & \ddots & \vdots \\ x_{N-1}^i & \dots & x_{N-1-dli} \end{bmatrix}.$$

Тогда оценки:

$$\theta = \begin{bmatrix} Y^T Y & Y^T X_1 & \dots & Y^T X_m \\ X_1^T Y & X_1^T X_1 & \dots & X_1^T X_m \\ \vdots & \vdots & \ddots & \vdots \\ X_m^T Y & X_m^T X_1 & \dots & X_m^T X_m \end{bmatrix} \times [Y X_1 \dots X_m]^T y.$$

4. Прогноз

Процедура прогноза состоит из нескольких этапов. Вначале осуществляется прогноз КФП параметров модели естественного движения населения. Исходя из прогнозных значений КФП строится прогноз параметров модели. Затем осуществляется прогноз межрегиональной миграции. В результате формируется поток прибывающего и выбывающего из региона населения. По полученным данным осуществляется прогноз половозрастных структур городского и сельского населения региона.

4.1. Прогноз КФП параметров модели

Прогноз КФП параметров модели можно проводить разными способами:

1. Экспертным методом, когда дальнейшее поведение во времени какого-либо параметра назначается экспертом (экспертами) на основании его (их) личного опыта.
2. Трендовым методом, когда по предыстории поведения КФП определяется тенденция его поведения, и эта тенденция распространяется на перспективу.
3. Статистическими методами, когда последовательность значений во времени КФП рассматривается как случайный процесс, строятся статистические оценки характеристик этого процесса, по которым и оцениваются наиболее вероятные значения выбранного КФП в будущем времени.

Анализ алгоритмов прогноза КФП параметров модели произведем на примере прогноза КФП рождаемости. Имеющиеся в нашем распоряжении выборки были разбиты на две части: обучающую (до 1975 года) и проверочную (после 1970 года). По обучающей выборке строились линейные факторные экстраполяторы:

$$R_j(t+1) = \sum_{i=1}^5 a_{ji} F_i(t+1),$$

где R_j – один из КФП рождаемости, $\{F_i\}$ – множество факторов, вошедших в лучшую модель для j -го фактора, a_{ji} – коэффициенты фильтра. Затем на проверочной выборке оценивалось качество работы экстраполятора. Выяснилось, что при чисто факторном прогнозе работоспособным является экстраполятор для КФП R_1

$$R_1 = 27,9779 - 0,08516 \times F_9 - 0,002346 \times F_5 + 0,00100 \times F_4.$$

Результаты его работы приведены на рис. 10. Этот фильтр дает хороший прогноз на проверочной выборке

$$\max \delta(t) = \max \left(\text{abs} \left(R_1(t) - \hat{R}_1(t) \right) / R_{1(t)} \right) = 0,13 / 25,5 \cong 0,005.$$

Для КФП $R_2 - R_5$ были построены следующие авторегрессионные экстраполяторы:

