

Л. С. НИКОЛАЕВА, И. В. ТИМОХИНА

**ОПРЕДЕЛЕНИЕ ХИМИЧЕСКОГО СОСТАВА КАРБОНАТОВ  
ИЗ ГРУППЫ КАЛЬЦИТА ПО КОМПЛЕКСУ ИХ ФИЗИЧЕСКИХ  
СВОЙСТВ С ПОМОЩЬЮ МЕТОДОВ МАТЕМАТИЧЕСКОЙ  
СТАТИСТИКИ**

(Представлено академиком Д. С. Коржинским 3 III 1971)

При определении химического состава минералов по комплексу их физических свойств все более широкое применение получают методы математической статистики и, в частности, классический регрессионный анализ, реализующий для построения оценок коэффициентов регрессии метод наименьших квадратов (м.н.к.) (<sup>1-3</sup>, <sup>5</sup>, <sup>6</sup>). Одной из основных посылок классического регрессионного анализа является предположение, что независимые переменные  $x$  определяются с пренебрежимо малой ошибкой по сравнению с ошибкой определения зависимой переменной  $y$  (<sup>7</sup>). Если же ошибки определения независимых переменных сравнимы с ошибкой определения зависимой переменной, то при статистической обработке экспериментального материала пренебрежение ошибками определения  $x$  может привести к занижению значений дисперсий оценок коэффициентов регрессии и дисперсий оценок поверхности отклика (дисперсии предсказанного значения изучаемой функции) в заданной точке, а также к фиктивной неадекватности исследуемой модели. Это объясняется тем, что в данном случае среднее значение  $E(y, x_i)$  и дисперсия результатов наблюдения  $\sigma^2(y, x_i)$  при заданном  $x_i$ , как это показано в (<sup>8</sup>, <sup>9</sup>), равны соответственно

$$E(y, x_i) \simeq \eta(\theta, x_i) + \alpha_i, \quad \alpha_i = \sum_{j=1}^k \frac{\partial^2 \eta(\theta, x)}{\partial x_j \partial x_j} \Big|_{x=x_i} \sigma^2 x_{ij}, \quad (1)$$

$$\sigma^2(y, x_i) \simeq \sigma^2 y_i + \beta_i^2, \quad \beta_i^2 = \sum_{j=1}^k \left[ \frac{\partial \eta(\theta, x)}{\partial x_j} \right]^2 \Big|_{x=x_i} \sigma^2 x_{ij}, \quad (2)$$

или для случая линейной зависимости поверхности отклика от независи-

$$\text{мых переменных} \quad \eta(\theta, x_i) = \theta_0 + \sum_{j=1}^k \theta_j x_{ij} \\ E(y, x_i) \simeq \eta(\theta, x_i), \quad (3)$$

$$\sigma^2(y, x_i) \simeq \sigma^2 y_i + \beta_i^2, \quad (4)$$

$$\beta_i^2 = \sum_{j=1}^k \theta_j^2 \sigma x_{ij}, \quad (5)$$

$\theta$  — коэффициент регрессии,  $\sigma^2 y_i$  — дисперсия измерения величины  $y_i$  при фиксированных значениях  $x_{ij}$ ,  $\sigma^2 x_{ij}$  — дисперсия определения  $j$ -й координаты точки  $x_{ij}$   $k$ -мерного факторного пространства.

Из практики известно, что при определении химического состава минералов обычно допускаются довольно значительные ошибки; вопрос о соотношении ошибок определения компонентов состава и физических характеристик минералов пока никем не обсуждался. Представлялось интерес-

