

WYKORZYSTANIE METOD ANALIZY SKUPIEŃ W INTERPRETACJI PIONOWYCH SONDOWAŃ ELEKTROOPOROWYCH (PSE)

UKD 550.837.311.053:519.2(084.3+084.21)(438-15Braszkowice)

Podstawową metodą analizy wyników pionowych sondowań elektrooporowych (PSE) jest interpretacja ilościowa. Dostarcza ona cennych informacji o budowie geologicznej badanego terenu w postaci przekroju geoelektrycznego, czyli podziału ośrodka skalnego na wiele warstw o określonych miąższościach i opornościach właściwych. Przy dużej liczbie punktów pomiarowych na danym terenie taka interpretacja wszystkich krzywych PSE po kolei jest zajęciem bardzo żmudnym i czasochłonnym. Tymczasem interpretację można znacznie ułatwić stosując pewne wstępne techniki opracowania pomiarów, polegające na urzeczonym porównaniu ze sobą wszystkich krzywych PSE. Wydziela się w ten sposób grupy krzywych o podobnym charakterze, dla których wyniki interpretacji będą podobne. Sama interpretacja – przeprowadzona takimi grupami – jest znacznie szybsza, łatwiejsza i pewniejsza. Ponadto uwidocznienie tych grup na mapie punktów pomiarowych daje obraz budowy geologicznej badanego obszaru (podobne sondowania odpowiadają terenom o podobnej budowie geologicznej). Takie wstępne opracowanie wykonane w terenie pozwala na wprowadzanie w trakcie badań pewnych zmian w pierwotnie założonym schemacie obserwacji.

Niektóre techniki grupowania krzywych PSE są znane od dawna i stosowane (4, 7). Należą do nich tzw. mapy typów krzywych – krzywe PSE dzieli się wizualnie ze względu na ich typ. Można tu też włączyć tzw. mapy izoomów, czyli mapy oporności pozornej wykreślane dla danego rozstawu układu pomiarowego. Z takich map można określić ogólne cechy budowy geologicznej badanego terenu, a następnie – interpretować osobno krzywe z obszarów o podobnej budowie geologicznej.

Ostatnio dla podziału PSE na grupy zaczęto stosować bardziej dokładne metody matematyczne. G.M. Habberjam (2, 3) zajmował się problemem określenia budowy geologicznej terenu prac elektrooporowych przy braku danych wiertniczych. Wprowadził w tym celu jako miarę podobieństwa krzywych oporności pozornej „parametr asocjacyjny”, określony wzorem:

$$d_{i,j} = \frac{1}{n+1} \sum_{k=1}^n [\lg \rho_k^{(i)} - \lg \rho_k^{(j)}]^2 \quad [1]$$

gdzie:

$d_{i,j}$ – wartość miary podobieństwa krzywej i z krzywą j ,
 n – liczba pomiarów oporności pozornej dla jednej krzywej,

$\rho_k^{(i)}$ – pomierzona oporność pozorna (k -ty pomiar w i -tej krzywej PSE).

G.M. Habberjam proponuje wykorzystanie wyników obliczeń parametru asocjacyjnego w formie „diagramu podobieństwa”, będącego graficznym przedstawieniem ma-

ciery podobieństwa zestawionej z wartością $d_{i,j}$. Takie podejście oznacza jednak stratę wielu informacji, uzyskanych przy wykonywaniu obliczeń i praktycznie jest możliwe tylko dla małej liczby sondowań.

Metodę Habberjama wykorzystali do oceny budowy geologiczno-tektonicznej terenu I. Landa, O. Mazac (5), używając zamiast parametru asocjacyjnego innego współczynnika – „powierzchni korelacyjnej”, czyli obliczanej planimetrem powierzchni między dwiema krzywymi PSE naniesionymi razem w jednym układzie współrzędnych bilogarytmicznych. Wyznaczenie powierzchni korelacyjnej jest prostsze i szybsze niż obliczenie parametru asocjacyjnego Habberjama.

