ГЛАДКИЕ МОДЕЛИ БИОЛОГИЧЕСКИХ ПОПУЛЯЦИЙ

Е.В. Лобанова, Н.Б. Медведева

Предложен метод построения моделей, выражающих численность биологических популяций, на основе временных рядов. На первом этапе строится сглаженный набор эмпирических данных, который отражает общие черты реального временного ряда. Это достигается посредством построения оптимизационного сплайна – кусочнополиномиальной функции, имеющей минимальное отклонение от эмпирических данных по методу наименьших квадратов. Далее строится система дифференциальных уравнений, правая часть которой имеет наименьшее отклонение по методу наименьших квадратов от производной оптимизационного сплайна на некоторой более частой сетке. Решение задачи Коши для построенной системы на тестовом промежутке времени берется в качестве прогноза модели. Метод применяется к конкретным временным рядам, делается оценка погрешности прогноза, исследуется зависимость погрешности от параметров метода. Кроме того метод применяется к искусственному временному ряду, содержащему случайные возмущения. Исследуется зависимость погрешности прогноза от величины возмущения.

Ключевые слова: временной ряд; модель популяции; сплайн; метод наименьших квадратов.

Введение

В данной работе строятся модели, выражающие численность биологических популяций, на основе анализа временных рядов [1 – 6].

Особенность временных рядов в экологии состоит в том, что, во-первых, для исследования доступно незначительное число значений, поскольку, как правило, представленные эмпирические данные – это результаты измерений в течение 30 – 50 лет. Во-вторых – малая точность, потому как практически невероятно измерить численность всей популяции, вплоть до отдельной особи, на большом ареале ее обитания. В-третьих – реакция биосферы и ее объектов на внешние воздействия носит избирательный характер, т.е. объекты биосферы не реагирует на все воздействия одновременно, при этом один и тот же объект биосферы в разные интервалы времени может реагировать на одинаковые воздействия по-разному, характер и сила наблюдаемой реакции могут не соответствовать параметрам внешних воздействий. Отсюда – возникновение сложностей при построении вида модели.

В связи с вышеизложенным, прежде чем приступить к моделированию по временному ряду, строится сглаженный набор эмпирических данных, который отражает лишь общие черты реального временного ряда. Это достигается посредством построения оптимизационного сплайна: весь временной интервал, в пределах которого лежат измерения наблюдаемой величины, делится на несколько равных промежутков, на каждом промежутке строится свой полином заданной степени, исходя из следующих условий: на границах временных интервалов значения соседних полиномов, а также их последовательных производных до некоторого фиксированного порядка должны совпадать. Более того, кусочный сплайн, составленный из этих полиномов, должен быть экстремальным в своем классе, исходя из условия оптимального отклонения от эмпирических данных на всем временном промежутке по методу наименьших квадратов. Далее, к сглаженному временному ряду применяется аналог метода Безручко – Смирнова [7], заключающегося в построении системы дифференциальных уравнений, решение задачи Коши для которой моделирует зависимость, выражаемую данным временным рядом. В данной статье описанные методы применяются к конкретным временным рядам, выражающим численность биологических популяций, вычисляется погрешность прогноза, исследуется зависимость погрешности от некоторых параметров метода. Кроме того, метод применяется к искусственному временному ряду, содержащему случайные возмущения, исследуется зависимость погрешности прогноза от величины возмущения.

1. Сглаживание эмпирических данных

Перечислим основные элементы процедуры построения оптимизационного сплайна. Пусть задан скалярный временной ряд $\{\nu_i\}_{i=0}^M$, задающий последовательность значений размера популяции на протяжении определенного количества лет, $t_i = i$ $(i = 0, \ldots, M)$ – временная сетка, M – натуральное число. Предположим, что M = lk, где $l \ge 1, k \ge 1$ – натуральные числа. Тем самым весь временной интервал I = [0, M], в пределах которого лежат измерения наблюдаемой величины, делится на k равных промежутков $I = \bigcup_{i=1}^k I_i$, на каждом из которых доступны, следовательно, l + 1 значений. По этим значениям строятся полиномы $\sigma_{ni}(t) = \sum_{j=0}^n c_{ij} \cdot t^j$ заданной степени n, исходя из следующих соображений: в каждой граничной точке $t_{i\cdot l}$ промежутков I_i соседние полиномы должны принимать одина-ковые значения вместе со всеми своими производными до заданного порядка d:

$$\sigma_i(t_{i\cdot l}) = \sigma_{i+1}(t_{i\cdot l}), \quad \sigma'_i(t_{i\cdot l}) = \sigma'_{i+1}(t_{i\cdot l}), \dots, \sigma_i^{(d)}(t_{i\cdot l}) = \sigma_{i+1}^{(d)}(t_{i\cdot l}), \quad i = 1, \dots, k-1.$$

