Geological Quarterly, 2005, **49** (3): 347–354

Compositional Data Analysis as a tool for interpretation of rock porosity parameters

Małgorzata LABUS

Labus M. (2005) — Compositional Data Analysis as a tool for interpretation of rock porosity parameters. Geol. Quart., **49** (3): 347–354. Warszawa.

This work examines the possibility of implementing the statistical method, Compositional Data Analysis (CDA), to rock porosity measurement data. The rock samples analysed are different types of sandstones from Poland. Porosimetric measurements were carried out using the mercury injection capillary pressure method, together with computer image analysis. In this paper, compositions of data concerning pore distributions are analysed. Rock pores were distributed in 4 classes of pore dimensions: transitive pores, submacropores, real macropores and over-capillary pores. Based on the CDA methods it was revealed that: there is no constant proportion between transitive and macropores in the sandstones analyzed; the variability of the real macropores, over-capillary and transitive pores and the skeletal grains fraction is mainly one-dimensional. The relative variation between transitive pores and real macropores determines the variability in the sandstone stratigraphical groups; however, these groups are not very distinct as regards the pore sizes and skeletal grains proportion. The rock parameters: median pore diameter; threshold pore diameter and total pore area might be used as independent variables in equations that explain reasonable parts of variability of pore sizes and skeletal grains fraction log-ratios.

Małgorzata Labus, Institute of Applied Geology, Silesian Technical University, Akademicka 2, PL-44-100 Gliwice, Poland, e-mail: mlabus@polsl.pl *(received: October 26, 2004; accepted: February 22, 2005).*

Key words: Compositional Data Analysis, porosity, pore sizes, sandstones.

INTRODUCTION

This study examines the use of Compositional Data Analysis implementation for rock porosity measurement data. The rock samples analysed represent different sandstones from Poland (regarding geological region or stratigraphical position). Out of the 50 samples 12 are from Palaeozoic or Mesozoic rocks of the Holy Cross Mts., 8 are from the Sudetan Mountains, mainly Cretaceous sandstones, 22 are from the Late Carboniferous sandstones of Upper Silesia, 7 represent the Cretaceous or Older Tertiary System of the Beskidy region, and 1 sample is a Cretaceous sandstone from the Silesian-Kraków Upland.

Porosimetric measurements were carried out using the mercury injection capillary pressure method, at the Oil and Gas Institute in Kraków. The cumulative intrusion curves obtained of pore volumes versus diameter enabled the determination of percentages in different pore classes. The porosimeter penetrates pores from 0.01 to 100 µm. Hence, the pore classes have been determined as follows (Pazdro, 1983; Hobler, 1977):

Transitive pores 10^{-8}–10^{-7} m
Submacropores 10^{-7}–10^{-6} m } macropores
Real macropores 10^{-6}–10^{-4} m } macropores
Over capillary pores $>10^{-4}$ m } macropores

In a few samples the effective porosity was 0, and those samples have not been taken into consideration. Some other rock parameters were also determined: skeletal density, bulk density, median pore diameter, total pore area = specific surface, hysteresis and threshold pore diameter.

COMPOSITIONAL DATA ANALYSIS

CDA method is a relatively new statistical method introduced by Aitchison in 1986 (Aitchison, 1986). Compositional data consists of vectors x with non-negative parts $x_1,...,x_D$, representing proportions of some whole. Therefore they are subject to the constraint: $x_1+...+x_D = 1$. This condition is referred to as x being a composition of D parts summing to 1 (some other constant can be used, e.g. 100 when using percentages). As a consequence, the components of the above equation cannot be independent since they sum to a constant. To simplify, it is stated that the data are "closed". Such data are very popular in geochemistry, petroleum chemistry, sedimentology, palynology, environmetrics, *etc*., as well as in other fields, i.e. medical statistics, ecology, zoology, sociometrics, economics, *etc*.

The characteristic features of a compositional data set are (Reyment and Savazzi, 1999):

— a compositional data set (samples of a population) may be represented in the form of a matrix;

— each row of the data-matrix corresponds to a single specimen (i.e. rock sample); this is known as a replicate (= a single experimental or observational unit);

— each column of the data matrix represents a single chemical element, a mineral species, in short, a part;

— each entry in the data-matrix is non-negative;

— each row of the data-matrix sums to 1 (proportions), or respectively, 100 (%), (sometimes another row-constant can be found, e.g. owing to some manipulation on the part of the analyst);

— correlation coefficients change if one of the variables is removed from the data-matrix and the rows made to sum to 1 or 100 again. The same effect is also produced if a new component is added to the study.

The last property means that deleting (or adding) one or more variables from (to) the data-set might have a significant numerical effect on the correlations between the remaining variables.

DATA VISUALISATION

One of the most widely used visualization methods in CDA is the ternary diagram. Ternary diagrams used in geology are a useful visual representation of the variability of 3-part compositions. For compositions with more than four parts there is not a satisfactory way of obtaining a visual representation of the variability. In CDA it is possible to construct subcompositions of parts in order to present data on ternary diagrams (Aitchison, 2003*a*, *b*).

Apart from ternary diagrams, very popular in CDA is the biplot (Aitchison and Greenacre, 2002). It was introduced by Gabriel (1971) and is now a widely applied method to visualize rows and columns of many different kinds of data matrices.

