

Stanisław WYBRANIEC

Statystyczno-informacyjna analiza wyników badań geofizycznych dla określenia perspektywiczności obszarów poszukiwań

WSTĘP

Przy poszukiwaniu surowców mineralnych coraz szerzej stosuje się badania geofizyczne. W wyniku przeprowadzenia badań powierzchniowych otrzymuje się obraz zmienności mierzonych parametrów. Zadaniem interpretatora jest ustalenie, jakim fragmentom obrazu geofizycznego przyporządkowane mogą być szukane ciała złożowe. W niektórych przypadkach zadanie to jest proste, np.: ciała o wysokiej pobudliwości magnetycznej dają wysokie anomalie pola magnetycznego; złoża masywne siarczków metali dają wybitne anomalie na obrazie oporności elektrycznej itd. Z reguły jednak związek między parametrami geofizycznymi a budową geologiczną nie jest prosty i jednoznaczny. W takich przypadkach potrzebna jest szczegółowsza analiza zależności parametrów geofizycznych od czynników geologicznych, wymagająca jednoczesnego wykorzystania wielu parametrów, czyli interpretacja kompleksowa.

Niżej przedstawiono jeden ze sposobów jednoczesnego wykorzystania wielu parametrów do oceny perspektywiczności obszaru przy poszukiwaniu surowców mineralnych. Jako przykład podano opracowanie wyników badań geoelektrycznych metodą pionowych sondowań elektrycznych, prowadzonych przy poszukiwaniu surowca kaolinowego w strefie wietrzenia masywów granitowych.

Analizując wyniki z rejonu Strzegom-Sobótka wykorzystano dane z interpretacji sondowań przeprowadzonej przez magistrów: Przemysława Stenzla, Roberta i Stanisławę Kucharskich oraz Sławomira Branieckiego z PPG. Mgr O. Gawrońskiemu, prowadzącemu prace poszukiwawcze w tym regionie, składam podziękowanie za pomoc przy korzystaniu z materiałów geologicznych.

ZASADA METODY

Znane są liczne przykłady zastosowania wieloparametrowej analizy do oceny perspektyw złożowych badanego obszaru. Jako typowe można podać prace: R. C. Batesa (1959), J. B. Wysokoostrowskiej i D. S. Zielenieckiego (1968), A. A. Nikitina (1970) oraz M. A. Ajzenberga, T. L. Zacharowej i in. (1972). Wszystkie stosowane sposoby analizy opierają się na metodach statystycznych. Poszczególni badacze przyjmują dla swoich obliczeń różne kryteria statystyczne. R. C. Bates (1959) zastosował następujący sposób. Dla każdego parametru wyróżniał kilka przedziałów wartości. Na podstawie analizy statystycznej istniejących danych geologicznych ustalał, który przedział dla poszczególnych parametrów jest charakterystyczny dla obszarów złożowych. Każdy punkt obszaru opisany został szeregiem wartości kolejnych parametrów. Im więcej wartości należy do wyznaczonych przedziałów charakterystycznych, tym bardziej perspektywiczny jest dany punkt. J. B. Wysokoostrowska i D. S. Zieleniecki (1968) wykorzystali metodę opartą na teorii informacji. Szersze jej omówienie podane będzie niżej. A. A. Nikitin (1970) zastosował współczynnik wiarygodności, a M. A. Ajzenberg (M. A. Ajzenberg, T. L. Zacharowa i in. 1972) wielowymiarową analizę regresji i korelacji. W większości metod za punkt wyjścia bierze się twierdzenia Bayesa dla obliczania prawdopodobieństwa *a posteriori*. Twierdzenie to odnosi się do następującego problemu (M. Fisz, 1969):

Jeżeli mamy zbiór (populację) A pewnych elementów scharakteryzowanych parametrem (zmienną losową) X w przedziałach wartości X_k , $k = 1, \dots, N$, a zbiór ten dzieli się na pewną ilość wzajemnie wykluczających się klas A_i , $i = 1, \dots, n$ (tzn. poszczególny element może należeć tylko do jednej z klas) i jeżeli mamy prawdopodobieństwa wystąpienia każdej klasy $P(A_i)$, czyli prawdopodobieństwa *a priori* oraz rozkłady prawdopodobieństwa dla każdej z klas (prawdopodobieństwa warunkowe) $P(X|A_i)$, to twierdzenie Bayesa pozwala określić prawdopodobieństwo *a posteriori* wystąpienia danej klasy $A_{i=j}$ $P(A_{i=j}|X_{k=l})$ przy pojawieniu się elementu o wartości parametru z przedziału $X_{k=l}$. Inaczej mówiąc prawdopodobieństwo *a posteriori* określa szansę przynależności rozpatrywanego elementu do wybranej klasy. Brzmi ono:

$$P(A_{i=j}|X_{k=l}) = \frac{P(A_{i=j}) \cdot P(X_{k=l}|A_{i=j})}{\sum_{i=1}^n P(A_i) \cdot P(X_{k=l}|A_i)} \quad (1)$$

dla dwóch klas jest wzór:

$$P(A_1|X_{k=l}) = \frac{P(A_1) P(X_{k=l}|A_1)}{P(A_1) P(X_{k=l}|A_1) + P(A_2) P(X_{k=l}|A_2)} \quad (2)$$

Ten przypadek jest najczęściej stosowany, przy czym klasa A_1 stanowi złożo surowca, a klasa A_2 — przestrzeń płonna. Wzór (2) można napisać też w postaci następującej:

$$P(A_1 | X_{k=1}) = \frac{P(A_1) P(X_{k=1} | A_1)}{P(A | X_{k=1})} \quad (3)$$

gdzie $P(A|X)$ — jest rozkładem prawdopodobieństwa parametru X dla całej populacji.

Ponieważ z reguły nie są znane prawdopodobieństwa *a priori* wystąpienia poszczególnych klas, tzn. nie wiadomo, jaką część badanego obszaru stanowić może złoże, przyjmuje się zwykle, że prawdopodobieństwo obu klas jest równe i wynosi 0,5. Wtedy wzory (2) i (3) przybierają postać:

$$P(A_1 | X_{k=1}) = \frac{P(X_{k=1} | A_1)}{P(X_{k=1} | A_1) + P(X_{k=1} | A_2)} \quad (4)$$

$$P(A_1 | X_{k=1}) = \frac{0,5 P(X_{k=1} | A_1)}{P(A | X_{k=1})} \quad (5)$$

Niektórzy badacze uważają, że przyjmowanie równości prawdopodobieństwa *a priori* dla obu (czy wielu) klas jest niesłuszne. Stanowisko takie zawiera dużo racji, ale ponieważ zwykle obliczone prawdopodobieństwa *a posteriori* traktuje się jakościowo, to przyjęcie wspomnianego założenia jest uzasadnione. Z wzoru (3) wynika, że prawdopodobieństwo *a priori* stanowi tylko stały mnożnik i jego wielkość nie wpłynie na obraz rozkładu powierzchniowego obliczonych prawdopodobieństw, czyli można go pominąć. Tak więc wzór ten w końcowej formie będzie brzmiał:

$$P(A_1 | X_{k=1}) = \frac{P(X_{k=1} | A_1)}{P(A | X_{k=1})} \quad (6a)$$

oraz analogicznie dla drugiej klasy:

$$P(A_2 | X_{k=1}) = \frac{P(X_{k=1} | A_2)}{P(A | X_{k=1})} \quad (6b)$$

Gdy przy korzystaniu z wzorów (4) i (5) uzyska się wartości prawdopodobieństwa $P(A_1 | X_{k=1})$ większe od 0,5, uważa się, że wskazuje to na obecność klasy A_1 , czyli złoża; wartości mniejsze mówią o obecności klasy A_2 , czyli o braku złoża. Pamiętać jednak należy, że są to dane orientacyjne i z reguły zawyżone z uwagi na nieuwzględnienie prawdziwych prawdopodobieństw *a priori*. Podobnie rzecz się ma przy korzystaniu z wzoru (6a).

Oprócz wyżej wymienionego sposobu stosuje się także nieco inne (A. A. Nikitin, 1970). Wykorzystuje się mianowicie pojęcie współczynnika wiarygodności $L(X)$, który określanany jest jako iloraz prawdopodobieństw *a posteriori* dwóch różnych klas:

$$L(X_{k=1}) = \frac{P(A_1 | X_{k=1})}{P(A_2 | X_{k=1})} = \frac{P(A_1) \cdot P(X_{k=1} | A_1)}{P(A_2) \cdot P(X_{k=1} | A_2)} \quad (7)$$

Przy założeniu równości prawdopodobieństw *a priori* otrzymuje się:

$$L(X_{k=i}) = \frac{P(X_{k=i} | A_1)}{P(X_{k=i} | A_2)} \quad (8)$$

Gdy współczynnik wiarygodności jest większy od 1, to przyjmuje się, że element należy do klasy A_1 , a gdy jest mniejszy od 1 — to do klasy A_2 .