Кусочный сплайн $\sigma_n(t)$, составленный из функций $\sigma_{n1}(t), \ldots, \sigma_{nk}(t)$:

$$\sigma_n(t) = \begin{cases} \sigma_{n1}(t), & t \in I_1, \\ \sigma_{n2}(t), & t \in I_2, \\ \dots & \\ \sigma_{nk}(t), & t \in I_k, \end{cases}$$
(1)

должен быть экстремальным в своем классе, исходя из условия оптимального отклонения сплайна от эмпирических данных и его производных от нуля на всем временном промежутке:

$$\sum_{i=0}^{M} (\sigma_n(t_i) - \nu_i)^2 + \sum_{s=1}^{d} \sum_{i=0}^{M} (\sigma_n^{(s)}(t_i))^2 \to \min$$

Такой способ построения позволяет однозначно определить все коэффициенты $c_{10}, ..., c_{kn}$ оптимизационного сплайна.

2. Реконструкция систем дифференциальных уравнений

Метод моделирования по временным рядам с помощью построения системы обыкновенных дифференциальных уравнений, разработанный Безручко – Смирновым [7], состоит в следующем. В случае, если исходный временной ряд – векторный размерности *D*, то вычисляются разностные производные первого порядка компонент этого ряда. Модель строится в виде задачи Коши для системы дифференциальных уравнений

$$\begin{cases} \dot{x}_1(t) = F_1(x_1, x_2, ..., x_D), \\ \dot{x}_2(t) = F_2(x_1, x_2, ..., x_D), \\ ... \\ \dot{x}_D(t) = F_D(x_1, x_2, ..., x_D), \end{cases}$$
(2)

в которой F_j , $j = 1 \dots D$ – многочлены некоторой фиксированной степени K с неопределенными коэффициентами, которые ищутся из условия минимального отклонения разностных

производных компонент временного ряда от соответствующих компонент правой части системы (2) по методу наименьших квадратов.

В случае же, если исходный временной ряд – скалярный, модель строится в виде задачи Коши для системы обыкновенных дифференциальных уравнений некоторой размерности D>1

$$\begin{array}{l}
x_1 = x_2, \\
\dot{x}_2 = x_3, \\
\dot{x}_3 = x_4, \\
\dots, \\
\dot{x}_D = F(x_1, \dots, x_D),
\end{array}$$
(3)

где $F(x_1, \ldots, x_D)$ – многочлен степени K, коэффициенты которого ищутся методом наименьших квадратов из условия минимального отклонения разностной производной временного ряда порядка D от значений многочлена F на наборе разностных производных временного ряда до порядка D-1 включительно. Вопрос о выборе размерности D требует специального обсуждения.

Метод Безручко – Смирнова показал свою эффективность при моделировании временных рядов, возникающих при измерениях в технических устройствах, содержащих сотни и даже тысячи измерений.

3. Схема построения модели

Первый этап построения модели популяции по временному ряду – это нахождение оптимизационного сплайна, который может быть как векторным, так и скалярным. Далее применяется аналог метода Безручко – Смирнова к дисретизированному сплайну, для чего временной промежуток делится на большое количество частей N – порядка сотен или даже тысяч, таким образом получается новый сглаженный временной ряд, производные которого в точках сетки вычисляются в отличие от метода Безручко-Смирнова аналитическим дифференцированием оптимизационного сплайна в системе MAPLE.

