MINERALOGIA POLONICA
Vol. 14, No 1—2 — 1983

*Wojciech NARĘBSKI**

# A PROPOSAL OF SIMPLIFIED STATISTICAL EVALUATION OF SMALL AMOUNT OF GEOCHEMICAL DATA

UKD 550.84.08 małe ilości danych :519.272 uproszczona :519.241.5/.6:518.4

Abstract. The proposed procedure is applicable in geochemical characteristics of natural objects on the ground of 8 to 20 analytical data. It consists in graphical estimation of frequency distribution types and of basic statistical parameters of individual elements using linear and logarythmic probability papers without preliminary grouping the data into classes. The obtained values for normal distributions can be checked by calculation of arithmetic means and estimation of standard deviations from the range values using Tippet's formula. The estimation of parameters of lognormal distributions is carried out by combination of the graphical method with determining variation coefficients from the values of variance by means of Gorlitski's diagram.

## INTRODUCTION

Rapid development of analytical techniques results in considerable increase of the number of data in geochemical studies. Nevertheless, there are some geological materials which, because of objective reasons, cannot be characterized by sufficient amount of analytical data for applying generally used electronic computation of statistical parameters of frequency distribution of individual elements and of their mutual interrelations expressed by means of correlation coefficients.

In the present author's opinion, even if the number of data amounts to approximately 10, the attempts to interpret the element abundances and correlations not only by means of diagrams but also by at least approximate estimations of their frequency distribution types and parameters are very important and useful. In geochemistry and petrology these parameters are usually equally or even petrogenetically more important than average contents of individual rock components. Besides, some authors, neglecting frequency distribution types, often present merely arithmetic means of contents of elements which, in the case of lognormal frequency distributions, are not correct estimates of their abundances.

The aim of this communication is to present a simple method of evaluation of small amount of analytical data describing a given geological population already applied by the present author, with successive modifications, in solving several geochemical and petrological problems (Narębski 1974, Narębski and Wichrowski

* Polish Academy of Sciences, Museum of the Earth. Mailing address: Collegium Geologicum UJ ul. Oleandry 2a, 30-063 Kraków.

1975, 1979). The proposed procedure is based mainly on estimating fundamental statistical parameters by means of graphical and simple calculation methods, resembling in some respects those presented by Lepeltier (1969).

## STATISTICAL PROCEDURE

The frequency distribution types of the elements studied are estimated using enlarged content scale normal and lognormal probability papers. The analytical data should be, as usual, preliminarily ordered according to increasing values and the latters (without grouping them into classes) plotted first on linear-scale probability paper according to percent cummulative frequencies corresponding to successive cell midpoints, depending on the amount of analytical data available. In the case of 10 data these midpoints are successively: 5, 15, 25 etc. If the projection points locate approximately along a straight line — the frequency distribution of a given element is considered to be normal. If, however, they are arranged along a curve distinctly convex upwards (indicating positive assymetry of frequency distribution curve resulting from concentration of data within low-value range), we have to plot them on probability paper with possibly enlarged logarythmic scale. In such cases we usually obtain approximately straight lines indicating lognormal frequency distribution type of the elements studied.

It should be emphasized that in geochemical papers we find both correct and rather erroneous applications of the graphical method in question. Some authors do not take into account the relation between Gaussian curve and the plot on probability paper and sometimes consider partial distributions (eg. caused by high sensitivity threshold of a given analytical method) to be representative for an element, without extrapolating the straight line towards the lower values, undeterminable by the method used (Tkachev, Judovich 1975). Besides, some errors are due to quite artificial preliminary grouping into classes of very small number of data showing distinctly asymmetric frequency distributions (Jaworski 1972, Sachanbiński 1980).

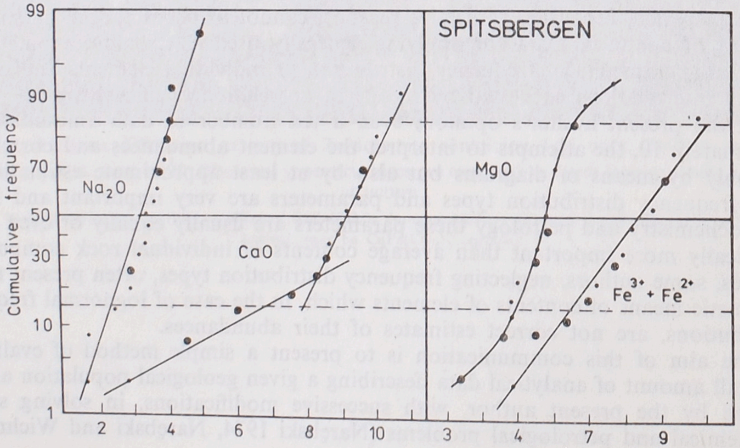


Fig. 1. Linear scale cumulative frequency distributions for $Na_2O$, CaO, MgO and total Fe in Precambrian amphibolites of Spitsbergen (after Narębski, 1974). Positive excess (*S*-shaped line) of MgO frequency distribution and bimodal of CaO are observed