$$R_2(t+1) = 0,5411 \times R_2(t) + 0,4355 \times R_2(t-1) + 0,1771,$$

$$R_3(t+1) = 1,4281 \times R_3(t) - 0,4350 \times R_3(t-1),$$

$$R_4(t+1) = 0,7626 \times R_4(t) + 0,171 \times R_4(t-1) + 0,1203,$$

$$R_5(t+1) = 0,7126 \times R_5(t) + 0,2453 \times R_5(t-1) + 0,0430.$$

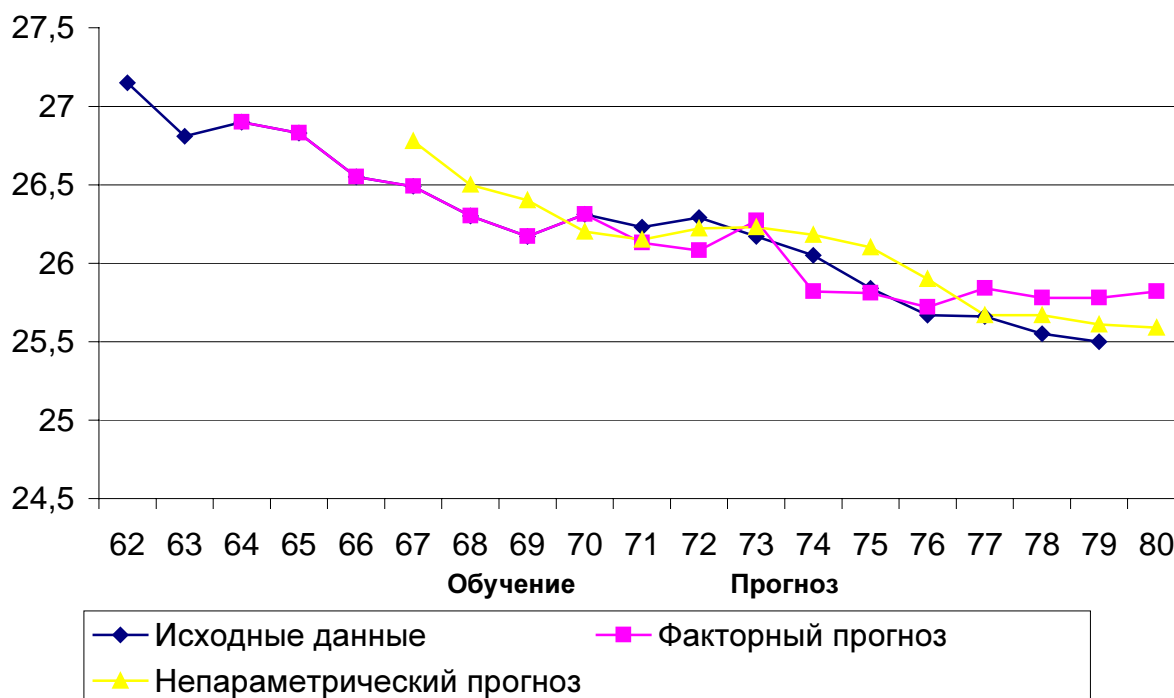


Рис. 10. Динамика КФП функции рождаемости R_1 . Реальные данные и прогноз.

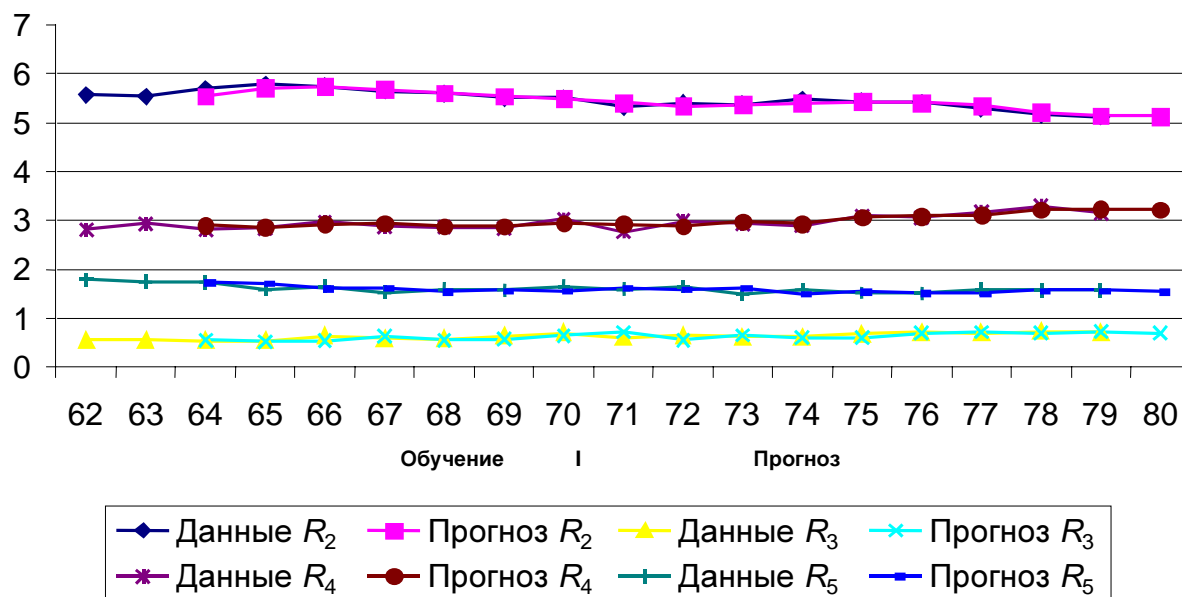


Рис. 11. Реальные данные и прогноз КФП функции рождаемости.

