

Д. т. н. Ю. А. ДОЛГОВ, к. т. н. А. Ю. ДОЛГОВ

Молдова, г. Тирасполь, Приднестровский гос. университет им. Т. Г. Шевченко  
E-mail: dolax@mail333.com

Дата поступления в редакцию  
21.05—03.12 2003 г.

Оппонент к. т. н. О. Л. СМЕРНОВ  
(ГУАП, г. С.-Петербург)

## МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ТЕХНОЛОГИЧЕСКОГО ПРОЦЕССА ПО ВЫБОРКАМ МАЛОГО ОБЪЕМА

*Применение метода в условиях производства позволяет получить математическую модель технологического процесса, пригодную для управления этим процессом.*

Разработанные к настоящему времени методы математического моделирования технологических процессов по пассивным многомерным данным с доказательством адекватности полученной модели, такие как модифицированный метод случайного баланса (ММСБ) и метод наименьших квадратов с предварительной ортогонализацией факторов (МНКО), описанные в [1], требуют достаточно длинной таблицы данных (иногда до сотен строк), что не всегда возможно. Напротив, в современных производственных условиях при выборочном контроле все чаще прибегают к выборкам малого объема. Одновременно с этим для оперативного управления технологическим процессом требуется математическая модель (ММ), полученная по результатам контроля как можно меньшего количества изготовленных партий продукции. Другими словами, возможна ситуация (например, при пуско-наладочных работах), когда требуется получить ММ по таблице, у которой число столбцов (контрольных факторов) значительно превышает число строк (изготовленных партий изделий).

Такая ситуация с математической точки зрения неразрешима в общем виде, однако в частном слу-

чае ее можно свести к одной из наработанных схем расчета — МНКО. Для этого необходимы два условия: таблица экспериментальных данных представляет собой сверхнасыщенный план, а целевая функция представлена величинами, более или менее равномерно покрывающими значительную часть диапазона ее существования. Первое условие означает, что столбцы таблицы данных (факторы) в большинстве коррелированы между собой, и после устранения дублирующей информации (например, методом корреляционных плеяд) число столбцов в таблице становится меньше числа строк. Второе условие обеспечивает достаточный разброс координат экспериментальных точек с относительно равномерным покрытием поверхности целевой функции.

Проиллюстрируем сказанное на производственном примере.

При наладке технологического процесса (ТП) производства кристаллов одного типа ИМС были получены данные, приведенные в **табл. 1**, где  $Y$  — среднее значение выхода годных ИМС по всем пластинам соответствующей партии в процентах;  $X_i$  — результаты контрольных измерений по ходу ТП, очищенные от грубых промахов по методу Е. Н. Львовского [2, с. 37]. (Конкретные названия контролируемых параметров  $X_i$  и их размерность для предлагаемых методов расчета не существенны.)

Так как число столбцов таблицы существенно превышает число строк (предполагается наличие

Таблица 1

Процент выхода годных ИМС ( $Y$ ) и результаты контроля промежуточных операций ( $X_i$ ) по 9 опытным партиям

Номер партии, $j$	$Y$	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$	$X_6$	$X_7$	$X_8$	$X_9$	$X_{10}$	$X_{11}$	$X_{12}$	$X_{13}$	$X_{14}$	$X_{15}$	$X_{16}$	$X_{17}$	$X_{18}$
1	30	31,0	3,5	20,3	54,8	11,6	140	7	0,56	36	26,5	68,8	4,6	22,5	0,73	5,62	0,86	41,0	13,7
2	25	32,0	4,6	29,8	66,4	11,4	148	7	0,56	36	54,7	55,1	5,4	6,2	0,73	3,62	0,63	44,3	16,8
3	37	33,3	4,7	32,5	70,5	12,4	153	6	0,46	35	31,1	73,5	5,6	15,7	1,45	1,25	0,74	38,4	16,4
4	45	28,2	3,9	16,7	49,4	11,7	127	5	0,73	25	57,0	69,3	6,2	13,5	0,58	2,50	0,85	35,8	14,8
5	67	29,6	3,7	18,5	51,8	11,1	123	4	0,76	26	17,2	83,4	3,3	3,5	0,87	0,50	0,43	26,0	13,3
6	70	32,0	4,6	29,8	66,4	10,6	126	3	0,80	22	21,7	47,6	6,0	0,9	1,16	2,20	0,93	26,9	12,9
7	73	35,9	2,2	24,6	62,7	11,8	132	4	0,95	14	46,4	47,6	6,0	2,7	1,66	1,93	0,58	22,6	8,7
8	76	26,8	3,2	32,6	62,9	10,9	117	5	0,70	26	35,6	75,6	5,8	1,3	2,32	2,11	0,65	26,1	10,3
9	94	34,9	4,8	35,6	75,3	10,0	105	3	0,96	22	7,9	12,5	2,4	5,7	1,58	0,87	0,42	20,4	6,2
$\bar{X}_j$	57,44	31,52	3,91	26,74	62,24	11,28	130,1	4,9	0,720	26,9	33,12	59,27	5,03	8,0	1,231	3,40	0,677	31,28	12,57
$S_j$	23,849	2,99	0,867	6,92	8,68	0,715	15,10	1,54	0,172	7,51	16,91	21,56	1,34	7,52	0,570	3,31	0,184	8,68	3,54

