

KRZYSZTOF LABUS\*, MAŁGORZATA LABUS\*

## Zastosowanie analizy danych złożonych (CDA) w geologii

### Słowa kluczowe

Analiza danych złożonych, chemizm wód podziemnych, porowatość skał

### Streszczenie

W artykule przedstawiono podstawowe założenia metody danych złożonych — *Compositional Data Analysis*. Metoda ta pozwala na statystyczną interpretację danych, w których wartości zmiennych sumują się do pewnej całości. Pokazano przykłady zastosowania metody CDA dla danych hydrogeologicznych oraz petrologicznych — w tym interpretację tych danych i ich wizualizację za pomocą diagramów trójkątnych. Specyfiką metody jest możliwość stosowania szeregu procedur przekształceń, np. amalgamacji danych, perturbacji, uwalniania i centrowania.

### Wprowadzenie

Metoda danych złożonych lub „składowych” (*Compositional Data Analysis*) jest stosowaną od niedawna metodą statystycznego opracowania danych, wprowadzoną przez Aitchinsona w 1986 roku (Aitchison 1986). Jest ona ciągle rozwijana i znajduje zastosowanie w coraz to nowych dziedzinach nauki (Aitchison, Barceló-Vidal 2002; Aitchison, Kay 2003). Stosunkowo szeroko stosowana jest w naukach o Ziemi, m.in. w mineralogii i petrografii (Tolosana-Delgado i in. 2002; Martín-Fernández i in. 2003), paleontologii (Buccianti, Monechi 2002) i geochemii (Buccianti 2003; Otero i in. 2003; Eynatten 2003). W węższym zakresie analiza CDA wykorzystywana jest także w archeometrii (Baxter i in. 2003),

---

\* Dr inż., Politechnika Śląska, Wydział Górnictwa i Geologii, Instytut Geologii Stosowanej, Gliwice.

Recenzował prof. dr hab. inż. Marek Nieć

genetyce (Gasparini, Di Gaetano 2003) i innych dziedzinach. Poniżej zostaną przedstawione nowe, oryginalne przykłady zastosowania procedury CDA w hydrogeologii i petrologii.

### 1. Założenia metody Analizy Danych Złożonych

Analiza danych złożonych (CDA) obejmuje zespół procedur, po raz pierwszy przedstawiony w pracy Aitchisona (1986). Założeniem *Compositional Data* jest to, iż składają się z  $x$  wektorów o nieujemnych elementach  $x_1, \dots, x_D$ , stanowiących pewną całość:

$$x_1 + \dots + x_D = 1$$

Ponieważ składowe w równaniu sumują się do jedności, nie są zmiennymi niezależnymi. Nazywa się je (w dużym uproszczeniu) danymi zamkniętymi (*closed data*). Jest to rodzaj danych najczęściej występujących w geologii — np. geochemii (np. skład tlenkowy), sedimentologii (udział ziaren w poszczególnych klasach frakcji), petrografii, palinologii, paleoekologii, pomiarach środowiskowych itp.

Informacja zawarta w wektorach (kompozycjach o nieujemnych elementach  $x_1, \dots, x_D$ ) jest związana ze względną, a nie absolutną, zawartością składnika. Stąd też zależności pomiędzy składnikami mogą być wyrażane w formie proporcji. Transformacja przestrzeni wektorów kompozycyjnych — simplex — w rzeczywistość przestrzeń euklidesową może być dokonywana poprzez zespół transformacji logarytmicznych (Aitchison 2003b).

Poniżej wymieniono charakterystyczne cechy zbioru danych składowych (*Compositional Data*) (Reyment, Savazzi 1999):

- każdy rząd macierzy danych odnosi się do pojedynczej próbki (np. wody, skały),
- każda kolumna macierzy danych przedstawia pojedynczy składnik (*part*), np. pierwiastek chemiczny, minerał, jon itp.,
- każdy składnik macierzy jest nieujemny,
- każdy rząd macierzy sumuje się do 1 (np. proporcje) lub do 100 (np. udział procentowy). Można znaleźć inną stałą wartość, np. dzięki przekształceniom (tzw. perturbacjom),
- współczynnik korelacji pomiędzy składnikami zmienia się, jeśli jedną ze zmiennych usuniemy z macierzy i znowu zsumujemy rzędy do 100 (poprzez operację subkompozycji), jest to właściwość korelacji zależnej od zmiennych. Taki sam efekt uzyskuje się przez dodanie do rozważań nowego składnika.

Zgodnie z ideą danych złożonych przedstawioną przez Aitchisona (1986), zakłada się co następuje: Jeśli  $W = (w_1 \ w_2 \ \dots \ w_D)$  jest niezerowym wektorem, który nie jest kompozycją ( $w_1 \ w_2 \ \dots \ w_D$  oznaczają np. zawartości podstawowych składników cieczy wyrażone w  $\text{mg}/\text{dm}^3$ ), to ze zbioru  $W$  można zawsze utworzyć pewien zbiór danych złożonych  $X$ . Operacja ta polega na podzieleniu wartości poszczególnych składników przez ich sumę:  $x_i = w_i / (w_1 + w_2 + \dots + w_D)$ , a  $X = (x_1 \ x_2 \ \dots \ x_D)$  jest wynikającym z tego działaniem zbiorem

(kompozycją — obejmującą dla podanego przykładu udziały procentowe składników cieczy). Występująca wtedy zależność:

$$\frac{w_i}{w_j} = \frac{x_i}{x_j}$$

jest zawsze prawdziwa dla  $i, j = 1, 2, \dots, D$ , przy założeniu, że  $w_j$  i  $x_j$  nie są zerowe. Ta właściwość podkreśla fakt, że dane złożone zawierają jedynie informację o względnych wielkościach — relacjach pomiędzy składnikami, a nie o wielkościach absolutnych — wartościach (Pawłowsky-Glahn, Olea 2004).