Generally speaking, a biplot enables a graphical display of observations and variables on the same chart, in a way that approximates their correlation. In a biplot the observations are usually marked with points, and variables by rays (vectors) emanating from the origin. Both their lengths and directions are important to the interpretation (Fig. 1).

A biplot consists of an origin, *O*, which represents the centre of the compositional data set, a vertex *i* for each of the parts, and a case marker, c_n , for each of the cases. The join of *O* to a vertex *i* is called a ray *Oi*, and the join of two vertices *i* and *j* is termed the link *ij*. Depending on the amount of variability explained by the biplot, the links and rays provide information on the covariance structure of the data set. For compositional data, the biplot is obtained using transformed parts; i.e. for each sample the parts are divided by the geometric mean of the row and logarithms are taken. The main properties of the biplot and its interpretation are shown below, after Aitchison (2003*a*) and Aitchison and Greenacre (2002).

1. Distances between column points in the biplot approximate to the standard deviation of the corresponding log-ratios (logarithms of ratios between relevant pairs of components). A short link between column points indicates that the component ratio is relatively constant in the data, while a large link indicates a large relative variation:

$$|ij|^2 \approx \mathrm{var}\left\{\log(x_i / x_j)\right\}$$

$$\text{and also: } |Oi|^2 \approx \mathrm{var}\left[\log\left\{x_1 / g(x)\right\}\right]$$

where: $g(x)$ is the geometric centre.

2. Angle cosines between links in the biplot approximate to correlations between log-ratios. If links *ij* and *kl* intersect in a point *M* then:

$$\cos|iMk| = corr\left\{\log\left(x_i / x_j\right), \log\left(x_k / x_l\right)\right\}$$

This means that in a case where two links are at a right angle, $\cos|iMk| \approx 0$, and there is zero correlation of the two log-ratios. This feature is useful in investigation of possible independence of subcompositions involved.

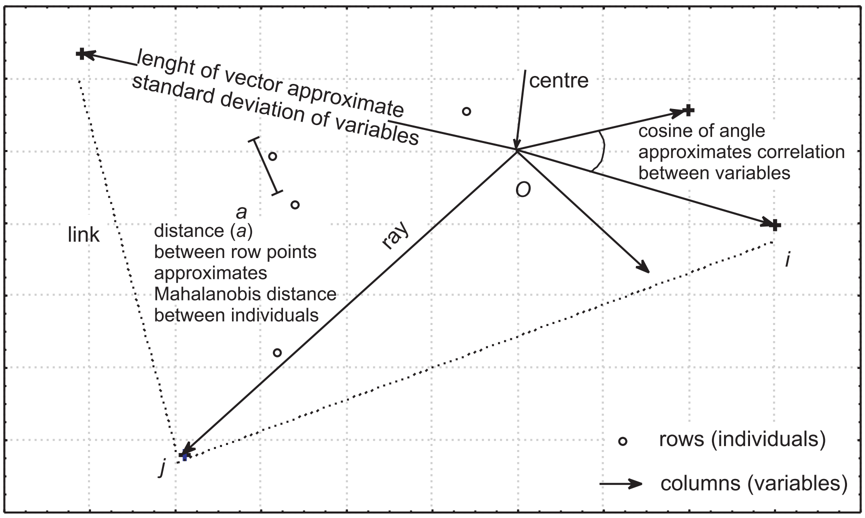

Fig. 1. Interpretation parameters of a general biplot

T — transitive pores, *S* — submacropores, *R* — real macropores, *O* — over-capillary pores, *Re* — skeletal grains

3. The centre O is the centre of gravity of the D vertices 1,…,D. The biplot for any subcomposition is formed by selecting the vertices corresponding to the parts of the subcomposition and taking the centre O, of the subcompositional biplot as the centroid of the vertices. This operation is possible since ratios are preserved under the formation of subcompositions.

4. If a subset of vertices (1,…,C) is co-llinear, the associated subcomposition has a biplot that is one-dimensional, and this leads to the conclusion that the subcomposition has one-dimensional variability. In case the column points are located along a straight line, a model describing this interdependency can be deduced from the relative lengths of their links. This means that if three transformed components A, B and C lie in an approximate straight line with distances AB and BC equal to α and β, respectively, then the constant log-contrast is of the form:

$$\beta \log(A) + \alpha \log(C) - (\alpha + \beta)\log(B) = \text{const};$$
$$\text{thus} \quad (A/B)^{\beta} \sim (B/C)^{\alpha}$$

where: A, B, and C stand for the corresponding original, non-transformed parts.

5. Case markers have the property that the inner product $Oc_n ji$ represents the departure of $\log(x_i/x_j)$ for case c_n from the average of this log-ratio over all cases. For illustration, consider Figure 2. Let be P the orthogonal projection of the centre O on the link ji and P_n the corresponding projection of the compositional marker c_n. Furthermore, consider the link ji as divided into positive and negative parts by the point P, the positive part being in the direction of the vertex j from P. If P_n falls on the positive (or negative) side of this line then the log-ratio of $\log(x_{ni}/x_{nj})$ of the n-th composition exceeds (falls short of) the average value of this log-ratio over all cases. The further P_n is from P, the greater is this exceedance (shortfall). If P_n coincides with P then the compositional log-ratio coincides with the average.