Таблица 1

Модель	Компоненты	$\delta$	$\delta_\epsilon$	$\hat{\theta}_a$	$\hat{\theta}_c$	$\hat{\theta}_d$	$\hat{\theta}_{n_e}$	$\hat{\theta}_{n_0}$	$\hat{\theta}(n_0 - n_e)$	$\sigma_{\max} \hat{y}_i$ в форм. ед.
I	CaCO <sub>3</sub>	$\theta$	-0,2796	3,4999	-	-1,8587	14,4797	-	-	$\pm 0,040$
	MgCO <sub>3</sub>	$\sigma$	0,0576	0,5337	0,8723	0,8723	6,8809	-	-	$\pm 0,049$
	FeCO <sub>3</sub>	$\sigma$	0,0558	-3,3713	0,4210	0,4210	-7,6654	-	-	$\pm 0,093$
II	CaCO <sub>3</sub>	$\theta$	0,0284	0,2812	0,4416	0,4416	3,4911	-	-	$\pm 0,011$
	MgCO <sub>3</sub>	$\sigma$	-0,7201	2,7038	-4,6977	-4,6977	46,9084	-	-	$\pm 0,015$
	FeCO <sub>3</sub>	$\sigma$	0,1756	1,6130	2,6836	2,6836	21,1816	-	-	$\pm 0,033$
III	CaCO <sub>3</sub>	$\theta$	-0,0736	-3,2599	1,0675	1,0675	-	-	-	$\pm 0,011$
	MgCO <sub>3</sub>	$\sigma$	0,0373	0,9337	0,4287	0,4287	-	-	-	$\pm 0,015$
	FeCO <sub>3</sub>	$\sigma$	0,0405	-1,4366	-0,3445	-0,3445	-	-	-	$\pm 0,015$
	CaCO <sub>3</sub>	$\theta$	0,0344	0,7750	0,4083	0,4083	-	-	-	$\pm 0,033$
	MgCO <sub>3</sub>	$\sigma$	-0,1268	-17,8905	3,3064	3,3064	-	-	-	$\pm 0,011$
	FeCO <sub>3</sub>	$\sigma$	0,1186	3,0709	0,4281	0,4281	-	-	-	$\pm 0,015$
	CaCO <sub>3</sub>	$\theta$	-	-4,7344	1,2513	1,2513	-	-	-	$\pm 0,015$
	MgCO <sub>3</sub>	$\sigma$	-	0,3157	0,0364	0,0364	-	-	-	$\pm 0,015$
	FeCO <sub>3</sub>	$\sigma$	-	-1,2660	-0,3970	-0,3970	-	-	-	$\pm 0,037$
		$\sigma$	26,6721	4,5508	0,3004	0,3004	-	-	-	
		$\sigma$	1,2704	0,2243	0,0458	0,0458	-	-	-	

Примечание. Содержание MnCO<sub>3</sub> определяется по формуле  $y_4 = 1,000 - (y_1 + y_2 + y_3)$ .

ным рассмотреть влияние ошибок определения физических характеристик на точность предсказания компонентов химического состава на довольно простом примере карбонатов из группы кальцита, а также получить интерполяционные уравнения для возможно более точного предсказания химического состава карбонатов по комплексу их физических свойств.

Особенностью карбонатов из группы кальцита является склонность к образованию изоморфных соединений; подавляющая часть образцов карбонатов представляет изоморфные смеси карбонатов Ca, Mg, Fe и Mn. Математические модели, рассматриваемые в настоящей работе, составлены на основе выборки из 11 карбонатов: трех кальцитов, двух магнезитов, трех сидеритов и трех родохрозитов, у которых измерены диэлектрическая проницаемость  $\epsilon$ , параметры элементарной ячейки  $a$  и  $c$ , удельный вес  $d$  и показатель преломления  $n_0$  и  $n_e$ . Точность измерения  $\epsilon \pm 0,03$ ;  $a \pm 0,001$  Å;  $c \pm \pm 0,01$  Å;  $d \pm 0,01$  г/см<sup>3</sup>,  $n_0 \pm 0,002$ ;  $n_e \pm 0,001$ . Точность определения химического состава зависит от содержания элементов в карбонате и изменяется в диапазонах для CaCO<sub>3</sub> от 0,0002 до 0,0136; для MgCO<sub>3</sub> от 0,0004 до 0,0167; для FeCO<sub>3</sub> от 0,0004 до 0,0067 и для MnCO<sub>3</sub> от 0,0002 до 0,0265 молекулярных количеств.

Поскольку каждый образец карбоната охарактеризован шестью физическими величинами, представлялась возможность рассмотреть различные сочетания физических