Operacje na danych złożonych odbywają się w ograniczonej części przestrzeni, jaką jest simplex. Najważniejszym założeniem jest to, iż pomimo, że wektor  $x$  składa się z  $D$  elementów, jego składniki są w pełni opisane przez  $d$  komponentów subwektora złożonego z  $d$  części, gdzie  $(d = D - 1)$ . Stąd  $x_D = 1 - x_1 - \dots - x_d$ . Tak więc zbiór  $D$ -elementowy jest wektorem  $d$ -wymiarowym.  $d$ -wymiarowy simplex, ograniczony  $D$ -wymiarową przestrzenią, jest zbiorem zdefiniowanym przez:

$$S^d = \{(x_1, \dots, x_D): x_1 > \dots > x_D > 0; x_1 + \dots + x_D = 1\}$$

W ten sposób simplex  $S^d$  jest zdefiniowany jako ograniczona część przestrzeni o wyższym wymiarze  $R^D$ .

W sytuacji gdy kompozycja składa się z dwóch składowych ( $S^1$ ), ich reprezentacją graficzną jest linia 1-2 (rys. 1) o długości równej 1. Punkt  $P$  znajduje się bliżej  $i$ -tego końca linii, jeśli odpowiadająca mu składowa  $x_i$  ma dużą wartość.

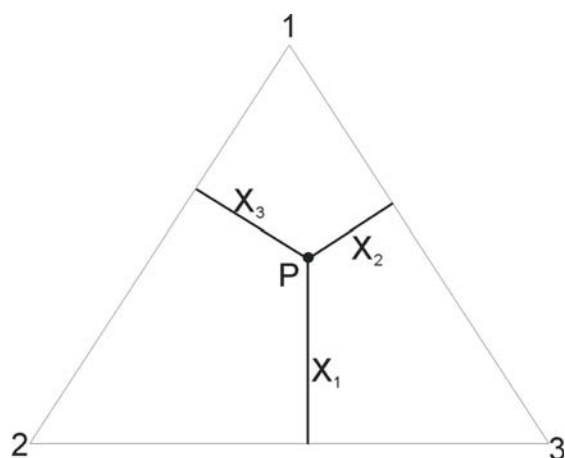


Rys. 1. Przedstawienie graficzne kompozycji dwuskładnikowej  $(x_1, x_2)$  na linii 1-2

Fig. 1. Representation of a 2-part composition  $(x_1, x_2)$  on the line 1-2

Simplex  $S^2$  odpowiada diagramowi trójkątnemu o wysokości równej jedności. Dla każdego punktu  $P$  w trójkącie o wierzchołkach 1-2-3 (rys. 2) promienie  $x_1, x_2, x_3$ , prostopadłe do odpowiednich boków trójkąta (2-3, 1-3, 1-2) spełniają warunek  $x_1 + x_2 + x_3 = 1$ . Im wyższa jest wartość  $x_i$ , tym większa odległość punktu  $P$  od boku przeciwnego wierzchołkowi  $i$ .

W przypadku kompozycji 4-elementowej należy rozpatrywać przestrzeń 3-wymiarową (simplex  $S^3$ ), odpowiadającą tetraedrowi regularnemu 1-2-3-4, o wysokości jednostkowej (rys. 3). Składowa  $x_i$  przedstawiona jest jako promień prostopadły do ściany tetraedru, przeciwległej w stosunku do wierzchołka  $i$ . Dla kompozycji składających się z większej liczby składowych nie ma wystarczająco dobrego sposobu graficznego przedstawienia,

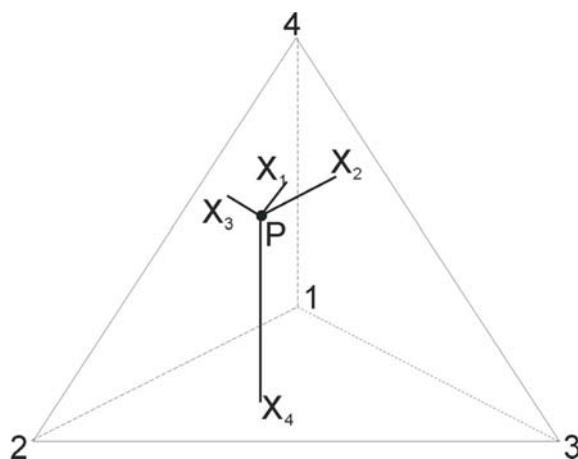


Rys. 2. Przedstawienie graficzne kompozycji trójskładnikowej  $(x_1, x_2, x_3)$  w trójkącie 1-2-3

Fig. 2. Representation of a 3-part composition  $(x_1, x_2, x_3)$  in the reference triangle 1-2-3

dlatego dąży się do zmniejszenia liczby składowych poprzez tworzenie subkompozycji (Aitchison 2003b).