сверхнасыщенного плана), то необходимым этапом на пути моделирования является сокращение размерности факторного пространства до числа, меньшего, чем число строк таблицы. Наиболее простым способом это можно сделать через корреляционные плеяды [3, с. 64], которые позволяют выявить группы (плеяды) коррелированных между собой факторов, фактически несущих одну и ту же информацию.

С этой целью необходимо предварительно построить таблицу коэффициентов корреляции по данным табл. 1 по принципу «каждый с каждым». Сами коэффициенты корреляции находились по обычной классической формуле

$$r = \frac{\sum_{j=1}^n (X_{ij} - \bar{X}_i)(X_{kj} - \bar{X}_k)}{(n-1)S_i \cdot S_k}, \quad (1)$$

однако в силу малого объема выборок ( $n = 9$ ) следует иметь в виду, что ниже порогового значения, вы-

численного с помощью табличного критерия Стьюдента при доверительной вероятности  $P_{\text{дов}} = 95\%$  и числе степеней свободы  $\nu = n - 2$ ,

$$|r|_{\text{порог}} = \frac{t_{\text{табл}}(q; \nu)}{\sqrt{[t_{\text{табл}}(q; \nu)]^2 + (n-2)}} = \frac{2,3646}{\sqrt{2,3646^2 + 7}} = 0,666$$

достоверность правильного вычисления коэффициента корреляции становится весьма сомнительной. К сожалению, то же пороговое значение, вычисленное по другой формуле, —

$$|r|_{\text{порог}} = \frac{\sqrt{36+n} - \sqrt{n}}{6} = 0,618,$$

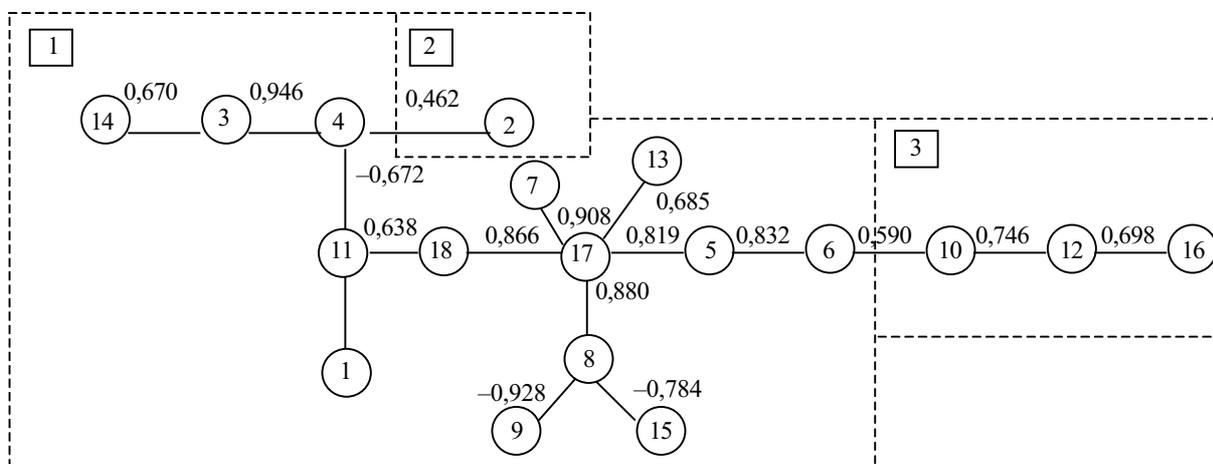
дает другое числовое значение, и нам придется делать выбор. Выберем  $|r|_{\text{порог}} = 0,618$  как наиболее жесткое при  $n < 50$ .

