

ПІДХІД ДО ІДЕНТИФІКАЦІЇ ПАРАМЕТРІВ МОДЕЛЕЙ ТЕХНОЛОГІЧНИХ ПРОЦЕСІВ ГІРНИЧОГО ВИРОБНИЦТВА

Анотація. Розглянуто питання побудови математичної моделі за допомогою регресійного аналізу. Досліджено можливість застосування імітаційного моделювання в умовах гірничого виробництва. Встановлено залежність точності апроксимації лінійної регресійної моделі від числа незалежних предикторних змінних та від кількості експериментів навчальної вибірки.

Ключові слова: технологічний процес, регресійний аналіз, метод найменших квадратів, ідентифікація, математична модель

Постановка проблеми

Сьогодні інформаційні технології (ІТ) знаходять застосування майже у всіх галузях промисловості. Вони стали невід'ємною ланкою автоматизації технологічних процесів, розвитку систем підтримки прийняття рішень та оптимізації складних технічних систем. Прогрес сучасних ІТ сприяє не тільки підвищенню продуктивності праці, надійності обладнання та його техніко-економічних показників, але й відкриває нові можливості для наукового дослідження. Зараз з'являються та оновлюються методи аналізу технологічних об'єктів, застосування яких раніше стримувалося малопродуктивними обчислювальними машинами. Однією із таких задач є побудова математичної моделі багатofакторного технологічного процесу.

Як показують раніше проведені дослідження [1], існують суттєві передумови оптимізації параметрів технологічних процесів розробки залізної руди відкритим способом. Проте ефективне керування підготовкою гірської маси в кар'єрі гірничо-збагачувального комбінату (ГЗК) можливе лише за умови синтезу математичної моделі з найбільш актуальними параметрами. Необхідність оптимізації в режимі часу близькому до реального вимагає врахування якомога більшої кількості технологічних показників. Так, наприклад, розглядаючи пе-

реділи кар'єру як об'єкти керування, можна виділити близько 15-20 вхідних змінних, кожна з яких має досить велику дисперсію. Відношення середньоквадратичного відхилення до математичного очікування для деяких з цих параметрів може сягати 20-40%. Крім того, наявність стохастичних збурень, значні похибки вимірювання дуже ускладнюють процедуру ідентифікації. Тому створення багатофакторної математичної моделі кар'єру у межах буро-вибухових робіт є актуальним завданням, оскільки допоможе знайти метод мінімізації сумарних питомих витрат на розробку 1 т руди.

Аналіз публікацій по темі дослідження

Питанням побудови регресійних математичних моделей присвячено багато робіт [2-4]. Значний внесок у розвиток теорії обробки статистичних даних внесли вчені: Айвазян С.А., Гайдишев І.П., Івахненко О.Г., Днейпер Н., Джонсон Л., Сміт Г. та ін. Достатньо глибокі дослідження у цій сфері слугують гарним підґрунтям для розробки ефективних методів машинного навчання. Останні особливо корисні при застосуванні інтелектуальних інформаційних технологій у гірничій галузі промисловості. Проведений аналіз показав, що існуючі техніко-економічні моделі [4, 5] сьогодні потребують вдосконалення та актуалізації, оскільки змінилася не тільки вартість енергоресурсів, ринкові умови та деякі переділи гірничого виробництва, але й підвищилися вимоги до якості продукції підприємства. Крім того, використання сучасних засобів обчислювальної техніки тепер дозволяє урахувати додаткові фактори та уточнити математичну модель об'єкта.

Формулювання цілей статті

Виходячи з вищезазначеного, було вирішено побудувати регресійну математичну модель за статистикою технологічної бази даних підприємства. Необхідно визначити залежність точності такої моделі від обсягу вибірки, ідентифікувати тип регресійного рівняння та його порядок шляхом імітаційного моделювання. Додатково необхідно дослідити вплив статистичного шуму на результати імітаційного моделювання.

Викладення матеріалу та результати

Особливістю обраного авторами технологічного об'єкта є його висока складність, багатофакторність та значна часова інерційність [1]. Зважаючи на труднощі проведення експериментів у кар'єрі ГЗК

та надзвичайно високу їх вартість, аналіз було здійснено за допомогою імітаційного моделювання, послідовність операцій якого представлено на рис.1. На рисунку прийнято такі умовні позначення: $i=1..N$ – номер експерименту; N – загальна кількість експериментів у навчальній вибірці; f_i – міцність породи за шкалою М.М. Протод'яконова; γ_i – щільність породи, т/м³; z_i – тріщинуватість гірського масиву, 1/м; a_i – ступінь обводненості свердловин; W_i – лінія найменшого опору за подошвою уступу, м; p_i, l_{3i} – довжина перебуру та забивки відповідно, м; q_i – питомі витрати вибухівки, кг/м³; D_{Cpi} – середній розмір шматка породи, мм; η_{Hi} – вміст фракції негабариту у висадженій гірській масі, %; $\bar{k}_i = \{k_{j|i}\}, j=1..4$ – вектор, що характеризує коефіцієнт використання обладнання в часі (буровий станок, екскаватор, транспорт та дробарку); $S_{\Sigma i}$ – сумарні питомі витрати, грн./т.