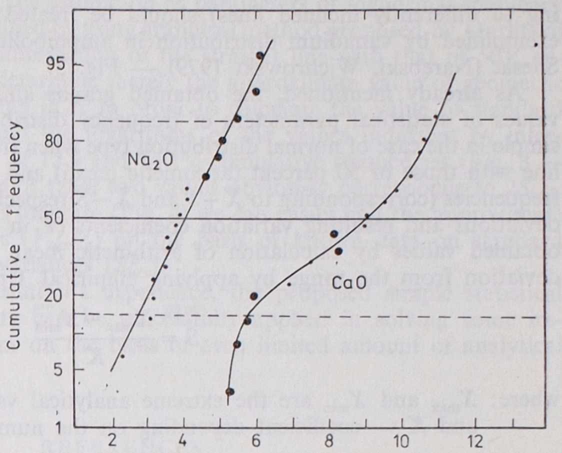


Fig. 2. Linear scale cumulative frequency distributions for $Na_2O$ and CaO in Cambrian pillow lavas of the G. Kaczawskie Mts (after Narębski 1974). Reversed *S*-form of CaO line indicates negative excess of its frequency distribution curve

The graphical method in question, when properly applied, allows not only to estimate the approximate values of statistical parameters of distributions but also to show their eventual deviations from normal or lognormal one. So eg. when the resulting line is slightly S-shaped (MgO graph in Fig. 1) the corresponding frequency distribution curve shows positive excess (is more sharp in shape than theoretical Gaussian one) whereas negative excess (too flat frequency distribution curve) is expressed by reversed S-form of this line on probability paper (CaO graph in Fig. 2).

If no matter of the kind of probability paper used, the resulting line is broken (eg. CaO graph in Fig. 1), the element in question shows polymodal distribution and in statistical interpretation each of the populations distinguished (correspond-

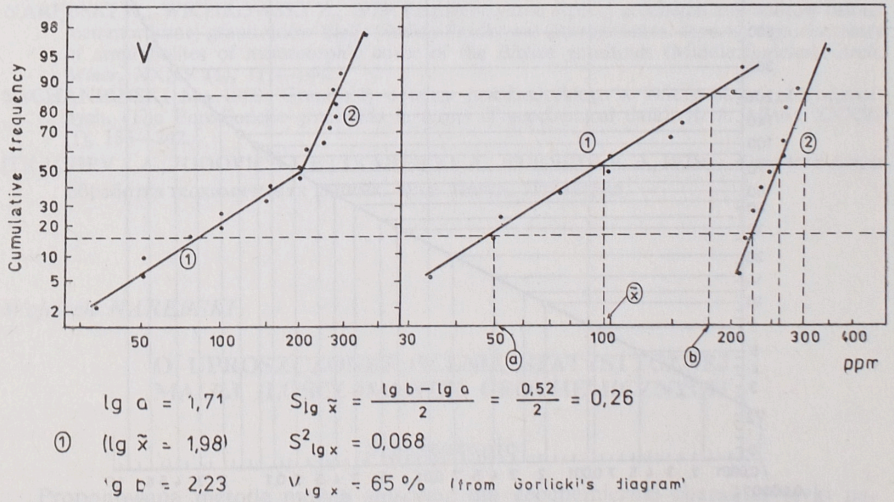


Fig. 3. Log-scale cumulative frequency distribution diagram for *V* in amphibolites of the Bielice Mts showing distinct bimodality. Calculation of statistical parameters for lognormal distribution of the lower-content populations is illustrated

ing to differently inclined lines) should be treated separately. This case can be exemplified by vanadium distribution in amphibolites of the Bielice Mts., Lower Silesia (Narębski, Wichrowski 1979) — Fig. 3.

As already mentioned, the obtained graphs allow to determine approximate values of statistical parameters of frequency distributions. The procedure is very simple in the case of normal distribution type when from intersections of the straight line with those of 50 percent (arithmetic mean) and of 84.1 and 15.9 cummulative frequencies (corresponding to $\overline{X}+S$ and $\overline{X}-S$ respectively) we can estimate standard deviations and resulting variation coefficients ($V$ in %). It is advised to check the obtained values by calculation of arithmetic mean and to estimate the standard deviation from the range by applying empirical Tippet's formula (Kaplan 1970):

$$S=\frac{X_{max}-X_{min}}{K}$$

where: $X_{max}$ and $X_{min}$ are the extreme analytical values
and $K$ — coefficient depending on the number of data (n) (Table 1).