Результаты всех вычислений представлены в табл. 2.

Таблица 2

Матрица коэффициентов корреляции

	Y	X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>	X <sub>8</sub>	X <sub>9</sub>	X <sub>10</sub>	X <sub>11</sub>	X <sub>12</sub>	X <sub>13</sub>	X <sub>14</sub>	X <sub>15</sub>	X <sub>16</sub>	X <sub>17</sub>	X <sub>18</sub>
Y	1	0,079	-0,169	0,345	0,284	-0,719	-0,898	-0,870	0,863	-0,820	-0,387	-0,501	-0,277	-0,737	0,633	-0,689	-0,459	-0,945	-0,860
X <sub>1</sub>	0,079	1	0,077	0,346	0,618	0,015	0,086	-0,016	0,337	-0,273	-0,335	-0,685	-0,217	-0,039	-0,081	0,152	-0,261	-0,209	-0,333
X <sub>2</sub>	-0,169	0,077	1	0,425	0,465	-0,274	0,084	0,010	-0,309	0,463	-0,304	-0,281	-0,284	0,128	0,215	0,279	0,097	0,291	0,330
X <sub>3</sub>	0,345	0,346	0,425	1	0,946	-0,375	-0,128	-0,170	0,039	0,015	-0,302	-0,516	-0,108	-0,348	0,670	0,184	-0,204	-0,222	-0,315
X <sub>4</sub>	0,284	0,618	0,465	0,946	1	-0,316	-0,046	-0,213	0,116	-0,038	-0,330	-0,672	0,061	-0,272	0,524	0,224	-0,238	-0,216	-0,328
X <sub>5</sub>	-0,719	0,015	-0,274	-0,375	-0,316	1	0,832	0,633	-0,609	0,365	0,59	0,599	0,557	0,552	-0,237	0,711	0,340	0,601	0,629
X <sub>6</sub>	-0,898	0,088	0,084	-0,128	-0,046	0,832	1	0,764	-0,788	0,641	0,483	0,565	0,481	0,514	-0,381	0,779	0,415	0,819	0,803
X <sub>7</sub>	-0,870	-0,016	0,010	-0,170	-0,213	0,633	0,764	1	-0,854	0,842	0,498	0,487	0,275	0,666	-0,343	0,602	0,303	0,908	0,683
X <sub>8</sub>	0,863	0,337	-0,309	0,039	0,116	-0,609	-0,788	-0,854	1	-0,928	-0,268	-0,629	-0,294	-0,505	0,317	-0,784	-0,432	-0,880	-0,852
X <sub>9</sub>	-0,820	-0,273	0,463	0,015	-0,038	0,365	0,941	0,842	-0,928	1	0,093	0,421	-0,021	0,646	-0,399	0,646	0,266	0,867	0,737
X <sub>10</sub>	-0,387	-0,335	-0,304	-0,302	-0,330	0,590	0,483	0,498	-0,268	0,093	1	0,299	0,746	0,292	-0,225	0,120	0,319	0,503	0,554
X <sub>11</sub>	-0,501	-0,685	-0,281	-0,516	-0,672	0,599	0,565	0,487	-0,629	0,421	0,299	1	0,370	0,266	-0,178	0,318	0,264	0,424	0,638
X <sub>12</sub>	-0,277	-0,217	-0,284	-0,108	0,061	0,557	0,481	0,275	-0,294	-0,021	0,746	0,370	1	0,007	-0,047	0,300	0,698	0,334	0,426
X <sub>13</sub>	-0,737	-0,039	0,128	-0,348	-0,272	0,552	0,514	0,666	-0,505	0,646	0,292	0,266	0,007	1	-0,481	0,658	0,431	0,685	0,447
X <sub>14</sub>	0,633	-0,081	0,215	0,670	0,524	-0,237	-0,381	-0,343	0,317	-0,399	-0,225	-0,178	0,047	-0,481	1	-0,047	-0,306	-0,599	-0,616
X <sub>15</sub>	-0,689	0,152	0,279	0,184	0,224	0,711	0,779	0,602	-0,784	0,646	0,120	0,318	0,300	0,658	-0,047	1	0,405	0,626	0,573
X <sub>16</sub>	-0,459	-0,261	0,097	-0,204	-0,2380	0,340	0,415	0,303	-0,432	0,266	0,319	0,264	0,698	0,431	-0,306	0,405	1	0,499	0,501
X <sub>17</sub>	-0,945	-0,209	0,291	-0,222	-0,216	0,601	0,819	0,908	-0,880	0,867	0,503	0,424	0,334	0,685	-0,599	0,626	0,499	1	0,866
X <sub>18</sub>	-0,860	-0,333	0,330	-0,315	-0,328	0,629	0,803	0,683	-0,852	0,737	0,554	0,638	0,426	0,447	-0,616	0,574	0,501	0,866	1