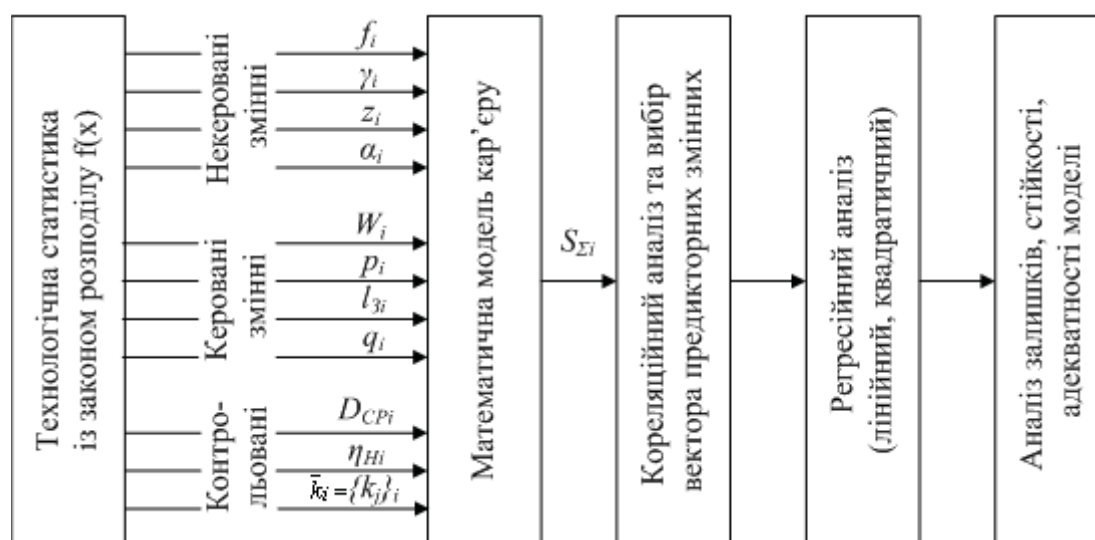


Рисунок 1 - Схема імітаційного моделювання

Слід зазначити, що крім традиційного лінійного та нелінійного регресійного аналізу існує чимало підходів до математичного опису об'єктів керування, таких як метод групового урахування аргументів (МГУА), нейронні мережі, методи на базі нечіткої логіки, інтерполяція на основі сплайнів тощо. Проте, як показав аналіз, застосування останніх із зазначених методів пов'язане з деякими труднощами. Так, наприклад, інтерполяція поліномом чи сплайнами при наявності випадкової складової у статистиці призводить до значних амплітуд коливань інтерпольованої функції та «викидів». У цьому випадку на графіку залежності простежується значна кількість локальних екст-

ремумів, що ускладнює процес пошуку оптимуму. МГУА, як і нейромережеві підходи передбачають велику кількість обчислень, а у випадку спрощення, перетворюються у звичайне регресійне рівняння.

У якості лінійного регресійного рівняння було взято наступний поліном

$$y(X) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m, \quad (1)$$

де x_1, x_2, \dots, x_m – предикторні змінні; m – кількість змінних.

Аналіз параметрів технологічного процесу показав, що для побудови якнайповнішої регресійної моделі потрібно врахувати близько 14 параметрів. Згідно з методом найменших квадратів вектор коефіцієнтів V можна знайти, розв'язавши матричне рівняння