W niniejszej pracy do podziału sondowań na grupy proponuje się bezpośrednie zastosowanie statystycznych metod analizy skupień (zwanych też metodami hierarchicznej analizy skupień lub metodami taksonomicznymi), dobrze znanych i szeroko stosowanych w naukach przyrodniczych – również w geologii (1). Liczba rozpatrywanych krzywych nie jest tu ograniczona, a podział na grupy jest dokonywany automatycznie. Wymagana jest jednak duża liczba obliczeń i w praktyce stosowanie metod analizy skupień wymaga użycia elektronicznych maszyn cyfrowych. Obliczenia dla prezentowanej pracy wykonano w Centrum Obliczeniowym Uniwersytetu Wrocławskiego, według programów napisanych przez autora.

KRÓTKI OPIS METOD ANALIZY SKUPIEŃ

Metody analizy skupień stosuje się do zespołu obiektów, z których każdy jest charakteryzowany wieloma cechami liczbowymi. W rozpatrywanym wypadku obiektami są krzywe PSE, a cechami – pomiary oporności pozornej dla różnych rozstawów układu pomiarowego. Obiekty (oraz ich grupy) podlegają łączeniu w pary, od szczebla najniższego, gdzie łączy się pojedyncze obiekty, do najwyższego, w którym wszystkie obiekty zostały już podzielone na dwie grupy łączone w ostatnią parę. Łączenie w pary odbywa się na podstawie wyliczonej uprzednio macierzy podobieństw (czyli zestawu miar podobieństw między poszczególnymi obiektami). Po przeprowadzeniu analizy skupień otrzymuje się hierarchiczną strukturę powiązań między obiektami, uwidacznianą graficznie w formie dendrogramu.

Obliczenia składają się z dwu etapów. W pierwszym etapie dla wszystkich par obiektów wylicza się pewien parametr (miarę) podobieństwa $d_{i,j}$. Zestaw wszystkich parametrów podobieństwa tworzy symetryczną macierz podobieństwa. W drugim etapie następuje połączenie obiektów – a potem ich grup – w grupy wyższego rzędu, tworząc w efekcie hierarchiczny system skupień. Istnieją

różne miary podobieństwa i różne sposoby łączenia. W niniejszej pracy rozpatrzono trzy różne miary i dwa sposoby łączenia, co w sumie daje sześć możliwości powstania różnych dendrogramów.

Jako miary podobieństwa wykorzystano następujące wielkości:

1) współczynnik korelacji

$$d_{i,j} = \frac{\text{cov}_{i,j}}{\sigma_i \sigma_j} = \frac{\sum_{k=1}^n \rho_{i,k} \rho_{j,k} - \frac{1}{n} \sum_{k=1}^n \rho_{i,k} \sum_{k=1}^n \rho_{j,k}}{\sqrt{\left[\sum_k \rho_{i,k}^2 - \frac{1}{n} \left(\sum_k \rho_{i,k} \right)^2 \right] \left[\sum_k \rho_{j,k}^2 - \frac{1}{n} \left(\sum_k \rho_{j,k} \right)^2 \right]}} \quad [2]$$

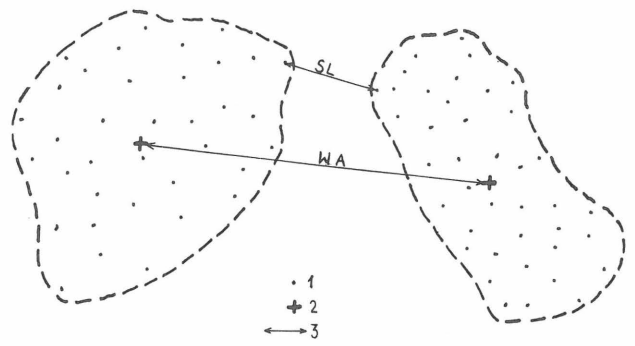
gdzie:

$\text{cov}_{i,j}$ – kowariancja między obiektami i i j ,
 σ_i – standardowe odchylenie pomiarów dla obiektu i ,

n – liczba cech w danym obiekcie,

$\rho_{i,k}$ – wartość cechy – oporności pozornej (k -ty pomiar w i -tym obiekcie);

2) cosinus kąta między obiektami traktowanymi jako wektory w przestrzeni n -wymiarowej

