ISSN 0002-306X. Изв. НАН РА и ГИУА. Сер. ТН. 2008. Т. LXI, № 1.

УДК 658.564:519.83

АВТОМАТИЗАЦИЯ И
СИСТЕМЫ УПРАВЛЕНИЯ

С.Г. КЮРЕГЯН, С.В. АБГАРЯН, А.К. ДАВТЯН, С.Р. ПИЛУНЦ

ПРОГНОЗИРУЮЩИЕ СВОЙСТВА МОДЕЛЕЙ, ПОСТРОЕННЫХ МЕТОДОМ ГРУППОВОГО УЧЕТА АРГУМЕНТОВ

Исследованы математические модели технологических процессов, построенные методом группового учета аргументов. На основании анализа среднеквадратической ошибки модели предложены рекомендации к построению адекватной модели, обладающей прогнозирующими свойствами.

Ключевые слова: случайный процесс, моделирование, дисперсия, адекватность, управление.

Современные технологические процессы представляют собой сложные объекты, входные, выходные переменные и внутренние параметры которых зависят от многочисленных, порой трудно учитываемых случайных факторов. Для управления такими процессами необходимо иметь достаточно адекватное описание взаимных зависимостей, характеризующих данный процесс. При неполной информации о механизме процесса математическое описание можно получить статистическими методами на основании соответствующей обработки экспериментальных данных наблюдений за выходной y и входными $x=\{x_j\}$ ($j=\overline{1,m}$) переменными, собранными непосредственно на действующем объекте в режиме нормальной работы.

Производственные технологические процессы обладают следующими специфическими особенностями:

- сравнительно низкая точность промышленных методов контроля и регистрации данных;
- узкие диапазоны изменения технологических переменных, зачастую соизмеримые с погрешностями контрольно-измерительной аппаратуры;
- значительные транспортные запаздывания в объекте, а также инерционный характер самих процессов;
- влияние сезонных и суточных изменений параметров окружающей среды;
- неконтролируемые изменения качественного состава сырья и переключение режимов при переходе с одного сорта продукции на другой.

В простейших случаях для стационарного случайного процесса зачастую математическую модель представляют уравнением множественной линейной или квазилинейной регрессии, коэффициенты полиномов которых вычисляют с помощью метода наименьших квадратов (МНК) [1-3], причем для квазилинейной модели требуется более длинная выборка данных наблюдений ($i=\overline{1,n}$).

Однако перечисленные особенности накладывают дополнительные требования к построению математических моделей, в особенности предназначенных для решения задач управления технологическим процессом. Подобные модели должны обладать достаточными прогнозирующими свойствами, иначе непосредственное применение их может привести к неудовлетворительным результатам.

Один из наиболее приемлемых методов, при котором степень полинома получаемой модели мало зависит от длины выборки, основан на применении алгоритмов, реализующих многорядную селекцию моделей по критерию регулярности методом группового учета аргументов (МГУА). Получаемые при этом модели, согласно рекомендациям [3], пригодны для решения задач прогнозирования, идентификации структуры и параметров сложных объектов по результатам наблюдения их работы, оптимального управления с оптимизацией прогноза.

МГУА предписывает разбивать данные наблюдений на две части: проверочную $n_{пр}$ и обучающую $n_{об}$ последовательности. Для этого вначале выборка ранжируется по значениям дисперсий (отклонений) вектора входных переменных, вычисленных для каждого i-го наблюдения:

$$D_{xi} = \frac{1}{m}\sum_{j=1}^{m}\left(\frac{x_{ji} - \overline{x}_j}{\overline{x}_j}\right)^2, \qquad (1)$$

где $\overline{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ji}$ - среднее значение j-й входной переменной.

Затем в качестве обучающей последовательности отбирается оптимальное количество реализаций с большими значениями отклонений, а остальные реализации образуют проверочную последовательность. Обучающая последовательность используется для оптимизации коэффициентов уравнения регрессии по МНК, как и в обычном регрессионном анализе, а проверочная – для оценки степени регулярности полученного уравнения регрессии на каждом уровне селекции по величине относительного значения среднеквадратической ошибки (СКО):

$$\sigma_p = \left(\sum_{r=1}^{n_{пр}}\left(y_r - y_{мг}\right)^2 / \sum_{r=1}^{n_{пр}} y_{мг}^2\right)^{1/2}, \qquad (2)$$

где y_r, $y_{мг}$ – соответственно значения выхода r-го наблюдения и модели.