Таблица 2

№№ образцов	CaCO <sub>3</sub>			MgCO <sub>3</sub>			FeCO <sub>3</sub>		
	а	б	в	а	б	в	а	б	в
1000	0,997	0,946	1,036	0,002	0,002	0,001	0,001	0,001	0,012
11423	0,975	0,855	0,942	0	0	0	0,016	0,077	-0,054
1034	0,958	0,860	0,945	0,033	0,038	0,036	0,007	0,051	-0,058
59510	0,003	0,009	0,028	0,994	0,975	0,978	0,003	0,002	0,051
58963	0,007	0,024	0,013	0,981	0,998	0,999	0,010	0,037	0,031
0391	0,008	0,003	0,001	0,022	0,025	0,032	0,934	0,461	0,894
7999	0,017	0,007	0,001	0,140	0,135	0,143	0,804	0,290	0,758
0393	0,008	0,016	0,017	0,222	0,200	0,206	0,658	0,370	0,705
48908	0,003	0,014	0,033	0,004	0,006	0,001	0,004	0,027	0,010
0490	0,012	0,011	0,031	0,007	0,004	0,001	0,001	0	0
65278	0,013	0,038	0,018	0,049	0,066	0,059	0,031	0,055	-0,020
$\sigma_{\max} \hat{y}_i$		±0,008	±0,040		±0,014	±0,019		±0,004	±0,093
$\hat{\chi}^2$		1777,0	4,88		47,79	2,96		16194,79	3,95

Примечание. Способ получения данных: а — химический анализ, б — м. н. к., в — итерационный метод.

свойств (различные многомерные линейные модели  $\hat{y}_i = \hat{\eta}(\theta, x_1, \dots, x_n)$ , значок  $\wedge$  употребляется для обозначения рассчитанных оценок рассматриваемых случайных величин) и выбрать из них оптимальные, позволяющие получать наиболее точные оценки компонентов состава. К каждой модели применены м.н.к. и итерационный метод построения оценок коэффициентов регрессии (8, 9). Адекватность полученных уравнений проверялась путем сравнения значений, рассчитанных по формуле  $\hat{\chi}_i^2 = \sum_{i=1}^n [y_i - \eta(\hat{\theta}, x_i)]^2 w_i$ ,  $w_i = 1 / (\sigma^2 y_i + \beta_i^2)$ , с табличным  $\chi_{(7)}^2 = 14,067$  для 95% уровня значимости. Необходимо отметить, что для всех рассмотренных моделей только итерационный метод позволил получить адекватные решения. Обработка этого же самого материала с помощью обычного м.н.к. привела к получению неадекватных моделей вследствие того, что величины  $\hat{\chi}_i^2 = \sum_{i=1}^n [y_i - \eta(\hat{\theta}, x_i)]^2 w_i$ ,  $w_i = 1 / (\sigma^2 y_i)$ , оказались завышенными по сравнению с их реальными значениями и поэтому попали в область маловероятных значений для суммы взвешенных квадратичных отклонений, имеющих  $\chi^2$ -распределение. Оценки коэффициентов регрессии  $\hat{\theta}_i$ , полученные для всех рассмотренных моделей с помощью итерационного метода, ошибки этих оценок  $\sigma_{\hat{\theta}_i}$ , максимальные ошибки предсказания компонентов состава  $\sigma_{\max} \hat{y}_i$  приведены в табл. 1.

В качестве исходной выбрана модель I, содержащая четыре физические характеристики:  $\hat{y}_i = \eta(\hat{\theta}, \varepsilon, a, d, n_e)$ . В табл. 2 для примера проведено сравнение результатов вычисления компонентного состава карбонатов, полученных без учета (по м.н.к.) и с учетом ошибок определения физических величин (с помощью итерационного метода). Из анализа табл. 2 видно, что каждая величина, рассчитанная итерационным методом, согласуется с экспериментальным значением (с точностью до  $\sigma_{\max} \hat{y}_i$ ), а их оценки по м.н.к. получены с ошибками, заниженными на 1—2 порядка по сравнению с реальными. Влияние ошибок определения физических величин на точность предсказания химического состава карбонатов может быть рассмотрено на примере модели I. Подсчет поправок к дисперсиям результатов наблюдения  $\sigma^2(y, x_i)$  по формуле (5) с привлечением данных табл. 1 и ошибок определения физических величин показывает, что наибольший вклад в рассматриваемую величину внесут ошибки определения

Таблица 3

Образец	Модель	CaCO <sub>3</sub>			MgCO <sub>3</sub>			FeCO <sub>3</sub>		
		$\hat{y}_i$	$\Delta y$	$\sigma_{\max} \hat{y}_i$	$\hat{y}_i$	$\Delta y$	$\sigma_{\max} \hat{y}_i$	$\hat{y}_i$	$\Delta y$	$\sigma_{\max} \hat{y}_i$
0392	Хим. ан.	0,013			0,177			0,718		
	I	0,037	-0,024	0,040	0,170	-0,007	0,020	0,798	-0,080	0,093
	II	0,017	-0,004	0,011	0,185	-0,008	0,015	0,724	-0,006	0,033
	III	0,023	-0,010	0,011	0,184	-0,007	0,015	0,761	-0,043	0,037
1047	Хим. ан.	0,904			0,003			0,008		
	I	0,897	0,007	0,040	0,013	-0,010	0,020	0,000	-0,008	0,093
	II	0,894	0,010	0,011	0,011	0,012	0,015	0,000	0,008	0,033
	III	0,892	0,012	0,011	0,011	-0,008	0,015	0,036	-0,028	0,037
	IV	0,886	0,018	0,158	0,046	-0,042	0,210	0,000	0,008	0,300

Примечание.  $\Delta y = y_{\text{анал}} - \hat{y}_i$ .