Jeżeli element charakteryzowany jest przez szereg parametrów X_1, X_2, \dots, X_n , i jeżeli parametry te są zmiennymi losowymi niezależnymi, to wtedy

$$P(\vec{X} | A_1) = P(X_1 | A_1) \cdot P(X_2 | A_1) \cdot \dots \cdot P(X_n | A_1) \quad (9)$$

oraz

$$P(\vec{X} | A_2) = P(X_1 | A_2) \cdot P(X_2 | A_2) \cdot \dots \cdot P(X_n | A_2) \quad (10)$$

Stąd

$$L(\vec{X}) = \frac{P(X_1 | A_1) \cdot P(X_2 | A_1) \cdot \dots \cdot P(X_n | A_1)}{P(X_1 | A_2) \cdot P(X_2 | A_2) \cdot \dots \cdot P(X_n | A_2)} = L(X_1) \cdot L(X_2) \cdot \dots \cdot L(X_n) \quad (11)$$

gdzie $\vec{X} = (X_1, X_2, \dots, X_n)$ wektor losowy parametrów (dla uproszczenia opuszczono dalsze wskaźniki dolne przy parametrach).

Prawdopodobieństwo *a posteriori* istnienia klasy A_1 wylicza się ze wzoru:

$$P(A_1 | \vec{X}) = \frac{L(\vec{X})}{L(\vec{X}) + 1} \quad (12)$$

Przy czym, podobnie jak poprzednio, prawdopodobieństwo większe od 0,5 przemawia za klasą A_1 , a mniejsze od 0,5 — za klasą A_2 . J. B. Wysokostrowska i D. S. Zieleniecki (1968) do obliczenia wskaźników perspektywiczności obszaru wykorzystują pojęcia z teorii informacji. Określają oni ilość informacji zawartą w pewnym przedziale wartości parametru X dla klasy A_1 w porównaniu z całą populacją w postaci:

$$I_{A_1 \rightarrow A} = \log \frac{P(X_{k=i} | A_1)}{P(A | X_{k=i})} \quad (13)$$

Jeżeli wzór (13) porówna się ze wzorem (6), to można zauważyć, że część logarytmowana jest prawdopodobieństwem *a posteriori* wystąpienia klasy A_1 w oparciu o rozkład całej populacji. Logarytm współczynnika wiarygodności autorzy nazywają przyrostem informacji zawartym w przedziale $X_{k=i}$ klasy A_1 w stosunku do klasy A_2 .

$$\Delta I_{A_1/A_2} = \log \frac{P(X_{k=i} | A_1)}{P(X_{k=i} | A_2)} \quad (14)$$

W przypadku, gdy rozporządza się wieloma niezależnymi parametrami, całą informację dotyczącą wystąpienia klasy A_1 przy pojawieniu się danego elementu otrzyma się przez sumowanie informacji dla poszczególnych parametrów. Jak wynika z cytowanej pracy, takie ujęcie ilości

informacji stwarza pewne kłopoty przy obliczeniach, dlatego w niniejszym artykule proponuje się inny wariant obliczania wskaźników perspektywiczności.

Podstawowy wzór C. E. Shannona na średnią ilość informacji brzmi (vide M. Mazur, 1970).

$$H = - \sum P(X_i) \cdot \log_2 P(X_i) \quad (15)$$

gdzie: H — średnia ilość informacji wyrażona w bitach,
 $P(X_i)$ — prawdopodobieństwo zdarzenia X_i ,
 \log_2 — logarytm o podstawie 2.

W dalszym ciągu stosowana będzie terminologia wprowadzona przez M. Mazura (1970), który H określa jako średnią liczbę informacji identyfikujących. W naszym przypadku wzór (15) dla klasy A_1 będzie miał postać.

$$H_{X_k|A_1} = - \sum_{k=1}^N P(X_k|A_1) \cdot \log_2 P(X_k|A_1) \quad (16)$$

gdzie: N — ilość przedziałów wartości parametru X .

