УДК 681.51:622.73

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ РЕГРЕССИИ И МЕТОДА ГРУППОВОГО УЧЕТА АРГУМЕНТОВ ПРИ МОДЕЛИРОВАНИИ ПРОЦЕССОВ ПЕРЕРАБОТКИ ПОЛЕЗНЫХ ИСКОПАЕМЫХ

Баласанян Сейран Шамирович[1],
seyran@sunicom.net

Геворгян Эрмине Михайловна[1],
hermine79@rambler.ru

[1] Капанский филиал Национального политехнического университета Армении,
Армения, 3307, г. Капан, ул. Багаберд, 28.

Актуальность работы *обусловлена необходимостью проведения сравнительного анализа методов регрессии и метода группового учета аргументов с целью оценки эффективности их применения при моделировании процессов переработки полезных ископаемых.*

Цель работы: *оценка эффективности применения методов регрессии и метода группового учета аргументов при построении статистических моделей технологических процессов переработки полезных ископаемых на основании результатов теоретического исследования, компьютерных имитационных экспериментов и практического применения указанных методов.*

Методы исследования: *методы математической статистики, метод имитационного моделирования, метод индуктивного моделирования.*

Результаты. *Проведен сравнительный анализ методов регрессии и метода группового учета аргументов путем их теоретического исследования, компьютерных имитационных экспериментов и практического применения при построении модели, описывающей статистическую зависимость прибыли от выходных интервальных характеристик технологической системы измельчения руды Зангезурского медно-молибденового комбината. В результате логического анализа вышеупомянутых методов сделан вывод, что сравнительно высокое прогнозирующее свойство моделей, построенных методом группового учета аргументов, обеспечивается как выбором оптимальной структуры модели, так и за счет описания случайной ошибки. С помощью компьютерных имитационных экспериментов исследовано влияние объёма выборки, уровня статистического шума на прогнозирующих способностях моделей, построенных с использованием обоих методов. Установлено, что сравнительно высокое прогнозирующее свойство моделей, построенных методом группового учета аргументов, проявляется особенно при умеренном статистическом шуме и малых выборках, соизмеримых с числом входных переменных. Исследованы также возможности рассмотренных методов с точки зрения выявления физических и системных закономерностей различных объектов с заданными постулированными функциями. Эффективность практического применения рассмотренных методов оценена по результатам построения технико-экономической модели технологической системы измельчения руды. Применение шагового регрессионного метода позволило построить наилучшую с точки зрения компромисса между адекватностью и сложностью модель, что свидетельствует о целесообразности применения методов регрессии при построении статистических моделей технологических процессов переработки полезных ископаемых.*

Ключевые слова:
Регрессионная модель, имитационный эксперимент, селекция, многорядный полиномиальный алгоритм, полезные ископаемые, измельчение руды.

Введение

Построение математических моделей технологических процессов (ТП) добычи и переработки полезных ископаемых часто осуществляется статистическими методами [1–7]. Одной из основных проблем построения статистических моделей является выбор наилучших входных переменных и определение оптимальной структуры модели. Наряду с широко применяемыми на практике регрессионными методами выбора наилучшего состава входных переменных и структуры модели [8–13] используется также метод группового учета аргументов (МГУА) [14–23], представляющий собой дальнейшее развитие метода регрессионного анализа.

По утверждениям авторов МГУА и его пользователей, модели, построенные с применением этого метода, по своим прогнозирующим свойствам превосходят регрессионные модели, в силу того, что МГУА, благодаря применению эвристических принципов самоорганизации, обеспечивает автоматический отбор информативных входных переменных и выбор структуры регрессионной модели оптимальной сложности, особенно при зашумленных малочисленных статистических данных.

Регрессионные методы [8–13] также позволяют выбрать наилучший состав входных переменных и структуру регрессионной модели с применением статистических критериев при рассмотрении в качестве дополнительных входных переменных преобразованные исходные переменные и использовании комбинаторного метода порождения моделей. Заметим, что с точки зрения возможности и эффективности применения построенных моделей для целей управления, наряду с прогнозирующим свойством, большое значение имеет степень их соответствия системным закономерностям моделируемого процесса, что, как правило, обеспечивается при использовании регрессионных методов.

В данной работе проведен сравнительный анализ методов регрессии и МГУА с целью оценки эффективности их применения при построении статистических моделей ТП переработки полезных ископаемых.

Теоретическое исследование методов регрессии и МГУА

При классическом регрессионном анализе постулированная модель представляется в следующем виде [8–13]:

$$\tilde{y} = G(\tilde{x}, \beta) + \tilde{\varepsilon}, \qquad (1)$$

где $\tilde{x}=(\tilde{x}_1,\tilde{x}_1,...,\tilde{x}_m)$ – вектор входных (независимых) переменных; $\tilde{y}$ – случайная выходная (зависимая) переменная; $\beta=(\beta_0,\beta_1,\beta_2,...,\beta_m)$ – неизвестный вектор коэффициентов модели; $\tilde{\varepsilon}$ – случайная величина (случайное возмущение, ошибка, шум), формально учитывающая влияние случайных факторов. Предполагается что, как переменные модели, так и случайное возмущение распределены нормально с параметрами $M[\tilde{x}]=0$, $\sigma_{\varepsilon}^2=\sigma^2=\text{const}$.

Очевидно, что в силу действия случайных факторов постулированная модель (1) не может точно прогнозировать значение выходной переменной при заданных значениях входных переменных. Следовательно, нет оснований говорить об «истинной» модели в полном смысле слова. Обычно под «истинным» значением выходной переменной понимают его условное математическое ожидание при заданных значениях входных переменных:

$$M[\tilde{y} / x] = M[G(x, \beta)] + M[\tilde{\varepsilon}] = G(x, \beta), \qquad (2)$$

где $G(x,\beta)$ – линейная модель или модель, которую можно свести к линейной по коэффициентам $\beta_0,\beta_1,\beta_2,...,\beta_m$ с помощью соответствующего преобразования.

Соотношение (2) представляет собой теоретическую регрессионную модель (уравнение регрессии y относительно x). Основной задачей регрессионного анализа является выявление и приближенное математическое описание причинно-следственной зависимости $G(x,\beta)$ между выходной и входными переменными. С этой целью на основании статистических данных (выборки $\{x_i,y_j\}_1^n$) методом наименьших квадратов строится статистический аналог соотношения (2) – эмпирическая регрессионная модель

$$\hat{y} = \widehat{G}(x, b), \qquad (3)$$

адекватность которой, т. е. степень ее соответствия статистическим данным $\{x_i,y_j\}_1^n$, оценивается с помощью выборочного коэффициента детерминации (коэффициента множественной корреляции)

$$\widehat{R}^2 = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \Big/ \sum_{i=1}^{n} (y_i - \bar{y})^2. \qquad (4)$$

Показанно [8–10], что при больших выборках статистических данных ($n>>m$) вместо (4) можно использовать следующую приближенную формулу:

$$\widehat{R}^2 \approx 1 - \widehat{\sigma}^2 \Big/ \widehat{\sigma}_y^2.$$

Коэффициент $\hat{R}^2$ показывает, какая часть вариации $\tilde{y}$ обусловлена регрессионной моделью. Из соотношения (4) следует, что максимально возможное значение коэффициента $\hat{R}^2$ ограничено погрешностью измерений, наличием неучтенных или неконтролируемых факторов и никак не зависит от состава входных переменных и структуры модели. Например, если 10 % вариации выходной переменной модели обусловлено погрешностью измерения переменных, то максимально возможное значение $\hat{R}^2$ не может превышать 0,9 при включении в модель всех входных переменных и любой структуре модели.

