

Н.П. Третьяков  
(Российский государственный социальный университет;  
e-mail: trn11@rambler.ru)

## К ВОПРОСУ О МЕТОДОЛОГИИ ПРОГНОЗИРОВАНИЯ МИРОВОЙ СТАТИСТИКИ ПОЖАРОВ

**Аннотация.** Исследуются показатели пожарной обстановки: среднее число пожаров (абсолютное и в расчёте на 1000 чел.) и среднее число жертв (абсолютное, в расчёте на 100000 чел. и в расчёте на 100 пожаров) в год. Используются статистические данные Центра Пожарной Статистики Международной Ассоциации Пожарно-спасательных Служб (СТИФ).

**Ключевые слова:** статистика пожаров, прогнозирование, регрессионные уравнения, коллокации, лаговые переменные.

## N.P. Tretyakov METHODOLOGY OF FORECASTING OF WORLD FIRE STATISTICS

**Abstract.** Under consideration parameters of fire statistics: mean number of fires (total and per 1000 inh.) and mean number of fire deaths (total, per 100000 inh. and per 100 fires). Statistical data provided by Center of Fire Statistics (International Association of Fire and Rescue Services – STIF) is used.

**Key words:** secure cooperation, CIS, efficiency of cooperation, forecasting.

Статья поступила в редакцию 4 октября 2010 г.

В рамках развития концепции использования лаговых переменных для совершенствования методики прогнозирования временных рядов показателей пожарной обстановки, перспективной является *модель средней квадратичной коллокации* [5]. Под *коллокацией*, с математической точки зрения, понимается определение функции путем подбора аналитической аппроксимации к определенному числу заданных линейных функционалов.

Коллокационная модель прогнозирования сохраняет основные преимущества классических регрессионных моделей – инвариантность по отношению к линейным преобразованиям исходных данных и результатов, оптимальность решения (в смысле наиболее точного прогноза из всех возможных вариантов линейных решений на основе заданных исходных данных). В то же время, по сравнению с классическими регрессионными моделями, она имеет дополнительные *достоинства*:

1) как наблюдаемые, так и оцениваемые величины могут быть разнородными (иметь различную физическую, экономическую или математическую природу), что особенно важно при прогнозировании показателей пожарной обстановки;

2) коллокационная модель может быть использована не только для построения оптимального прогноза однородных данных, но и для оценивания лю-

бых интересующих характеристик по неоднородной исходной информации.

Основная идея метода коллокаций состоит в том, что, имея аппроксимацию автоковариационной функции в аналитическом виде (2), можно сделать ее прогноз для нескольких временных тактов вперед и затем вычислить соответствующее прогнозное значение исследуемого показателя. Интервал прогноза можно получить, если в формуле (6) воспользоваться доверительным интервалом для математического ожидания  $E(\vec{Y})$ .

### Математический алгоритм метода коллокаций

Вектор наблюдений показателя  $Y$ , имеющий размер  $N$  (число наблюдений), предполагается центрированным (среднее значение равно нулю):  $\vec{Y} = \vec{Y}_{центр} = \|Y_j\|, j = 1, \dots, N$ .

Далее вычисляются автоковариационные функции вектора  $\vec{Y}$

$$L_{yy}(r) = E(Y_t Y_{t+R}), \quad R = 0, \dots, N-1, \quad (1)$$

где  $Y_t$  – значение переменной в момент времени  $t$ ,  $R$  – лаг.

Подобным образом в данном методе вычисляются автоковариации всех возможных лаговых значений. Конструктивно это удобно делать, генерируя вспомогательную квадратичную матрицу (размера  $N \times N$ )  $\hat{Q} = \vec{Y} \cdot \vec{Y}^T$ , тогда сумма элементов  $\hat{Q}$  по главной диагонали дает значение  $L_{yy}(0)$ , сумма по второй диагонали даст  $L_{yy}(1)$  и т.д. Подобным образом получают значения  $L_{yy}(R)$  для всех  $R$ .

Далее приступают к ключевому шагу в методе коллокаций: аппроксимации автоковариационной функции  $L_{yy}(R)$  подходящей нелинейной функцией. В случае отсутствия какой-либо дополнительной информации о ней, рекомендуется использовать следующую функцию:

$$L(R) = L(0) e^{-\alpha R} \cos(\beta R). \quad (2)$$

Подбор параметров  $\alpha$  и  $\beta$  производится путем стандартной процедуры регрессии (нелинейной по параметрам). Имея значения параметров, по формуле (2) вычисляется массив "теоретических" значений  $\hat{L}(R)$ . По ним строится еще одна вспомогательная матрица:

$$\hat{K}_{yy} = \begin{bmatrix} \hat{L}(0) & \hat{L}(1) & \hat{L}(2) & \hat{L}(3) & \dots & \hat{L}(N-1) \\ 0 & \hat{L}(0) & \hat{L}(1) & \hat{L}(2) & \dots & \hat{L}(N-2) \\ 0 & 0 & \hat{L}(0) & \hat{L}(1) & \dots & \hat{L}(N-3) \\ \cdot & \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \cdot & \dots & \cdot \\ 0 & 0 & 0 & 0 & \dots & \hat{L}(0) \end{bmatrix}. \quad (3)$$

Иными словами,  $\hat{K}_{yy}$  представляет собой верхнюю треугольную матрицу с циклической структурой, т.е. все элементы одной диагонали одинаковы, а элементы нижних диагоналей равны нулю.