Liczbę informacji identyfikujących dla przedziału $X_{k=l}$ klasy A_1 można przedstawić jako:

$$H_{X_{k=l}|A_1} = - P(X_{k=l}|A_1) \cdot \log_2 P(X_{k=l}|A_1) \quad (17)$$

Analogicznie dla przedziału $X_{k=l}$ z całej populacji wzór (17) ulegnie modyfikacji i uzyska postać następującą:

$$H_{X_{k=l}|A} = - P(A|X_{k=l}) \cdot \log_2 P(A|X_{k=l}) \quad (18)$$

Liczbę informacji identyfikujących klasę A_1 w odniesieniu do całej populacji A dla parametru X można przedstawić w postaci:

$$H_{X(A_1|A)} = \sum_{k=1}^N (H_{X_k|A_1} - H_{X_k|A}) \quad (19)$$

gdzie: N — ilość przedziałów wartości parametru X .

Dla poszczególnego przedziału $X_{k=l}$ jest więc

$$H_{X_{k=l}(A_1|A)} = H_{X_{k=l}|A_1} - H_{X_{k=l}|A} \quad (20)$$

Wielkość ta będzie w skrócie nazywana informacją identyfikującą. Suma informacji identyfikujących klasę A_1 wyrażona wzorem (19) mówi o informacyjności rozpatrywanego parametru w odniesieniu do klasy A_1 (sumuje się tylko wartości dodatnie).

Informację identyfikującą można obliczyć także w stosunku do innej klasy danej populacji, np. A_2 , określając ją jako różnicę między informacjami identyfikującymi dla dwóch rozpatrywanych klas A_1 i A_2 w odniesieniu do całej populacji A .

$$H_{X_{k=l}(A_1|A_2)} = (H_{X_{k=l}|A_1} - H_{X_{k=l}|A}) - (H_{X_{k=l}|A_2} - H_{X_{k=l}|A}) = H_{X_{k=l}|A_1} - H_{X_{k=l}|A_2} \quad (21)$$

Jak więc widać, w rezultacie otrzymuje się wzór analogiczny do wzoru (20).

Jeżeli rozpatrywane klasy scharakteryzowane są przez kilka parametrów, a parametry te są zmiennymi losowymi niezależnymi, to informacje identyfikujące otrzymane dla różnych parametrów można sumować. W praktyce jednak spotykamy się na ogół z parametrami skorelowanymi, czyli zależnymi. Dla celów orientacyjnych i w tym przypadku sumuje się informacje identyfikujące, co też uczyniono w przykładzie podanym niżej. Pamiętać należy o tym, że w ten sposób wynik końcowy będzie obciążony pewnym błędem, tym większym im ściślejsza będzie korelacja między parametrami.

Z wzorów (19), (20) i (21) wynika, że jeżeli rozkłady prawdopodobieństw klas będą takie same, to informacja identyfikująca będzie równa zero. Wartości dodatnie mówią o występowaniu klasy A_1 , a wartości ujemne w pierwszym przypadku (wzór 20) wskazują na brak klasy A_1 , a w drugim (wzór 21) — klasę A_2 .

PRZYKŁAD ZASTOSOWANIA

Opisaną metodę opartą na informacjach identyfikujących zastosowano do oceny perspektywności rejonu Strzegom — Sobótka pod kątem występowania surowca kaolinowego. Korzystano z parametrów uzyskanych z badań geoelektrycznych metodą pionowych sondowań elektrycznych. Na omawianym obszarze istnieje intruzja granitowa przykryta osadami trzecio- i czwartorzędu. Powierzchniowe partie granitu ulegały w trzeciorzędzie kaolinizacji. Strefy kaolinizacji nie zachowały się wszędzie, a i stopień kaolinizacji nie był zapewne jednakowy na całym obszarze. Powierzchniowe zdjęcie geoelektryczne metodą sondowań miało za zadanie określić morfologię niezwiertzałego podłoża granitowego oraz pomóc przy poszukiwaniu stref kaolinizacji. Jakkolwiek pierwszy problem został rozwiązany bez większych trudności, to jednak nie udało się jednoznacznie wyodrębnić stref kaolinów. Wynika to z tego, że oporności rzeczywiste kaolinów nie różnią się od oporności niektórych innych skał tego obszaru. Mimo to wyniki badań geoelektrycznych okazały się bardzo przydatne. Znalaziono tu i udokumentowano kilka złóż kaolinów. Jednak duża część obszaru nie została jeszcze przebadana i dlatego wydawało się celowe przedstawić ocenę perspektywności całego obszaru opartą na dotychczasowych wynikach.

