

Ю. ХАЛДНА

**ОБ ИСПОЛЬЗОВАНИИ КОЭФФИЦИЕНТОВ ПАРНОЙ,
ЧАСТНОЙ И МНОЖЕСТВЕННОЙ КОРРЕЛЯЦИИ
ДЛЯ ОЦЕНКИ СТОХАСТИЧЕСКИХ СВЯЗЕЙ
МЕЖДУ КОМПОНЕНТАМИ В ГОРЮЧИХ СЛАНЦАХ**

Неорганическая часть горючих сланцев имеет сложный, многокомпонентный состав. Кроме макрокомпонентов (SiO_2 , CaO , MgO , Fe_2O_3 , Al_2O_3 , Na_2O , K_2O , керогена и разных форм серы) в ней представлена широкая гамма микрокомпонентов (Ti, Mn, V, Ni, Cu, Co, Pb, Zn, Ba и др.). Представляет интерес выяснение стохастических связей между указанными составными частями горючих сланцев, и в первую очередь между микро- и макрокомпонентами. С одной стороны, это позволит глубже понять процессы и условия формирования горючих сланцев. С другой стороны, данные о стохастических связях между компонентами могут быть полезными при разработке новых методов переработки горючих сланцев. Поэтому стохастические связи между микро- и макрокомпонентами в горючих сланцах из разных месторождений изучались многими исследователями [1—4]. Во всех указанных работах в этих целях вычислялись значения коэффициента парной корреляции

$$R_{1,2} = \frac{s_{1,2}}{s_1 \cdot s_2}, \quad (1)$$

где

$$s_{1,2} = \frac{1}{N-1} \sum_{i=1}^{i=N} (x_{i,1} - x_1)(x_{i,2} - x_2), \quad (2)$$

$$s_j = \left[\frac{1}{N-1} \sum_{i=1}^{i=N} (x_{i,j} - x_j)^2 \right]^{1/2} \quad (3)$$

и

$$\bar{x}_j = \frac{1}{N} \sum_{i=1}^{i=N} x_{i,j}. \quad (4)$$

Через $x_{i,1}$ и $x_{i,2}$ обозначены содержания рассматриваемых компонентов (1 и 2) в пробах $1, 2, \dots, i, \dots, N$.

Однако в горючих сланцах имеется, кроме компонентов 1 и 2, и ряд других компонентов (3, 4, ..., P), содержания которых в изученных образцах ($x_{i,3}$, $x_{i,4}$, ..., $x_{i,P}$) также варьируют. Между компонентами 1 и 2, с одной стороны, и 3, 4, ..., P, с другой — могут иметь место стохастические связи; например, изменение содержания компонента 3 может влиять на содержание компонента 1. Если это так, то выводы, сделанные на основе значений коэффициента парной корреляции ($R_{1,2}$) не однозначны, так как при вычислении $R_{1,2}$ не учтено влияние изменений $x_{i,3}$, $x_{i,4}$, ..., $x_{i,P}$ в образцах $1, 2, \dots, N$. Более подробно этот вопрос рассмотрен в монографиях [5—8].

Учитывая это, при изучении стохастических связей между компонентами горючих сланцев следует отдать предпочтение коэффициенту частной корреляции

$$R_{1,2,3,4,\dots,P} = \frac{-c_{1,2}}{\sqrt{c_{1,1} \cdot c_{2,2}}}, \quad (5)$$

где $c_{1,2}$, $c_{1,1}$ и $c_{2,2}$ — элементы матрицы

$$s^{-1} = \begin{vmatrix} c_{1,1} & c_{1,2} & c_{1,3} & \dots & c_{1,P} \\ c_{2,1} & c_{2,2} & c_{2,3} & \dots & c_{2,P} \\ \dots & \dots & \dots & \dots & \dots \\ c_{P,1} & c_{P,2} & c_{P,3} & \dots & c_{P,P} \end{vmatrix}, \quad (6)$$

являющейся обратной по отношению к ковариационной матрице

$$s = \begin{vmatrix} s_{1,1}^2 & s_{1,2} & s_{1,3} & \dots & s_{1,P} \\ s_{2,1} & s_{2,2}^2 & s_{2,3} & \dots & s_{2,P} \\ \dots & \dots & \dots & \dots & \dots \\ s_{P,1} & s_{P,2} & s_{P,3} & \dots & s_{P,P}^2 \end{vmatrix}, \quad (7)$$

составленной из элементов, вычисленных по уравнениям (2) и (3).

В коэффициенте частной корреляции элиминировано влияние варьирования содержаний всех остальных компонентов на стохастическую связь между рассматриваемыми компонентами (1 и 2) [5]. Поэтому выводы, сделанные на основе значений коэффициента частной корреляции более содержательны, чем выводы, сделанные на основе значений коэффициента парной корреляции.

Следует отметить, что $1 \geq R_{1,2} \geq -1$ и $1 \geq R_{1,2,3,4,\dots,P} \geq -1$. Однако между значениями $R_{1,2}$ и $R_{1,2,3,4,\dots,P}$ отсутствуют жесткие соотношения типа $R_{1,2} \geq R_{1,2,3,4,\dots,P}$ (или наоборот). Какой из этих двух коэффициентов корреляции больше или меньше, зависит от $x_{i,3}$, $x_{i,4}$, ..., $x_{i,P}$ [6].