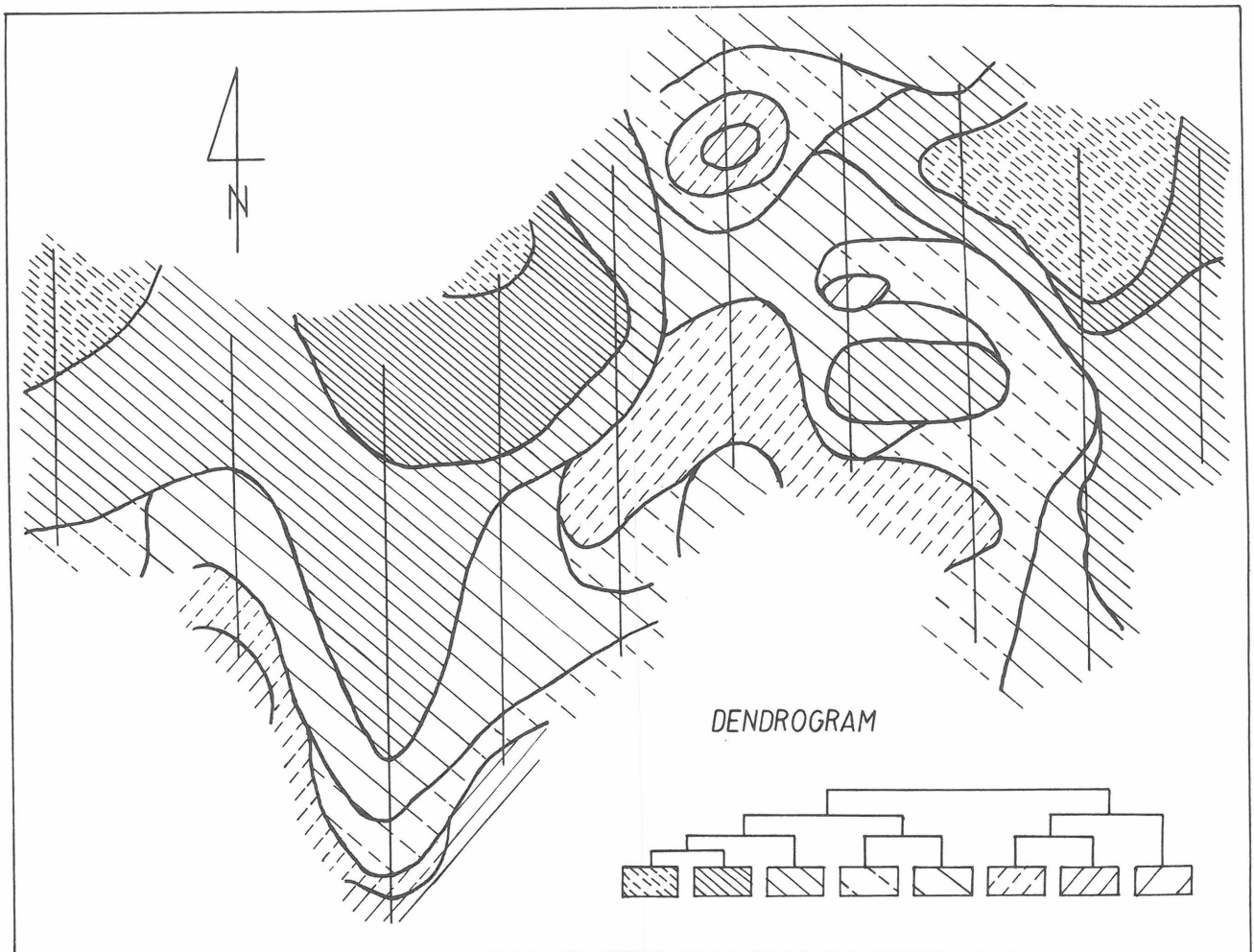


Ryc. 1. Ilustracja sposobu łączenia grup obiektów

1 – obiekty, 2 – obiekty zastępcze, 3 – sposoby łączenia, WA – sposób średnich ważonych, SL – sposób łączenia pojedynczego

Fig. 1. The mode of clustering of groups of objects

1 – objects, 2 – substitution objects, 3 – modes of clustering, WA – weighted average mode, SL – single linkage mode