Если временной ряд – векторный размерности *D*, то модель ищется в виде задачи Коши для системы (2), где

$$F_j(x_1, x_2, ..., x_D) = \sum_{l_1, l_2, ..., l_D=0}^K c_{j, l_1, l_2, ..., l_D} \prod_{k=1}^D x_k^{l_k}, \quad \sum_{k=1}^D l_k \le K,$$

— полиномы степени не выше K, коэффициенты которых ищутся для каждого j из условия минимума

$$\sum_{i=0}^{N} (F_j(\sigma_n^1(h \cdot i), \dots, \sigma_n^D(h \cdot i)) - (\sigma_n^j)'(h \cdot i))^2 \to \min,$$

где σ_n^j – оптимизационный сплайн для *j*-й компоненты временного ряда, h = M/N.

Когда наблюдению доступен лишь скалярный временной ряд, процедура построения системы (3) включает в себя еще один предварительный этап (по сравнению с моделированием по векторному ряду) – выбор размерности модели D, который в простейшем случае можно осуществить на основе перебора. В данной статье мы исследуем зависимость погрешности прогноза от величины D. Модель в этом случае ищется в виде задачи Коши для системы (3), где

$$F(x_1, x_2, \dots, x_D) = \sum_{l_1, l_2, \dots, l_D=0}^{K} c_{l_1, l_2, \dots, l_D} \prod_{k=1}^{D} x_k^{l_k}, \quad \sum_{k=1}^{D} l_k \le K,$$
(4)

– многочлен степени К, коэффициенты которого находятся из условия минимума

$$\sum_{i=0}^{N} (F(\sigma_n(h \cdot i), \sigma'_n(h \cdot i), \dots, \sigma_n^{(D-1)}(h \cdot i)) - \sigma_n^{(D)}(h \cdot i))^2 \to \min$$

Качество построенной модели исследуется с точки зрения адекватности прогноза. Промежуток определения временного ряда делится на две части – тренировочную и тестовую: $M = M_{train} + M_{test}$. Для исследования погрешности прогноза оптимизационный сплайн строится не на всем промежутке I = [0, M], а на промежутке $[0, M_{train}]$, далее на этом же промежутке строится система дифференциальных уравнений (2) или (3). Далее ищется численное решение задачи Коши для этой системы дифференциальных уравнений на всем промежутке I = [0, M]. Численное решение сравнивается с исходными эмпирическими данными на тестовом промежутке $[M_{train}, M]$. Ошибка прогноза в случае скалярного временного ряда вычисляется по формуле

$$\varepsilon = \max_{t_i \in [M_{train}, M]} |x_1(t_i) - \nu_i|, \tag{5}$$

а в случае векторного – по аналогичной формуле для каждой компоненты.

Программа построения оптимизационного сплайна, модельной системы дифференциальных уравнений, решения соответствующей задачи Коши, построения графиков и вычисления погрешности прогноза была реализована в системе MAPLE 14.

4. Реализация метода в случае одновидовой популяции

Приведем ряд демонстрационных примеров реализации вышеизложенного метода с использованием математического пакета Maple 14. В качестве одновидовой системы была выбрана популяция цапли Ardea cineria на юге Ланкашира, Англия, с указанными в табл. 1 показателями на протяжении 30 лет [8].

Таблица 1

Год	1912	1913	1914	1915	1916	1917	1918	1919	1920	1921
Число	20	21	22	25	29	34	37	38	37	37
пар										
Год	1922	1923	1924	1925	1926	1927	1928	1929	1930	1931
Число	37	40	43	45	46	46	46	47	49	51
пар										
Год	1932	1933	1934	1935	1936	1937	1938	1939	1940	1941
Число	52	52	51	52	53	55	56	57	57	58
пар										

Показатели численности популяции цапли

На основании приведенных данных был построен оптимизационный сплайн, сглаживающий эмпирические данные, на обучающем промежутке длины $M_{train} = 22$ года. В качестве степеней полиномов, образующих данный сплайн, была выбрана степень n = 9, уровень гладкости сплайна d = 3. Весь обучающий временной интервал был разделен на k = 7 равных частей по l + 1 = 4 измерения в каждом. График полученного сплайна $\sigma_9(t)$ изображен



Рис. 1. График оптимизационного сплайна $\sigma(t)$ для популяции цапли Ardea cineria на юге Ланкашира, Англия, 1912 – 1941 гг.