It must be underlined that that the links are the fundamental elements of compositional biplot interpretation, contrary to the case of variation diagrams for unconstrained data, where the role of rays is more meaningful. The complete set of links, by specifying all the relative variances, determines the compositional covariance structure and provides information about subcompositional variability and independence.

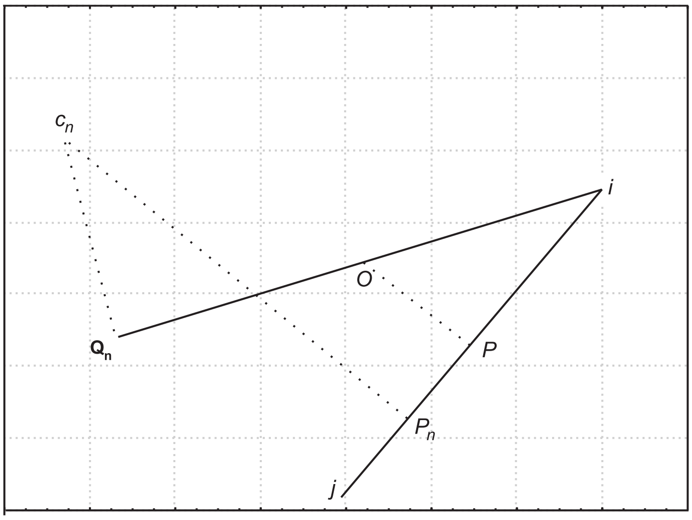

Fig. 2. Case markers interpretation

For explanations see text

STATISTICAL INTERPRETATION

In this paper compositions concerning pore distribution (4 classes of pore dimensions and the share of skeletal grains) are analysed, as well as subcompositions of the data matrix. Subcomposition means composition of a subset of selected parts. An important property of compositional data is that the ratio of any two components of the subcomposition is the same as the ratio of the corresponding two components in the full, original composition.

Figure 3 presents the biplot of the available compositions (47 samples).

A rough interpretation of the biplot for 47 compositions (Fig. 3) might be the following:

1. The T–R link is the longest, which indicates that probably the greatest relative variation in the ratios of components is between the transitive and real macropores. i.e. the large distance between the vertices T and R is indicative of significant variability of the log-ratio — $\log(T/R)$, which means that there is no constant proportion between transitive pores and macropores.

2. The near-orthogonality of the S–O link and the T–R link implies that the log-ratio of transitive to real macropores might be independent of the ratio S–O (submacropores and over-capillary). This feature can be expressed also in an equivalent way: the subcompositions T–R and S–O might be independent. However, this hypothesis of subcompositional independence could not be assessed, as the normality of the $\log(S/O)$ distribution was not confirmed by a formal test (discussed later in the paper).

3. The approximate co-linearity of the vertices R, O, T, Re, indicates that probably the variability of the corresponding subcomposition is mainly one-dimensional, and suggests the possibility of a log-contrast principal component analysis to determine the form of the constant log-contrast (discussed later).

4. In Figure 3 “Cum. prop. expl.” (cumulative proportion explained) is the cumulative variance explained by the (D-1) principal components (for D analyzed parts), given that all the principal components together explain 100% of the sample variation. In the example, the first of the components explains 67% of variance, whereas in the second it is 22%, adding up to 89% of total variability explained, which is also the variability explained by the biplot.

5. Four compositions of pore sizes (samples 11, 26, 33 and 35, marked as outliers in Fig. 3) have atypicality indices greater than 0.95. From the position of the case marker for composition 11, it is obvious that this atypicality is due to the combination of unusually high ratios of over-capillary pores to submacropores (O–S) and transitive to real macropores (T–R). In the case of the remaining outliers

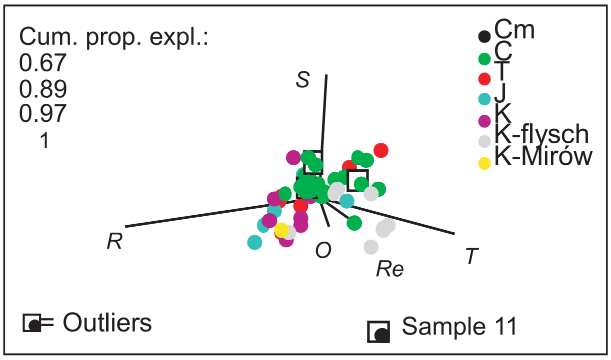

Fig. 3. Biplot of pore and skeletal grains distribution — 47 compositions

Cm — Cambrian quartzite sandstone (Holy Cross Mts.), C — late Carboniferous sandstones (Upper Silesia), T — Triassic sandstones (Holy Cross Mts.), J — Jurassic sandstones (Holy Cross Mts.), K — Cretaceous sandstones (Lower Silesia), K-flysch — late Cretaceous flysch sandstones (Beskidy region), K-Mirów — Cretaceous sandstone from Mirów (Silesian-Kraków Upland); symbols as in Figure 1