С точки зрения регрессионного анализа коэффициенту частной корреляции можно дать весьма ясное объяснение. В рамках этого подхода $R_{1,2,3,4,\dots,P}$ определяется как коэффициент парной корреляции между остатками [5, 8]

$$e_{i,1} = x_{i,1} - \hat{x}_{i,1} \quad (8)$$

и

$$e_{i,2} = x_{i,2} - \hat{x}_{i,2}, \quad (9)$$

где $\hat{x}_{i,1}$ и $\hat{x}_{i,2}$ — оценки, соответственно, $x_{i,1}$ и $x_{i,2}$, вычисляемые по соответствующим регрессионным уравнениям [8]:

$$\hat{x}_{i,1} = B_0 + \sum_{j=3}^{j=P} B_j x_{i,j}, \quad (10)$$

$$\hat{x}_{i,2} = B'_0 + \sum_{j=3}^{j=P} B'_j x_{i,j}. \quad (11)$$

Наряду с изучением взаимных стохастических связей между любыми парами компонентов в сланцах на основе коэффициентов частной корреляции, в ряде случаев можно ставить вопрос несколько иначе: в какой мере содержание какого-то одного компонента ($x_{i,1}$) может быть описано

линейными зависимостями содержания этого компонента от содержаний всех остальных компонентов ($x_{i,2}, \dots, x_{i,P}$)? Например, можно задать вопрос: в какой мере имеющийся в сланцах ванадий обнаруживает линейные зависимости со всеми макрокомпонентами? Ответ может быть получен путем вычисления коэффициента множественной корреляции [7]

$$R_{1.2,3,\dots,P} = \sqrt{1 - \frac{1}{s_1^2 \cdot c_{1,1}}}, \quad (12)$$

где s_1^2 и $c_{1,1}$ — соответствующие элементы матриц s (7) и s^{-1} (6).

С точки зрения регрессионного анализа, $R_{1.2,3,\dots,P}$ представляет собой коэффициент парной корреляции между $x_{i,1}$ и его линейными оценками по уравнению регрессии [5]:

$$\hat{x}_{i,1} = \sum_{j=2}^{j=P} c_j x_{i,j}. \quad (13)$$

Можно показать, что $R_{1.2,3,\dots,P} \geq |R_{1,j}|$, где $j=2,3,\dots,P$ [6]. Доля изменений рассматриваемого компонента (1), обусловленная линейными связями со всеми остальными компонентами, определяется величиной $R_{1.2,3,\dots,P}^2$ [7]. По ней можно судить о том, насколько приемлема точка зрения, согласно которой концентрация каждого отдельно взятого микрокомпонента в сланце рассматривается как сумма соответствующих вкладов, пропорциональных содержаниям макрокомпонентов.

Наконец следует отметить, что прежде, чем сделать какие-либо выводы по коэффициентам корреляции, необходимо оценить их статистическую значимость. Для этого вычисляется критерий Стьюдента

$$t = \frac{R\sqrt{F}}{\sqrt{1-R^2}}, \quad (14)$$

где R — коэффициент корреляции (парной, или частной, или множественной) и F — число степеней свободы. В случае парной корреляции $F=N-2$ [7], для частной и множественной корреляции $F=N-P$ [7]. Практически можно вычислить при заданной доверительной вероятности (например, при вероятности 0,95) критическое значение коэффициента корреляции (R_{KR}) с учетом F и использованием уравнения (14). Если вычисленный $R \geq R_{KR}$, то значение R следует считать статистически значимым.

На основе вышеприведенных уравнений нами (с участием Р. Юга) составлена программа ЭВМ для вычисления коэффициентов парной, частной и множественной корреляции. Программа составлена для ЭВМ ЕС1052 на языке Фортран IV; интересующиеся указанной программой могут обратиться к автору настоящей статьи.

1. *Калмурзаев К. Е., Дегыненко Л. А.* О корреляционных связях химических элементов в породах черных сланцев. — В кн.: Проблемы осадочной геологии докембрия. VII, кн. 2, М., 1982, с. 62—68.
2. *Armands G.* Geochemical Studies of Uranium, Vanadium and Molybdenum in a Swedish Alum Shale. — In: Stockholms Contributions in Geology. Stockholm, 1972—73.
3. *Петерсель Д., Лоог А.* О минералогии и геохимии оболовых песчаников и диктионемовых сланцев Северной Эстонии. — Уч. зап. Тартуского государственного университета, вып. 561. Труды по геологии IX. Тарту, 1981, с. 30—49.
4. *Палвадре Р., Лоог А., Халдна Ю., Ахелик В., Талкоп Р.* Корреляционные связи между содержаниями компонентов в эстонских аргиллитах. — Горючие сланцы, 1984, 1, № 2.
5. *Краммер Г.* Математические методы статистики. М., 1975, с. 336—341.
6. *Лукомский Я. И.* Теория корреляции и ее применение к анализу производства. М., 1961, с. 193.
7. *Jahn W., Vahle H.* Die Factoranalyse und Ihre Anwendung. Berlin, 1970, с. 39.
8. *Рао С. Р.* Линейные статистические методы и их применение. М., 1968, с. 240.

*Институт химии
Академии наук Эстонской ССР
г. Таллин*

*Представил А. Аарна
Поступила в редакцию
11. 11. 1983*

Ü. HALDNA

**THE USE OF LINEAR, PARTIAL AND MULTIPLE
CORRELATION COEFFICIENTS FOR ESTIMATING
STOCHASTIC RELATIONSHIPS BETWEEN
COMPONENTS IN OIL SHALES**

The use of partial correlation coefficient instead of the linear correlation coefficient is recommended.

This allows to eliminate the influence of changes in other concentrations of variable components of oil shale. A computer programme algorithm for the calculation of the partial correlation coefficient is presented.

*Academy of Sciences of the Estonian SSR,
Institute of Chemistry
Tallinn*