на рис. 1. Затем выбранный промежуток был разбит на N = 300 частей, и полученные значения легли в основу нахождения методом наименьших квадратов коэффициентов системы вида (3) размерности D = 3 со степенью полинома (4) в правой части, равной K = 2:

$$\begin{cases} \dot{x}_1 = x_2, \\ \dot{x}_2 = x_3, \\ \dot{x}_3 = F(x_1, x_2, x_3), \end{cases}$$
(6)

где $F(x_1, x_2, x_3) = c_0 + c_1 x_1 + c_2 x_2 + c_3 x_1 x_2 + c_4 x_1^2 + c_5 x_2^2 + c_6 x_1 x_3 + c_7 x_2 x_3 + c_8 x_3^2$.

После того, как все коэффициенты $c_0, ..., c_8$ были найдены из условия:

$$\sum_{i=0}^{N} (F(\sigma_9(h \cdot i), \sigma'_9(h \cdot i), \sigma''_9(\frac{M}{N} \cdot i)) - \sigma'''_9(\frac{M}{N} \cdot i))^2 \to \min,$$
(7)

где h = M/N = 0, 1, численным методом было получено решение системы (6) с начальными условиями $x_1(0) = 20, x_2(0) = \sigma'_9(0), x_3(0) = \sigma''_9(0)$. Функция $x_1(t)$, является на тестовом промежутке прогнозом численности популяции английской цапли (рис. 2). Подсчитано, что в данном случае погрешность прогноза ε , вычисленная по формуле (5) не превышает двух особей.

Как можно увидеть из рис. 2 графики оптимизационного сплайна и найденного решения системы (6) практически совпадают на тренировочном промежутке, а график прогноза визуально хорошо совпадает со значениями исходной выборки на тестовом промежутке.

Для указанных выше начальных данных и значений n, d, l и D была исследована зависимость погрешности прогноза от частоты разбиения на N отрезков, по которым осуществлялось нахождение коэффициентов в условии (7). Из полученных данных (рис. 3) можно сделать вывод, что стабильное уменьшение ошибки прогноза и, следовательно, увеличение его точности достигается уже при значениях N > 500.



Рис. 2. Прогноз численности популяции цапли Ardea cineria на юге Ланкашира, Англия, 1912 – 1941 гг.

Как показали также аналогичные исследования для двухвидовой популяции, существуют небольшие значения N, при которых достоверность прогноза достаточно велика (здесь это достигается, к примеру, при N = 100, N = 275, N = 400).



Рис. 3. Зависимость погрешности прогноза от частоты разбиения N для численности цапли Ardea cineria на юге Ланкашира, Англия, 1912 – 1941 гг.

Наконец, реализация метода для различных размерностей D системы вида (3) при n = 9, d = 3, l + 1 = 4, N = 300 привела к следующему результату: самой оптимальной в плане

точности прогноза стала размерность D = 3, что соответствует теоретическим результатам Такенса о вложениях [9].

Проведенные эксперименты со степенью полиномов n, составляющих оптимизационный сплайн $\sigma_n(t)$ и построенных при наличии l + 1 = 2 измерений на одном отрезке разбиения всего временного промежутка, показали, что с ростом степени n растет точность прогноза численности исследуемой популяции.

5. Исследование метода с помощью искусственной выборки



Рис. 4. График сплайна $\sigma_9(t)$ изображен тонкой линией, жирной линией изображено решение системы (6) – прогноз для искусственного временного ряда $x(t) = \sin(0, 4 \cdot t)$ с уровнем шума, не превосходящим 4%

Для исследования зависимости точности прогноза от величины вводимых шумов мы рассмотрели искусственный временной ряд, получаемый из функции $x(t) = \sin(0, 4 \cdot t)$ путем добавления случайных добавок величины δ на целочисленной сетке $t_i = i, i = 0, ..., 30$, (ромбики на рисунке 4). Аналогично случаю одновидовой системы был построен сплайн на тренировочном промежутке [0, 22], затем система вида (6). Исследование влияния добавленных шумов в данную конструкцию выявило характерное снижение неточности прогноза при уменьшении уровня зашумленности временного ряда. Зависимость погрешности прогноза ε от шумов δ представлена на рис. 5.