Граф корреляционных плеяд (границы плеяд обозначены пунктиром, номера плеяд даны в квадратах, а номера факторов — в кружочках)

Анализ табл. 2 приводит к выводу, что величина  $Y$  тесно коррелирована с факторами  $X_5, X_6, X_7, X_8, X_9, X_{13}, X_{14}, X_{15}, X_{17}, X_{18}$  по отдельности, а от остальных факторов зависит слабо, но пренебрегать ими без дополнительных исследований нельзя. В свою очередь, все факторы  $X_1—X_{18}$  находятся между собой в сложном взаимодействии, расшифровать которое поможет граф корреляционных плед (см. рисунок), построенный по матрице коэффициентов корреляции.

Из рисунка следует, что всего плед оказался три, а значит, в будущую модель могут войти без потери исходной информации не более трех факторов из первоначальных 18. Так как в исходной табл. 1 имеется всего 9 строк, то можно утверждать, что выполнено одно из фундаментальных требований теории планирования эксперимента: число исходной таблицы должно быть меньше числа эффектов, вошедших в модель. Сверхнасыщенный план превратился в обычный за счет устранения избыточной информации.

Выбирая по одному представителю из каждой плеяды по принципу максимальной тесноты связи с выходным параметром (степени влияния) и располагая факторы в порядке убывания этой меры, получим, что будущую модель следует искать в виде зависимости

$$\tilde{Y} = f(X_{17}, X_{16}, X_2). \quad (2)$$

Что касается корреляционных связей между самими факторами  $X_1—X_{18}$ , то их можно легко построить в виде линейных или квадратичных уравнений регрессии по методу Чебышева [4]. Например, для пары  $X_3—X_4$  это выглядит так:

$$\tilde{X}_4 = \bar{X}_4 + r_{3,4} \cdot \frac{X_3 - \bar{X}_3}{S_3} \cdot S_4 = 62,24 + 0,946 \frac{X_3 - 26,74}{6,92} \cdot 8,68 = 1,187 X_3 + 30,51 \quad (3)$$

с коридором существования

$$\Delta X_4 = \pm Z_{\text{дов}} \cdot S_3 \sqrt{1 - r_{3,4}^2} = \pm 1,96 \cdot 8,68 \sqrt{1 - 0,946^2} = \pm 5,52.$$

Как указывалось ранее, расчет модели (2) следует вести методом МНКО, при котором для достижения наибольшей эффективности отобранные факторы следует расположить в порядке убывания их влияния на выходную величину (табл. 3).

Суть расчетов заключается в том, чтобы перейти из декартового пространства в пространство специ-

Таблица 3

Факторы — представители плед, упорядоченные по степени влияния на выходную величину

j	Z <sub>1</sub>	Z <sub>2</sub>	Z <sub>3</sub>	Z <sub>4</sub>	Z <sub>5</sub>	Z <sub>6</sub>	Y
	X <sub>17</sub>	X <sub>16</sub>	X <sub>2</sub>	X <sub>17</sub> <sup>2</sup>	X <sub>16</sub> <sup>2</sup>	X <sub>2</sub> <sup>2</sup>	
1	41,0	0,86	3,5	1681,00	0,7396	12,25	30
2	44,3	0,63	4,6	1962,49	0,3969	21,16	25
3	38,4	0,74	4,7	1474,56	0,5476	22,09	37
4	35,8	0,85	3,9	1281,64	0,7225	15,21	45
5	26,0	0,43	3,7	676,00	0,1849	13,69	67
6	26,9	0,93	4,6	723,61	0,8649	21,16	70
7	22,6	0,58	2,2	510,76	0,3364	4,84	73
8	26,1	0,65	3,2	681,21	0,4225	10,24	76
9	20,4	0,42	4,8	416,16	0,1764	23,04	94
$\bar{Z}_K$	31,28	0,677	3,91	1045,97	0,4880	15,964	—