$$M_Z V = V_Y \Rightarrow V = M_Z^{-1} V_Y, \quad (2)$$

$$\text{де } V = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \dots \\ \beta_m \end{bmatrix}, M_Z = \begin{bmatrix} \sum_{i=1}^N 1 & \sum_{i=1}^N x_{1i} & \sum_{i=1}^N x_{2i} & \dots & \sum_{i=1}^N x_{mi} \\ \sum_{i=1}^N x_{1i} & \sum_{i=1}^N (x_{1i})^2 & \sum_{i=1}^N x_{1i} x_{2i} & \dots & \sum_{i=1}^N x_{1i} x_{mi} \\ \sum_{i=1}^N x_{2i} & \sum_{i=1}^N x_{1i} x_{2i} & \sum_{i=1}^N (x_{2i})^2 & \dots & \sum_{i=1}^N x_{2i} x_{mi} \\ \dots & \dots & \dots & \dots & \dots \\ \sum_{i=1}^N x_{mi} & \sum_{i=1}^N x_{1i} x_{mi} & \sum_{i=1}^N x_{2i} x_{mi} & \dots & \sum_{i=1}^N (x_{mi})^2 \end{bmatrix}, V_Y = \begin{bmatrix} \sum_{i=1}^N y_i \\ \sum_{i=1}^N x_{1i} y_i \\ \sum_{i=1}^N x_{2i} y_i \\ \dots \\ \sum_{i=1}^N x_{mi} y_i \end{bmatrix}.$$

Слід зауважити, що не для всякої статистики можна розв'язати рівняння (2). Причиною цього може бути погано обумовлена матриця M_Z та близькість її визначника до нуля. З точки зору даних це пояснюється мультиколінеарністю – тісним кореляційним взаємозв'язком між предикторними змінними, які разом впливають на апроксимовану функцію. Для відбору найбільш значущих змінних розраховується матриця кореляцій за формулою

$$R_{X,Y} = \frac{M[XY] - M[X]M[Y]}{\sqrt{(M[X^2] - M^2[X])(M[Y^2] - M^2[Y])}}, \quad (5)$$

де $M[X], M[Y]$ – математичні очікування змінних X та Y .

Якщо для деяких предикторних змінних $R_{XiXj} \rightarrow 1$, то потрібно один із них виключити зі статистики. Після встановлення кореляційного зв'язку між кожною із незалежних змінних зі значенням цільової функції Y визначається порядок, згідно з яким регресійні змінні потрібно включати в модель (табл. 1). До обраного предиктора із найбільшим за модулем значенням коефіцієнта кореляції послідов-

но додаються залишені предиктори і розраховується коефіцієнт детермінації для кожної моделі. Процес приєднання предикторів закінчується, коли значення коефіцієнта детермінації стає меншим, ніж одержане на попередньому кроці.

Таблиця 1

Коефіцієнти кореляції та порядок включення змінних у модель

X_i	x_1	x_2	x_3	x_4	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
$R_{x_i,y}$	0,27	0,05	-0,53	0,19	-0,18	-0,06	-0,11	0,27	0,56	0,52	0,33	-0,21	-0,53	-0,17
№ з/п	6	14	2	9	10	13	12	7	1	4	5	8	3	11

Ступінь адекватності регресійної моделі статистичним даним перевіряється за допомогою коефіцієнта детермінації R^2 , що показує, яка частка дисперсії апроксимованої функції пояснюється впливом предикторних змінних, та середньоквадратичної похибки $\epsilon_{СКВ}$, яка відображає точність апроксимації. Ці показники визначаються згідно з формулами відповідно

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - y(x_{1i}, x_{2i}, \dots, x_{mi}))^2}{\sum_{i=1}^N \left(y_i - \frac{1}{N} \sum_{i=1}^N y_i \right)^2}, \epsilon_{СКВ} = \sqrt{\frac{\sum_{i=1}^N (y_i - y(x_{1i}, x_{2i}, \dots, x_{mi}))^2}{N}}. \quad (6)$$

Залежність коефіцієнта детермінації від кількості врахованих в моделі предикторних змінних представлена на рис. 2. На графіку зображено криві при різній кількості експериментів. Із характеру залежностей видно, що модель стає тим точнішою, чим більшу кількість предикторів до неї включити.