6. Применение метода к двухвидовой популяции

Были проведены эксперименты с данными о численности популяций рысей и зайцев [8], представленными в табл. 2. В качестве модели взаимодействия двух популяций мы рассмотрели задачу Коши для системы обыкновенных дифференциальных уравнений вида (2)



Рис. 5. Зависимость погрешности прогноза ε для искусственной системы $x(t) = \sin(0, 4 \cdot t)$ от уровня вводимого шума δ

Таблица 2

Показатели численностей популяций рысей и зайцев

Год	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959
Число пар										
рысей / зайцев	35/20	33/21	35/18	21/25	11/26	10/28	9/38	8/39	9/49	14/67
Год	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969
Число пар										
рысей / зайцев	18/86	24/96	22/90	31/93	37/82	34/65	30/59	24/51	19/49	20/54
Год	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979
Число пар										
рысей / зайцев	14/57	15/63	18/80	23/86	27/96	29/94	30/81	33/75	35/64	33/62

размерности D=2

$$\begin{cases} \dot{x}_1 = F_1(x_1, x_2), \\ \dot{x}_2 = F_2(x_1, x_2), \end{cases}$$
(8)

где $F_1(x_1, x_2)$ и $F_2(x_1, x_2)$ – многочлены степени K = 3 от двух переменных:

$$F_j(x_1, x_2) = \sum_{l_1+l_2 \le 3} c_{j,l_1,l_2} x_1^{l_1} x_2^{l_2}, \ j = 1, 2.$$

Аналогично одномерному случаю мы взяли для каждой компоненты временного ряда степень полиномов, образующих оптимизационный сплайн, равной n = 9 с уровнем гладкости d = 3, разбиение обучающего временного интервала на равные части по l + 1 = 4измерения в каждом. Для построения системы (8) тренировочный промежуток был разбит на N = 300 частей. Численно было построено решение системы (8) с начальными условиями $x_1(0) = 20, x_2(0) = 35$. Полученные результаты отражены на рис. 6.

Исследование зависимости погрешности прогноза от степени полиномов K в правой части (8) показало, что возрастание степени правой части К ведет, во-первых, к сильному росту времени счета, делая его иногда невозможным, а во-вторых, вообще говоря, не ведет



Рис. 6. Прогноз численности популяций рысей и зайцев. Оптимизационные сплайны обозначены толстыми линиями, прогнозы – тонкими линиями

к увеличению точности прогноза. В обоих случаях наилучший результат был достигнут при K = 3. Исследование зависимости погрешности прогноза, рассчитанной по формуле (5), от степени оптимизирующего сплайна n и в этом случае показало, что, как правило, чем выше его степень, тем надежнее прогноз.

Работа проводилась при финансовой поддержке РФФИ, грант 13-01-00512.

Литература

- 1. Crutchfield, J.P. Equations of Motion from a Data Series / J.P. Crutchfield , B.S. McNamara // Complex Systems. 1987. V. 1. P. 417–452.
- Cremers, J. Construction of Differential Equations from Experimental Data / J. Cremers, A. Hubler // Z. Naturforschung A. – 1987. – V. 42. – P. 797–802.
- Gouesbet, G. Construction of Phenomenological Models from Numeri-Cal Scalar Time Series / G. Gouesbet, J. Maquet // Physica D. – 1992. – V. 58. – P. 202–215.
- Gouesbet, G. Global Vector-Field Approximation by Using a Multi-Variate Polynomial Approximation on Nets / G. Gouesbet , C. Letellier // Phys.Rev. E. – 1994. – V. 49. – P. 4955–4972.
- Восстановление структуры динамической системы по временным рядам /Д.А. Грибков, В.В. Грибкова, Ю.А. Кравцов, Ю.И. Кузнецов и др. // Радиотехника и электроника. – 1994. – Т. 39, вып. 2. – С. 269–277.
- Янсон, Н.Б. Моделирование динамических систем по экспериментальным данным / Н.Б. Янсон, В.С. Анищенко // Изв. ВУЗов. Прикладная нелинейная динамика. – 1995. – Т. 3, № 3. – С. 112–121.
- Безручко, Б.П. Реконструкция обыкновенных дифференциальных уравнений по временным рядам / Б.П. Безручко, Д.А. Смирнов. – Саратов: ГосУНЦ «Колледж», 2000. – 40 с.
- 8. Уильямсон, М. Анализ биологических популяций /М. Уильямсон. Москва: Мир, 1975.