Klasę A_1 (czyli strefę skaolinizowaną) reprezentowały dane z 57 wierceń i z sondowań wykonanych w ich bezpośrednim sąsiedztwie. Całą populację stanowiło 860 sondowań, którymi pokryto obszar badań. Analizowano trzy parametry: głębokość załęgania stropu podłoża wysokooporowego, oporność rzeczywistą warstwy leżącej bezpośrednio nad podłożem wysokooporowym oraz przewodnictwo podłużne całego nadkładu. Parametry te są skorelowane. Np. współczynnik korelacji między przewodnictwem podłużnym nadkładu i głębokością podłoża wynosi 0,60 z odchyleniem standardowym 0,079, a między przewodnictwem podłużnym nadkładu i opornością rzeczywistą strefy kaolinowej — 0,52 z odchyleniem standardowym 0,081. Współczynniki te nie okazały się na tyle duże, by zniekształcić otrzymane wyniki w sposób wyraźny.

Na rysunku 1 przedstawiono histogramy rozkładów poszczególnych parametrów dla otworów stwierdzających strefę kaolinizacji oraz dla

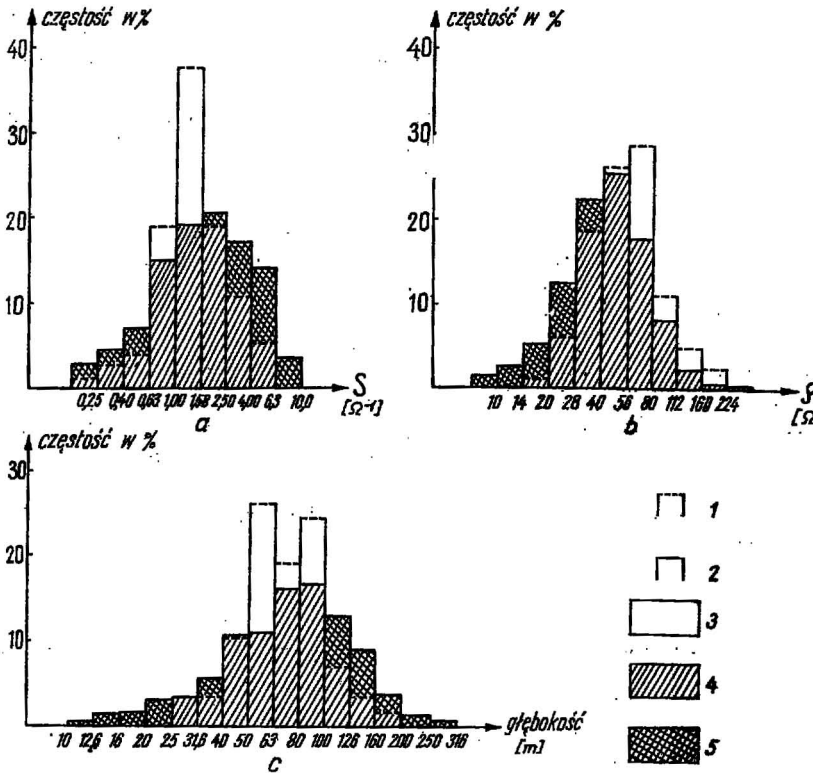


Fig. 1. Histogramy rozkładów parametrów geoelektrycznych: a — przewodnictwa podłużnego nadkładu (S); b — oporności rzeczywistej warstwy leżącej na podłożu (ρ_{n-1}); c — głębokości zalegania podłoża wysokooporowego (H_n)

Distribution histograms of geoelectric parameters: a — longitudinal conductance of the cover (S); b — true resistivity of the layer directly overlying the basement (ρ_{n-1}); c — depth to the high resistivity basement (H_n)

1 — dla rejonów ze strefami kaolinizacji (klasa A_1); 2 — dla całości obszaru (populacja A); 3 — różnica częstości między klasą A_1 a populacją A większa od zera; 4 — częstości dla klasy A_1 , jeśli są one wyższe od częstości dla A oraz częstości dla A , jeśli są one wyższe od częstości dla A_1 ; 5 — różnica częstości między klasą A_1 a populacją A mniejsza od zera