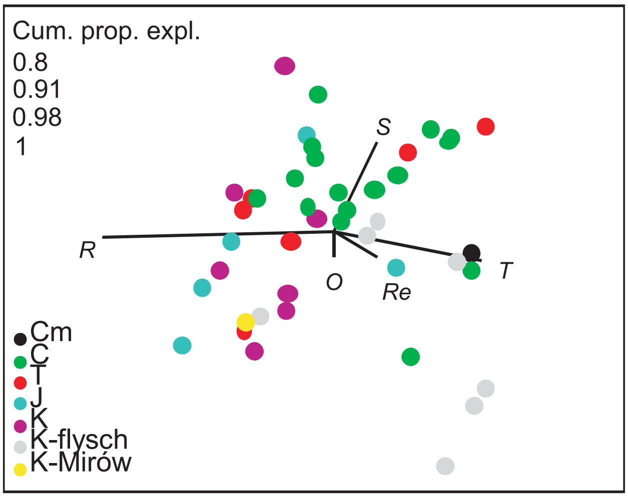

Fig. 4. Biplot of pore and the skeletal grains — 43 compositions set

Symbols as in Figure 1

(three samples), their position on the biplot is less apparent; however, their atypicality is due to high ratios of skeletal grains and transitive pores (*Re*–*T*), and over-capillary and skeletal grains (*O*–*Re*), respectively. These outliers were excluded, prior to the next stages of interpretation.

The interpretation of the 43 compositions set, that was left after the exclusion of outliers, is presented below in detail:

The graphical results of the biplot (i.e. the length of links) should be compared to the numerical results of the compositional variation array that provides a useful descriptive summary of the variability pattern of compositions (Aitchison, 2003*b*). Table 1 presents the variation array of the analyzed composition; its upper triangle displays the log-ratio variances: $\hat{\tau}_{ij} = \mathrm{var}\left\{\log\left(x_{ri}/x_{rj}\right)\right\}$, while the lower one a shows the log-ratio means: $\hat{\xi}_{ij} = E\left\{\log\left(x_{ri}/x_{rj}\right)\right\}$.

It can be seen from Table 1 that the most significant variation is between transitive (*T*) and real macropores (*R*), with $\hat{\tau}_{TR} = 5.68$. A similar conclusion could be drawn from the analysis of the biplot in Figure 4, where the longest link, *TR*, indicates that the greatest relative variation in the ratios of components might be between the transitive and real macropores. The negative value of $\hat{\xi}_{TR} = -0.79$ suggests that the *T* values are smaller than the *R* values in an average sense. The inequality: $\sqrt{\hat{\tau}_{ij}} > \hat{\xi}_{TR}$ proves that for a substantial number of samples the log(*T*/*R*) value is positive with the corresponding percentage of transitive (*T*) exceeding that of real macropores (*R*).

The relative variation between *Re* (skeletal grains) and *S* (submacropores) is one of the smallest: $\hat{\tau}_{ReS} = 0.583$, see also the *S*–*Re* link in the biplot (Fig. 4). The positive values of $\hat{\xi}_{TR} = 3.46$, and $\sqrt{\hat{\tau}_{ij}} > \hat{\xi}_{TR}$, are consistent with the obvious domination of the skeletal grains (*Re*) over the pore volumes in any of the samples.

In the biplot of the set of 43 compositions (Fig. 4), the first of the components explains 80% of variance, whereas the second explains 11%, adding up to 91% of total variability explained, which is a rather high value. The *S*–*Re* link and the *T*–*R* link are nearly orthogonal. This geometrical relationship suggests that the log-ratio of transitive to real macropores might be independent of the ratio *S*–*Re* (submacropores and skeletal grains). In other words: subcompositions *T*–*R* and *S*–*Re* might be independent. This possible subcompositional independence requires confirmation by a formal test, based on procedures assuming multivariate normality of transformed compositions (alr-transformed compositions). In the first step, the tests of logistic normality were performed in a manner suggested by Pawlowsky-Glahn and Buccianti (2002) and Aitchison (2003*b*). Marginal and bivariate angle test statistics were applied. The relevant formulae and the critical values are shown in Tables 2 and 3.

Table 1

Compositional variation array for the 43 compositions set

	Re	*T*	*S*	*R*	*O*
Re	–	0.71	0.58	3.12	0.23
T	4.18	–	1.11	5.68	1.32
S	3.46	–0.72	–	3.36	0.73
R	3.39	–0.79	–0.07	–	2.42
O	5.51	1.33	2.05	2.12	–

Upper triangle: Variances; lower triangle: Means

Table 2

Marginal test statistics and their critical values

Significance level [%]	10	5	2.5	1
Anderson-Darling $Q_A = \{-\frac{1}{N}\sum_{i=1}^{N}(2i-1)[\log z_i + \log(1-z_{N+1-i})] - N\}(1+\frac{4}{N}-\frac{25}{N^2})$	0.656	0.787	0.918	1.092
Cramer-von Mises $Q_C = [-\sum_{i=1}^{N}(z_i - \frac{2i-1}{2N})^2 + \frac{1}{12N}](1+\frac{1}{2N})$	0.104	0.126	0.148	0.178
Watson $Q_W = Q_C - N(z-\frac{1}{2})^2(1+\frac{1}{2N})$, where $z = \frac{1}{N}\sum_{i=1}^{N} z_i$	0.096	0.116	0.136	0.163