ально подобранных ортогональных осей координат, превратив все величины  $Z_{kj}$  в  $\psi_{kj}(Z)$  для каждого  $k$ -го фактора и  $j$ -й строки исходной таблицы (в нашем случае — табл. 3), найти в этом ортогональном пространстве адекватную математическую модель (в виде уравнения регрессии), а затем совершить обратный переход в декартово пространство. Расчеты следует вести по формулам

$$\psi_{kj}(Z) = Z_{kj} - \sum_{i=0}^{k-1} \xi_{ki} \cdot \psi_{ij}(Z) \text{ для всех } i < k, \quad (4)$$

$$\xi_{kj} = \frac{\sum_{j=1}^N Z_{kj} \cdot \psi_{ij}(Z)}{\sum_{j=1}^N [\psi_{ij}(Z)]^2}$$

где  $\xi_{kj}$  — вспомогательный коэффициент, подсчитанный для всех  $N$  строк исходной таблицы данных.

Для удобства представим результаты расчетов в табл. 4 и 5, которые сопряжены друг с другом.

Таблица 4

Поправочные коэффициенты  $\xi_{ki}$

k	i					
	0	1	2	3	4	5
1	31,28					
2	0,677	0,01055				
3	3,91	0,02891	-0,34306			
4	1045,27	64,6286	-194,28	9,0227		
5	0,4880	0,01282	1,3708	0,01980	0,001186	
6	15,964	0,18235	-2,8295	7,3340	0,002081	-0,16812

Таблица 5

Ортогональные полиномы и проверка адекватности модели

j	$\psi_{0j}(Z)$	$\psi_{1j}(Z)$	$\psi_{2j}(Z)$	$\psi_{3j}(Z)$	$\psi_{4j}(Z)$	$\psi_{5j}(Z)$	$\psi_{6j}(Z)$	$Y_j$	$\hat{Y}_{onj}$	$(Y_j - \hat{Y}_{onj})^2$
1	+1	+9,72	+0,080	-0,664	29,0735	-0,00401	-0,4515	30	28,81	1,409
2	+1	+13,02	-0,184	+0,251	37,7434	-0,05552	0,3725	25	23,22	3,168
3	+1	+7,12	-0,012	+0,580	-38,4302	0,01887	0,6231	37	40,33	11,080
4	+1	+4,52	+0,125	-0,098	-30,5820	0,04341	-0,4349	45	44,89	0,012
5	+1	-5,28	-0,191	-0,122	-64,0377	0,10478	-0,8060	67	71,24	18,012
6	+1	-4,38	+0,299	+0,920	11,2021	-0,00832	0,0687	70	72,59	6,686
7	+1	-8,68	-0,005	-1,461	38,6770	-0,05041	1,0766	73	75,57	6,619
8	+1	-5,18	+0,028	-0,550	-18,8815	-0,00419	-0,6279	76	69,43	43,225
9	+1	-10,88	-0,142	+1,156	36,0312	-0,04309	0,0979	94	89,20	23,079
$\Sigma(\bullet)^2$	9	602,74	0,203	5,485	12058,952	0,02080	3,1373	—	$\Sigma(\bullet)$	113,286

Затем по формуле

$$A_k = \frac{\sum_{j=1}^N Y_j \cdot \psi_{kj}(Z)}{\sum_{j=1}^N [\psi_{kj}(Z)]^2} \quad (5)$$

находятся оценки коэффициентов регрессии (искомой модели в пространстве ортогональных полиномов — осей координат), которые проверяются на значимость по критерию Стьюдента

$$t_k = \frac{|A_k|}{S_k} \leq t_{\text{табл}}(q; \nu = N - 1). \quad (6)$$

При выполнении условия (6) оценка  $A_k$  признается незначимой (статистически равной нулю) и в модель не включается.

При расчетах этого примера был принят  $q = 5\%$ -ный уровень значимости (возможная ошибка в выводах), число степеней свободы  $\nu = 27$ , тогда  $t_{\text{табл}} = 2,052$ . Среднеквадратичное отклонение  $S_k$  оценки коэффициента  $A_k$  рассчитывалось по формуле

$$S_k = \sqrt{\frac{S_{cp}^2 \{Y\}}{\sum_{j=1}^N [\psi_{kj}(Z)]^2}}, \quad (7)$$

где  $S_{cp}^2 \{Y\}$  — средняя дисперсия опытов (в нашем случае  $S_{cp}^2 \{Y\} = 10,35$ ).