1 — for areas with kaolinization zones (A_1 class); 2 — for the whole area (population A); 3 — difference between frequencies for A_1 class and those of population A , higher than zero; 4 — frequencies for A_1 class if these are higher than those of A or frequencies for A if these are higher than those of A_1 ; 5 — difference between frequencies for A_1 class and those of population A , lower than zero

całego obszaru. W tabelach 1, 2 i 3 zilustrowano sposób obliczania informacji identyfikującej klasę A_1 (strefa kaolinizacji) dla poszczególnych przedziałów kolejnych parametrów. Za prawdopodobieństwa wzięto częstości dla danych przedziałów. Granice przedziałów przyjęto w skali logarytmicznej, gdyż zauważono, że rozkład parametrów zbliżony jest do rozkładu logarytmiczno-normalnego. Poza tym błąd w określaniu poszczególnych parametrów także rośnie zgodnie ze skalą logarytmiczną. Jak

Tabela 1

Obliczanie informacji identyfikującej klasę A_1 według przewodnictwa podłużnego nadkładu S

| Przedziały X_k (w Ω^{-1}) | Prawdop. (częstość) $P(X A_1)$ | Liczba inf. identyfik. $H_{X A_1}$ | Prawdop. (częstość) $P(A X)$ | Liczba inf. identyfik. $H_{X A}$ | Inform. identyfik. $H_{X(A_1 A)}$ |
|--|--------------------------------------|--|------------------------------------|--|---|
| —0,25 | 0,014 | 0,08 | 0,030 | 0,15 | —0,07 |
| 0,25—0,40 | 0,027 | 0,14 | 0,046 | 0,20 | —0,06 |
| 0,40—0,63 | 0,040 | 0,19 | 0,071 | 0,27 | —0,08 |
| 0,63—1,00 | 0,189 | 0,45 | 0,151 | 0,41 | +0,04 |
| 1,00—1,58 | 0,378 | 0,53 | 0,192 | 0,46 | +0,07 |
| 1,58—2,50 | 0,189 | 0,45 | 0,207 | 0,47 | —0,02 |
| 2,50—4,00 | 0,108 | 0,35 | 0,173 | 0,44 | —0,09 |
| 4,00—6,30 | 0,055 | 0,23 | 0,092 | 0,31 | —0,08 |
| 6,30—10,00 | 0,000 | 0,00 | 0,038 | 0,18 | —0,18 |

Liczebność klasy $A_1: N_{A_1} = 74$; liczebność całej populacji — $N_A = 860$; informacyjność parametru — $\Sigma H_X = +0,11$

Tabela 2

Obliczanie informacji identyfikującej klasę A_1 według oporności rzeczywistej warstwy leżącej na podłożu (ρ_{n-1})

| Przedziały X_k , (w Ωm) | Prawdop. (częstość) $P(X A_1)$ | Liczba inf. identyfik. $H_{X A_1}$ | Prawdop. (częstość) $P(A X)$ | Liczba identyfik. $H_{X A}$ | Inform. identyfik. $H_{X(A_1 A)}$ |
|---------------------------------------|--------------------------------------|--|------------------------------------|-----------------------------------|---|
| —10 | 0,000 | 0,00 | 0,017 | 0,11 | —0,11 |
| 10—14 | 0,000 | 0,00 | 0,029 | 0,15 | —0,15 |
| 14—20 | 0,012 | 0,07 | 0,055 | 0,23 | —0,15 |
| 20—28 | 0,062 | 0,24 | 0,126 | 0,38 | —0,14 |
| 28—40 | 0,188 | 0,45 | 0,224 | 0,48 | —0,03 |
| 40—56 | 0,262 | 0,51 | 0,256 | 0,50 | +0,01 |
| 56—80 | 0,288 | 0,52 | 0,178 | 0,44 | +0,08 |
| 80—112 | 0,112 | 0,36 | 0,082 | 0,29 | +0,07 |
| 112—160 | 0,050 | 0,22 | 0,024 | 0,13 | +0,08 |
| 160—224 | 0,025 | 0,13 | 0,008 | 0,06 | +0,07 |
| >224 | 0,000 | 0,00 | 0,002 | 0,02 | —0,02 |

$N_{A_1} = 80$; $N_A = 876$; $\Sigma H_X = +0,31$

widać z tablicy, informacyjność parametrów jest różna. Największą informacyjność wykazuje oporność rzeczywista warstwy nad podłożem — +0,31 bity, na drugim miejscu stoi parametr głębokości podłoża — +0,23 bity, wreszcie przewodnictwo podłużne nadkładu — +0,11 bity.