Однако применение МГУА при аналогичных условиях позволяет обеспечить значение $\hat{R}^2$, превышающее его максимально возможное значение, как утверждают авторы метода, за счет селективного выбора информативных входных переменных и оптимального усложнения модели. На наш взгляд, при наличии статистического шума, обусловленного погрешностью измерений и неучтенными факторами, большое значение $\hat{R}^2$, превышающее его максимально возможное значение, обеспечивается за счет излишнего усложнения модели, которая вместо существующей закономерности описывает случайные ошибки. Поясним сказанное на следующем простом примере. С целью установления зависимости напряжения участка цепи от сопротивления и силы тока производились эксперименты, в результате которых с определенной точностью измерялись значения силы тока и падения напряжения. Результаты экспериментов приведены на рис. 1.

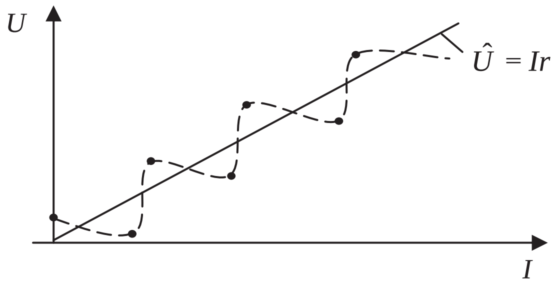

Рис. 1. *Результаты экспериментов*

Fig. 1. *Results of the experiments*

Если построить регрессионную модель на основании экспериментальных данных, то полученная зависимость с коэффициентом $\hat{R}^2<1$ приблизительно будет описывать известный закон Ома для участка цепи $\hat{U}=Ir$. Если же усложнить модель, например, используя в качестве аппроксимирующей функции ряд Фурье, то можно увеличить значение $\hat{R}^2$ до 1. Однако такая модель уже вместо существующей линейной физической закономерности будет описывать ошибки измерений. Примерно то же самое происходит при использовании МГУА, когда простые по своей физической природе объекты, описываемые линейными или квадратичными регрессионными моделями, с целью повышения прогнозирующего свойства аппроксимируются многочленами высокого порядка (до 10–20).

В связи с вышесказанным заметим, что основной задачей моделирования, в частности идентификации объектов, является выявление и математическое описание закономерности при наличии случайных факторов и статистического шума, а не обеспечение только прогнозирующего свойства модели, на что уделяется основное внимание в МГУА.

МГУА представляет собой дальнейшее развитие регрессионного анализа и является типичным методом индуктивного моделирования, основанным на некоторых принципах теории обучения и самоорганизации, в частности на принципе «селекции» или направленного отбора. Этот метод включает семейство индуктивных алгоритмов, обеспечивающих рекурсивный селективный отбор моделей, на основе которых строятся более сложные модели. Точность моделирования на каждом следующем шаге рекурсии увеличивается за счет усложнения модели. Как правило, исходную выборку разбивают на обучающую (объёмом $n_{об}$) и проверочную (объёмом $n_{пр}$) выборки соотношением $n_{об}/n_{пр}=3/2$. Первая выборка используется для оценки коэффициентов модели методом наименьших квадратов (МНК), а вторая – для определения качества модели с помощью коэффициента детерминации $\hat{R}^2$ или среднеквадратического отклонения ошибки (СКО) σ^2 (остаточной дисперсии). С каждым шагом итерации СКО уменьшается, но после достижения определенного уровня сложности, зависящего от характера и количества данных, а также общего вида модели, СКО начинает расти.

Рассмотрим более подробно процесс синтеза модели оптимальной сложности, широко используемого на практике многорядного полиномиального алгоритма МГУА [20, 22].

1. Выбирается общий вид перебираемых моделей (опорных функций) с m независимыми переменными, используя обобщенный полином Колмогорова–Габора:

$$Y = a_0 + \sum_{i=1}^{m} a_i x_i + \sum_{j=1}^{m}\sum_{i\le j} a_{ij} x_i x_j + \sum_{j=1}^{m}\sum_{i\le j}\sum_{k\le j} a_{ijk} x_i x_j x_k + \ldots,$$

который с высокой точностью может аппроксимировать любую непрерывную на конечном интервале функцию в виде полинома определенной степени (согласно теореме Вейерштрасса). Сложность модели в таком случае определяется количеством коэффициентов a_{ij} ($i,j=\overline{1,m}$, $i\le j$).

2. Используя опорные функции, строятся различные варианты моделей, включающие попарные комбинации исходных переменных, из которых составляются уравнения решающих функций, как правило, не выше второго порядка:

$$y_1 = f(x_1, x_2),\ \ y_2 = f(x_1, x_3), \ldots,\ \ y_s = f(x_i, y_i).$$

Обычно в качестве функции f выбираются простые зависимости:

$$y(x_i, x_j) = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j$$

или

$$y(x_i, x_j) = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j + a_4 q_i^2 + a_5 x_j^2.$$

Используя обучающую выборку, методом наименьших квадратов для каждой модели определяются коэффициенты a_0, a_1, a_2, a_3, a_4, a_5. Набор полученных моделей составляет первый ряд селекции. Среди всех моделей первого ряда селекции выбираются несколько (от 2 до 10) наилучших. Качество моделей определяется коэффициентом детерминации, или СКО ошибки σ^2, с помощью проверочной выборки.

3. Отобранные частные модели формируют множество новых переменных, которые являются исходными переменными для частных моделей второго ряда селекции: $z_1=f(y_i,y_j)$, $z_2=f(y_i,y_k)$,….

Коэффициенты новых моделей находятся по МНК, используя обучающую последовательность. Качество моделей оценивается с использованием проверочной последовательности, и среди них выбираются наилучшие, которые используются в качестве переменных моделей следующего, третьего ряда и т. д. Сложность общей модели возрастает от ряда к ряду. Если в качестве опорных функций используются полиномы второй степени, то на каждом шаге итерации степень результирующего полинома удваивается.

Дальнейшее усложнение модели прекращается по достижении минимума СКО σ^2 (или максимума коэффициента детерминации $\hat{R}^2$), определяемого с помощью проверочной выборки, или когда дальнейшее улучшение критерия селекции не будет превышать некоторого числа ε (параметр алгоритма).

4. На заключительном этапе, делая последовательную замену переменных, получают модель, включающую исходные переменные.

Сравнительный анализ метода шаговой регрессии и МГУА с помощью имитационных экспериментов

С целью сравнения метода шаговой регрессии и МГУА с точки зрения их возможности выявления и описания существующих зависимостей между выходным и входными переменными различных объектов с заданными функциями $G(x,\beta)$ проводились компьютерные имитационные эксперименты.

Моделирующий алгоритм, блок-схема которого приведена на рис. 2, работает следующим образом. Блок 1 осуществляет ввод исходных данных моделирования: значения параметров μ_j, σ_j ($j=\overline{1,m}$), корреляционную матрицу $\|r_{jl}\|$ ($j,l=\overline{1,m}$). Блоки 2–4 осуществляют имитацию функционирования исследуемого объекта с заданной функцией $G(x,\beta)$, т. е. формируют возможные реализации вектора входных переменных $\tilde{x}=(\tilde{x}_1,\tilde{x}_1,\ldots,\tilde{x}_m)$ методом невырожденного многомерного нормального распределения [26, 27], случайного возмущения (шума) $\tilde{\varepsilon}$ и в соответствии с соотношением (1) формируют значения выходной переменной $y^{\sim}$.