Table 1

Values of $K$ coefficients (after Kaplan 1970)

| n | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | — | — | 1.13 | 1.69 | 2.06 | 2.33 | 2.53 | 2.70 | 2.85 | 2.97 |
| 10 | 3.08 | 3.17 | 3.26 | 3.34 | 3.41 | 3.47 | 3.53 | 3.59 | 3.64 | 3.69 |
| 20 | 3.73 | 3.78 | 3.82 | 3.86 | 3.90 | 3.93 | 3.96 | 4.00 | 4.03 | 4.06 |
| 30 | 4.09 | 4.11 | 4.14 | 4.16 | 4.19 | 4.21 | 4.24 | 4.26 | 4.28 | 4.30 |

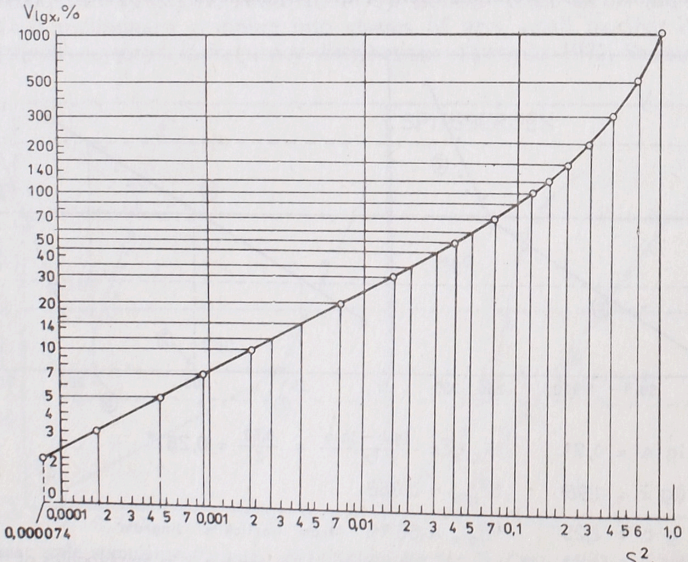


Fig. 4. Gorlitski's diagram for estimation of variation coefficient ($V_{lgx}$) for lognormal distributions on the ground of its variance ($S^2_{lgx}$)



Much more complicated is the estimation of parameters of lognormal frequency distributions but it can be simplified by using graphical method proposed by Gorlitskij (1970). In this case from the intersection of the straight line with that of 50% cumulative frequencies we may determine merely the approximate value of geometric mean and that of corresponding $\overline{\lg X}$. In order to calculate the value of standard deviation ($S_{lgx}$) we have to find the logarithms of the values indicated by intersections of the line with those of 15.9 and 84.1% cumulative frequencies (Fig. 3 — points a and b). The difference of these two values obviously corresponds to $2S_{lgx}$. Finally, by applying Gorlitski's diagram (Fig. 4) we can easily find the approximate values of variation coefficients ($V_{lgx}$%) on the basis of known data on standard deviations and the resulting variances ($S^2_{lgx}$).

According to the present author's experience, the proposed simple statistical evaluation of geochemical data can be successfully applied in solving some important petrogenetical problems on the basis of even limited amount of analytical material.

## REFERENCES


[GORLITSKIJ B. A.] ГОРЛИЦКИЙ Б. А., 1970: Распределеине малых элементов и проблемы металлогении осадочно-вулканогенных формации докембрия Украинского Щита. Наукова Думка Киев.

JAWORSKI A., 1972: Interpretacja statystyczna w zastosowaniu do badań geochemicznych. *Kwart. Geol.* 16 (2), 383—404.

[KAPLAN B. G.] КАПЛАН Б. Г., 1970: Экспресс-расчет основных математико-статистических показателей Изд. Maариф, Баку.

LEPELTIER C., 1969: A simplified statistical treatment of geochemical data by graphical representation. *Econ. Geol.* 64 (5), 538—550.

NARĘBSKI W., 1974: A statistical study of specific geochemical features of some spilitic rock series. In: Spilites and spilitic rocks. Edit. G. C. Amstutz, Springer Verl. pp. 127—159.

NARĘBSKI W., WICHROWSKI Z., 1975: Statistical-geochemical approach to the problem of origin of amphibolites of the Polish part of the West Tatra Mts. Proc. X Congr. CBGA sec. IV Bratislava, pp. 158—175.

NARĘBSKI W., WICHROWSKI Z., 1979: Petrogenetyczne aspekty geochemii amfibolitów osłony metamorficznej granitoidów Bielic (Sudety Środkowe) (Petrogenetical aspects of geochemistry of amphibolites of metamorphic cover of the Bielice granitoids (Middle Sudetes). *Arch. Miner.* XXXV (1), 111—144.