2a

Ryc. 2. Mapy korelacyjne i dendrogramy zredukowane dla łączenia pojedynczego

— linia profilu PSE, a – odległość euklidesowa, b – cosinus kąta, c – współczynnik korelacji

Fig. 2. Correlation maps and reduced dendrograms for single joining mode

— line of electric resistance profiling, a – Euclid distance, b – cosine of angle, c – correlation coefficient

$$d_{i,j} = \frac{\sum_k \rho_{i,k} \rho_{j,k}}{\sqrt{\sum_k \rho_{i,k}^2 \sum_k \rho_{j,k}^2}} \quad [3]$$

3) odległość euklidesowa między obiektami traktowanymi jako punkty w przestrzeni n -wymiarowej

$$d_{i,j} = \sqrt{\sum_k (\rho_{i,k} - \rho_{j,k})^2} \quad [4]$$

Ten ostatni parametr jest najbliższy parametrowi asocjacyjnemu Habberjama, z tym że wartości oporności pozornej nie są tu zlogarytmizowane, a sama wartość nie jest normalizowana.

Po wyznaczeniu macierzy podobieństw zawierającej wszystkie wartości $d_{i,j}$ następuje kolejne łączenie obiektów i ich grup w struktury hierarchiczne. Dla pojedynczych obiektów problem sposobu łączenia nie istnieje – łączy się w pary obiekty o najwyższym stopniu podobieństwa (czyli maksymalnej wartości $d_{i,j}$ dla współczynnika korelacji i cosinusa kąta oraz minimalnej wartości dla odległości euklidesowej). Łączenie grup obiektów można natomiast realizować w różny sposób. W niniejszym opracowaniu wykorzystano dwa warianty, uwidocznione schematycznie na ryc. 1 (dla wypadku dwuwymiarowego i odległości euklidesowej):

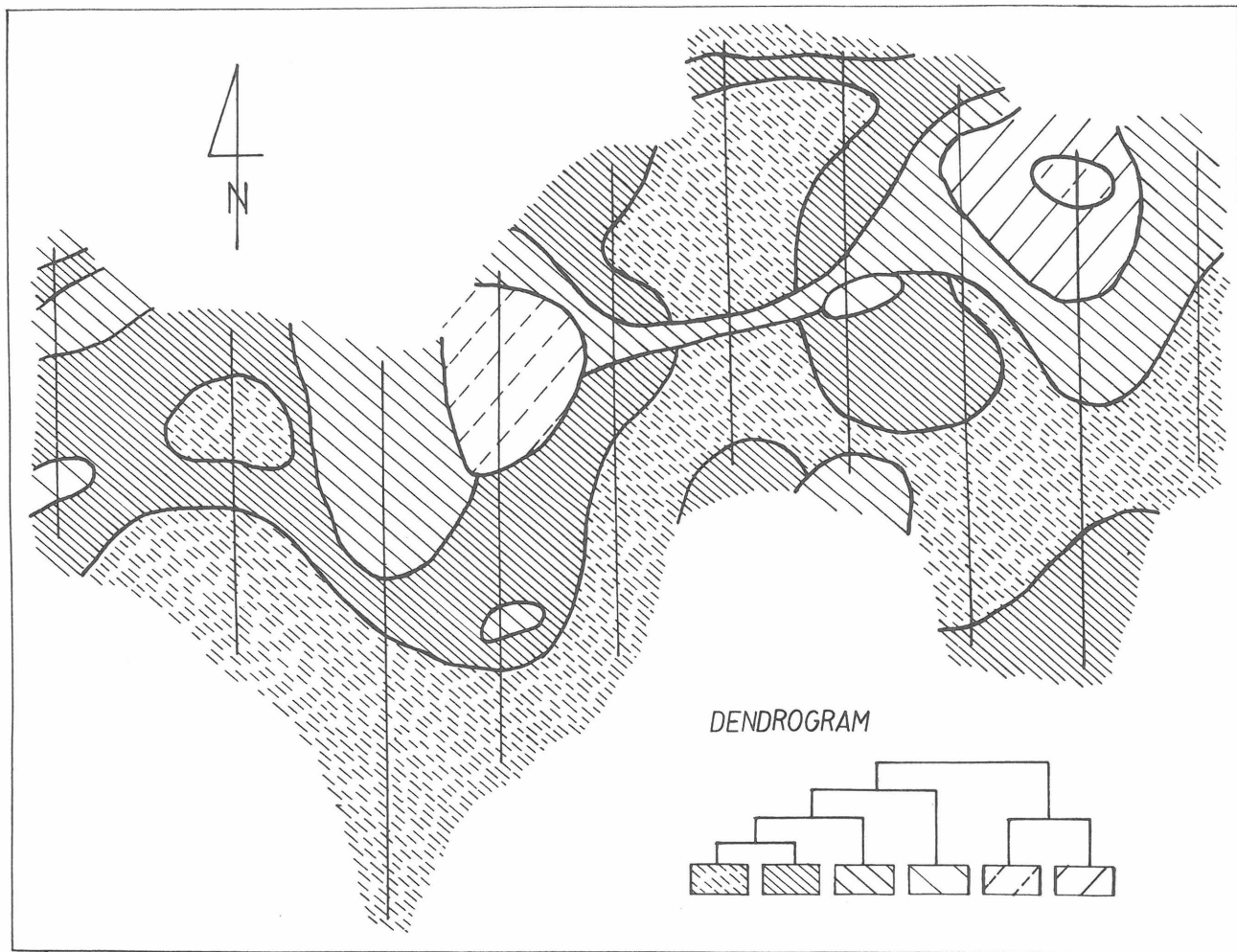
1) sposób średnich ważonych (weighted average – WA na ryc. 1). Dla każdego skupienia obiektów wylicza się średnie wartości ze wszystkich cech wszystkich obiektów w skupieniu. Tak wyliczone parametry traktuje się jako