Table 3

Bivariate angle test statistics and their critical values

Significance level [%]	10	5	2.5	1
Anderson-Darling $Q_A = -\frac{1}{N}\sum_{i=1}^{N}(2i-1)[\log z_i + \log(1-z_{N+1-i})] - N$	1.933	2.492	3.070	3.857
Cramer-von Mises $Q_C = \{[-\sum_{i=1}^{N}(z_i - \frac{2i-1}{2N})^2 + \frac{1}{12N}] - \frac{0.4}{N} + \frac{0.6}{N^2}\}(1+\frac{1}{N})$	0.347	0.461	0.581	0.743
Watson $Q_W = [-\sum_{i=1}^{N}(z_i - \frac{2i-1}{2N})^2 - N(z+\frac{1}{2})^2 - \frac{0.2}{12N} + \frac{0.1}{N^2}](\frac{N+0.8}{N})$	0.152	0.187	0.221	0.267

For the *i*-th marginal distribution of the log-ratio composition the observations are: y_{ri}=log(x_{ri}/x_{rN}), with r=1,…,N. The value of z_i is calculated from the following formula:

$$\Phi\left[\left(y_{i1} - y_1 / s_1\right)\right] = z_{i1}$$

where: s stands for the standard deviation, and Φ(.) is the cumulative distribution function of N(0;1).

Values of z_i used in the bivariate angle distribution tests are calculated as $z_i=\theta_i/(2\pi)$, where $\theta_i=\arctan(u_{i2}/u_{i1})$ + 0.5{1–$sign(u_{i1})$}π) + 0.5 {1+$sign(u_{i1})$}{1–sign(u_{i2})}π, and:

$$u_{i1} = \frac{\left(y_{i1} - \bar{y}_1\right)s_2}{\sqrt{s_1^2 s_2^2 - s_{12}^2}} - \frac{\left(y_{i2} - \bar{y}_2\right)s_{12}}{s_2\sqrt{s_1^2 s_1^2 - s_{12}^2}};\ u_{i2} = \frac{\left(y_{i2} - \bar{y}_2\right)}{s_2}$$

The signum function (*sign*) is defined as:

$$sign(x) = \begin{cases} -1 & if \quad x < 0 \\ 0 & if \quad x = 0 \\ +1 & if \quad x > 0 \end{cases}$$

The sequences of z_i values (different for the marginal and for bivariate angle test), rearranged in ascending order of magnitude, are used in the expressions of Q_A for the Anderson-Darling, Q_C for the Cramer-von Mises, and Q_W for the Watson test, and compared with the critical values. The larger the values of the statistics are, the smaller is the significance level.

In the next step, the dependence between the transformed compositions — log (T/R) and log(S/Re), assuming their logistic normality, was tested by means of the test for subcompositional independence (Aitchison, 2003*b*).

This test compares:

$$N\left\{\ln\left(\left|\hat{\Sigma}_{11}\right|\left|\hat{\Sigma}_{22}\right|\right) - \ln\begin{vmatrix}\hat{\Sigma}_{11} & \hat{\Sigma}_{12} \\ \hat{\Sigma}_{21} & \hat{\Sigma}_{22}\end{vmatrix}\right\}$$

against upper percentage points of the $x^2\{(c–1)(d–c)\}$ distribution, where: $\hat{\Sigma}_{ij}$ — sample covariation matrices; d — stands for the number of dimensions of the whole of the composition; c — is the number of the first subcomposition dimensions.

Results of the tests, applied to the composition without outliers, are shown in the following table (Table 4).

Comparison of the computed values of modified distribution function test statistics with the corresponding critical values (Tables 2 and 3) shows no significant departure from additive logistic normality at the significance level of over 5 percent, in all of the tests. This means, that there is no reason to reject the hypothesis of multivariate normality for the variables analyzed: log (T/R) and log(S/Re). The significance probability of 0.776 obtained from the subcompositional independence test (Table 4) allows rejection of the hypothesis of dependence between log(T/R)–log(S/Re), i.e. the subcompositions (T, R) and (S, Re) can be assumed to be independent. This can be interpreted as follows: the proportion of T to R is independent of the proportion of S to Re.

Table 4

Tests on multivariate normality and the compositional independence test of log(*T*/*R*) and log(*S*/*Re*)

Tests	Anderson-Darling	Cramer-von Mises	Watson
log(*T*/*R*) — marginal distribution	0.776	0.116	0.114
log(*S*/*Re*) — marginal distribution	0.709	0.115	0.099
Bivariate angle test statistics	0.862	0.170	0.089
log(*T*/*R*) — log(*S*/*Re*) compositional independence test			0.776

Apart from the biplot, the ternary diagrams of centred subcompositions of the data matrix are interpreted. The lines inside the graphs (Figs. 5 and 6) represent axes of principal components — log-contrast principal axes in the sense of Aitchison (2003*b*), which may be understood as "regression lines" in a regression model for compositions (Billheimer *et al*., 1998). Numbers (summing to 1) near the explanations of each of the principal components are the values of the regression parameter vectors.

In Figure 3, the link *TR* was almost perpendicular to the link *SO*. This suggested that the subcomposition formed by submacropores and over-capillary pores might be independent of the subcomposition formed by transitive pores and macropores. The correlation between log-ratios log(*T*/*R*) and log(*S*/*O*) equals approximately to 0. The link *TR* goes nearly through *O*; the log-contrast: $\beta \log(T) + \alpha \log(R) - (\alpha+\beta)\log(O) = \text{const}$, has the following form: $1.857 \log(T) + \log(R) - 2.857 \log(O) = \text{const.}$ or, after rounding the coefficients in such a way that the sum of all of them is zero, $2\log(T) + \log(R) - 3\log(O) = \text{constant}$. This is equivalent to $(T^2 R / O^3) = \text{constant}$.