SACHANBIŃSKI M., 1980: Granitoidy obszaru przedsudeckiego w świetle badań geochemicznych. (The Fore-Sudetic granitoids in terms of geochemical data). *Arch. Miner.* XXXVI (1), 135—242.

[TKACHEV J. A., JUDOVICH J. E.] ТКАЧЕВ Ю. А., ЮДОВИЧ Я. Э., 1975: — Статистическая обработка геохимических данных. Изд. Йаука, Ленинград.


*Wojciech NARĘBSKI*

## O UPROSZCZONEJ OCENIE STATYSTYCZNEJ MAŁEJ ILOŚCI DANYCH GEOCHEMICZNYCH

### Streszczenie

Proponowaną metodę można stosować dla geochemicznej charakterystyki naturalnych obiektów na podstawie 8—20 danych analitycznych. Polega ona na graficznym oznaczeniu typów rozkładów poszczególnych pierwiastków i ich pod-

stawowych parametrów statystycznych przy pomocy normalnej i lognormalnej siatki probabilistycznej bez wstępnego grupowania danych w klasy. Otrzymane wartości dla rozkładów normalnych należy sprawdzić obliczeniem średnich arytmetycznych oraz odchyleń standardowych z rozstępu za pomocą wzoru Tippet'a. Parametry rozkładów lognormalnych uzyskuje się wspomnianą metodą graficzną oraz przy pomocy wykresu Gorlickiego, pozwalającego na wyznaczenie współczynnika zmienności na podstawie wariancji.

## OBJAŚNIENIA FIGUR

Fig. 1. Normalne rozkłady liczebności kumulatywnych $Na_2O$, CaO, MgO i całk. Fe w prekambryjskich amfibolitach Zachodniego Spitsbergenu (wg Narębskiego 1974). Widoczny jes dwumodalny charakter rozkładu CaO i dodatnia kurtoza — MgO (s-kształtna forma lini dla tego pierwiastka)

Fig. 2. Normalne rozkłady liczebności kumulatywnych $Na_2O$ i CaO w kambryjskich lawach sferoidalnych G. Kaczawskich (wg Narębskiego 1974). Widoczna jest odwrócona s-kształtna forma linii CaO, wskazująca na spłaszczenie odpowiadającej jej krzywej Gauss'a

Fig. 3. Lognormalny rozkład liczebności kumulatywnych wanadu w amfibolitach Bielic, wykazujący wyraźną dwumodalność. Zilustrowano podział całości na dwie populacje i obliczenie parametrów statystycznych dla jednej z nich

Fig. 4. Wykres Gorlickiego do oznaczania współczynnika zmienności dla rozkładu lognormalnego ($V_{\lg x}$ %) na podstawie wariancji ($S^2_{\lg x}$)

*Войцех НАРЕМБСКИ*

# О УПРОЩЕННОЙ СТАТИСТИЧЕСКОЙ ХАРАКТЕРИСТИКЕ НЕБОЛЬШОГО КОЛИЧЕСТВА ГЕОХИМИЧЕСКИХ ДАННЫХ

## Резюме

Предложенный метод может быть использован для статистически-геохимической характеристики натуральных объектов на основе 8—20 аналитических данных. Он основан на графическом определении типов распределения химических элементов и их основных статистических параметров при помощи нормальной и логнормальной вероятностной бумаги без предварительной группировки данных в классы. Полученные параметры нормальных распределении проверается арифметическим счетом средних и определением стандартных отклонений по размаху при помощи формулы Типпета. Статистические параметры логнормальных распределений получается предложенным графическим методом и при помощи палетки Горлицкого для определения коэфициента вариации логарифмов по величине их дисперсии.

## [ОБЪЯСНЕНИЯ К ФИГУРАМ

Фиг. 1. Нормальные распределения кумулятивных численностей натрия, кальция, магния и железа в докембрийских амфиболитах зап. Спитсбергена (по Нарембскому 1974). Видны двумодальный характер распределения кальция и положительная куртоза распределения магния.

Фиг. 2. Нормальные распределения кумулятивных численностей натрия, и кальция в кембрийских шаровых лавах Качавских Гор (по Нарембскому 1974). Видна негативная куртоза кривой распределения кальция.

Фиг. 3. Логнормальное распределение кумулятивных численностей ванадия в амфиболитах Белиц, показывающие отчетливую бимодальность. Показано разделение обоих совокупностей и определение статистических параметров одной из них.

Фиг. 4. Палетка Горлицкого (1970) для определения коэфициента вариации логарифмов по величине их дисперсии.