cechy nowego obiektu, zastępującego całe skupienie. Podobieństwo między grupami wyliczane jest jako miara podobieństwa dwu takich obiektów zastępczych;

2) sposób łączenia pojedynczego (single linkage – SL na ryc. 1). Jako miarę podobieństwa dwu skupień przyjmuje się podobieństwo dwu najbliższych sobie (najbardziej podobnych) obiektów z różnych skupień. Dendrogramy, będące wynikiem graficznym tych działań, pokazano na ryc. 2a, b, c.

PRAKTYCZNE WYKORZYSTANIE METOD ANALIZY SKUPIEŃ

Wyniki działania opisanych metod sprawdzono praktycznie – na pomiarach PSE, wykonanych w okolicach Braszowic na Dolnym Śląsku (6). Wykonano tam 142 punkty pionowych sondowań elektrooporowych usytuowanych w 10 profilach. Wszystkie sondowania wykonano układem Schlumbergera o długości linii prądowej AB 130 m i składały się one z 13 pomiarów oporności pozornej. Rozmieszczenie profili pokazano na mapach korelacyjnych (ryc. 2a, b, c) i na mapie geologicznej (ryc. 3). Po wykonaniu obliczeń uzyskano sześć dendrogramów, których jednak nie było można przytoczyć w całości, ze względu na ich długość – dotyczą bowiem 142 obiektów. Przy interpretacji ilościowej wskazane jest korzystanie z całych dendrogramów, można w ten sposób znaleźć zespoły praktycznie identycznych krzywych, pochodzących często z różnych profili. Dla celów wstępnej oceny



2b

budowy geologicznej obszaru brano pod uwagę większe grupy sondowań, redukując dendrogramy do 6–8 grup obiektów (ryc. 2a, b, c). Taką liczbę grup można łatwo zobrazować w formie mapy korelacyjnej, nadającej się do interpretacji strukturalnej.

Mapy i dendrogramy dla łączenia pojedynczego przedstawiono na ryc. 2a (odległość euklidesowa), 2b (cosinus kąta) i 2c (współczynnik korelacji). Okazało się, że dla tej samej miary różne sposoby łączenia nie wpłynęły znacząco na zróżnicowanie dendrogramów. Różnice oczywiście istnieją, ale są bardzo małe w porównaniu z różnicami wynikającymi ze stosowania różnych miar podobieństwa. Dlatego też ograniczono się do podania map i dendrogramów tylko dla łączenia pojedynczego. Dendrogramy dla sposobu średnich ważonych zredukowane do 6–8 obiektów są praktycznie identyczne, ze względu na wykorzystanie tylko kilku najwyższych szczebli struktury hierarchicznej, a mapy korelacyjne różnią się co najwyżej dla pojedynczych punktów PSE.