Результаты работы этих экстраполяторов приведены на рис. 11. Видно, что они достаточно хорошо отслеживают эволюцию КФП. Самые большие ошибки получились при прогнозе КФП R_3 :

$$\max \delta_{R_3} = \max |(R_3 - \hat{R}_3) / R_3| = 0,125 / 0,6 \cong 0,28.$$

На рис. 12 показано сравнение реальных и прогнозных кривых $f(t, y)$ и $\hat{f}(t, y)$. Видно, что ошибка при $t = 1979$ незначительна, а для $t = 1974$ (один из

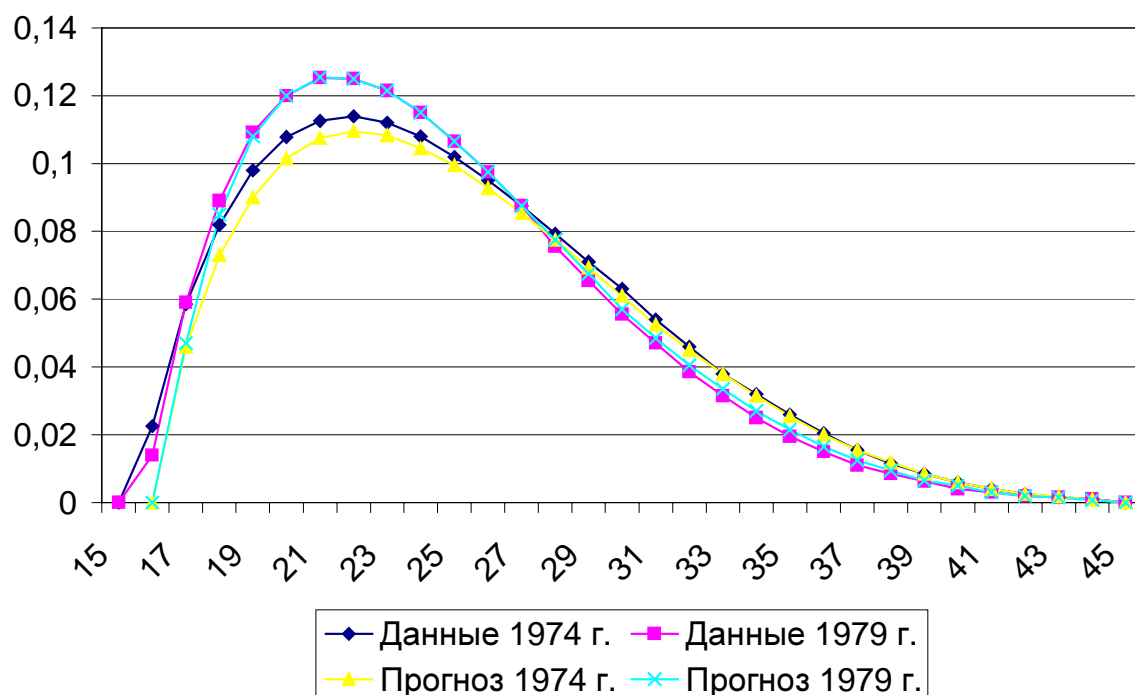


Рис. 12. Сравнение фактических данных и прогноза функции рождаемости за 1974 и 1979 гг.

плохих прогнозов, судя по рис. 12) относительная ошибка для возрастов 18-45 не превышает 0,1.

Прогноз других демографических показателей производится аналогично.