Ważną cechą danych złożonych jest to, że stosunek jakichkolwiek dwóch składowych subkompozycji jest taki sam jak stosunek odpowiadających im dwu składowych w pełnej, pierwotnej kompozycji. Ta własność „zachowania zależności” oznacza, iż jeśli zachodzi potrzeba ograniczenia rozmiaru analizy, redukcja ta musi być przeprowadzona zgodnie z procedurą tworzenia subkompozycji. Nie polega to na trywialnym „odcięciu” części danych, lecz na ponownym przekształceniu wektora  $W = (w_1 \ w_2 \ \dots \ w_{D-n})$  (teraz o ograniczonej liczbie składników wynoszącej  $D - n$ ) na kompozycję  $X = (x_1 \ x_2 \ \dots \ x_{D-n})$ , elementów typu  $x_i = w_i / (w_1 + w_2 + \dots + w_{D-n})$ .

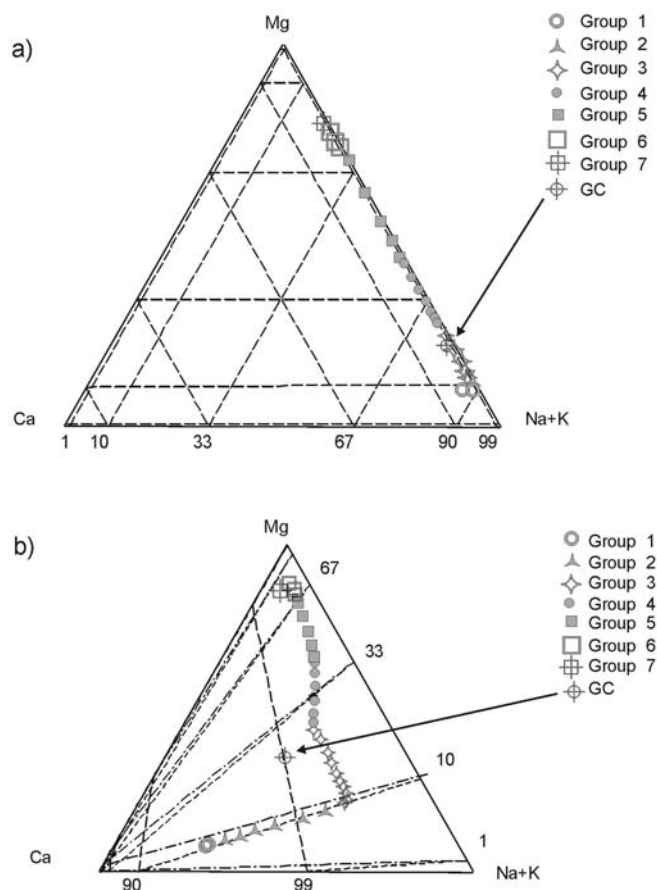


Rys. 3. Przedstawienie graficzne kompozycji czteroskładnikowej  $(x_1, x_2, x_3, x_4)$  w tetraedrze 1-2-3-4

Fig. 3. Representation of a 4-part composition  $(x_1, x_2, x_3, x_4)$  in the reference tetrahedron 1-2-3-4

## 2. Metody wizualizacji danych

W analizie danych złożonych, do graficznego przedstawienia zależności między zmiennymi stosowane są najczęściej wykresy typu biplot oraz diagramy trójkątne. Metoda biplot, wprowadzona przez Gabriela (1971) (Aitchison, Greenacre 2002), pozwala na przedsta-



Rys. 4. Diagramy trójkątne przedstawiające pierwotny (a) i poddany perturbacji (b) udział procentowy składników roztworu (subkompozycja kationów) na kolejnych etapach symulowanej ewaporacji wody morskiej

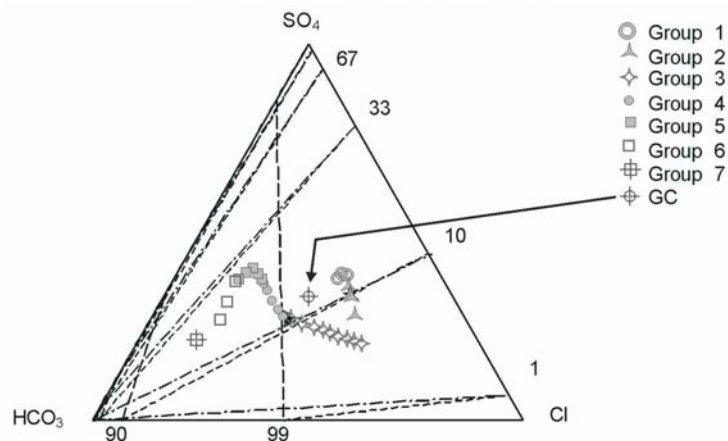
GC — środek geometryczny subkompozycji (0,01; 0,22; 0,77). Roztwory na etapach poprzedzających krystalizację: Group 1 — gipsu, Group 2 — halitu, Group 3 — anhydrytu, Group 4 — polihalitu, Group 5 — epsomitu, Group 6 — karnalitu, Group 7 — punkt krystalizacji karnalitu

Fig. 4. Ternary diagrams of original (a) and perturbed (b) cations subcomposition of sea water at different stages of evaporation

GC — subcomposition geometrical centre (0,01; 0,22; 0,77). Solutions at different stages before crystallization: Group 1 — gypsum, Group 2 — halite, Group 3 — anhydrite, Group 4 — polyhalite, Group 5 — epsomite, Group 6 — carnallite, Group 7 — carnallite crystallization point

wienie obserwacji i zmiennych na tym samym wykresie w sposób, który opisuje ich wzajemne zależności. Poniżej szerzej zostanie omówione zastosowanie diagramów trójkątnych. W standardowym ujęciu, stosowanym szeroko w naukach o Ziemi, diagramy te pozwalają na jednoczesne przedstawienie trzech składników — np. w przypadku typowych diagramów Peipera można operować trzema zestawami anionów i trzema zestawami kationów.