Dalszy tok postępowania wygląda następująco: Rozpatrując kolejne sondowanie z danych interpretacyjnych wybiera się trzy wspomniane parametry. Na podstawie tablic 1, 2 i 3 ustala się, jakie ilości informacji

Tabela 3

Obliczanie informacji identyfikującej klasę A_1 według głębokości zalegania podłoża wysokooporowego (H_n)

| Przedziały X_k , (w m) | Prawdop. (częstość) $P(X A_1)$ | Liczba inf. identyfik. $H_{X A_1}$ | Prawdop. (częstość) $P(A X)$ | Liczba inf. identyfik. $H_{X A}$ | Inform. identyfik. $H_{X(A_1 A)}$ |
|-----------------------------|--------------------------------------|--|------------------------------------|--|---|
| 10—12,6 | 0,000 | 0,00 | 0,006 | 0,04 | -0,04 |
| 12,6—16 | 0,000 | 0,00 | 0,014 | 0,09 | -0,09 |
| 16—20 | 0,000 | 0,00 | 0,017 | 0,10 | -0,10 |
| 20—25 | 0,000 | 0,00 | 0,032 | 0,16 | -0,16 |
| 25—31,6 | 0,035 | 0,17 | 0,034 | 0,17 | 0,00 |
| 31,6—40 | 0,035 | 0,17 | 0,058 | 0,24 | -0,07 |
| 40—50 | 0,105 | 0,34 | 0,107 | 0,34 | 00,0 |
| 50—63 | 0,263 | 0,51 | 0,111 | 0,35 | +0,16 |
| 63—80 | 0,193 | 0,46 | 0,163 | 0,43 | +0,03 |
| 80—100 | 0,246 | 0,50 | 0,168 | 0,43 | +0,07 |
| 100—126 | 0,070 | 0,27 | 0,131 | 0,38 | -0,11 |
| 126—160 | 0,035 | 0,17 | 0,090 | 0,31 | -0,14 |
| 160—200 | 0,018 | 0,10 | 0,037 | 0,18 | -0,08 |
| 200—250 | 0,000 | 0,00 | 0,014 | 0,09 | -0,09 |
| 250—316 | 0,000 | 0,00 | 0,007 | 0,05 | -0,05 |

$$N_{A_1} = 57; N_A = 949; \Sigma H_X = +0,24$$

identyfikującej odpowiadają przedziałom wartości, do których należą dane z sondowania. Następnie sumuje się poszczególne ilości informacji z uwzględnieniem znaku. Wartości sumaryczne informacji nanosi się na mapę i rysuje się izolinie ilości informacji identyfikującej klasę A_1 . Powierzchnie o najwyższych wartościach dodatnich są najbardziej perspektywiczne pod kątem szukanego surowca, w naszym przypadku kaolinu. Na fig. 2 przedstawiono wynik takiej procedury. Dla uproszczenia narysowano tylko dwie izolinie: zerową i o wartości $+0,15$ bita. Podano także elementy budowy: wychodnie skał krystalicznych, dyslokacje zlokalizowane badaniami geoelektrycznymi oraz znane wystąpienie surowca kaolinowego. Te ostatnie grupują się wokół maksymalnych wartości ilości informacji identyfikującej.

Dla dokładniejszej oceny przydatności metody przeanalizowano ilości informacji identyfikującej dla 57 wyjściowych wierceń napotykalających strefę kaolinizacji oraz dla 17 wierceń, które jej nie nawierciły. Na fig. 3 przedstawiono histogramy rozkładu ilości informacji identyfikującej dla poszczególnych parametrów i dla sumarycznej ilości informacji dla trzech parametrów. Z tych obliczeń wynika, że przy zastosowaniu proponowanej metody błąd uznania strefy kaolinowej za płonną wynosi 29,8%, a błąd uznania obszaru płonnego za złożowy — 23,5%. Jak więc widzimy, błędy te są do siebie zbliżone. Przy rozpatrywaniu oddzielnych parametrów są one nierówne i większe. Z fig. 3 wynika też, jakie korzyści daje jednoczesna analiza kilku parametrów. Analizując histogramy 3 a, b, c, obra-