Mapy dla różnych miar podobieństwa znacznie różnią się między sobą. Ogólny obraz strukturalny jest podobny i możliwy do skorelowania z mapą geologiczną (ryc. 3) i mapą izoomów tego obszaru (por. J. Peroń – 6, ryc. 3). Najlepsze wyniki daje zastosowanie jako miary podobieństwa odległości euklidesowej. Przy korzystaniu z cosinusa kąta i współczynnika korelacji otrzymuje się mapy mniej czytelne i trudniejsze do porównania z mapą geologiczną. Miary te dają też znacznie mniej wygodne w interpretacji dendrogramy schodkowe (szczególnie w wypadku współ-

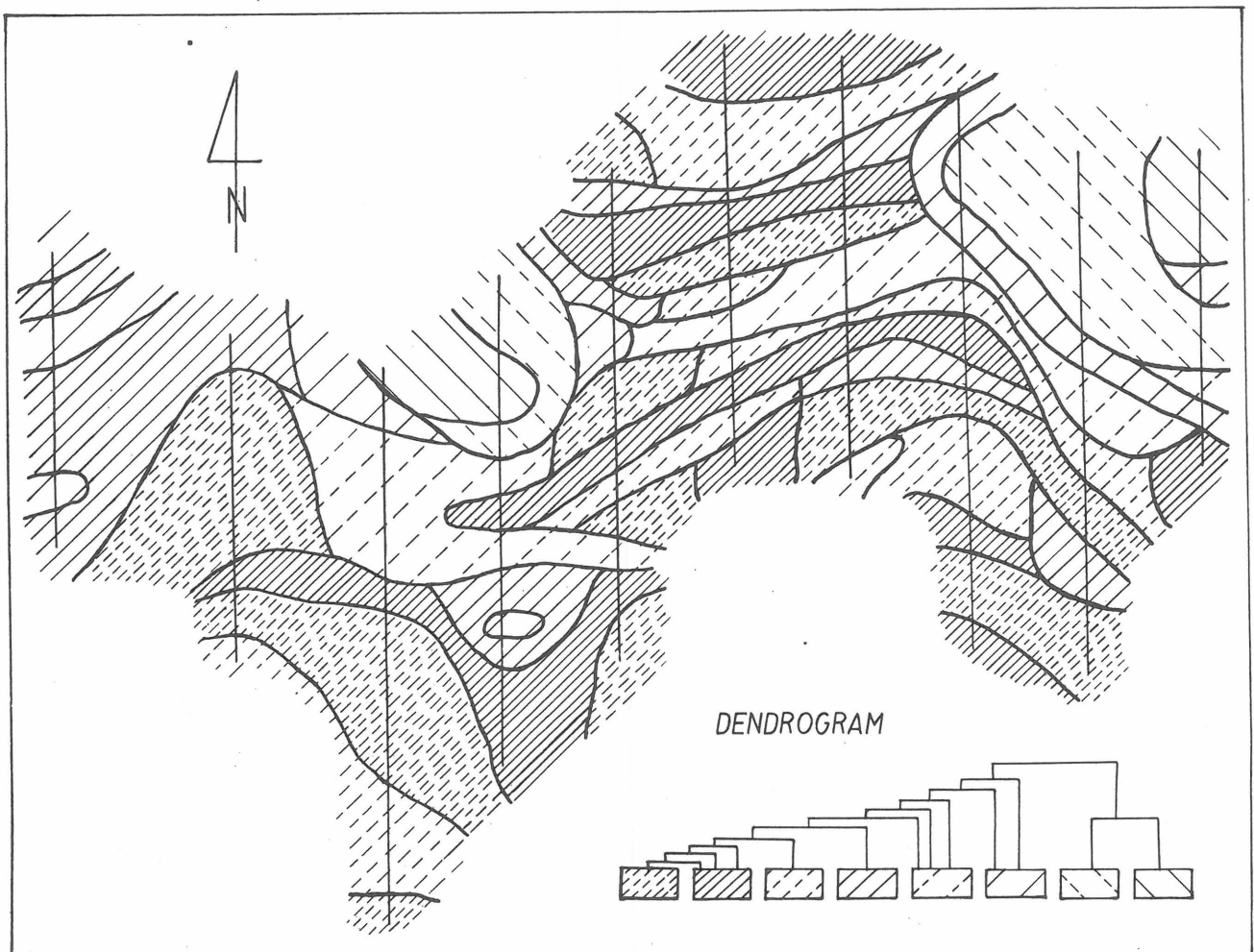
czynnika korelacji), co powoduje, że podział na grupy staje się sztuczny. Pożądane jest natomiast, aby powstawały skupienia o podobnej liczebności, łączone na zbliżonym poziomie. Temu warunkowi najlepiej odpowiada dendrogram uzyskany przy zastosowaniu odległości euklidesowej.