показателя преломления  $n_e$  и удельного веса  $d$ . Была рассмотрена модель II  $\hat{y}_i = \eta(\hat{v}, \varepsilon, a, c, d)$ , в которой показатель преломления  $n_e$  заменен на параметр элементарной ячейки  $c$ . Замена  $n_e$  на  $c$  позволила получить значительно более точные интерполяционные уравнения. Ошибки предсказания содержания CaCO<sub>3</sub> и FeCO<sub>3</sub> практически в три раза меньше соответствующих ошибок предсказания этих компонентов в первой модели (табл. 1).

При анализе данных табл. 1 также привлекает внимание, что в рассмотренных уравнениях коэффициенты при  $\varepsilon$  имеют значения, сопоставимые в некоторых случаях с ошибками ее определения. Для проверки значимости коэффициентов при  $\varepsilon$  рассмотрена модель III  $\hat{y}_i = \eta(\hat{\theta}, a, c, d)$ . Исключение из уравнений регрессии значений  $\varepsilon$  (см. табл. 1) не оказывает влияния на точность предсказания CaCO<sub>3</sub> и MgCO<sub>3</sub> и только несколько увеличивает ошибки предсказания FeCO<sub>3</sub>.

Из практических соображений представлялось интересным рассмотреть модель IV  $y_i = \eta[\hat{\theta}, n_0, (n_0 - n_e), d]$  и модель V, состоящую только из рентгеновских данных  $\hat{y}_i = \eta(\hat{\theta}, a, c, d_{101})$ . Ошибки предсказания компонентов состава в модели IV во много раз превосходят таковые в моделях II и III; при решении модели V получены вырожденные уравнения.

Из анализа данных, приведенных в табл. 1, следует, что среди рассмотренных моделей наименьшие ошибки предсказанных значений компонентов состава получены для моделей II и III, содержащих по два параметра решетки. Они позволяют предсказывать содержание CaO с ошибкой 0,62, MgO — с ошибкой 0,72 и FeO — с ошибками (2,04 — 2,29) вес. %.

Применимость полученных интерполяционных уравнений для определения химического состава карбонатов проверена на контрольных образцах. Из табл. 3 видно, что интерполяционные уравнения, полученные по моделям II и III, позволяют рассчитывать химический состав карбонатов с ошибками, не превышающими ошибки предсказания с вероятностью 95 %.

Всесоюзный научно-исследовательский институт синтеза минерального сырья  
Александров Владимирской обл.

Поступило  
25 II 1971

#### ЦИТИРОВАННАЯ ЛИТЕРАТУРА

- <sup>1</sup> А. Б. Вистеллус, Зап. Всесоюз. мин. общ., 85, в. 1 (1956). <sup>2</sup> Н. Л. Добрецов, Зап. Всесоюз. мин. общ., 88, в. 6 (1959). <sup>3</sup> М. И. Ильвицкий, З. М. Танатар-Барац, Зап. Всесоюз. мин. общ., 95, в. 2 (1966). <sup>4</sup> Инструкция по внутривлабораторному контролю точности (воспроизводимости) результатов количественных анализов рядовых проб полезных ископаемых, выполняемых в лабораториях Министерства геологии СССР, М., 1968. <sup>5</sup> Е. П. Калинин, Зап. Всесоюз. мин. общ., 96, в. 2 (1967). <sup>6</sup> К. Е. Кецежипская, ДАН, 185, № 4 (1969). <sup>7</sup> В. В. Палимов, Н. А. Чернова, Статистические методы планирования экстремальных экспериментов, «Наука», 1965. <sup>8</sup> Л. С. Николаева, В. В. Федоров, Зав. лаб. № 1 (1970). <sup>9</sup> В. В. Федоров, Анализ экспериментов при наличии ошибок в определении контролируемых переменных, Препринт № 2, МГУ, 1968.