Zaletą metod CDA jest możliwość uwzględniania większej ilości zmiennych i przedstawiania ich na diagramach trójkątnych, po sprowadzeniu do przestrzeni trójwymiarowej (dzięki wyodrębnieniu subkompozycji). Przestrzenią danych złożonych jest simplex  $S^D$  (Aitchison 1986), jeśli  $D = 3$  — simplex można graficznie przedstawić w formie diagramu trójkątnego. Dla przestrzeni simplex możliwa jest operacja zwana perturbacją. Perturbacją wektora  $x = (x_1, x_2, x_3)$  w  $S^3$ , przez wektor  $p = (p_1, p_2, p_3)$  w  $S^3$ , jest nowy wektor  $p \oplus x = C(p_1x_1, p_2x_2, p_3x_3)$  w  $S^3$ , gdzie  $C$  oznacza operację tzw. zamykania (ang. „closure”). W trakcie tej operacji każda składowa wektora  $p \oplus x$  jest dzielona przez sumę wszystkich jego składowych. Perturbacja wektora  $x$  przez jego odwrotność  $x^{-1} = (1/x_1, 1/x_2, 1/x_3)$  daje w rezultacie element neutralny  $e = C(1, 1, 1) = (c/3, c/3, c/3)$ , reprezentowany na diagramie trójkątnym przez punkt leżący w jego środku geometrycznym. Opisane właściwości perturbacji mają znaczenie praktyczne. Perturbacja zestawu danych złożonych przez odwrotność jej środka geometrycznego ( $g^{-1}$ ) doprowadza do przeskalowania, prowadzącego do zgrupowania danych wokół środka geometrycznego w diagramie trójkątnym. Operacja ta pozwala ponadto na transformację linii prostych (np. linii siatki w pierwotnym diagramie trójkątnym) również na linie proste w diagramie przeskalowanym. Ilustracją graficznych rezultatów perturbacji są rysunki 4, 5 i 6.

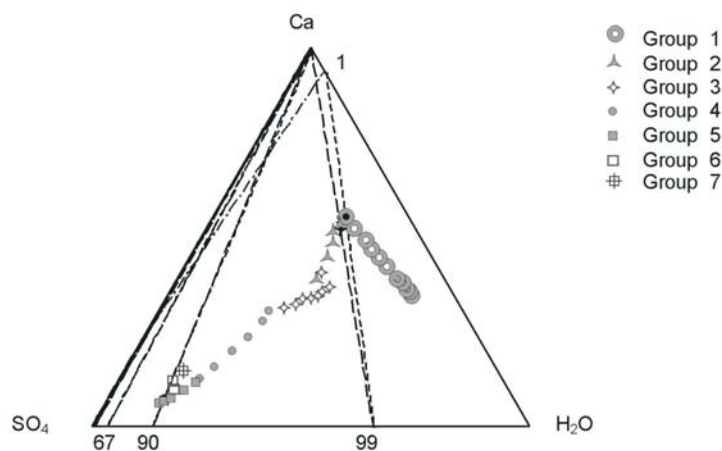


Rys. 5. Poddany perturbacji diagram udziału [%] składników roztworu (subkompozycja anionów) na kolejnych etapach symulowanej ewaporacji wody morskiej

GC — środek geometryczny subkompozycji (0,12; 0,01; 0,87), pozostałe objaśnienia jak na rys. 4

Fig. 5. Ternary diagram — after perturbation — of water composition (anions subcomposition), at different stages of sea water simulated evaporation

GC — subcomposition geometrical centre (0,12; 0,01; 0,87), groups explanations — as in fig. 4



Rys. 6. Poddany perturbacji diagram udziału [%] składników roztworu (subkompozycja Ca-SO<sub>4</sub>-H<sub>2</sub>O), na kolejnych etapach symulowanej ewaporacji wody morskiej  
Objaśnienia jak na rys. 4

Fig. 6. Ternary diagram — after perturbation — of water composition (subcomposition Ca-SO<sub>4</sub>-H<sub>2</sub>O), at different stages of sea water evaporation  
Explanations as in fig. 4

### 3. Przykład zastosowania analizy danych złożonych w hydrogeologii

Zastosowanie metody CDA w interpretacji zagadnień hydrogeochemicznych zostanie przedstawione na przykładzie zestawów danych, opisujących skład przeciętnej wody oceanicznej (wg Nordstrom i in. 1979) ulegającej ewaporacji. Skład wody obliczono na kolejnych etapach parowania (1500 etapów), symulowanego za pomocą programu PHREEQC (Parkhurst, Appelo 1999), zakładając, iż krystalizacja kolejnych soli (faz) następuje po osiągnięciu nasycenia nimi roztworu ( $SI = 0$ ). Otrzymany model, mimo uproszczeń, jest porównywalny z danymi prezentowanymi przez np. Fontes, Matray (1993). Masy składników w roztworze stanowią zmienne, natomiast przypadkami są ich kompozycje na kolejnych etapach ewaporacji. Na poniższych diagramach (rys. 4a i b) przedstawiono udziały procentowe składników roztworu (kationów) na kolejnych etapach symulowanej ewaporacji wody morskiej (Labus K. 2005).