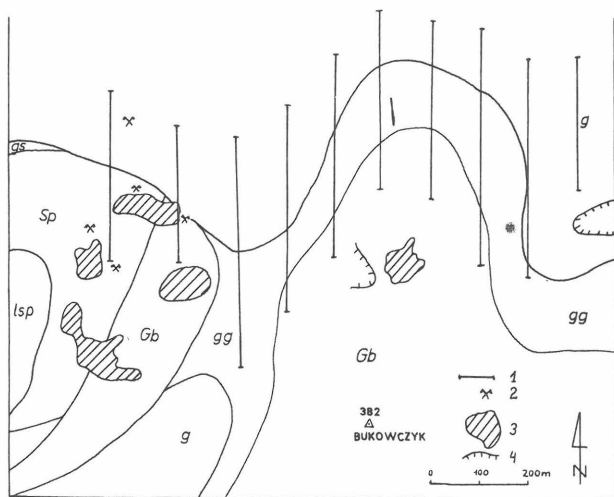
Wartości miar podobieństwa w dendrogramach przedstawiają się następująco: dla odległości euklidesowej wartość minimalna (czyli najwyższy osiągnięty stopień podobieństwa, nie uwidoczny na ryc. 2a) wynosi 6,65; ukazane na dendrogramie poziomy połączeń odpowiadają wartościom od 130 do 338. Dla współczynnika korelacji i cosinusa kąta zakres możliwych wartości wynosi od 1,0 (najwyższe podobieństwo) do $-1,0$. Wyliczone wartości maksymalnego podobieństwa w obu wypadkach wynoszą 1,000, a poziomy uwidocznione na dendrogramach z ryc. 2b i 2c odpowiadają wartościom od 0,969 do 0,869 dla cosinusa kąta i od 0,658 do $-0,0742$ dla współczynnika korelacji.

WNIOSKI

Wykorzystanie analizy skupień jako wstępnego etapu interpretacji krzywych PSE pozwala na:

- przeprowadzenie wstępnej oceny budowy geologiczno-tektonicznej badanego regionu nawet przy słabym rozpoznaniu wiertniczym,
- szybki i obiektywny podział krzywych na grupy według kształtu, co ułatwia i przyspiesza interpretację ilościową, szczególnie przy dużej ilości sondowań.





Ryc. 3. Mapa geologiczna obszaru

1 – linia profilu, 2 – otwory wiertnicze, 3 – hałdy, 4 – łomy i urwiska, Gb – gabbro, Sp – serpentynit, lsp – zwietrzelnina serpentynitu, gg – gliny deluwialne gabbro, gs – gliny deluwialne serpentynitów, g – gliny zwałowe

Fig. 3. Geological map of the studied area

1 – line of profile, 2 – boreholes, 3 – dumps, 4 – quarries and cleaves, Gb – gabbro, Sp – serpentinite, lsp – weathering cover of serpentinite, gg – gabbro deluvial loams, gs – serpentinite deluvial loams, g – tills

Przy wykorzystaniu trzech miar podobieństwa (współczynnik korelacji, cosinus kąta, odległość euklidesowa) i dwu sposobów łączenia (średnie ważone, łączenie pojedyncze) najlepsze rezultaty uzyskano przy zastosowaniu odległości euklidesowej z dowolnym sposobem łączenia.