4.2. Прогноз численности населения региона

С целью оценки качества работы модели были осуществлены ретроспективные прогнозы населения Томской области. Основой их являлись данные текущей статистики населения и восстановленные возрастные структуры населения городской местности. В качестве базовых были взяты оценки демографических показателей за годы $t_b = 1951-1996$. От базовых лет осуществлялся прогноз на $\tau = 1, 2, \dots, 15$ лет. Прогнозные ВС сравнивались с восстановленными и вычислялись ошибки прогноза для каждой группы населения возраста x : $\varepsilon(t_b, x, \tau)$. Затем вычислялись статистические оценки ошибок: $m(\varepsilon(x, \tau))$ – среднее, $D(\varepsilon(x, \tau))$ – дисперсия, $\max(\varepsilon(x, \tau))$ – максимальное значение. Эти характеристики использовались при анализе качества прогноза.

Подробно было исследовано поведение $m(\varepsilon(x, \tau))$ и $D(\varepsilon(x, \tau))$ для модели, в которой предполагалась неизменность во времени параметров f, μ, ξ, ζ, I, Q . Эта модель дала хорошие результаты при краткосрочном (1-5 лет) прогнозе всех возрастных групп (ошибка находится в пределах 5%) и среднесрочном прогнозе возрастных групп в диапазоне 30-70 лет. Самые большие ошибки на-

блюдались для возрастных групп 0-10, 16-20, 90 лет и старше. Для младших возрастных групп большие ошибки наблюдаются, по-видимому, из-за неизменности во времени функции f . Возрастные группы 16-20 лет обладают наибольшей миграционной активностью, для них необходимо отслеживать изменение во времени функций ξ , ζ . Относительно большие ошибки в численности населения старческих возрастов объясняются их малочисленностью.

Таким образом, можно считать, что эта простая модель может быть использована для краткосрочных прогнозов длительностью не более одной пятилетки.

5. Оценивание расселения

При решении задачи распределения резидентного населения [25] по территории города будем считать, что нам известно соответствующее территориальное размещение зданий различных типов. Такая информация легко задается, например, при стандартном определении темы «здания» в геоинформационной системе типа Arc/Info, ArcView, когда контуры зданий с их координатной привязкой объединены в одном слое системы, а их характеристики сведены в соответствующей данному слою таблице атрибутивных данных. Атрибутивные описания зданий обычно включают этажность, материал стен, год постройки и другие данные. Каждому зданию соответствует один контур, а ему – строка базы данных. Для темы «планировочные элементы» геометрическая составляющая объектов включает их контуры, а в атрибутивной хранятся их названия, даты образования и другие. Обозначим K_i – контур i -го объекта, а a_{ij} – значение j -й характеристики i -го объекта. Несколько упрощена тема $T = \langle T_i \rangle$, где $T_i = \langle K_i, A_i \rangle$ – составная строка, K_i – строка таблицы геометрических характеристик, $A_i = (a_{i1}, a_{i2}, \dots, a_{in})$ – строка таблицы атрибутивного описания. Для геометрических компонентов тем определены операторы пространственного анализа [26]. Нам в дальнейшем понадобятся операторы объединения: $P_o(K^1, K^2) = \langle K | K^1 \cup K^2 \rangle$ и пересечения: $P_p(K^1, K^2) = \langle K | K^1 \cap K^2 \rangle$.

В указанных ГИС обязательными атрибутами таблицы геометрических характеристик полигонной темы являются периметр и площадь замкнутого контура. Предположим, что для каждого K_i темы «здания» определены его общая площадь $s_i = a_{i2}$, этажность, тип здания. Будем предполагать, что плотность заселения или обеспеченность (средний размер площади, приходящейся на одного человека) различаются по типам жилых зданий. Действительно, средняя плотность заселения объективно различна для таких групп жилых помещений, как общежития, бараки, индивидуальные дома, многоэтажные дома массовой застройки (здесь также возможна дифференциация), дома улучшенной планировки, коттеджи. Подобные данные широко используются в зарубеж-

ной статистике населения для текущего локального статистического учета [27]. Некоторые сводные цифры обеспеченности зафиксированы переписью 1989 года [28].

Пусть $p(l)$ – неизвестная для нас плотность заселения жилых помещений группы l .