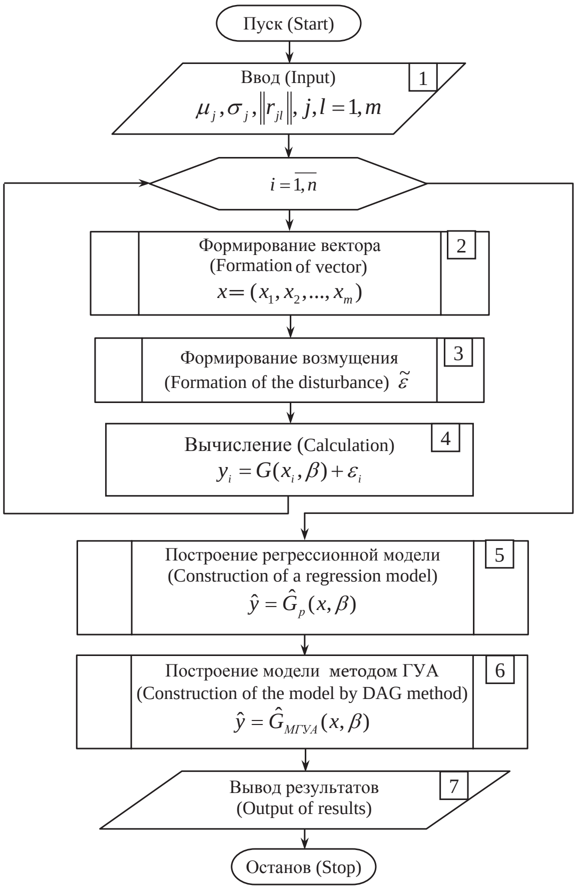

Рис. 2. *Блок-схема имитационной модели*

Fig. 2. *Block diagram of simulation model*

Далее на основании статистических данных $\{(x_{1i},x_{2i},...,x_{mi},y_i)$: $i=\overline{1,n}\}$, полученных в результате имитационного моделирования исследуемого объекта, шаговым регрессионным методом и МГУА строятся регрессионная модель M_p, иерархическая многоуровневая модель $M_{МГУА}$ и вычисляются коэффициенты детерминации R_p^2 и $R^2_{МГУА}$ полученных моделей. В процессе имитационных экспериментов были рассмотрены линейные и квадратичные функции $G(x,\beta)$, при различных значениях соотношения $\hat{\sigma}^2/\hat{\sigma}_y^2$=(0–0,5) и количестве реализаций n=20–200.

При построении регрессионных моделей объектов с квадратичными функциями $G(x,\beta)$ были рассмотрены дополнительные переменные, представляющие собой квадратичные формы исходных переменных первого порядка.

Построенные в результате имитационных экспериментов регрессионные модели всегда совпадали с заданными функциями $G(x,\beta)$, как первого, так и второго порядка. Значения коэффициента R_p^2 варьировались в пределах 0,5–0,81 в зависимости от значений коэффициентов корреляции между входными и выходным переменными и величины соотношения $\hat{\sigma}^2/\hat{\sigma}_y^2$.

Построенные с помощью МГУА модели совпали с заданными функциями $G(x,\beta)$ лишь при отсутствии или незначительном уровне случайного возмущения $\tilde{\varepsilon}$ ($\hat{\sigma}^2/\hat{\sigma}_y^2$=0–0,05). При больших значениях соотношения $\hat{\sigma}^2/\hat{\sigma}_y^2$ модели, построенные методом ГУА, несмотря на достаточно большие значения $R^2_{МГУА}$ (>0,9), не совпали с рассматриваемыми функциями $G(x,\beta)$, имели иерархическую многоуровневую (до 3–5 уровней) структуру и довольно большой порядок (6–10) исходных входных переменных. Значение коэффициента $R^2_{МГУА}$ уменьшается с увеличением значения $\hat{\sigma}^2/\hat{\sigma}_y^2$, оставаясь при этом всегда больше значения коэффициента R_p^2 регрессионных моделей.

Заметим, что сравнительно высокое прогнозирующее свойство моделей, построенных с помощью МГУА, проявляется особенно при умеренном статистическом шуме и малых выборках, соизмеримых с числом входных переменных.

Практическое применение метода шаговой регрессии и МГУА на примере системы измельчения руды

Возможности и особенности применения шаговой регрессии и МГУА рассмотрим на примере построения технико-экономической регрессионной модели, описывающей статистическую зависимость прибыли, получаемой от производства медно-молибденовых концентратов в течение 8-часового интервала (смены), от выходных интервальных характеристик технологической системы измельчения руды Зангезурского медно-молибденового комбината (ЗАО «ЗММК», Армения).

Прибыль, получаемую от производства концентратов в течение смены, можно представить в следующем виде:

$$P = D - q_1C - L, \tag{5}$$

где D – доход от реализации концентратов, выработанных в течение смены; C – условно-переменная часть затрат на переработку 1 т руды; q_1C – условно-переменная часть затрат, зависящая от массы (q_1) переработанной в течение смены руды; L – условно-постоянная часть затрат, не зависящая от объема переработки руды.

Согласно [2, 3, 5–7], к условно-переменным затратам можно отнести расходы на реагенты, шары, футеровку и другие технологические материалы, часть энергических и транспортных (внутрирудничных) затрат.

Условно-постоянные затраты включают амортизационные отчисления, расходы на капитальный ремонт, зарплату, часть энергетических затрат, некоторые статьи цеховых расходов и прочие расходы, которые не вошли в состав условно-переменных затрат.

Строго говоря, условно-переменные затраты, помимо объема переработки руды, зависят (хотя слабо) также от таких факторов, как крупность и твердость дробленой руды, крупность помола, минералогический состав и содержание металлов в руде и т. п. Однако учесть влияние этих факторов в настоящее время не представляется возможным [8]. Поэтому допускается, что условно-переменные затраты зависят только от сухой массы переработанной руды.

Таким образом, построение технико-экономической модели сводится к определению коэффициентов C, L и отысканию уравнения, описывающего зависимость дохода от реализации концентратов, выработанных в течение смены, от интервальных выходных характеристик ТСИР (q_1,q_2,q_3):

$$D=\varphi(q_1,q_2,q_3). \qquad (6)$$

В настоящее время единственно возможным подходом к построению указанной выше модели является статистический, который основывается на обработке экспериментальных данных, собранных непосредственно на действующем объекте.

С целью построения статистической модели (6) совместно с работниками ЗММК в течение 180 смен было произведено апробирование процесса измельчения и параметров конечного продукта. При этом регистрировались:

- сухая масса руды (q_1), переработанной в течение смены (тонн);
- среднесменное процентное содержание класса < 80 мкм в измельченном продукте (q_2);
- среднесменная плотность объединенной пульпы (г/л) (q_3);
- вес готового молибденового концентрата γ_{Mo} (тонн), выработанного в течение смены, и среднесменное процентное содержание молибдена (β_{Mo}) в нем;
- вес готового медного концентрата γ_{Cu}, выработанного в течение смены (тонн), и среднесменное процентное содержание меди (β_{Cu}) в нем.

Доход от реализации концентратов, выработанных в течение смены, вычислялся по формуле

$$D=\frac{\gamma_{\mathrm{Mo}}\beta_{\mathrm{Mo}}}{100}C_{\mathrm{Mo}}+\frac{\gamma_{\mathrm{Cu}}\beta_{\mathrm{Cu}}}{100}C_{\mathrm{Cu}},$$

где C_{Mo}, C_{Cu} – соответственно рыночные цены на молибден и медь.

Применение метода шаговой регрессии

До того как приступить к построению модели, было проверено выполнение следующих условий, обеспечивающих корректность применения корреляционно-регрессионного анализа:

- нормальность и стационарность входного процесса $q(t)=(q_1(t),q_2(t),q_3(t))$;
- нормальность и стационарность выходного процесса $D(t)$.

При проверке гипотезы о нормальном распределении переменных модели был использован критерий χ^2 Пирсона, а проверка стационарности входного и выходного процессов была выполнена с помощью критерия Шефе [26].

Следующим важным вопросом при построении регрессионной модели $D=\varphi(q_1,q_2,q_3)$ является выбор ее структуры. Опыт практического применения корреляционно-регрессионного анализа для математического описания производственных, в частности обогатительных процессов [1–7], свидетельствует о том, что в качестве статистической модели нормально функционирующего объекта достаточно рассмотреть алгебраический полином не выше второй степени.

На основании вышесказанного, а также учитывая сравнительно невысокие значения коэффициентов вариации переменных (ν_{q_1}=8,7 %, ν_{q_2}=1 %, ν_{q_3}=0,8 %, ν_D=8,3 %), максимальная степень аппроксимирующего полинома была выбрана равной двум.