Diagram na rysunku 4a jest zbliżony do trójkąta kationów w diagramie Pipera, z tą jednak różnicą, że diagram Pipera przedstawia względne udziały miliwali substancji w roztworze, demonstrowany wykres prezentuje udziały procentowe mas jonów głównych. Zmienność proporcji mas kationów w roztworze natomiast jest tutaj słabo widoczna, wskutek nieznacznego udziału Ca (poniżej 1%). Perturbacja oryginalnego diagramu (rys. 4b) pozwala uzyskać bliższy wgląd w analizowane relacje, zwłaszcza że układ odniesienia również uległ przeskalowaniu. Niewątpliwą zaletą tego typu operacji jest obiektywizacja

stałej przekształcenia — uzależnionej od rozproszenia danych wokół wartości centralnej, jaką jest średnia geometryczna kompozycji.

Rysunki 4b i 5 obrazują trend względnych zmian udziału składników roztworu wraz z postępującym procesem ewaporacji. Kolejne fragmenty trajektorii parowania wody morskiej (między punktami krystalizacji kolejnych soli) mogą być w przybliżeniu reprezentowane przez odcinki proste. Jednakże np. dla etapu poprzedzającego krystalizację gipsu, analizowane diagramy mogą nie dostarczać informacji na temat zróżnicowania składu roztworu (punkty reprezentujące skład wód nakładają się na siebie). W tej sytuacji dopiero przeprowadzenie analizy kolejnej subkompozycji, w tym przypadku uwzględniającej udział wody w roztworze (rys. 6), pozwala na śledzenie zmian także i na tym etapie ewaporacji.

W trakcie prac prowadzonych w Instytucie Geologii Stosowanej Politechniki Śląskiej, zweryfikowano także możliwości przedstawiania zmian proporcji składowych danej kompozycji w formie funkcji względem zmiennej nie należącej do tej kompozycji (np. zależności udziałów jonów głównych w pompowanych wodach względem wielkości uzyskiwanej depresji).

#### 4. Przykład zastosowania analizy danych złożonych w petrologii

W poniższym przykładzie (Labus M. 2005) danymi są wyniki analiz porozymetrycznych wykonanych dla 50 próbek skał. Próbki reprezentowały piaskowce różnowiekowe, z różnych jednostek geologicznych Polski. Pochodziły one z Gór Świętokrzyskich (skały paleozoiczne i mezozoiczne), z niecki śródsudeckiej (kredowe piaskowce ciosowe), z Górnego Śląska (piaskowce górnokarbońskie) i z Beskidu Śląskiego (piaskowce fliszowe górno-kredowe i trzeciorzędowe).

Celem pracy było wykrycie trendów relacji pomiędzy udziałem wydzielonych przeździałów średnic porów:

- pory przejściowe  $10^{-8}$ — $10^{-7}$  m,
- submakropory  $10^{-7}$ — $10^{-6}$  m,
- makropory właściwe  $10^{-6}$ — $10^{-4}$  m,
- pory nadkapilarne  $>10^{-4}$  m.

W analizie CDA stosuje się tworzenie podzbiorów („subkompozycji”) lub amalgamację (łączenie) danych (Aitchison 2003a,b; Buccianti i in. 1999).

W pracy analizowano kompozycje danych opisujących rozkład porów o różnych wielkościach (4 klasy) oraz subkompozycje tego zbioru, a także zestawy danych amalgamowanych. Subkompozycja oznacza kompozycję złożoną z podzbioru kompozycji sprowadzoną do 100%. Amalgamacja polega na sumowaniu udziału kilku podzbiorów w kompozycji.

W przedstawionym na rysunku 7 diagramie trójkątnym zastosowano amalgamację podzbioru porów przejściowych i makroporów (w górnym narożniku trójkąta). Widać wyraźnie, iż punkty odpowiadające poszczególnym próbkom koncentrują się na linii: pory przejścio-





Rys. 7. Diagram projekcji udziałów porów

W górnym wierzchołku — amalgamowane udziały porów przejściowych i makroporów;

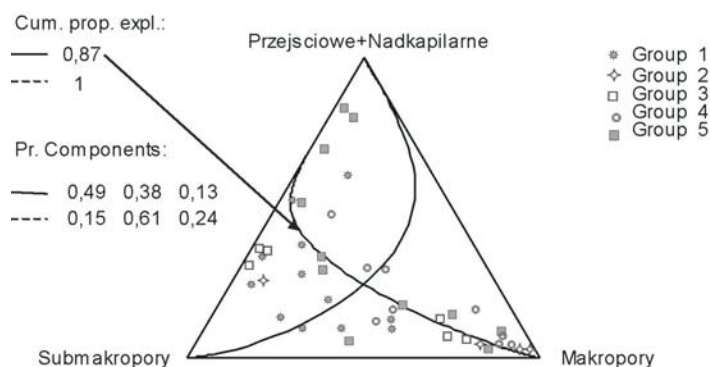
Group 1 — piaskowce górnokarbońskie, Group 2 — piaskowce triasu (Góry Świętokrzyskie),  
Group 3 — piaskowce jurajskie (Góry Świętokrzyskie), Group 4 — piaskowce kredowe (Sudety),  
Group 5 — górnokredowe piaskowce fliszowe Beskidu Śląskiego. Strzałka wskazuje oś pierwszej ze  
składowych głównych