Нам потребуется определить операции для СУБД и операторы пространственного анализа. Пусть a_{ki} – тип здания. Для выборки из таблицы $[a_{ij}]$ группы зданий типа l потребуется операция селекции:

$$A^l = \sigma_l(a) = \langle A_i | a_{ij} = l \rangle.$$

Каждой селектированной строке соответствует контур K_i^l .

Кроме того, для выделения зданий внутри контура некоторого планировочного элемента M^r потребуется пространственное пересечение контуров зданий и таких элементов:

$$\forall i: K_i^l(r) = \langle K_i^l \cap M^r \neq 0 \rangle,$$

$$K(r) = \langle i | K_i^l \cap M^r \neq 0 \rangle,$$

$$K^l(r) = P_p(K(l), M^r).$$

Тогда численность населения $N(r)$ микрорайона определяется:

$$N(r) = \sum_{i \in K(r)} p(l) s_i(l). \quad (25)$$

В действительности, довольно часто по моделям (1)-(21) может быть рассчитан прогноз общей численности населения города $N(t) = U(t) + V(t)$. Однако известна только общая площадь жилого фонда l -го типа, но неизвестны плотности $p(l, t)$ заселения интересующих нас групп зданий типа l .

Если эти плотности отсутствуют в явном виде, то пространственное распределение населения $N(t)$ может быть косвенным образом восстановлено по обучающей выборке. В простейшем случае предположим, что изменение со временем плотностей заселения пренебрежимо мало $p(l, t) = p(l)$.

Примем в качестве критерия приближения величину отклонения расчетной численности населения с учетом пространственного распределения от фактической:

$$\left| \sum_l p(l) \sum_i s_i(l, t) - N(t) \right| \leq \varepsilon,$$

где ε – принятая допустимая погрешность, S – общая площадь жилых помещений типа l , $t = 1, 2, \dots, T$ – перенумерованные временные отсчеты выборки.

Тогда для оценки плотностей можно использовать конечно-сходящиеся алгоритмы решения систем неравенств (КСА) [29], которые будут иметь вид

$$p[k+1, l] = p[k, l] + \gamma_k \frac{\left| \sum_l \sum p[k, l] \times S(k, l) - N(k) \right| \times S(k, l)}{\sum_l S(k, l) \times S(k, l)}.$$

Здесь $|f| = \text{abs}(f)$, $\gamma_k \leq 1$, k – номер итерации, $k = 1, 2, \dots$. При использовании (КСА) исходная выборка может быть размножена, временные отсчеты произвольно перемешаны и перенумерованы. В качестве начального приближения оценок $p(1, l)$ можно взять данные переписей [30]. В переписи 1989 года, например, присутствуют данные распределения семей по размеру жилой площади и по домам разного типа. Соответствующие плотности заселения жилых помещений могут быть рассчитаны аналогично и задействованы совместно с прогнозом семейной структуры, который рассчитывается на основании прогноза половозрастных структур населения [11].

Оценки плотностей, полученные (КСА), затем используются для расчета распределения населения по планировочным элементам.

Отметим, что в данной работе приведена схема получения оценок, а более тонкие вопросы, касающиеся, например, определения допустимой погрешности ε , вида стягивающей последовательности γ_k , учета условных неравенств, требуют отдельной публикации.

Помимо неравенств в данном случае можно воспользоваться и энтропийным методом [31].

ЛИТЕРАТУРА

1. Ходжаев Д.Г., Вишнякова В.С., Глабина Н.К. Эффективность расселения. – М.: Мысль, 1983. – 276 с.
2. Гейци И.И., Карпачев Г.И., Лавров Н.Г., Орлов А.П. Распределенные информационные системы территориального управления. – Новосибирск: ВЦ СО АН СССР, 1989. – 248 с.
3. Лексин В.Н., Швецов В.Н. Государство и регионы. – М.: УРСС, 1997. – 372 с.
4. Холл П. Городское и региональное планирование. – М.: Стройиздат, 1993. – 248 с.
5. Яргина З.Н., Хачатрянц К.К. Социальные основы архитектурного проектирования. – М.: Стройиздат, 1990. – 340 с.