Таким образом, для построения регрессионной модели имеем следующие исходные и преобразованные переменные: D – отклик, выходная переменная; q_1, q_2, q_3 – исходные входные переменные; $q_4=q_1\cdot q_2$, $q_5=q_1\cdot q_3$, $q_6=q_2\cdot q_3$, $q_7=q_2^2$, $q_8=q_3^2$, $q_9=q_1^2$ – преобразованные входные переменные.

Контрольные данные:

- число наблюдений –180;
- критическое значение общего (частного) F-критерия – $F_{0,975}(1;170)$=3,9;
- уровень риска для доверительных границ – 5 %.

При построении регрессионной модели были использованы расчетные формулы, приведенные в [8–13].

Шаговая регрессионная процедура начинается с построения простой корреляционной матрицы **R**. Проверка значимости полученных коэффициентов парной корреляции по критерию Стьюдента показала, что все они статистически значимы при уровне значимости α=0,05.

		D	q_1	q_2	q_3	q_4	q_5	q_6	q_7	q_8	q_9
	D	1	0,61	0,18	0,30	0,82	0,62	0,22	0,12	0,25	0,60
	q_1	0,61	1	–0,30	0,30	0,50	0,95	–0,25	–0,31	0,24	0,95
	q_2	0,18	–0,30	1	–0,62	0,45	–0,47	0,95	0,95	–0,62	–0,28
	q_3	0,30	0,30	0,62	1	–0,16	0,30	–0,29	–0,55	0,95	0,24
R =	q_4	0,82	0,50	0,45	–0,16	1	0,50	0,41	0,40	–0,15	0,51
	q_5	0,62	0,95	–0,47	0,30	0,50	1	–0,28	–0,41	0,45	0,95
	q_6	0,22	–0,25	0,95	–0,29	0,41	–0,28	1	0,95	–0,28	–0,25
	q_7	0,12	–0,31	0,95	–0,55	0,40	–0,41	0,95	1	–0,53	–0,25
	q_8	0,25	0,24	–0,62	0,95	–0,15	0,45	–0,28	–0,53	1	0,2 0
	q_9	0,60	0,95	–0,28	0,24	0,51	0,95	–0,25	–0,25	0,20	1

Шаг 1

1. В качестве первой независимой переменной для включения в регрессионную модель выбирается q_4, которая наиболее сильно коррелирована с откликом:

$$\max_i\{\left|r_{Dq_i}\right| : i=\overline{1,9}\}=\left|r_{Dq_4}\right|=0,82.$$

2. Вычисляется значение общего F-критерия для уравнения $\hat{D}=f_1(q_4)$ по формуле

$$\widehat{F}_{q_4}=\frac{R^2_{Dq_4}}{1-R^2_{Dq_4}}(n-m-1)=365,2,$$

где n – число наблюдений; m – число независимых переменных регрессионного уравнения; $R_{Dq_4}=r_{Dq_4}$.

Уравнение $\hat{D}=f_1(q_4)$ является статистически значимым, так как расчетное значение F-критерия превосходит выбранное критическое значение (3,9).

3. Вычисляются остаточная дисперсия $\hat{\sigma}^2$ и коэффициент вариации отклика $v_{\hat{D}}$ по формулам

$$\widehat{\sigma}^2=\sigma_{\hat{D}}^2(1-R_{Dq_4}^2);\;\; v_{\hat{D}}=\frac{\widehat{\sigma}}{\bar{D}}\cdot 100=5{,}05\;\%.$$

Шаг 2

1. Вычисляются частные коэффициенты корреляции (табл. 1) между каждой независимой переменной q_i ($i=\overline{1,9}$; $i\neq 4$) и откликом, при условии, что переменная q_4 включена в регрессионную модель, по формуле

$$r_{Dq_i\bullet q_4}=\frac{-D_{13}^{1i}}{\sqrt{D_{11}^{1i}D_{33}^{1i}}},$$

где D_{13}^{1i}, D_{11}^{1i}, D_{33}^{1i} – алгебраические дополнения соответствующих элементов корреляционной матрицы D^{1i}:

$$D^{1i}=\begin{vmatrix}1 & r_{Dq_4} & r_{Dq_i}\\ r_{Dq_4} & 1 & r_{q_4q_i}\\ r_{Dq_i} & r_{q_4q_i} & 1\end{vmatrix}.$$

Таблица 1. *Коэффициенты частной корреляции*

Table 1. *Coefficients of private correlation*

i	1	2	3	5	6	7	8	9
$\lvert r_{Dq_i\bullet q_4}\rvert$	0,146	0,136	0,611	0,464	0,052	0,157	0,433	0,22
$\lvert r_{Dq_i\bullet q_4q_3}\rvert$	0,0067	0,00335	–	0,163	0,00441	0,017	0,0016	0,0075

2. Выбирается

$$\max_i\{\left|r_{Dq_i\bullet q_4}\right|: i=\overline{1,9}; i\neq 4\}=\left|r_{Dq_3\bullet q_4}\right|=0{,}611.$$

В соотвтствии с шаговым методом в качестве следующей переменной, претендующей для включения в регрессионную модель, выбирается q_3, которая характеризуется наиболее высоким частным коэффициентом корреляции с откликом.

3. Вычисляется множественный коэффициент корреляции для регрессионной модели $\hat{D}=f_2(q_4,q_3)$ по формуле

$$R_{D\bullet q_4q_3}=\sqrt{1-\frac{D^{13}}{D^{11}}}=0{,}9339,$$

где

$$D^{13}=\begin{vmatrix}1 & r_{Dq_4} & r_{Dq_3}\\ r_{Dq_4} & 1 & r_{q_4q_3}\\ r_{Dq_3} & r_{q_4q_3} & 1\end{vmatrix}.$$

4. Вычисляется значение частного F-критерия для проверки значимости вклада q_3:

$$\widehat{F}_{q_3/q_4}=\frac{(R_{D\bullet q_4q_3}^2-r_{D\bullet q_4}^2)}{1-R_{D\bullet q_4q_3}^2}(n-m-1)=276{,}7.$$

Как видно, расчетное значение частного F-критерия превосходит табличное значение (3,9). Это свидетельствует о том, что новая переменная q_3 дает статистически значимое снижение остаточный суммы квадратов (см. табл. 3). Следовательно, переменная q_3 в регрессионной модели сохраняется.

5. Вычисляется значение частного F-критерия для проверки значимости вклада переменной q_4, который имел бы место, если бы в модель была введена сначала q_3, затем q_4 (вклад q_4/q_3), по формуле

$$\widehat{F}_{q_4/q_3}=\frac{(R_{D\bullet q_4q_3}^2-r_{D\bullet q_3}^2)}{1-R_{D\bullet q_4q_3}^2}(n-m-1)=1075.$$

Поскольку расчетное значение частного F-критерия больше критического, переменная q_4 в регрессионной модели сохраняется.

6. Вычисляется значение общего F-критерия для модели $\hat{D}=f_2(q_4,q_3)$:

$$\widehat{F}_{q_4q_3}=\frac{R_{D\bullet q_4q_3}^2(n-m-1)}{(1-R_{D\bullet q_4q_3}^2)m}=400{,}6.$$

Общий F-критерий показывает, что регрессионная модель $\hat{D}=f_2(q_4,q_3)$ является статистически значимой при уровне значимости α=0,05.