Если данные по дисперсии отсутствуют, ее придется заменять верхней предельной оценкой — максимально возможной дисперсией, равной дисперсии всей выборки  $Y$ , деленной на 4 (точнее, на 3,92) — ширину размаха квантилей при доверительной вероятности  $P_{\text{дов}} = 95\%$ . При этом модель получится с более широким коридором ошибок.

В результате расчетов получилось:

$A_0=57,44$					
$A_1=-2,698$	$S_1=0,131$	$t_1=20,558$	$A_4=0,0238$	$S_4=0,0293$	$t_4 < 1$
$A_2=-2,320$	$S_2=7,140$	$t_2 < 1$	$A_5=-22,482$	$S_5=22,307$	$t_5=1,008$
$A_3=+3,618$	$S_3=1,374$	$t_3=2,633$	$A_6=-3,145$	$S_6=1,816$	$t_6=1,732$

Таким образом, искомая математическая модель в пространстве ортогональных полиномов запишется в виде

$$\hat{Y}_{\text{оп}} = 57,44 - 2,698 \cdot \psi_1(Z) + 3,618 \cdot \psi_3(Z). \quad (8)$$

Проверка адекватности модели (ее соответствия опытными данным) осуществлялась путем сравнения дисперсии адекватности (здесь  $d=3$  — число членов модели)

$$S_{\text{ад}}^2 = \frac{\sum_{j=1}^N (Y_j - \hat{Y}_{\text{оп}j})^2}{N - d} = \frac{113,286}{9 - 3} = 18,881$$

с дисперсией опытов по критерию Фишера

$$F = \frac{S_{\text{ад}}^2}{S^2 \{Y\}} = \frac{18,881}{10,35} = 1,824 < F_{\text{табл}}(5\%; 6; 27) = 2,470.$$

Полученный результат свидетельствует о том, что найденная модель описывает экспериментальные данные с точностью, не превышающей ошибки самого эксперимента.

Для удобства работы с моделью ее необходимо перевести обратно в пространство декартовых координат, для чего в модель (8) подставим развернутые выражения полиномов  $\psi_1(Z) = Z_1 - \bar{Z}_1$  и  $\psi_3(Z) = Z_3 - \bar{Z}_3 - \xi_{31}(Z_1 - \bar{Z}_1) - \xi_{32}(Z_2 - \bar{Z}_2) - \xi_{21}(Z_1 - \bar{Z}_1)$ :

$$\begin{aligned} \hat{Y} &= 57,44 - 2,698(Z_1 - 31,28) + 3,618 \{ Z_3 - \\ &- 3,91 - 0,02891(Z_1 - 31,28) - \\ &- (-0,34306) \cdot [Z_2 - 0,677 - 0,01055(Z_1 - 31,28)] \} = \\ &= 130,5 - 2,816Z_1 + 1,241Z_2 + 3,618Z_3 \end{aligned}$$

или, переходя к первоначальным обозначениям,

$$\hat{Y} = 130,5 - 2,816X_{17} + 1,241X_{16} + 3,618X_2. \quad (9)$$

Для уверенности можно произвести повторную проверку адекватности модели (9), которая должна подтвердить ее аутентичность модели (8), что и произошло.

\*\*\*

Таким образом, доказана практическая возможность получения адекватных математических моделей в условиях многофакторной выборки малого объема при сверхнасыщенной таблице пассивных экспериментальных данных — даже при применении методов классической математической статистики. Анализ производственных ситуаций позволяет сделать вывод, что из двух условий, оговоренных в начале статьи, первое выполняется всегда, а второе — только в частном случае, который необходимо организовать.

Применение данного метода в цеховых условиях (особенно при пуско-наладочных работах) позволит получить математическую модель технологического процесса, пригодную для управления этим процессом, и не только для массового или крупносерийного производства, но и при производстве мелкосерийных изделий, чего до настоящего времени не было.

#### ИСПОЛЬЗОВАННЫЕ ИСТОЧНИКИ

1. Долгов Ю. А. Статистическое моделирование. — Тирасполь: ПГУ, 2002.
2. Львовский Е. Н. Статистические методы построения эмпирических формул. — М.: Высшая школа, 1988.
3. Дружинин Г. В. Методы и оценки прогнозирования качества. — М.: Радио и связь, 1982.
4. Митропольский А. К. Техника статистических вычислений. — М.: Наука, 1971.