6. Rainford P., Masser I. Population forecasting and urban planning practice: a case study // *Environment and planning A*, 1987, Vol. 19, p. 1463-1475.
7. Рутгайзер В.М. Социальная сфера. – М.: Экономика, 1989. – 238 с.
8. Рыбаковский Л.Л. Методологические вопросы прогнозирования населения. – М.: Статистика, 1978. – 207 с.
9. Курс демографии / Под ред. А.Я. Боярского. 2-е изд. – М.: Статистика, 1974. – 454 с.
10. Староверов О.В. Модели движения населения. – М.: Наука, 1979. – 342 с.
11. Староверов О.В. Воспроизводство структуры населения и браков. – Экономика и математические методы, 1977, т. 13, вып. 1, с. 72-82.
12. Матлин И.С. Моделирование размещения населения. – М.: Наука, 1975. – 167 с.
13. Пресса Р. Народонаселение и его изучение. – М.: Наука, 1966. – 444 с.
14. Тябаев Е.С. Идентификация и прогнозирование параметров региональной модели естественного движения населения // Модели и системы управления трудовыми ресурсами в территориальной АСУ. – Томск: Изд-во Том. ун-та, 1983, с. 75-83.
15. Зуев Г.М., Сороко Е.Л. Математическое описание миграционных процессов // *Автоматика и телемеханика*, 1978, № 7, с. 94-101.
16. Сухова Л.С. Корреляционный анализ факторов рождаемости // Проблемы демографии и народного благосостояния. – М.: Московский экономико-статистический институт, 1976, с. 21-25.
17. Тябаев Е.С. Региональная модель естественного движения населения // Модели и системы управления трудовыми ресурсами в территориальной АСУ. – Томск: Изд-во Том. ун-та, 1983, с. 63-74.
18. Рюмкин А.И. Динамическая модель межрайонной миграции с учетом факторов // Модели и системы управления трудовыми ресурсами в территориальной АСУ. – Томск: Изд-во Том. ун-та, 1983, с. 94-110.
19. Тябаев Е.С., Рюмкин А.И. Пакет программ демографических прогнозов // Управляющие системы и машины, 1983, № 4, с. 117-120.
20. Митропольский Е.М. Техника статистических вычислений. – М.: Наука, 1976. – 576 с.
21. Hoem Jan M., Berg E. Some problem in Hadwiger fertility graduation // *Scan. Actual J.*, 1975, No. 3, p. 129-144.
22. Periti P. Aging and Age-dependent Mortality of Human Populations // *Mathematical Biosciences*, 1973, Vol. 17, p. 371-374.

23. Тарасенко Ф.П. Непараметрическая статистика. – Томск: Изд-во Том. ун-та, 1976. – 292 с.
24. Епанечников В.А. Непараметрическая оценка многомерной плотности вероятностей // Теория вероятностей и ее приложения, 1969, т. 14, вып. 1, с. 156-160.
25. Мерфи Р. Американский город. – М.: Прогресс, 1972. – 320 с.
26. Кошкарев А.В., Тикунов В.С. Геоинформатика. – М.: Картгеоцентр, 1993. – 213 с.
27. Smith S.K. A Review and Evaluation of the Housing Unit Method of Population Estimation // Journal of American Statistical Association. 1986, Vol. 81, No. 394.
28. Жилищные условия населения Томской области: Статистический сборник № 5 по данным Всесоюзных переписей населения 1979, 1989 гг. – Томск, 1991. – 188 с.
29. Якубович В.А. Рекуррентные конечно–сходящиеся алгоритмы решения бесконечных систем неравенств // Доклады АН СССР. 1966, т. 166, № 6, с. 1308-1311.
30. Возрастной состав и состояние в браке населения Томской области: Статистический сборник № 1 по данным Всесоюзных переписей населения 1979, 1989 гг. – Томск, 1990. – 71 с.
31. Сафарова Г.Л., Макаров Н.Н., Чистякова Н.Е. Энтропийная модель семейной структуры населения // Проблемы прогнозирования. – 1986, № 6, с. 108-115.