9. Takens, F. Detecting Strange Attractors in Turbulence, in Dynamical Systems and Turbulence, Warwick. eds. D.Rang and L.S.Young / F. Takens // Lecture Notes in Mathematics. -1980. - V. 898. - P. 366-381.

Екатерина Валерьевна Лобанова, магистрант, кафедра «Вычислительная математика», Челябинский государственный университет (г. Челябинск, Российская Федерация), leo4prada@gmail.com.

Наталия Борисовна Медведева, доктор физико-математических наук, профессор, кафедра «Вычислительная математика», Челябинский государственный университет (г. Челябинск, Российская Федерация); кафедра «Дифференциальные и стохастические уравнения», Южно-Уральский государственный университет (г. Челябинск, Российская Федерация), medv@csu.ru.

Поступила в редакцию 11 мая 2013 г.

Bulletin of the South Ural State University. Series "Mathematical Modelling, Programming & Computer Software", 2014, vol. 7, no. 2, pp. 55–65.

MSC 92D25

DOI: 10.14529/mmp140205

Smooth Models of Biological Populations

E.V. Lobanova, Chelyabinsk State University, Chelyabinsk, Russian Federation, leo4prada@gmail.com,

N.B. Medvedeva, Chelyabinsk State University, South Ural State University, Chelyabinsk, Russian Federation, medv@csu.ru

We propose a method for constructing models to express the size of biological populations based on time series. At the first stage we construct a smoothed-out collection of empirical data reflecting the common features of an actual time series by using an optimizing spline, which is a piecewise polynomial function at the minimal distance from the empirical data in the sense of the least-squares method. Then we construct a system of ODEs which has the minimal least-squares distance from the derivative of the optimizing spline on certain finer lattice. We take the solution to the Cauchy problem for this system as the forecast by the model. We apply the method to concrete time series, estimate the error of the forecast and study its dependence on the parameters of the method. In addition, we apply the method to an artificial time series containing random perturbations. We study the dependence of error in the forecast on the size of random perturbation.

Keywords: time series; model of population; spline; least-squares method.

References

- Crutchfield J.P., McNamara B.S. Equations of Motion from a Data Series. Complex Systems, 1987, vol. 1, pp. 417–452.
- Cremers J. Hubler A. Construction of Differential Equations from Experimental Data. Z. Naturforschung A, 1987, vol. 42, pp. 797–802.
- Gouesbet G., Maquet J. Construction of Phenomenological Models from Numerical Scalar Time Series. *Physica D*, 1992, vol. 58, pp. 202–215.

- 4. Gouesbet G., Letellier C. Global Vector-Field Approximation by Using a Multi-Variate Polynomial Approximation on Nets. *Phys.Rev. E*, 1994, vol. 49, pp. 4955–4972.
- Gribkov D.A., Gribkova V.V., Kravtsov Yu.A., Kuznetsov Yu.I., Rzhanov A.G. [Restructuration of the Dynamic Dystem from Time Series]. *Radiotekhnika i elektronika*, 1994, vol. 39, issue 2, pp. 269–277. (in Russian)
- Yanson N.B., Anishchenko V.S. [Modelling Dynamical Systems from Experimental Data]. Izvestiya Vuzov. Prikladnaya nelineinaya dinamika, 1995, vol. 3, no. 3, pp. 112–121. (in Russian)
- 7. Bezruchko B.P., Smirnov D.A. *Rekonstruktsiya obyknovennykh differentsialnykh uravneniy po vremennym ryadam* [Reconstruction of Ordinary Differential Equations from Time Series]. Saratov, GosUnTs Kolledzh, 2000. 40 p.
- 8. Williamson M. The Analysis of Biological Populations. London, Edward Arnold, 1972. 272 p.
- Takens F. Detecting Strange Attractors in Turbulence, in Dynamical Systems and Turbulence, Warwick. *Lecture Notes in Mathematics*, 1980, vol. 898, pp. 366–381.

Received May 11, 2013