In Figure 4, however, showing the compositional biplot without the outliers, the link *TR* is almost perpendicular to the link *SRe*. This suggests that the subcomposition formed by submacropores and skeletal grains is independent of the

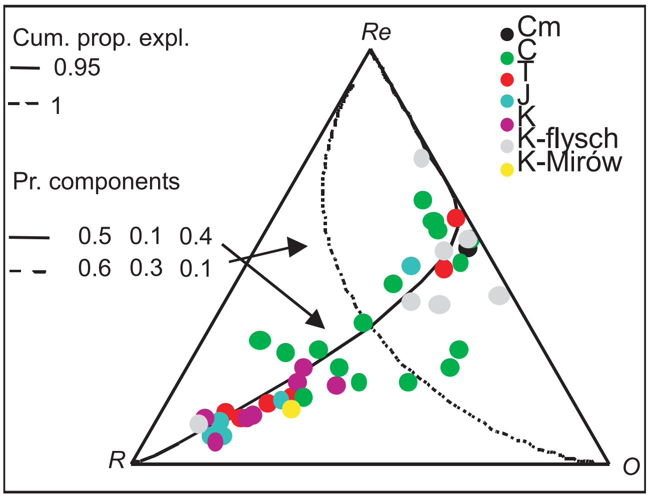

Fig. 6. Ternary diagram of pore sizes and skeletal grains distribution — subcomposition of skeletal grains, real macropores and over-capillary pores

Symbols as in Figures 1 and 4

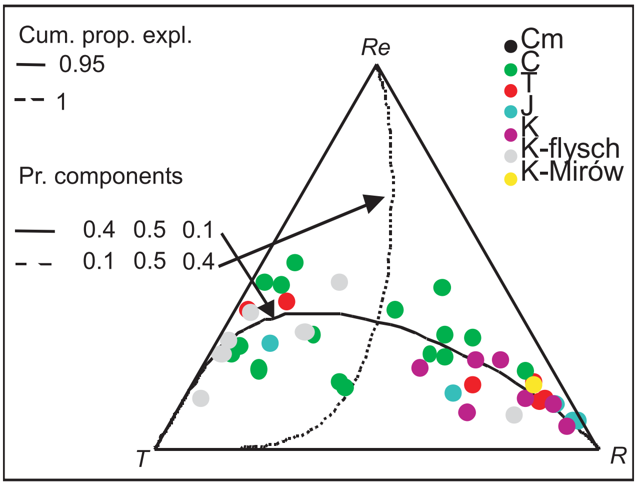

Fig. 5. Ternary diagram of pore sizes and skeletal grains distribution — subcomposition of skeletal grains, transitive and real macropores

Symbols as in Figures 1 and 4

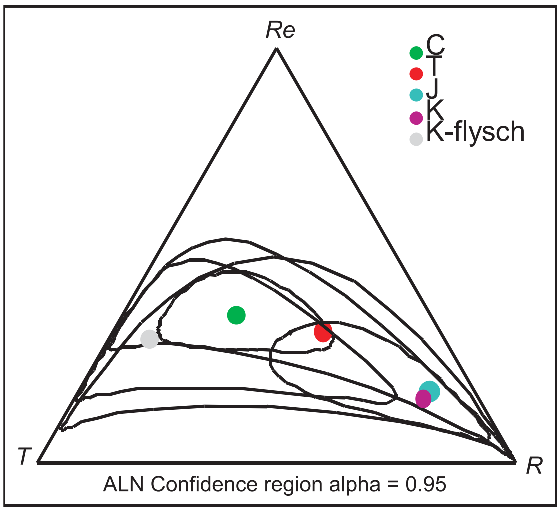

Fig. 7. Confidence regions (α=0.95) and geometric means for sandstone groups

Symbols as in Figures 1 and 4

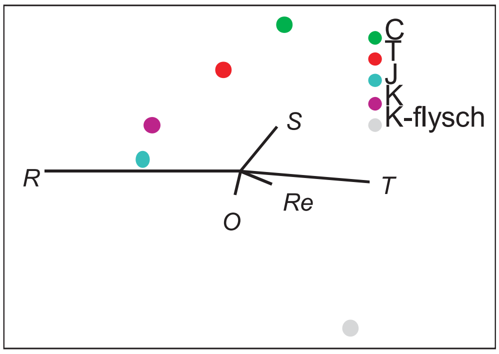

Fig. 8. Biplot of pore and skeletal grains distribution — showing geometrical centres of stratigraphical groups