Fig. 7. Ternary diagram of pore sizes distribution

top — amalgamated transitive and real macropores; Group 1 — late Carboniferous sandstones,  
Group 2 — Triassic sandstones (Holly Cross Mts.), Group 3 — Jurassic sandstones (Holly Cross Mts.),  
Group 4 — Cretaceous sandstones (Sudety Mts.), Group 5 — late Cretaceous flysh sandstones (Beskidy  
region). The arrow points the axis of the first principal component

we — makropory amalgamowane i submakropory. Oznacza to, iż zmienna: „pory nadkapilarne” niesie ze sobą najmniejszą ilość informacji.

Dla ułatwienia interpretacji dokonano amalgamowania tej zmiennej z pozostałymi. Przykład rezultatu takiej operacji przedstawia rysunek 8.



Rys. 8. Diagram projekcji udziałów porów

W górnym narożniku — amalgamowane udziały porów przejściowych i nadkapilarnych;

oznaczenia grup — jak na rys. 7. Strzałka wskazuje oś pierwszego z czynników głównych

Fig. 8. Ternary diagram of pore sizes distribution.

top — amalgamated transitive and over-capillary pores; groups — as in fig. 7

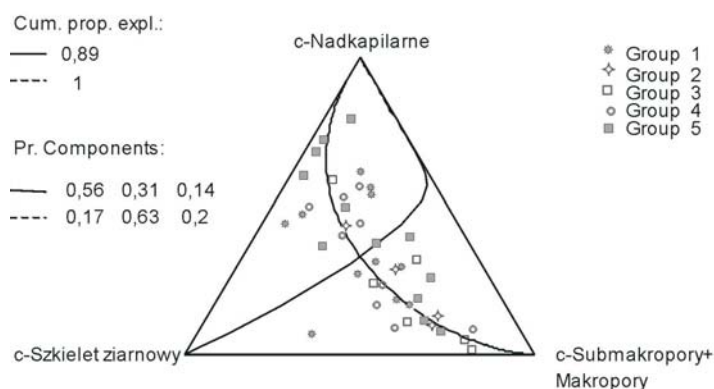
Na obu diagramach (rys. 7 i 8) przedstawiono osie wyodrębnionych składowych głównych (będących swego rodzaju liniami regresji w modelu regresji dla danej kompozycji) oraz wartości odpowiadające parametrom wektorów regresji obydwu składowych.

Oznaczenie „Cum. prop. expl” odnosi się do skumulowanej wariancji wyjaśnionej przez  $(n - 1)$  składowych głównych ( $n$  — liczba analizowanych zmiennych), przy założeniu, że składowe te wyjaśniają 100% wariancji. W przykładzie pierwszym (rys. 7) widać, iż pierwsza ze składowych wyjaśnia 73% wariancji, druga — pozostałą jej część (dopełnienie do 100%), tj. 27% wariancji.

Z drugiego diagramu (rys. 8) wynika, iż składowa pierwsza (odpowiadająca tym razem już za 87% wariancji) jest najsilniej skorelowana z udziałem porów przejściowych i nadkapilarnych. Innymi słowy, największy udział tej składowej obserwowany jest dla kształtowania zmienności porów przejściowych i nadkapilarnych (parametr wektora regresji = 0,49).

W dalszych obliczeniach zakładano, iż pełną kompozycję stanowiąc będą wszystkie stwierdzone przedziały porów oraz szkielet ziarnowy. Realizacja tego założenia wymagała proporcjonalnego podziału porowatości zmierzonej metodą rtęciową na części, proporcjonalnie do udziału kolejnych przedziałów porów (posłużono się operacją stosowaną dla CDA — „*unconstraining*” uwalniania). Dopełnienie do 100% stanowić miał udział szkieletu ziarnowego w objętości skały.

Analizowano populacje parametrów petrofizycznych w różnych konfiguracjach, kierując się zasadą wykorzystywania do analizy tych zmiennych, które charakteryzuje największa wariancja, tj. informatywność. Ostatecznie zbiór danych zoptymalizowano poprzez wyodrębnienie subkompozycji, uwzględniającej brak porów nadkapilarnych oraz zawierającej udziały szkieletu ziarnowego, porów przejściowych oraz amalgamowanych udziałów porów submakro i makro. Rezultat graficzny przedstawia rysunek 9.