Принципиальное отличие МГУА от обычного регрессионного анализа заключается в том, что целью первого является достижение минимума целесообразно выбранного критерия селекции (СКО на проверочной последовательности), а целью второго – достижение минимума СКО на всех экспериментальных точках при заранее заданном виде (что зачастую носит субъективный характер) уравнения регрессии. Подбором соотношения $n_{пр}/n_{об}$ добиваются минимума СКО модели (м. В результате получают иерархическую (многоуровневую) модель $y_м=f(x,n_{об}/n_{пр})$, параметрически зависящую от $n_{об}/n_{пр}$, с оценкой в виде относительного значения СКО:

$$\sigma_{м} = \left[\frac{1}{n-1} \sum_{i=1}^{n} \frac{(y_i - y_{мi})^2}{\bar{y}^2} \right]^{\frac{1}{2}} =$$

$$= \left[\frac{1}{n-1} \sum_{i=1}^{n_{пр}} \frac{(y_i - y_{мi})^2}{\bar{y}^2} + \frac{1}{n-1} \sum_{i=n_{пр}+1}^{n} \frac{(y_i - y_{мi})^2}{\bar{y}^2} \right]^{\frac{1}{2}}$$

или же:

$$\sigma_{м}^2 = \frac{n_{пр} - 1}{n-1} \sigma_{пр}^2 + \frac{n - n_{пр} - 1}{n-1} \sigma_{об}^2, \qquad (3)$$

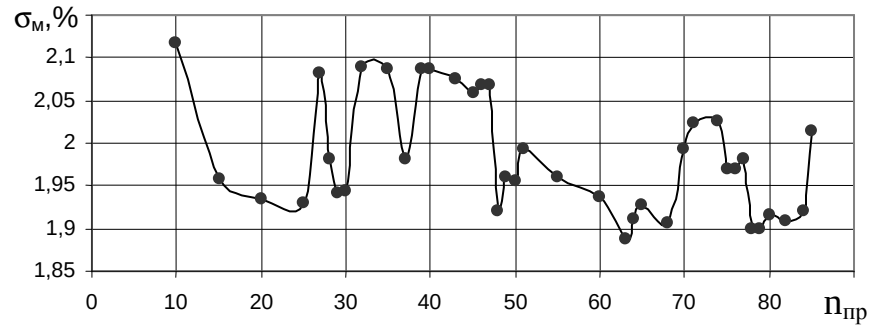
где $\bar{y}$ – среднее значение выходной переменной; $\sigma_{пр}$, $\sigma_{об}$ – относительные значения СКО, вычисленные соответственно на проверочной и обучающей последовательностях:

$$\begin{cases} \sigma_{пр}^2 = \dfrac{1}{n_{пр} - 1} \displaystyle\sum_{i=1}^{n_{пр}} \dfrac{(y_i - y_{мi})^2}{\bar{y}^2}, \\ \sigma_{об}^2 = \dfrac{1}{n_{об} - 1} \displaystyle\sum_{q=1}^{n_{об}} \dfrac{(y_q - y_{мq})^2}{\bar{y}^2}. \end{cases} \qquad (4)$$

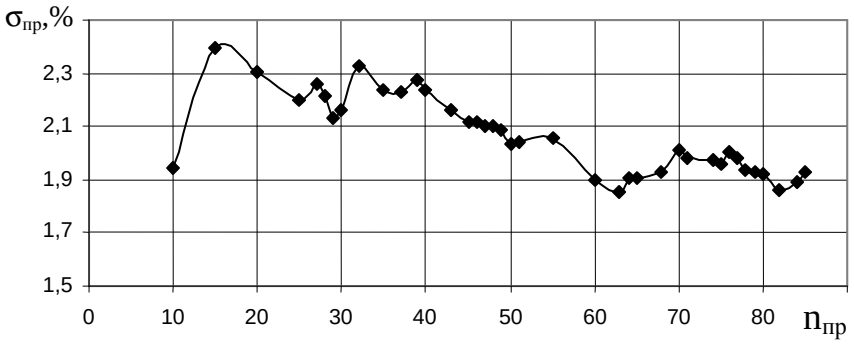
Из (4) ясно, что представленные СКО зависят от длины $n_{пр}$ проверочной последовательности. Если модель, построенная по алгоритму регулярности МГУА, единственно оптимальная, как утверждают ее авторы [3], то, варьируя длиной $n_{пр}$ в (3), можно получить минимум $\sigma_{м}$. Аналитическое решение новой задачи минимизации $\sigma_{м}$ по параметру $n_{пр}$ невозможно, поскольку $\sigma_{пр}(n_{пр})$ и $\sigma_{об}(n_{пр})$ в выражении (3) не являются априори аналитическими функциями $n_{пр}$ и вычисляются только в результате многорядной селекции при реализации алгоритма МГУА. Авторы МГУА в [4] обращают внимание на проблематичность разбиения базы данных на обучающую и проверочную последовательности. Поэтому проанализировать выражения (3) и (4) можно только на конкретных численных примерах.