7. Вычисляются остаточная дисперсия и коэффициент вариации отклика по формулам

$$\widehat{\sigma}^2=\sigma_D^2(1-R^2{}_{D\bullet q_4q_3});\;\; v_{\hat{D}}=\frac{\widehat{\sigma}}{\bar{D}}100=3{,}15\;\%.$$

Шаг 3

1. Вычисляются частные коэффициенты корреляции $r_{Dq_i\bullet q_4q_3}(i=\overline{1,9}; i\neq 3, i\neq 4)$ (табл. 2) между каждой переменной, не включенной в регрессионную модель, и откликом, при условии, что q_4 и q_3 уже включены в модель, по формуле

$$r_{Dq_i\bullet q_4q_3}=\frac{-D_{14}^{2i}}{\sqrt{D_{11}^{2i}\cdot D_{33}^{2i}}},$$

где

$$D^{2i}=\begin{vmatrix}1 & r_{Dq_4} & r_{Dq_3} & r_{Dq_i}\\ r_{Dq_4} & 1 & r_{q_4q_3} & r_{q_4q_i}\\ r_{Dq_3} & r_{q_4q_3} & 1 & r_{q_3q_i}\\ r_{Dq_i} & r_{q_3q_i} & r_{q_3q_i} & 1\end{vmatrix}.$$

Выбирается

$$\max_i\{\left|r_{Dq_i\bullet q_4q_3}\right|: i=\overline{1,9}; i\neq 3; i\neq 4\}=\left|r_{Dq_5\bullet q_4q_3}\right|=0{,}163.$$

Следовательно, следующей переменной, претендующей на включение в регрессионное уравнение, является q_5.

2. Вычисляется множественный коэффициент корреляции для регрессионной модели $\hat{D}=f_3(q_4,q_3,q_5)$:

$$R_{D\bullet q_4 q_3 q_5} = \sqrt{1 - \frac{D^{25}}{D^{25}_{11}}} = 0,935,$$

где

$$D^{25} = \left| \begin{array}{ccc:c} & D^{13} & & \begin{array}{c} r_{Dq_5} \\ r_{q_4 q_5} \\ r_{q_5 q_3} \end{array} \\ \hdashline r_{Dq_5} & r_{q_4 q_5} & r_{q_5 q_3} & 1 \end{array} \right|.$$

3. Вычисляется значение частного F-критерия для проверки значимости вклада:

$$\widehat{F}_{q_5/q_4 q_3} = \frac{(R^2_{D\bullet q_4 q_3 q_5} - R^2_{D\bullet q_4 q_3})}{(1 - R^2_{D\bullet q_4 q_3 q_5})}(n - m - 1) = 2,95.$$

Поскольку значение частного F-критерия меньше критического, т. е. вклад переменной q_5 является статистически незначимым при α=0,05, то она не включается в модель и шаговая регрессионная процедура заканчивается.

Таким образом, в качестве наилучших переменных регрессионной модели выбираются q_3 и q_4. Процесс выбора наилучших переменных компактно представлен с помощью табл. 2.

Таблица 2. *Значения F-критерия и коэффициента R*

Table 2. *Values of F-criterion and coefficient R*

Включаемая переменная Included variable	q_4	q_3/q_4	q_4/q_3	q_5/q_4q_3
Регрессионная модель на предыдущем шаге Regression model at the previous step	$\Delta=\bar{D}$	$\Delta=f_1(q_4)$	$\Delta=f_3(q_3)$	$\Delta=f_2(q_4,q_3)$
Регрессионная модель на данном шаге Regression model at this step	$\Delta=f_1(q_4)$	$\Delta=f_2(q_4,q_3)$	$\Delta=f_2(q_4,q_3)$	$\Delta=f_4(q_4,q_3,q_5)$
Общий $\widehat{F}$-критерий General $\widehat{F}$-criterion	365,3	400,6	400,6	408,3
Частный $\widehat{F}$-критерий Private $\widehat{F}$-criterion	365,3	276,7	1075	2,95
Коэффициент детерминации Coefficient of determination	0,8199	0,8722	0,8722	0,8742
Коэффициент вариации Coefficient of variation	8,83	5,05	5,05	3,5

Построение регрессионной модели $\widehat{D}=f_2(q_3,q_4)$ осуществляется следующим образом.

1. Методом наименьших квадратов определяются оценки стандартизованных коэффициентов выбранной регрессионной модели $\widehat{\alpha}_1$=0,8924 и $\widehat{\alpha}_2$=0,453, которые являются решением следующей системы линейных алгебраических уравнений:

$$\begin{cases} \widehat{\alpha}_1 - 0,16\widehat{\alpha}_2 = 0,82, \\ -0,16\alpha_1 + a_2 = 0,31. \end{cases}$$

2. Определяются 95%-е доверительные границы для коэффициентов $\widehat{\alpha}_1$ и $\widehat{\alpha}_2$ (табл. 3) по формулам:

$$\widehat{\alpha}_{i\text{н}} = \widehat{\alpha}_i - t(n - m - 1; 0,975)\widehat{\sigma}_{\widehat{\alpha}_i},$$

$$\widehat{\alpha}_{i\text{в}} = \widehat{\alpha}_i + t(n - m - 1; 0,975)\widehat{\sigma}_{\widehat{\alpha}_i},$$

где $\widehat{\sigma}_{\widehat{\alpha}_i} = (\sqrt{1 - R^2_{D\bullet q_4 q_3}}) / \sqrt{n(1 - r^2_{q_4 q_3})} = 0,0271.$

Таблица 3. *Значения коэффициентов регрессии и их доверительных границ*

Table 3. *Values of the regression coefficients and their confidence bounds*

Независ. перемен. Independ. variables	Стандартизованные коэффициенты регрессии Standardized regression coefficients	Стандартная ошибка коэффициентов регрессии Standard error of regression coefficients	Дов. границы для коэффициентов Confidence bounds for coefficients (α=0,05)		Относительная ошибка вычисления Relative error of calculations
			верхний upper	нижний lower	
q_4	0,6385	0,0553	0,7465	0,5305	17 %
q_3	0,30554	0,553	0,4134	0,1974	35 %

3. Вычисляются оценки натуральных коэффициентов уравнения регрессии по формулам:

$$\hat{b}_1 = \frac{\widehat{\sigma}_D}{\widehat{\sigma}_{q_4}}\widehat{\alpha}_1; \ \hat{b}_2 = \frac{\widehat{\sigma}_D}{\widehat{\sigma}_{q_3}}\widehat{\alpha}_2; \ \hat{b}_0 = \bar{D} - \hat{b}_1\bar{q}_4 - \hat{b}_2\bar{q}_3.$$

4. Определяются 95%-е доверительные границы для коэффициентов $\hat{b}_1$ и $\hat{b}_2$ по формулам:

$$\hat{b}_{1\text{в/н}} = \hat{b}_1 \pm t(n - m - 1; 0,975)\frac{\widehat{\sigma}_{\widehat{\alpha}_1}}{\widehat{\sigma}_{q_4}}\widehat{\sigma}_D,$$

$$\hat{b}_{2\text{в/н}} = \hat{b}_2 \pm t(n - m - 1; 0,975)\frac{\widehat{\sigma}_{\widehat{\alpha}_2}}{\widehat{\sigma}_{q_3}}\widehat{\sigma}_D.$$

5. Определяется ширина 95%-го доверительного интервала единичного предсказания для регрессионной модели при значениях $q_4=\bar{q}_4$ и $q_3=\bar{q}_3$ по формуле

$$\Delta D = t(n - m - 1; 0,975)\widehat{\sigma}\sqrt{\frac{n+1}{n}} \approx 2\widehat{\sigma}.$$

Относительная погрешность единичного предсказания по уравнению регрессии для $q_4=\bar{q}_4$ и $q_3=\bar{q}_3$, определяемая как $\frac{\Delta D}{\bar{D}}100\ \%$, составляет 6,315 %.

Теперь рассмотрим вопросы оценки адекватности и работоспособности построенной модели. Известно [9, 10, 13], что не всякая статистически значимая регрессионная модель обязательно пригодна для прогнозирования или управления. Действительно, если размах отклика $\hat{D}$ незначительно превосходит величину случайной ошибки, предсказание не будет иметь никакой ценности, так как, несмотря на значимость, такая модель будет описывать только ошибки. Поэтому, для того чтобы модель можно было считать удовлетворительной для целей предсказания, расчетное значение общего F-критерия должно не просто превышать выбранную процентную точку, а превосходить ее примерно в четыре раза [9].