Symbols as in Figures 1 and 4

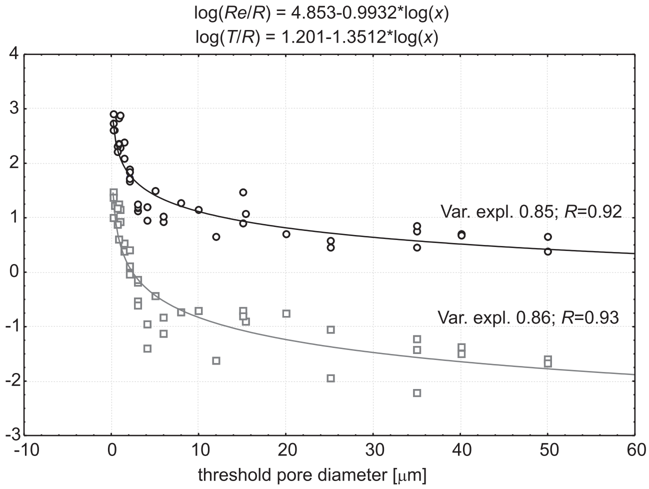

Fig. 9. Log-ratios log(*Re*/*R*) (above) and log(*T*/*R*) (below) as a function of the threshold pore diameter

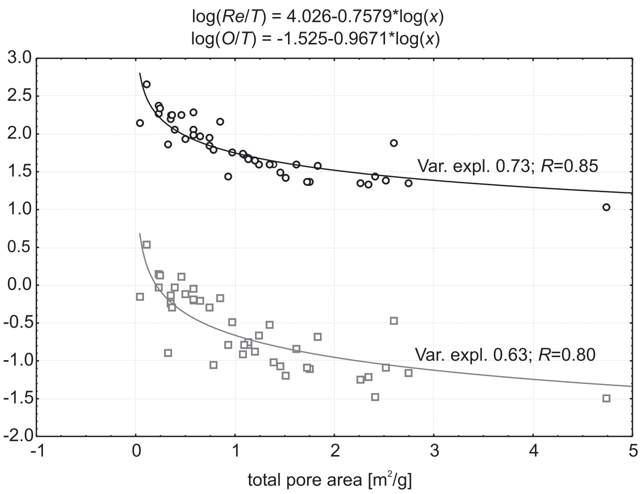

Fig. 10. Log-ratios log(*Re*/*T*) (above) and log(*O*/*T*) (below) as a function of the total pore area

subcomposition formed by transitive pores and macropores, as confirmed by the presented tests. The correlation between log-ratios log(T/R) and log(S/Re) equals approximately to 0. Now, the link TR goes through Re; the log-contrast: $\beta \log(T) + \alpha \log(R) - (\alpha + \beta)\log(Re)$ = const., has the following form: 2,71 log(T) + log (R) – 3.71 log (Re) = const., leading to ($T^3 R/ Re^4$) = const. This relationship is also presented in triangular coordinates for the 3-part composition in Figure 5, showing a good fit to the data, explaining 0.95% of the total variability in the subcomposition.

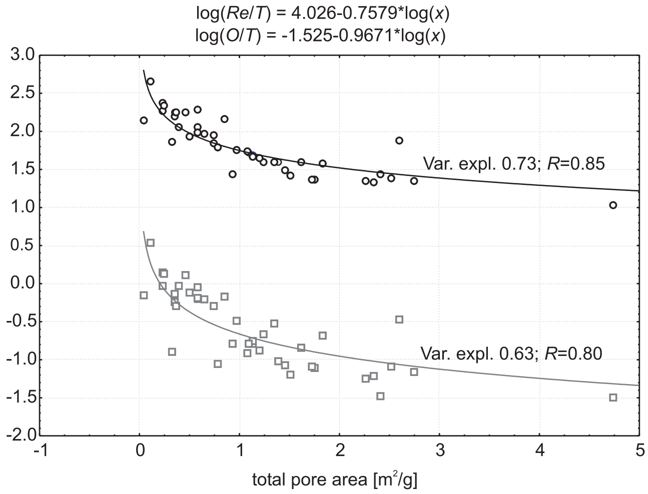

Fig. 11. Log-ratios log(*Re*/*R*) (above) and log(*S*/*R*) (below) as a function of the median pore diameter

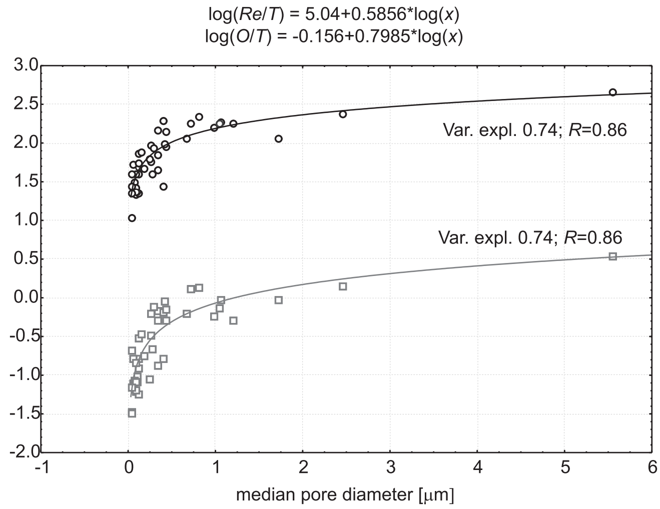

Fig. 12. Log-ratios log(*Re*/*T*) (above) and log(*O*/*T*) (below) as a function of the median pore diameter

At this point we are led to answer a question whether the presented log-contrast has any geological meaning. From the petrological point of view such an equation could not be explained by any of the existing laws and rules. But it would be worth testing, if log-contrasts may serve as characteristics ("codes"), for various clastic rocks (different in the sense of age, lithology, and diagenesis).