Прежде всего необходимо определить характер функций (3) и (4) в зависимости от длины $n_{пр}$ для моделей, построенных МГУА. В качестве примера выбран технологический процесс флотации молибденовой руды [5]. Исходные данные представляли собой $n_{и}$=122 средних сменных данных наблюдений за 12-ю независимыми входными x_j и 2-мя выходными y_k переменными. Проведенные байесовские оценки и статистический анализ показали [5,6], что исходные данные относятся к одной генеральной совокупности, имеют стационарный характер и подчиняются нормальному закону распределения. Для построения модели выделена выборка, состоящая из n=100 наблюдений, остальные $n_{э}$=22 приняты в качестве экзаменационных данных для проверки прогнозирующих свойств моделей. На рис. 1 и 2 представлены характеристики $\sigma_{мk}(n_{пр})$ и $\sigma_{прk}(n_{пр})$, рассчитанные по

(3) и (4) подробно с единичным шагом для следующих моделей выходных переменных: процентное содержание молибдена в концентрате $y_{м1}$ (k=1) и выход концентрата $y_{м2}$ (k=2). Как видно из графиков, функции $\sigma_{мk}$ и $\sigma_{прk}$ являются сложными функциями $n_{пр}$ со множеством локальных минимумов. Очевидно, что оптимальная модель соответствует глобальному минимуму графика $\sigma_{мk}$.

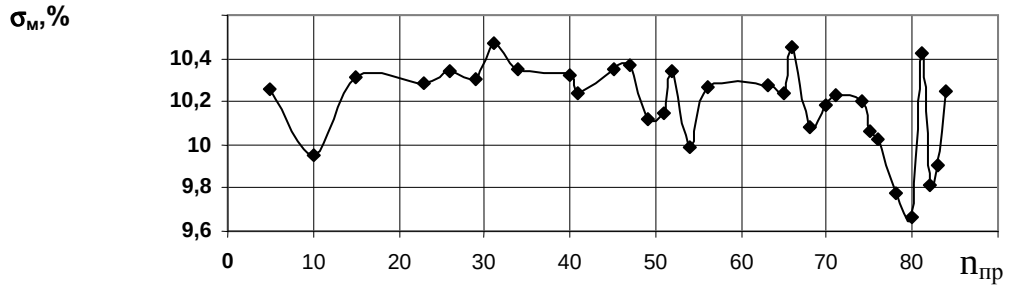


а)

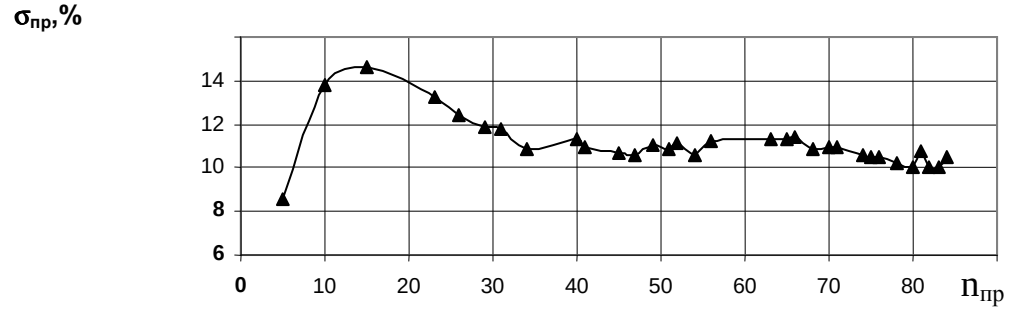


б)