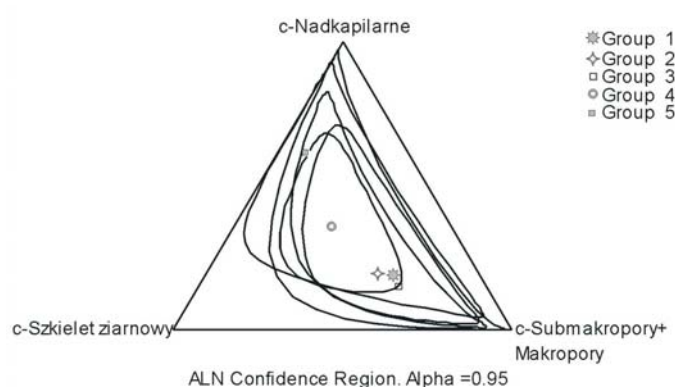


Rys. 9. Diagram projekcji udziałów porów i szkieletu ziarnowego  
Oznaczenie „c” przed każdą ze zmiennych wskazuje, że zostały one poddane operacji centrowania;  
oznaczenia grup — jak na rys. 7

Fig. 9. Ternary diagram of pore sizes and skeletal grains distribution. The letters „c” preceding the variables names means that they were centered. Explanations for groups — as in fig. 7

W tym przypadku stwierdzono bardziej znaczące zależności pomiędzy wyodrębnionymi składnikami a zmiennością udziałów porów w kolejnych przedziałach wielkości. Składnik 1 odpowiada za 89% zmienności i w największym stopniu tłumaczy wariancję porów nadkapilarnych — parametr wektora regresji = 0,56. W tym przypadku dla obu składowych uzyskano parametry wektora regresji o wartościach ponad 0,5, co świadczy o pełniejszym wyjaśnieniu wariancji zbioru danych niż w przykładzie zobrazowanym na rysunku 8.

Należy jednak stwierdzić, iż podział piaskowców na grupy, związany z ich pochodzeniem, nie jest wystarczająco wyraźny, o czym świadczą pokrywające się częściowo pola obszarów ufności dla kolejnych grup skał (rys. 10).



Rys. 10. Obszary ufności i średnie geometryczne zbiorów. Objasnienia — jak do rys. 7

Fig. 10. Confidence regions and geometric means for sandstone groups. Explanations as in fig. 7

Trzeba tutaj zaznaczyć, iż punkty reprezentujące średnie geometryczne poszczególnych grup piaskowców leżą na linii odpowiadającej trendowi zmian udziału porów należących do poszczególnych klas wymiarów. Udział amalgamowanych submakroporów wraz z makroporami maleje na korzyść porów przejściowych w kierunku: Grupa 3 > 4 > 2 > 1 > 5. Trend ten jest modyfikowany przez zmienność udziału szkieletu ziarnowego — przeciętnie największą w piaskowcach karbońskich (grupa 1).

Podsumowując powyższe rozważania należy stwierdzić, że

- Składowa pierwsza (wyjaśniająca 87% wariancji) jest najsilniej skorelowana z udziałem porów przejściowych i nadkapilarnych.
- W przypadku wyodrębnienia subkompozycji uwzględniającej brak porów nadkapilarnych oraz zawierającej udziały szkieletu ziarnowego, porów przejściowych oraz amalgamowanych udziałów submakroporów i makroporów, stwierdzono możliwość wyodrębnienia składowej odpowiadającej za 89% zmienności, tłumaczącej wariancję porów nadkapilarnych (składowa wektora regresji = 0,56).
- Podział piaskowców na grupy związany z ich pochodzeniem nie jest wystarczająco wyraźny, o czym świadczą pokrywające się częściowo pola przedziałów ufności dla kolejnych grup skał.

- Punkty reprezentujące średnie geometryczne poszczególnych grup piaskowców leżą na linii odpowiadającej trendowi zmian udziału porów należących do poszczególnych klas wymiarów.
- Udział amalgamowanych submakroporów wraz z makroporami maleje na korzyść porów przejściowych w kierunku: Grupa 3 > 4 > 2 > 1 > 5. Trend ten jest modyfikowany przez zmienność udziału szkieletu ziarnowego — przeciętnie największą w piaskowcach karbońskich (grupa 1).

### Podsumowanie

Zaprezentowana metoda danych złożonych (CDA) wprowadza nowe możliwości interpretacji danych geologicznych w przypadku gdy dane są danymi złożonymi, tj. wartości zmiennych sumują się do 100% lub do jedności. Zespół metod CDA:

- pozwala zobrazować wzajemne relacje między zmiennymi, a także pozwala na interpretację trendów tych relacji,
- umożliwia wizualizację wyników w porównaniu z interpretacją zmienności zbioru przy zastosowaniu metody składowych głównych na diagramach trójkątnych,
- pozwala na analizowanie związków pomiędzy większą liczbą zmiennych niż 3 (jak np. w tradycyjnych diagramach Piper, czy trójkątach klasyfikacyjnych skał), dzięki operacji amalgamowania,
- daje dodatkowe możliwości interpretacji, wynikające z zastosowania przekształceń (przeskalowania diagramów) w operacji perturbacji lub centrowania,
- umożliwia przedstawianie zmian proporcji składowych danej kompozycji w formie funkcji względem zmiennej nie należącej do tej kompozycji.