В том случае, когда необходимо построить прогноз на  $p$  временных тактов вперед, необходимо вычислить  $p$  значений  $L(N-1+1), L(N-1+2), \dots, L(N-1+p)$ . Далее эти значения присоединяются к концу массива  $L(R)$ , записанного в виде строки и с отброшенными первыми  $p$  значениями:

$$[L(p), L(p+1), \dots, L(N-1), L(N-1+1), \dots, L(N-1+p)] \quad (4)$$

(размер этого вектора-строки по-прежнему  $N$ ). Записывая эту строку в обратном порядке, получаем вектор  $\vec{L}_p^T$ :

$$\vec{L}_p^T = [L(N-1+p), \dots, L(N-1+2), L(N-1+1), L(N-1), L(N-2), \dots, L(p)] \quad (5)$$

Наконец, прогнозное значение показателя  $Y$  на  $p$  временных тактов вперед вычисляется по формуле

$$Y_{N+p} = \vec{L}_p^T \cdot \hat{K}_{YY}^{-1} \cdot \vec{Y}_{центр} + E(\vec{Y}) \quad (6)$$

(в последней формуле к полученному при матричном умножении значению прибавляется среднее *исходного* массива, т.к. метод коллокаций изначально дает предсказание для централизованного показателя).

### Апробация

Произведена апробация разработанного алгоритма прогнозирования на основе статистических данных Центра Пожарной Статистики Международной Ассоциации Пожарно-спасательных Служб (СТИФ). Исходные показатели приведены в табл. 1.

Таблица 1

Исходные показатели пожарной опасности

№	Страна	Население, тыс. чел.	Среднее число жертв			Среднее число пожаров в год	Среднее число пожаров на 1000 чел. в год
			в год	на 100 тыс. чел.	на 100 пожаров		
1	2	3	4	5	6	7	8
1	Китай	1321852	2206	0,17	0,18	251786	0,19
2	Индия	1129866	8500	0,75	4,25	200000	0,18
3	США	301140	3625	1,20	1,20	1613400	5,36
4	Россия	141378	18759	13,27	13,16	236698	1,67
5	Филиппины	91077	249	0,27	2,50	9877	0,11
6	Вьетнам	85262	90	0,11	0,11	2154	0,03
7	Германия	82401	479	0,58	0,26	184485	2,24
8	Турция	71159	340	0,48	0,48	59618	0,84
9	Франция	63714	455	0,71	0,13	357654	5,61
10	Великобритания	60776	532	0,87	0,11	489942	8,06
11	Италия	58148	112	0,19	0,05	211504	3,64
12	Украина	46300	3909	8,44	7,30	53546	1,16
13	Южная Африка	42880	1817	4,24	3,52	51620	1,20
14	Польша	38518	503	1,30	0,28	179815	4,67
15	Узбекистан	27780	184	0,66	1,20	15295	0,55
16	Малайзия	24821	62	0,25	0,23	27012	1,09
17	Тайвань	22859	169	0,74	2,23	7590	0,33
18	Румыния	21537	217	1,01	1,88	11957	0,56
19	Австралия	20434	122	0,60	0,60	113442	5,55
20	Казахстан	15285	125	3,43	3,03	17340	1,13
21	Греция	10706	70	0,66	0,26	27391	2,56
22	Португалия	10643	93	0,87	0,14	64560	6,07

Апробация проводилась методом ретропрогноза. Представляет интерес сравнение структуры взаимозависимости показателей, полученной по выше-описанной методике, с простейшими трендовыми оценками, использованными для прогноза показателей.

В табл. 2 приводятся значения коэффициентов детерминации  $R^2$ , являющихся основными показателями качества модели. Вычисления проводились на основе данных за период 2001-2009 гг.

Таблица 2

	БЫЛО	СТАЛО	БЫЛО	СТАЛО	Изменение
	Количество переменных (справа) в системе	Количество переменных (справа) в новой системе	$R^2$ средний по всем уравнениям системы	$R^2$ средний по всем уравнениям новой системы	$R^2$ %
Россия	46	55	77	89	15,6
Китай	44	39	75	86	14,7
Индия	45	43	74	84	13,5
США	43	45	68	79	16,2
Вьетнам	40	40	69	76	10,1
Германия	37	29	72	75	4,2
Турция	42	33	61	75	23,0
Украина	44	50	52	76	46,2
Узбекистан	44	32	72	79	9,7
<i>Среднее</i>	<i>42,8</i>	<i>40,7</i>	<i>68,9</i>	<i>79,9</i>	<i>17,0</i>

Из этих данных можно сделать однозначный вывод, что разработанная методика идентификации позволяет существенно повысить коэффициент детерминации регрессионной модели, и, что особенно важно, в большинстве случаев это сопровождается даже уменьшением общего количества переменных. В тех случаях, когда количество переменных не изменяется (Вьетнам), увеличение  $R^2$  означает, что вместо некоторых прежних переменных в уравнения вводятся новые, которые сильнее коррелируют с прогнозируемыми показателями. По Украине увеличение коэффициента детерминации составило 46 %, что говорит о слабой адекватности использованной ранее модели. Итак, разработанное усовершенствование методики позволяет улучшить качество прогнозных моделей в среднем по всем странам на 17 %.

#### Литература

1. **Брушлинский Н.Н.** Экономическая оценка борьбы с пожарами в современном мире. – М.: ВНИИПО, 1998.
2. **Брушлинский Н.Н.** Мировая пожарная статистика в конце XX века. – М.: Академия ГПС, 2000.
3. **Fire statistics.** STIF Report, № 14. Moscow-Berlin, June 2009.
4. **Бабешко Л.О.** "Коллокационные модели прогнозирования в финансовой сфере", М., Экзамен, 2001.