Рис. 1. СКО модели $У_1$ в зависимости от длины проверочной последовательности:
а – зависимость ($_{мk}(n_{пр})$; б – зависимость ($_{прk}(n_{пр})$



а)



б)

Рис. 2. СКО модели $У_2$ в зависимости от длины проверочной последовательности:
а – зависимость $(_{мk}(n_{пр})$; б – зависимость $(_{прk}(n_{пр})$

Не имея априорного представления о характере графика $\sigma_{прk}$, исследователь мог бы путем пробного пошагового расчета остановиться на выборе в качестве оптимального соотношения $n_{об}/n_{пр}$ точки локального минимума и построить соответствующую ей модель, которая, по-видимому, была бы неоптимальной. В нашем случае поле рассеяния СКО $\sigma_{мk}$ моделей достаточно узкое: от 1,89 до 2,10 – для $у_{м1}$ и от 9,6 до 10,4 – для $у_{м2}$. В таких узких пределах отпадает актуальность поиска оптимальной по СКО модели, поскольку погрешность крайне неоптимальной отличается всего на 10...11% от оптимальной. Более важное значение приобретает прогнозирующее свойство моделей.

Для оценки и сравнений моделей воспользуемся значениями дисперсий выходных переменных базовой выборки ($s^2_б$) и построенных моделей ($s^2_м$) по формулам

$$s^2_б = \frac{1}{n-1}\sum_{i=1}^{n}\frac{(y_i - \bar{y})^2}{\bar{y}^2}, \qquad s^2_м = \frac{1}{n-1}\sum_{i=1}^{n}\frac{(y_{мi} - \bar{y})^2}{\bar{y}^2},$$

(5)

на основании которых можно вычислить остаточную дисперсию модели, что для множественной линейной (или квазилинейной) регрессии совпадает с квадратом СКО модели, и индекс множественной корреляции R^2_{yx} [2]:

$$\sigma^2_м = s^2_б - s^2_м, \qquad (6)$$

$$R^2_{yx} = \frac{s^2_м}{s^2_б}. \qquad (7)$$

Построены альтернативные модели для различных значений отношений $n_{об}/n_{пр}$, соответствующих некоторым точкам локальных и глобальных минимумов графиков $\sigma_{мk}$, показанных на рис. 1 и 2. Результаты расчетов сведены в таблицу.

Таблица

Модель, y_{kv}	Параметры								
	$\bar{y}$	$s^2_б \times 10^{-4}$	$n_{об}/n_{пр}$	№ вх. перем.	Кол-во ур.	$s^2_м \times 10^{-4}$	$\sigma_м$,%	$\sigma_{пр}$,%	R_{yx}
y_{11}	48,80	5,819	37/63	1,2,4-6, 11,12	6	2,260	1,89	1,86	0,623
y_{12}	48,80	5,819	21/79	1-11	6	2,211	1,90	1,93	0,616
y_{13}	48,80	5,819	52/48	2-4,6-9, 11	5	2,127	1,92	2,10	0,605
y_{14}	48,80	5,819	75/25	1,2,4-6, 8,11	5	2,099	1,93	2,20	0,601
y_{21}	5288,6	305,3	20/80	1,2,6,8, 11	5	211,8	9,67	10,08	0,833
y_{22}	5288,6	305,3	46/54	1-4,6-12	6	205,5	9,99	10,59	0,821
y_{23}	5288,6	305,3	32/68	1,2,4-8, 10-12	6	203,6	10,08	10,87	0,817
y_{24}	5288,6	305,3	51/49	1,2,6,9,11	6	202,8	10,12	11,03	0,815

Как и следовало ожидать, СКО альтернативных моделей мало отличаются друг от друга и от СКО оптимальной модели. Однако при выборе модели необходимо обратить внимание на селекцию входных переменных. Например, для выходных переменных $y_{м1}$ и $y_{м2}$ важными входными переменными являются x_1 и x_2 – соответственно процентное содержание молибдена в исходной руде и количество поступающей руды. Как видно из таблицы, в модели y_{13} отсутствует входная переменная x_1, поэтому эта модель не может быть принята в качестве адекватной. При выборе модели определенную роль может сыграть и сложность, оцениваемая количеством уровней иерархии.