### LITERATURA

- Aitchison J., 1986 — *The Statistical Analysis of Compositional Data*, Monographs on Statistics and Applied Probability, Chapman and Hall, London.
- Aitchison J., 2003a — *A Concise Guide to Compositional Data Analysis*. CDA Workshop, Girona.
- Aitchison J., 2003b — *The Statistical Analysis of Compositional Data*, The Blackburn Press, New Jersey
- Aitchison J., Barceló-Vidal C., 2002 — *Compositional Processes: A statistical search for Understanding*. Terra Nostra, Schriften der Alfred-Wegener-Stiftung 04/2002, 8th Annual Conference of IAMG, pp. 381—386.
- Aitchison J., Greenacre M., 2002 — *Biplots of Compositional Data*. Appl. Statist., 51, pp. 375—382.
- Aitchison J., Kay J.W., 2003 — *Possible Solutions of some essential Zero Problems in Compositional data Analysis*. Proceedings of CODAWORK'03, The First Compositional data Analysis Workshop, Girona, Spain.
- Baxter M.J., Beardah C.C., Cool H.E.M., Jackson C.M., 2003 — *Compositional data Analysis in Archaeometry*. Proceedings of CODAWORK'03, The First Compositional data Analysis Workshop, Girona, Spain.

- Buccianti A., Monechi S., 2002 — Bi-plot analysis to discover relationships among species in micro-paleontological quantitative data. *Terra Nostra, Schriften der Alfred-Wegener-Stiftung 04/2002*, 8th Annual Conference of IAMG, pp. 399—404.
- Buccianti A., Pawlowsky-Glahn V., Barceló-Vidal C., Jarauta-Bragulat E., 1999 — Visualization and modelling of natural trends in ternary diagrams: A geochemical case study. *Proceedings of IAMG'99*, Trondheim, Norway.
- Buccianti A., Vaselli O., Nisi B., Minissale A., Tassi F., 2003 — Monitoring Procedures in Environmental Geochemistry and Compositional Data Analysis Theory. *Proceedings of CODAWORK'03, The First Compositional data Analysis Workshop*, Girona, Spain.
- Eynatten H., Barceló-Vidal C., Pawlowsky-Glahn V., 2003 — Modelling Compositional Change: The Example of Chemical Weathering of Granitoid Rocks. *Mathematical Geology*, 35 (3), pp. 231—251.
- Fontes J.C. and Matray J.M., 1993 — Geochemistry and origin of formation brines from the Paris Basin, France. 1. Brines associated with Triassic salts. *Chemical Geology*, 109: 149—175.
- Gasparini M., Di Gaetano C., 2003 — On the Use of Principal Components in Contemporary Population Genetics: A case study. *Proceedings of CODAWORK'03, The First Compositional data Analysis Workshop*, Girona, Spain.
- Labus M., 2005 — Compositional Data Analysis as a tool for interpretation of rock porosity parameters. *Geol.Quart.*, 49(3): 347—354.
- Labus K., 2005 — Possibilities of Compositional Data Analysis application in hydrogeology. *Proc. XII Narodni Hydrogeologicky Kongres. Česke Budejovice 19—22 IX 2005*.
- Martín-Fernández J.A., Palowsky-Glahn V., Barceló-Vidal C., Kovács L.Ó., Kovács G.P., 2003 — *Proceedings of CODAWORK'03, The First Compositional data Analysis Workshop*, Girona, Spain.
- Nordstrom D.K., Plummer L.N., Wigley T.M.L., Wolery T.J., Ball J.W., Jenne E.A., Bassett R.L., Crerar D.A., Florence T.M., Fritz B., Hoffman M., Holdren G.R., Jr., Lafon G.M., Matigod S.V., McDuff R.E., Morel F., Reddy M.M., Sposito G., and Thraillkill J., 1979 — A comparison of computerized chemical models for equilibrium calculations in aqueous systems: in *Chemical Modeling in aqueous systems, speciation, sorption, solubility, and kinetics*. Jenne, E.A., ed., Series 93, American Chemical Society, p. 857—892.
- Otero N., Tolosana-Delgado R., Soler A., 2003 — *Proceedings of CODAWORK'03. The First Compositional data Analysis Workshop*, Girona, Spain.
- Pawlowsky-Glahn V., Olea R.A., 2004 — *Geostatistical Analysis of Compositional Data*. Univ. Press, Oxford.
- Parkhurst D.L. and Appelo C.A.J., 1999 — User's guide to PHREEQC (Version 2) — A computer program for speciation, batch-reaction, one-dimensional transport, and inverse geochemical calculations: U.S. Geological Survey Water-Resources Investigations Report 99—4259, 310 p.
- Reyment R.A., Savazzi E., 1999 — *Aspects of Multivariate Statistical Analysis in Geology*. Elsevier.
- Tolosana-Delgado R., Palomera-Roman R., Gimeno-Torrente D., Pawlowsky-Glahn V., Thió-Henestrosa S., 2002 — A First Approach to Classification of Basalts Using Trace Elements, *Terra Nostra, Schriften der Alfred-Wegener-Stiftung 04/2002*, 8th Annual Conference of IAMG, pp. 435—440.

**IMPLEMENTATION OF COMPOSITIONAL DATA ANALYSIS INTO GEOLOGY****Key words**

Compositional Data Analysis, groudwater chemical composition, rock porosity

**Abstract**

The paper presents the principles of Compositional Data Analysis. This statistical method enables interpreting the data, being proportions of some whole. The two examples of implementation of this method were presented, with use of hydrogeological and petrographical data. The visaulisation by means of ternary diagrams is helpful for interpreting the results. The method is particularly featured by the possibility of applplication of a range of helpful manipulations, such as: amalgamation, perturbation, unconstraining and centering of data.