From Figure 7 it appears clear that stratigraphical groups of sandstones are not distinct, as regards the pore sizes and skeletal grains proportion. On the other hand the points which represent the geometrical means of the sandstone groups lay on the line corresponding to the trend of pore sizes distribution (the axis of the first principal component on Fig. 4). The share of transitive pores decreases in favour of real macropores in the following direction: J~>K>T>C>K-flysch. This trend is modified by the variance of skeletal grains, on average the greatest in Carboniferous sandstones. This interpretation, however, might be inaccurate owing to differences in the sample sizes of groups.

On the next biplot (Fig. 8), the geometrical centres of stratigraphical groups of sandstones are shown against the background of the biplot rays. (This picture is to some extent similar to the ternary diagram in Figure 7). The position of the group centres shows that the share of transitive pores (T) is decreasing, with growth of real macropores (R), arranging the samples in the known order: J~>K>T>C>K-flysch. The falling fraction of the tiny, transitive pores (T) may be connected with the decrease in the degree of diagenesis, which is dependent on the age of sandstones. The older rocks would contain less pores of large volume.

In order to describe the possible process by which the pore sizes variability may depend on rock parameters (skeletal density; bulk density, median pore diameter, total pore area = specific surface, hysteresis and threshold pore diameter) some log-ratio regression equations were defined. It was found that median pore diameter; threshold pore diameter and total pore area might be used as independent variables in these equations, explaining reasonable parts of the variability of the log-ratios examined. Testing of the goodness of fit (accepted at $\alpha = 0.05$; $p<0.01$) of the regression lines was performed using the least squares method.

The growth of the threshold pore diameter value leads to decrease of the log-ratios analyzed: $\log(Re/R)$ and $\log(T/R)$, seen in Figure 9, possibly due to the growth of the real macropores fraction in the subcomposition analyzed.

The growing value of the total pore area is connected with the decrease of the log-ratios analyzed — $\log(Re/T)$ and $\log(O/T)$, seen in Figure 10, This might be easily explained by the close relationship between the area parameter and the volume of the smallest (transitive) pores.

The increasing value of the median pore diameter value causes the decrease of the analyzed log-ratios — $\log(Re/R)$ and $\log(S/R)$ — relationship (Fig. 11), similar to the one already described in the cases of $\log(Re/R)$ and $\log(T/R)$.

The growth of the median pore diameter value leads to the increase of the analyzed log-ratios —$\log(Re/\mathrm{T})$ and $\log(O/T)$, by the increase of the smallest pores fraction as seen in Figure 12.

CONCLUSIONS

1. For the sandstones analyzed, the greatest relative variation in the ratios of components is between the transitive and real macropores; this means that there is no constant proportion between transitive pores and macropores.

2. Subcompositions of transitive and macropores, and submacropores and skeletal grains can be assumed to be independent. This hypothesis was confirmed by the test of compositional independence.

3. There is probably some relationship between the proportions of the real macropores, transitive pores and the skeletal grains fraction. Their variability is mainly one-dimensional.

4. Stratigraphical groups of sandstones are not distinct as regards the pore sizes and skeletal grains proportion; however, the points which represent geometric means of the sandstone groups lay on the line corresponding to the trend of pore sizes distribution. The share of transitive pores decreases in favour of real macropores in the following direction: J~>K>T>C>K-flysch. This trend is modified by the variance of skeletal grains, on average the greatest in Carboniferous sandstones.

5. The falling fraction of the tiny, transitive pores (T) — the order: J~>K>T>C>K-flysch — may be connected with the decrease in the degree of diagenesis, which is dependent on the age of sandstones.

The median pore diameter; threshold pore diameter and total pore area might be used as independent variables in equations that explain reasonable parts of the variability of log-ratios of pore sizes and skeletal grains fraction.

Acknowledgments. The author would like to express her thanks to the reviewers of the article. The paper greatly benefited from thorough, helpful and inspiring review by Prof. Vera Pawlowsky-Glahn.

REFERENCES

AITCHISON J. (1986) — The Statistical Analysis of Compositional Data, Monographs on Statistics and Applied Probability. Chapman and Hall, London.

AITCHISON J. (2003*a*) — A Concise Guide to Compositional Data Analysis. CDA Workshop, Girona.

AITCHISON J. (2003*b*) — The Statistical Analysis of Compositional Data. The Blackburn Press, New Jersey.

AITCHISON J. and GREENACRE M. (2002) — Biplots of Compositional Data. Appl. Statist., **51**: 375–392.

BILLHEIMER D., GUTTORP P. and FAGAN W. F. (1998) — Statistical Analysis and interpretation of discrete compositional data. NRCSE Techn. Rep. Ser., **011**.

GABRIEL K. R. (1971) — The biplot-graphic display of matrices with application to principal component analysis. Biometrica, **58**: 453–467.

HOBLER H. (1977) — Badania fizykochemicznych właściwości skał. PWN.

PAWLOWSKY-GLAHN V. and BUCCIANTI A. (2002) — Visualization and modelling of sub-populations of compositional data: statistical methods illustrated by means of geochemical data from fumarolic fluids. Int. J. Earth Sc., **91**: 357–368.

PAZDRO Z. (1983) — Hydrogeologia ogólna. Wyd. Geol.

REYMENT R. A. and SAVAZZI E. (1999) — Aspects of Multivariate Statistical Analysis in Geology. Elsevier.