Однако возвратимся к оценкам прогнозирующих свойств моделей. Если принять во внимание, что для множества распределений (равномерного, нормального, трапецеидальных, экспоненциальных и др.) с доверительной вероятностью 0,9 погрешность модели составляет $\Delta_м=1,6\sigma_м$ [1,2], то для оптимальной модели $y_{м1}$ эта погрешность составит 3,02%, а для $y_{м2}$ – 15,47%. В первом случае это значение выходит за пределы рассеяния базы данных ($s_{б1}=2,41\%$), а во втором - укладывается в пределах ($s_{б2}=17,47\%$). Надо полагать, что модель $y_{м2}$ обладает прогнозирующими свойствами, по крайней мере, для 90% наблюдений из базы данных, несмотря на бтльшую погрешность, а модель $y_{м1}$ – лишь для 79% наблюдений, распределенных по нормальному закону.

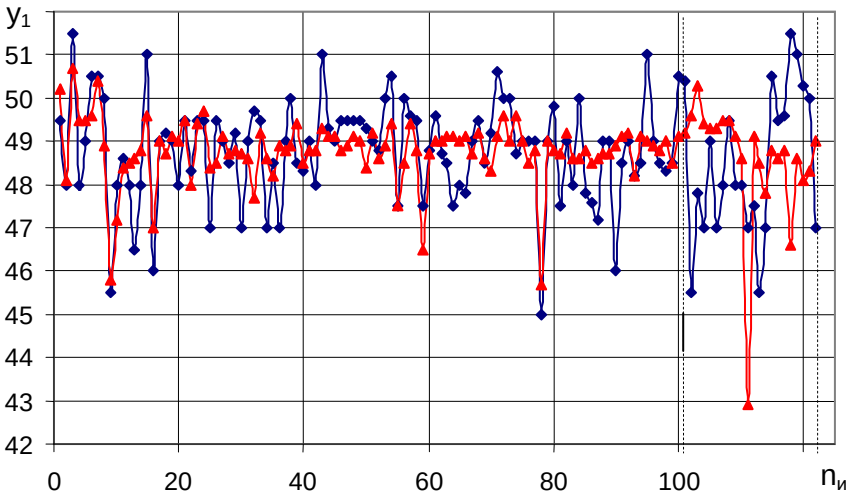
Проверим прогнозирующие свойства моделей на примере экзаменационной последовательности. На рис. 3 приведены сравнения изменений выходных переменных данных наблюдений с результатами, полученными с помощью построенных моделей. На участке базы данных, использованных для построения моделей (от 1-го до 100-го наблюдения), тенденция (направление нарастания и убывания), предсказываемая моделями, хорошо согласуется с данными наблюдений. На участке экзаменационной последовательности (от 101-го до 122-го наблюдения), зона которой отмечена пунктирными линиями, модель $y_{м2}$ (рис. 3б) дает хороший прогноз, поскольку правильно предсказывает тенденцию изменения процесса, хотя и с большой погрешностью - $\sigma_м=9,67\%$ (см. табл.), а модель $y_{м1}$ (рис. 3а), обладая меньшей погрешностью - $\sigma_м=1,89\%$, (см. табл.), не предсказывает тенденцию изменения процесса, как и следовало ожидать.

Таким образом, резюмируя проведенное исследование, можно предложить следующие основные рекомендации при построении моделей с использованием алгоритма МГУА.

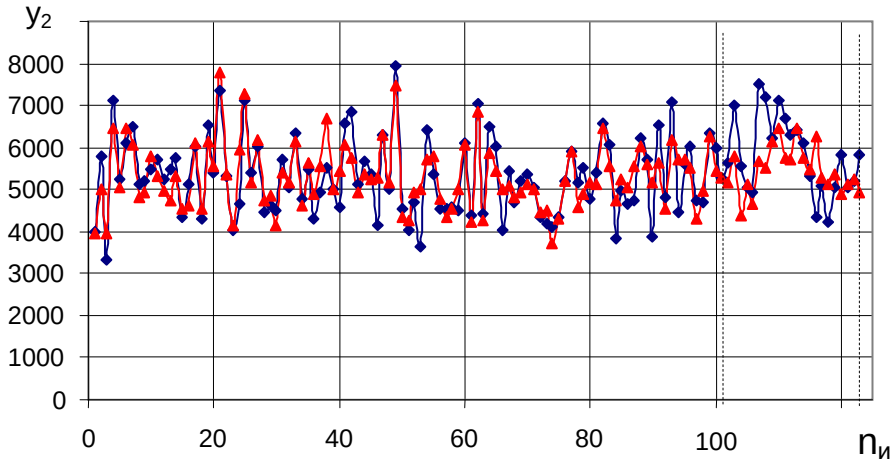
1. Поиск оптимальной модели по минимуму СКО актуален, когда поле рассеяния СКО модели колеблется в существенных для данной задачи пределах.