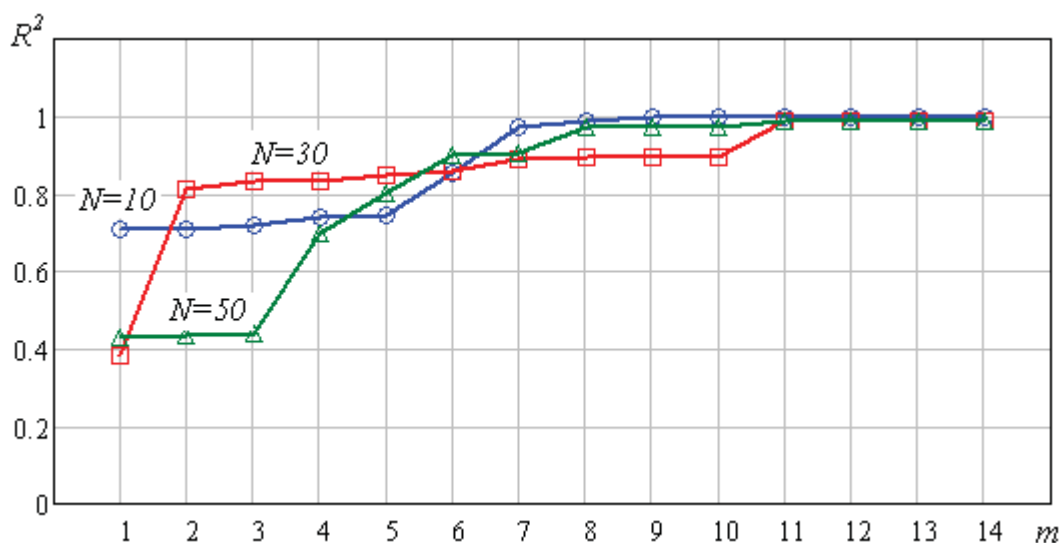


Рисунок 2 - Залежність точності моделі від кількості врахованих предикторів

Однією із головних задач, які постають при імітаційному моделюванні, є визначення числа реалізацій N , що необхідно провести для забезпечення потрібної статистичної точності результатів. Для цього можна скористатися відомою формулою [7]

$$N = \frac{t_{\alpha}^2 \sigma_E^2}{\varepsilon^2}, \quad (7)$$

де α – довірна ймовірність; t_{α} – квантиль нормального закону, який відповідає заданому значенню α і визначається за таблицями нормального розподілу; σ_E – дисперсія показника ефективності експерименту. У якості останнього можна застосувати критерій R^2 , проте наперед визначити його дисперсію досить складно, тому було прийняте рішення встановити число N емпіричним способом.

Статистична вибірка розбивається на дві частини: перша є навчальною і містить N_H записів, а друга – тестова складається із N_T записів. Аналіз якості результатів проводиться за допомогою двох коефіцієнтів детермінації: R^2_H – відображає точність апроксимації навчальної вибірки даних; R^2_T – тестової вибірки, тобто точність прогнозування математичною моделлю. Результати роботи моделі при об'ємі навчальної вибірки 20 записів представлено на рис. 3. Слід звернути увагу, що лінійна модель дає гарні результати на навчальних даних ($R^2_H = 0,996$), проте екстраполяцію виконує із значною похибкою ($R^2_T = 0,793$).

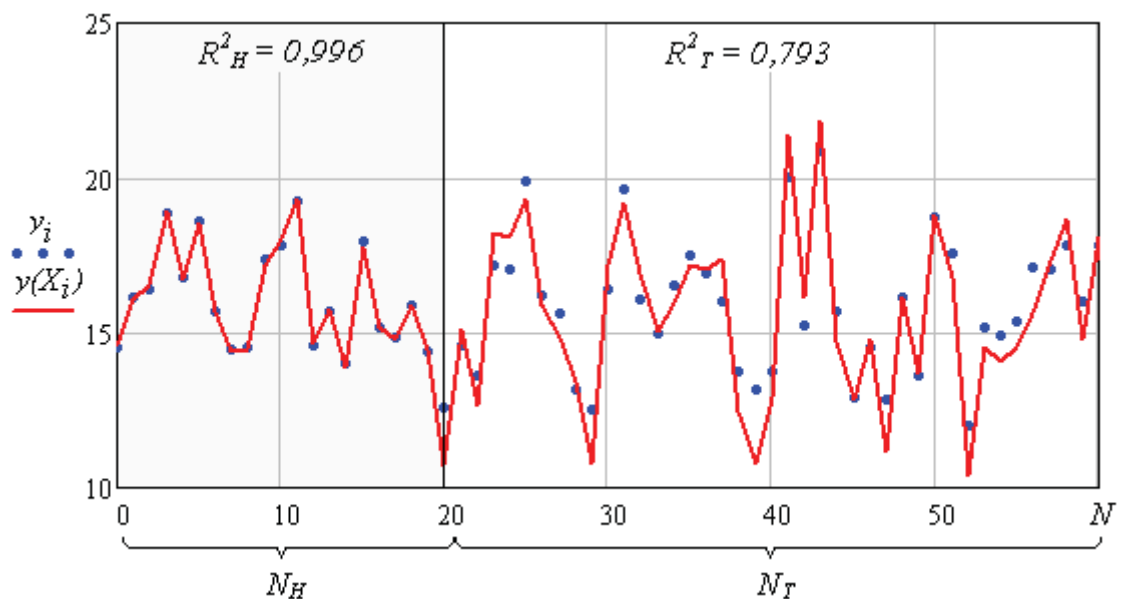


Рисунок 3 - Результати навчання лінійної регресійної моделі

Для встановлення необхідної кількості експериментів було проаналізовано декілька реалізацій статистики. Для кожного значення N_H визначалась точність прогнозування. Результати моделювання зображені на рис. 4.