В соответствии с вышесказанным полученная регрессионная модель адекватна, поскольку расчетное значение общего F-критерия в 200 раз превосходит процентную точку $F_{0,95}(2;177)=2,95$, обладает достаточно хорошей прогнозирующей способностью ($\hat{R}^2=0,8834$) и вполне пригодна для предсказания величины дохода D по заданным значениям интервальных выходных характеристик системы измельчения.

6. Составляется таблица дисперсионного анализа (табл. 4) регрессионной модели.

Таблица 4. *Показатели дисперсионного анализа*
Table 4. *Indicators of variance analysis*

Источник рассеяния Source of scattering	Степень свободы Degree of freedom	Сумма квадратов Sum of squares	Средний квадрат Mean square	Значение F-критерия The value of F-criterion
Общий Total	179	1	–	–
Обусловленный регрессией Due to regression $\hat{D}=f(q_4,q_3)$	3	0,8722	0,291	400,6
Обусловленный q_4 Due to variable q_4	1	0,6724	0,6724	365,3
Обусловленный q_3/q_4 Due to variable q_3/q_4	1	0,1998	0,1998	276,7
Обусловленный q_4/q_3 Due to variable q_4/q_3	1	0,776	0,776	1045
Обусловленный q_5/q_4q_3 Due to variable q_5/q_4q_3	1	0,021	0,0021	2,95
Остаток Residue	176	0,1278	0,00073	–

Определив по данным калькуляции себестоимости медного и молибденового концентратов значения C и L указанным ранее образом и подставив их в (5), получим технико-экономическую регрессионную модель для прогнозирования прибыли.

В нормированном масштабе эта модель имеет следующий вид:

$$P = q_1(0,039q_2 - 0,433) + 113,7q_3 - 170499. \quad (7)$$

Построенная регрессионная модель компактна и с точки зрения компромисса между адекватностью и сложностью обладает структурой оптимальной сложности, что позволяет использовать ее как для целей прогнозирования, так и для управления.

Применение МГУА

Теперь для разработки технико-экономической модели будем использовать описанный ранее многорядный полиномиальный алгоритм МГУА.

При построении модели имеем следующие исходные данные:

- D – отклик, выходная переменная;
- q_1, q_2, q_3 – входные переменные;
- число наблюдений – 180;
- объем обучающей выборки – 108;
- объем проверочной выборки – 72.

1. Выбирается общий вид перебираемых моделей (опорных функций) с m независимыми переменными, используя обобщенный полином Колмогорова–Габора второго порядка:

$$Y = a_0 + \sum_{i=1}^{m} a_i q_i + \sum_{j=1}^{m} \sum_{i \le j} a_{ij} q_i q_j .$$

Используя опорные функции, строятся возможные варианты моделей, включающие попарные комбинации исходных переменных, из которых составляются уравнения решающих функций не выше второго порядка:

$$y(q_i, q_j) = a_0 + a_1 q_i + a_2 q_j + a_3 q_i q_j + a_4 q_i^2 + a_5 q_j^2,$$

$$y_1 = f_1(q_1,q_2) = a_0^{(1)} + a_1^{(1)} q_1 + a_2^{(1)} q_2 + a_3^{(1)} q_1 q_2 + a_4^{(1)} q_1^2 + a_5^{(1)} q_2^2,$$

$$y_2 = f_2(q_1,q_3) = a_0^{(2)} + a_1^{(2)} q_1 + a_2^{(2)} q_3 + a_3^{(2)} q_1 q_3 + a_4^{(2)} q_1^2 + a_5^{(2)} q_3^2,$$

$$y_3 = f_3(q_2,q_3) = a_0^{(3)} + a_1^{(3)} q_2 + a_2^{(3)} q_3 + a_3^{(3)} q_2 q_3 + a_4^{(3)} q_2^2 + a_5^{(3)} q_3^2.$$

2. Используя обучающую выборку, методом наименьших квадратов для каждой модели определяются коэффициенты $a_0^{(k)}$, $a_1^{(k)}$, $a_2^{(k)}$, $a_3^{(k)}$, $a_4^{(k)}$, $a_5^{(k)}$, ($k=\overline{1,3}$).

Набор полученных моделей составляет первый ряд селекции. С помощью проверочной выборки определяются коэффициенты детерминации построенных моделей: $\hat{R}^2_{q_1q_2}=0,738$, $\hat{R}^2_{q_1q_3}=0,403$, $\hat{R}^2_{q_2q_3}=0,115$, по которым оценивается качество моделей. Среди указанных моделей первого ряда селекции выбираются два наилучшие.

3. Отобранные частные модели формируют множество новых переменных y_1 и y_2, которые являются исходными переменными для единственной модели второго ряда селекции:

$$D = \varphi(y_1, y_2) = a_0^* + a_1^* y_1 + a_2^* y_2 + a_3^* y_1 y_2 + a_4^* y_1^2 + a_5^* y_2^2.$$

Коэффициенты новой модели находятся методом наименьших квадратов, используя обучающую последовательность. Качество полученной модели оценивается коэффициентом детерминации, вычисленным с использованием проверочной выборки ($\hat{R}^2$=0,9285).

4. Делая последовательную замену переменных, получается модель с исходными переменными, которая описывает статистическую зависимость дохода от выходных интервальных характеристик ТСИР:

$$
\begin{aligned}
D = a_0 + a_1 q_1 + a_2 q_2 + a_3 q_3 + a_4 q_1 q_2 + a_5 q_1 q_3 + a_6 q_2 q_3 + \\
+ a_7 q_1^2 + a_8 q_2^2 + a_9 q_3^2 + a_{10} q_1 q_2 q_3 + a_{11} q_1^2 q_2 + a_{12} q_1^2 q_3 + \\
+ a_{13} q_1 q_2^2 + a_{14} q_1 q_3^2 + a_{15} q_1^3 + a_{16} q_2^3 + a_{17} q_1^2 q_2^2 + a_{18} q_1^2 q_3^2 + \\
+ a_{19} q_1^3 q_2 + a_{20} q_1^3 q_3 + a_{21} q_1 q_3^3 + a_{22} q_2 q_3^3 + a_{23} q_1^2 q_2 q_3 + \\
+ a_{24} q_1 q_2 q_3^2 + a_{25} q_1^4 + a_{26} q_2^4 + a_{27} q_3^4 .
\end{aligned}
$$

Подставляя построенную модель в соотношение (2), получим модель для прибыли, представляющую собой неполный полином четвертого порядка, структура которой намного сложнее по сравнению со структурой регрессионной модели (7).

Очевидно, что с точки зрения компромисса между адекватностью и сложностью построенная методом шаговой регрессии модель (7), безусловно, предпочтительнее модели, полученной с помощью МГУА, которая, обладая достаточно сложной структурой, обеспечивает небольшой прирост (5,1 %) коэффициента детерминации.

Несмотря на несложную структуру, регрессионная модель (7) обладает достаточно хорошей прогнозирующей способностью ($\hat{R}^2$=0,8834),что позволяет использовать ее в составе компьютерной модели для оценки и исследования эффективности функционирования ТСИР ЗММК с учетом влияния надежности измельчительного оборудования [6].