2. Зависимость СКО модели от длины проверочной последовательности представляет собой сложную функцию со множествами минимумов, построение которой следует проводить подробно на каждом шаге для выявления глобального минимума.

3. Прогнозирующие свойства модели определяются из дисперсионного анализа модели (5)-(7) и оценки доверительных интервалов; только минимум СКО модели не обеспечивает адекватность модели.



а)



б)

Рис. 3. Графики изменений выходных переменных в зависимости от сменных наблюдений; а – процентное содержание молибдена в концентрате (y_1); б – выход концентрата (y_2); ◆ - данные наблюдений, ▲ - выход модели

СПИСОК ЛИТЕРАТУРЫ

1. **Дрейпер Н., Смит Г.** Прикладной регрессионный анализ. – М.: Статистика, 1973. – 392с.
2. **Ф¸рстер Э., Р¸нц Б.** Методы корреляционного и регрессионного анализа. – М.: Финансы и статистика, 1983. – 302с.
3. **Ивахненко А.Г., Зайченко Ю.П., Димитров В.Д.** Принятие решений на основе самоорганизации. – М.: Сов. радио, 1976. – 280с.
4. **Ивахненко А.Г., Юрачковский Ю.П.** Моделирование сложных систем по экспериментальным данным. – М.: Радио и связь, 1987. – 120с.
5. **Кюрегян С.Г., Мамиконян Б.М., Шмелев В.К., Абгарян С.В., Баласанян С.Ш.** К вопросу об управлении флотационным процессом обогащения молибденовой руды // Изв. НАН РА и ГИУА. Сер. ТН. – 2003. – Т. LVI, № 1. – С. 107-113.
6. **Абгарян С.В., Кюрегян С.Г., Айрапетян В.Г., Овакимян Р.М.** Байесовская оценка параметров флотационного технологического процесса // Вестник Инженерной академии Армении. – Ереван, 2006. – Т. 3, № 3. – С. 456-461.

ГИУА. Материал поступил в редакцию 03.06.2007.

Ս.Գ. ԿՅՈՒՐԵՂՅԱՆ, Ս.Վ. ԱԲԳԱՐՅԱՆ, Ա.Կ. ԴԱՎԹՅԱՆ, Ս.Ռ. ՓԻԼՈՒՆՑ

ԱՐԳՈՒՄԵՆՏՆԵՐԻ ԽՄԲԱՅԻՆ ՀԱՇՎԱՌԿՄԱՆ ՄԵԹՈԴՈՎ ԿԱՌՈՒՑՎԱԾ ՄՈԴԵԼՆԵՐԻ ԿԱՆԽԱԳՈՒՇԱԿԻՉ ՀԱՏԿՈՒԹՅՈՒՆՆԵՐԸ

Հետազոտված են տեխնոլոգիական գործընթացների մաթեմատիկական մոդելները՝ կառուցված արգումենտների խմբային հաշվառման մեթոդով: Մոդելի միջին քառակուսային սխալի վերլուծության հիման վրա առաջարկված են երաշխիքներ ադեկվատ, կանխագուշակող հատկությամբ օժտված մոդելների կառուցման համար:

Առանցքային բառեր. պատահական գործընթաց, մոդելավորում, դիսպերսիա, համարժեքություն, ղեկավարում:

S.G. KYUREGHYAN, S.V. ABGARYAN, A.K. DAVTYAN, S.R. PILUNTS

PREDICTING PROPERTIES OF THE MODELS CONSTRUCTED BY THE METHOD OF ARGUMENT GROUP ACCOUNT

The mathematical models of technological processes constructed by the method of argument group account are investigated. On the basis of the analysis a model standard deviation is proposed for the recommendations to construct an adequate, possessing predicting properties, model.

Keywords: stochastic process, modelling, dispersion, adequacy, control.