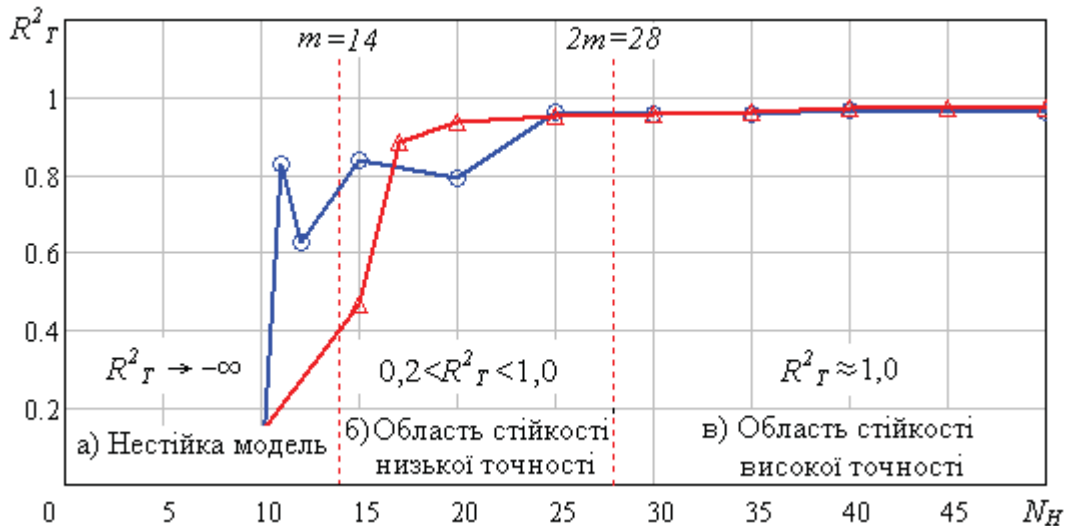


Рисунок 4 - Залежність точності прогнозування лінійної моделі від об'єму вибірки навчання

Графік умовно розбито на три частини: при об'ємі навчальної вибірки, меншої за кількість предикторних змінних, лінійна регресійна модель нестійка; при об'ємі більшому у два рази – коефіцієнт детермінації близький до одиниці; в середній області – математична модель стійка, проте точність коливається у значних межах.

Висновки та перспективи подальших досліджень

Таким чином, у межах даної роботи авторами було запропоновано підхід до ідентифікації математичної моделі кар'єру за допомогою лінійного регресійного аналізу. Особливості технологічного процесу, а також одержані результати доводять можливість та доцільність застосування імітаційного підходу. Встановлено, що збільшення кількості незалежних предикторних змінних у лінійній моделі підвищує її точність і вимагає збільшити об'єм навчальної вибірки до значення $N_H \approx 2m$.

У зв'язку із екстремальними характеристиками даного технологічного процесу подальші дослідження будуть спрямовані на застосування квадратичної регресії, яка дозволить проводити ідентифікацію та оптимізацію параметрів технологічних об'єктів у режимі часу близькому до реального.

ЛИТЕРАТУРА

1. Влияние дробления пород на эффективность технологических процессов открытой разработки / [Друкованый М.Ф., Тартаковский Б.Н., Вишняков В.С., Ефремов Э.И.]. – К.: Наукова думка, 1974. – 269 с.
2. Айвазян С.А. Прикладная статистика: основы моделирования и первичная обработка данных. Справочное изд. / Айвазян С.А., Енюков И.С., Мешалкин Л.Д. – М.: Финансы и статистика, 1983. – 471 с.
3. Гайдышев И.П. Анализ и обработка данных: специальный справочник / И.П. Гайдышев. – СПб.: Питер, 2001. – 752 с.
4. Дрейпер Н. Прикладной регрессионный анализ / Н. Дрейпер, Г. Смит. – М.: Вильямс, 2007. – 912 с.
5. Тангаев И.А. Энергоемкость процессов добычи и переработки полезных ископаемых / И.А. Тангаев. – М.: Недра, 1986. – 231 с.
6. Денисов А.А. Теория больших систем управления: учебное пособие / А.А. Денисов, Д.Н. Колесников. – Л.: Энергоатомиздат, 1982. – 288 с.