Выводы

На основании исследования методов регрессии и МГУА, сравнения полученных результатов проведенных имитационных экспериментов, а также учитывая опыт практического применения указанных методов при построении статистических моделей производственных процессов, в частности процессов переработки полезных ископаемых, можно сделать следующие выводы:

1. Входные и выходные параметры современных автоматизированных технологических процессов изменяются в небольших пределах, при которых вполне достаточно рассмотрение их математических моделей в виде многочлена $G(x,\beta)$ второго порядка. Как показывает практический опыт математического моделирования технологических процессов переработки полезных ископаемых, при указанных условиях более эффективно применение регрессионных методов, используемых во многих современных прикладных программных средствах статистического анализа производственных процессов (Data Mining, SPSS, Maple, StatGraphics, Statistica, MathCad, Mathlab и т. п.).
2. В результате имитационных экспериментов установлено, что, несмотря на сравнительно невысокое прогнозирующее свойство, регрессионные модели всегда совпадали с заданными функциями $G(x,\beta)$, независимо от уровня случайного возмущения и объёма выборки, а модели, полученные с помощью МГУА, совпадали с заданными функциями лишь при отсутствии или незначительном уровне случайного возмущения и малых выборках. Следовательно, регрессионные модели более адекватно отражают системные закономерности моделируемых процессов, обладают оптимальной структурой с точки зрения компромисса между сложностью и адекватностью и вполне пригодны для перспективного и оперативного управления технологическими процессами.
3. Применение шагового регрессионного метода позволило выбрать наилучший состав входных переменных и рациональную структуру технико-экономической модели, адекватно описывающей статистическую связь между прибылью и выходными интервальными характеристиками технологической системы измельчения руды, что свидетельствует об эффективности применения методов регрессии при построении статистических моделей технологических процессов переработки полезных ископаемых.
4. МГУА целесообразно использовать в таких случаях, когда невозможно использовать методы регрессионного анализа (при малочисленных выборках, соизмеримых с числом входных переменных, нарушении условия нормальности распределения выборки и т. п.).
5. Сравнительно высокое прогнозирующее свойство моделей, построенных с помощью МГУА, обеспечивается как выбором оптимальной структуры модели, так и за счет описания случайной ошибки. Последнее обстоятельство является основной причиной несовпадения этих моделей с постулированными функциями $G(x,\beta)$ в имитационных экспериментах.
6. Иерархические многоуровневые модели, построенные с помощью МГУА настолько громоздки и сложны, что, несмотря на их хорошее прогнозирующее свойство, практически малопригодны для использования с целью оперативного управления современными технологическими процессами. Чрезмерная сложность этих моделей, включающих переменные высокого порядка (до 14–20), исключают возможность их исследования аналитическими методами, а применение численных методов оптимизации требует чрезмерно большого времени компьютерной реализации.

СПИСОК ЛИТЕРАТУРЫ

1. King R.P. Modeling and Simulation of Mineral Processing Systems. 2nd ed. – Boston: Butterworth-Heinemann, 2015. – 416 p.
2. Jovanović I., Miljanović I., Jovanović T. Soft computing-based modeling of flotation processes (review) // Minerals Engineering. – 2015. – V. 84. – P. 34–63.
3. Integrated Optimization for the Automation Systems of Mineral Processing / T. Chai, J. Ding, G. Yu, H. Wang // IEEE Transactions on Automation Science and Engineering. – 2014. – V. 11. Iss. 4. – P. 965–982.
4. Пасечник И.А., Шек В.М. Анализ методов моделирования процессов обрушения горных пород при подземной добыче полезных ископаемых // Горный информационно-аналитический бюллетень. – 2010. – № 6. – С. 194–198.
5. Разработка и применение АСУ процессами обогащения полезных ископаемых / В.В. Морозов, В.П. Топчаев, К.Я. Улитенко, З. Ганбаатар, Л. Дэлгэрбат. – М.: Изд. дом «Руда и Металлы», 2013. – 512 с.
6. Баласанян С.Ш. Стратифицированное моделирование сложных технологических систем. – Саарбрукен, Германия: LAP Lambert Academic Publishing, 2016. – 385 с.
7. Авдохин В.М. Основы обогащения полезных ископаемых. Т. 1. Обогатительные процессы. – М.: Изд-во «Горная книга», 2014. – 417 с.
8. Weisberg S. Applied Linear Regression. 4th ed. – Hoboken, NJ: Wiley, 2014. – 368 p.
9. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. – М.: Изд. дом «Вильямс», 2007. – 912 с.
10. Seber G.A.F., Lee A.J. Linear Regression Analysis. 2nd ed. – Hoboken, NJ: Wiley, 2003. – 582 p.
11. Hoffmann Jh.P., Shafer K. Linear Regression Analysis. Assumptions and Applications. – Washington: NASW Press, 2015. – 240 p.
12. Bingham N.H., Fry J.M. Regression. – London: Springer, 2010. – 284 p.
13. Montgomery D.C., Peck E.A., Vining G.G. Introduction to Linear Regression Analysis. Fifth ed. – New York: Wiley, 2013. – 836 p.
14. Ivakhnenko A.G. Heuristic self-organization on problems of engineering cybernetics // Automatic. –1970. – V. 6 (3). – P. 207–219.
15. Godfrey Onwubolu. GMDH and Implementation in C-Methodology. – London: Imperial College Press, 2015. – 303 p.
16. Atashrouz S., Pazuki G., Alimoradi Y. Estimation of the viscosity of nine nanofluids using a hybrid GMDH-type neural network system // Fluid Phase Equilibria. – 2014. –V. 372. – P. 43–48.
17. Ahmadi M.H., Ahmadi M.A., Rosen M.A. Using GMDH Neural Networks to Model the Power and Torque of a Stirling Engine // Sustainability. – 2015. – V. 7 (2). – P. 2243–2255.
18. Mueller J-A., Lemke F. Self-Organizing Data Mining. – Hamburg: Libri, 2000. – 225 p.
19. Ivakhnenko A.G., Ivakhnenko G.A., Mueller J.A. Self-Organization of Neuronets with Active Neurons // Pattern Recogn. and Image Analysis. – 1994. – V. 4. – № 4. – P. 177–188.
20. Ивахненко А.Г., Савченко Е.А. Исследование эффективности метода доопределения выбора модели в задачах моделированиием с применением МГУА // Проблемы управления и информатики. – 2008. – № 2. – С. 65–76.
21. Сарычев А.П. Системный критерий регулярности в методе группового учета аргументов // Проблемы управления и информатики. – 2006. – № 6. – С. 25–37.
22. Степашко В.С., Булгакова А.С. Обобщенный итерационный алгоритм метода группового учета аргументов // Управляющие системы и машины. – 2013. – № 2. – С. 9–26.
23. Goharriz M., Marandi S.M. An Optimized Neuro-Fuzzy GMDH System Based on Gravitational Search Algorithm for Evaluation of Lateral Ground // Int. J. Optim. Civil Eng. – 2016. – V. 6 (3). – P. 385–403.
24. Ермаков С.М. Метод Монте-Карло в вычислительной математике. Вводный курс. – СПб.: Невский диалект, 2009. – 192 с.
25. Balasanyan S.Sh., Gevorgyan H.M. The Construction of Statistical Model of a Virtual Object in the LabVIEW Environment // Proc. of 11th International Conference on Remote Engineering and Virtual Instrumentation REV2014. – Porto, 26–28 Feb., 2014. – P. 291–293.
26. Кузнецова О.С. Теория вероятности и математическая статистика. Краткий курс. – М.: Изд-во «Окей-книга», 2013. – 191 с.
27. Баласанян С.Ш. Метод стратифицированной формализации сложных технологических систем со многими состояниями // Известия Томского политехнического университета. Инжиниринг георесурсов. – 2016. – Т. 327. – № 1. – С. 6–18.

Поступила 15.01.2016 г.

Информация об авторах

Баласанян С.Ш., доктор технических наук, доцент, заместитель директора по научной работе Капанского филиала Национального политехнического университета Армении.

Геворгян Э.М., преподаватель кафедры информационных технологий, информатики и автоматизированных систем Капанского филиала Национального политехнического университета Армении.

UDC 681.51:622.73

COMPARATIVE ANALYSIS OF THE REGRESSION METHODS AND DATA GROUP ACCOUNTING METHOD IN MODELING MINERAL PROCESSING

Seyran Sh. Balasanyan[1],
seyran@sunicom.net

Hermine M. Gevorgyan[1],
E-mail: hermine79@rambler.ru

[1] Kapan branch of National Polytechnic University of Armenia,
28, Baghaberd Street, Kapan, 3307, Armenia.