Istnieje jeszcze wiele innych miar podobieństwa i sposobów łączenia, dlatego nie można mówić o końcowym rozwiązaniu problemu optymalnego doboru tych wielkości.

LITERATURA

1. Davis J.C. – Statistics and data analysis in geology. Wiley N.Y. 1973.
2. Habberjam G.M. – The association of resistivity soundings. Geophysical Prospecting 1970 vol. 18, no 2.
3. Habberjam G.M. – The comparison of sounding results and their interpretation in the absence of borehole control. Geoexploration sp. iss. 1976, vol. 14.
4. Jakubowski J.W., Liachow L.L. – Elektro-razwiedka. Izd. Nedra Moskwa 1974.
5. Landa I., Mazac O. – Improvement of structural and tectonic interpretation by correlation of the shape characteristics of vertical electrical sounding curve. Proc. 24 Geoph. Symp. Cracow 1979.
6. Peroń J. – Zastosowanie metody elektrooporowej do określenia budowy geologicznej rejonu Braszowic (strefa Niemczy). Acta Univ. Wratisl. 1979 nr 407.
7. Stenzel P., Szymanko J. – Metody geofizyczne w badaniach hydrogeologicznych i geologiczno-inżynierskich. Wyd. Geol. 1973.

SUMMARY

The quantitative interpretation is the basic method in analysis of results of vertical electric resistance profiling. However, this requires interpretation of each curve

separately which becomes laborious and time-consuming task when the number of curves is large. This procedure may be facilitated by using a technique of preliminary processing of measurement data by clustering curves into groups of curves similar in character. The interpretation of curves put into such groups is less time-consuming, easier and more reliable. Moreover, such groups when shown in a map tend to reflect geological-tectonic structure of an area. The hitherto used techniques involve compilation of maps of curve types (4, 7) and the use of calculations of similarity of curves ("association parameters" – Habberjam, 2, 3, and Landa and Mazac, 5).

The author proposes the use of statistic method – hierarchical analysis of clusters (Davis, 1), widely used in natural sciences. The method involves calculations of similarity measures for individual measurements and groups of measurements. In the present case, its final product is a dendrogram with hierarchical structure, well showing similarities and correlations between individual curves of electric resistance profiling. Correlation coefficient (formula 2), cosine of angle (formula 3) and Euclid distance (formula 4) were here used as similarity measures, and the modes of weighted means and single joining – in clustering measurements into groups (Fig. 1).

A set of 142 curves of electric resistance profiling, obtained for 10 profiles in the vicinities of Braszowice in the Lower Silesia (Peroń, 6), was used for testing the above method. Correlation maps were drawn with reference to subdivision of the profilings in accordance to relevant dendrograms (Figs. 2a, b, c) and results of computer calculations. Maps and dendrograms calculated for individual measures markedly differ from one another whereas those calculated for the same similarity measure are practically identical, despite of differences in the mode of clustering. That is why only maps and dendrograms obtained for single joining are shown here (Fig. 2).

Visual assessment of correlation maps and dendrograms and their comparison with geological map (Fig. 3) make it possible to state that the best results are obtained using Euclid distance as similarity measure. The results are the best both in the case of subdivision of electric resistance profiling curves within a dendrogram and comparison of correlation maps with the geological once.

РЕЗЮМЕ

Основным методом анализа результатов вертикальных электроопорных зондирований ПСЭ является количественная интерпретация. В случае большого количества кривых отдельная интерпретация каждой кривой является кропотливым и трудоёмким занятием. Для облегчения этой процедуры применяется предварительная техника обработки результатов измерений, которая состоит в составлении в группы кривых имеющих подобный характер. Интерпретация проведенная для групп кривых менее трудоёмкая, более простая и надёжная. С другой стороны представление этих групп на карте даёт изображение геологически-тектонического строения данного района. Применяемые до сих пор техники состоялись в составлении карт отдельных видов кривых (4), (7), а также в использовании вычислений сходства кривых („ассоциативные параметры" – Хабберьям (2), (3) и Ланда Мазак (5)).

Автор предлагает применение известного в естественных науках статистического метода – иерархического анализа накоплений (Давис, 1). Он заключается