The relevance *of the discussed issue is caused by the need to conduct a comparative analysis of regression methods and the data group accounting method in order to evaluate the effectiveness of their use in modeling mineral processing.*
The main aim *of the study is to assess the effectiveness of application of the regression methods and the data group accounting method in constructing statistical models of minerals processing based on the results of theoretical studies, computer simulation experiments and practical application of these methods.*
The methods used in the study: *methods of mathematical statistics, simulation method, method of inductive modeling.*
The results. *The authors have carried out the comparative analysis of regression methods and the data group accounting method by theoretical investigations, computer simulations and practical applications at the construction of model describing the statistical dependence of profit on the output interval characteristics of the ore grinding technological system of Zangezur Copper and Molybdenum Combine. As a result of the logical analysis of the above mentioned methods the authors concluded that a relatively high predictive property of the models constructed by the data group accounting method is ensured both by the selection of the optimal model structure, and by description of the random error. Using the computer simulation experiments the authors investigated the influence of sample size, level of statistical noise on the predictive ability of the models built using both methods. It was ascertained that a relatively high predictive ability of models constructed by the data accounting group method occurs especially at moderate statistical noise and small samples, comparable to the number of input variables. The possibilities of these methods were studied as well in terms of identifying physical and systemic regularities of various objects with the specified postulated functions. The effectiveness of the practical application of the considered methods is evaluated by the results of construction of techno-economic model of ore grinding technological system. The application of a stepwise regression method allowed constructing the best possible model in terms of a compromise between adequacy and complexity. This indicates the feasibility of applying the regression methods at constructing the statistical models of minerals processing.*

Key words:
Regression model, simulation experiment, selection, multi-row polynomial algorithm, minerals, ore grinding.

REFERENCES

1. King R.P. *Modeling and Simulation of Mineral Processing Systems*. Boston, Butterworth-Heinemann, 2014. 404 p.
2. Jovanovich I., Miljanovich I., Jovanovich T. Soft computing-based modeling of flotation processes (review). *Minerals Engineering*, 2015, vol. 84, pp. 34–63.
3. Chai T., Ding J., Yu G., Wang H. Integrated Optimization for the Automation Systems of Mineral Processing. *IEEE Transactions on Automation Science and Engineering*, 2014, vol. 11, Iss. 4, pp. 965–982.
4. Pasechnik I.A., Shek V.M. Analysis methods for modeling the processes of rocks destruction in underground mining. *Mining informational and analytical bulletin*, 2010, no. 6, pp. 194–198. In Rus.
5. Morozov V.V., Topchaev V.P., Ulitenko K.Ya., Ganbaatar Z., Delgerbat L. *Razrabotka i primenenie avtomatizirovannykh sistem upravleniya protsessami obogashcheniya poleznykh iskopaemykh* [The development and application of automated control process of mineral processing]. Moscow, Ruda i Metally Publ., 2013. 512 p.
6. Balasanyan S.Sh. *Stratifitsirovannoe modelirovanie slozhnykh tekhnologicheskikh sistem* [Stratified modeling of complex technological systems]. Saarbruken, Germany, LAP Lambert Academic Publishing, 2016. 385 p.
7. Avdokhin V.M. *Osnovy obogashcheniya poleznykh iskopaemykh. T. 1. Obogatitelnye protsessy* [The fundamentals of mineral processing. Vol. 1. Mineral processing]. Moscow, Gornaya kniga Publ., 2014. 417 p.
8. Weisberg S. *Applied Linear Regression. 4th ed.* Hoboken, NJ, Wiley, 2014. 368 p.
9. Draper N.R., Smith H. *Prikladnoy regressionny analiz* [Applied regression analysis]. Moscow, Vilyams Publ. house, 2007. 912 p.
10. Seber G.A.F., Lee A.J. *Linear Regression Analysis. 2nd ed.* Hoboken NJ, Wiley, 2003. 582 p.
11. Hoffmann Jh.P., Shafer K. *Linear Regression Analysis. Assumptions and Applications*. Washington, NASW Press, 2015. 240 p.
12. Bingham N.H., Fry J.M. *Regression*. London, Springer, 2010. 284 p.
13. Montgomery D.C., Peck E.A., Vining G.G. *Introduction to Linear Regression Analysis. Fifth ed.* New York, Wiley, 2013. 836 p.
14. Ivakhnenko A.G. Heuristic self-organization on problems of engineering cybernetics. *Automatic*, 1970, vol. 6 (3), pp. 207–219.
15. Godfrey Onwubolu. *GMDH and Implementation in C-Methodology*. London, Imperial College Press, 2015. 303 p.
16. Atashrouz S., Pazuki G., Alimoradi Y. Estimation of the viscosity of nine nanofluids using a hybrid GMDH-type neural network system. *Fluid Phase Equilibria*, 2014, vol. 372, pp. 43–48.
17. Ahmadi M.H., Ahmadi M.A., Rosen M.A. Using GMDH Neural Networks to Model the Power and Torque of a Stirling Engine. *Sustainability*, 2015, vol. 7 (2), pp. 2243–2255.
18. Mueller J-A., Lemke F. *Self-Organizing Data Mining*. Hamburg, Libri, 2000. 225 p.
19. Ivakhnenko A.G., Ivakhnenko G.A., Mueller J.A. Self-Organization of Neuronets with Active Neurons. *Pattern Recogn. and Image Analysis*, 1994, vol. 4, no. 4, pp. 177–188.

20. Ivakhnenko A.G., Savchenko E.A. Issledovanie effektivnosti metoda doopredeleniya vybora modeli v zadachakh modelirovaniem s primeneniem MGUA [Studying the effectiveness of the method of regularization of a model selection in modeling problems using GMDH]. *Problems of control and Informatics*, 2008, no. 2, pp. 65–76.
21. Sarychev A.P. Sistemny kriteriy regulyarnosti v metode gruppovogo ucheta argumentov [Regularity system criterion in the data group accounting method]. *Problems of control and Informatics*, 2006, no. 6, pp. 25–37.
22. Stepashko V.S., Bulgakova A.S. Obobshchenny iteratsionny algoritm metoda gruppovogo ucheta argumentov [Generalized iterative algorithm of the data group accounting method]. *Control systems and machines*, 2013, no. 2, pp. 9–26.
23. Goharriz M., Marandi S.M. An Optimized Neuro-Fuzzy GMDH System Based on Gravitational Search Algorithm for Evaluation of Lateral Ground. *Int. J. Optim. Civil Eng.*, 2016, vol. 6 (3), pp. 385–403.
24. Ermakov S.M. *Metod Monte-Karlo v vychislitelnoy matematike* [Monte Carlo Method in computational mathematics]. St-Petersburg, Nevsky dialekt Publ., 2009. 192 p.
25. Balasanyan S.Sh., Gevorgyan H.M. The Construction of Statistical Model of a Virtual Object in the LabVIEW Environment. *Proc. of 11th International Conference on Remote Engineering and Virtual Instrumentation REV2014. Porto*, 26–28 Feb., 2014. pp. 291–293.
26. Kuznetsova O.S. *Teoriya veroyatnosti i matematicheskaya statistika* [Probability theory and mathematical statistics]. Moscow, Okey-kniga Publ., 2013. 191 p.
27. Balasanyan S.Sh. Stratified formalization method of complex multistate systems. *Bulletin of the Tomsk Polytechnic University. Geo Assets Engineering*, 2016, vol. 327, no. 1, pp. 6–18. In Rus.

Received: 15 January 2016.

Information about the authors

Seyran Sh. Balasanyan, Dr. Sc., associate professor, Kapan branch of National Polytechnic University of Armenia.

Hermine M. Gevorgyan, professor, Kapan branch of National Polytechnic University of Armenia.