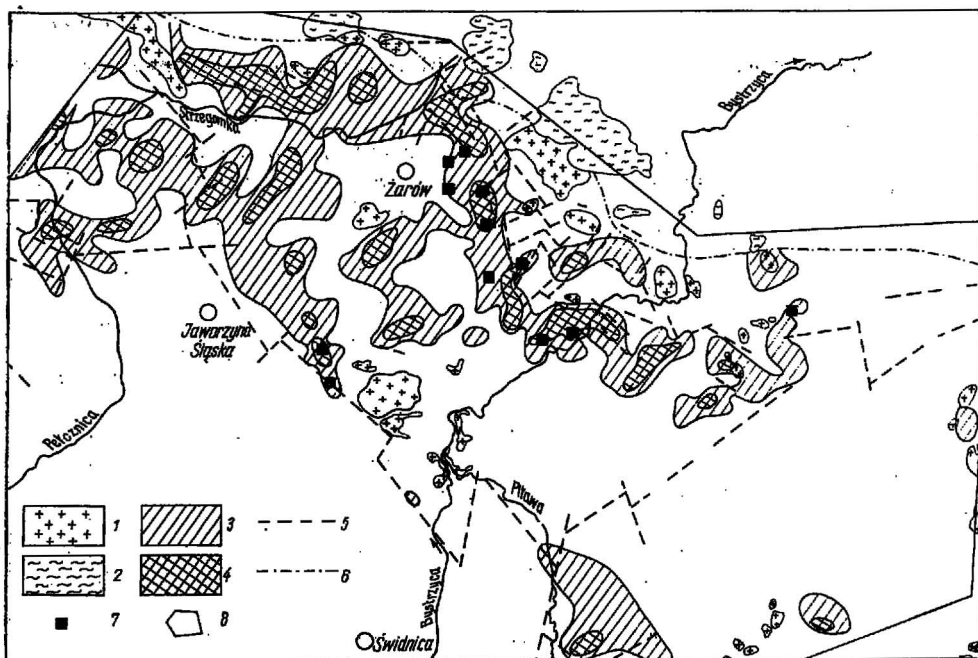


Fig. 2. Mapa perspektywiczności obszaru pod kątem poszukiwania stref kaolinizacji na podstawie informacji identyfikujących zawartych w parametrach geoelektrycznych

Prospect map of kaolinization zones occurrence constructed on the basis of the identifying information from geoelectrical parameters

1 — granit; 2 — skały okrywy metamorficznej; 3 — obszary perspektywiczne o wartości informacji identyfikującej od 0 do +0,15 bita; 4 — obszary wysoko perspektywiczne o wartości informacji identyfikującej powyżej +0,15 bita; 5 — uskoki przesłedzone przy pomocy profilowania elektrycznego; 6 — granica między granitami a okrywą metamorficzną; 7 — wystąpienia surowca kaolinowego; 8 — obszar pokryty zdjęciem geoelektrycznym

1 — granite; 2 — metamorphic rocks of the cover; 3 — prospect areas with the value of identifying information between 0 and +0,15 bits; 4 — highly prospect areas with the value of identifying information above +0,15 bits; 5 — dislocations localised by electrical profiling; 6 — boundary between granites and metamorphic cover; 7 — known kaoline deposits; 8 — geoelectrical survey area

zujące informacje identyfikujące klasę A_1 na podstawie jednego parametru, można zauważyć, że z wyjątkiem histogramu 3 b wszystkie przedziały wartości spotyka się zarówno w przypadku stref skaolinizowanych, jak i płonnych. Należy tylko dodać, że wartości dodatnie częstsze są dla stref skaolinizowanych, a wartości ujemne — dla płonnych. Tylko w przypadku drugiego parametru (H_n) poniżej wartości $-0,10$ bita istnieje pewność, że nie natrafi się na strefę kaolinizacji. Taki przypadek, gdy te same wartości obrazują nam dwie różne klasy, bywa nazywany (M. Mazur, 1970) pseudoinformowaniem, a informacje tego typu — pseudoinformacjami.

Na histogramie 3 d obraz jest inny. Tutaj oprócz pseudoinformacji występują także informacje jednoznaczne, czyli transinformacje. I tak przedziały powyżej $+0,10$ bita są charakterystyczne tylko dla klasy A_1 , czyli stref kaolinizacji, a przedziały poniżej $-0,20$ bita są reprezentowane prawie wyłącznie przez strefy płonne.

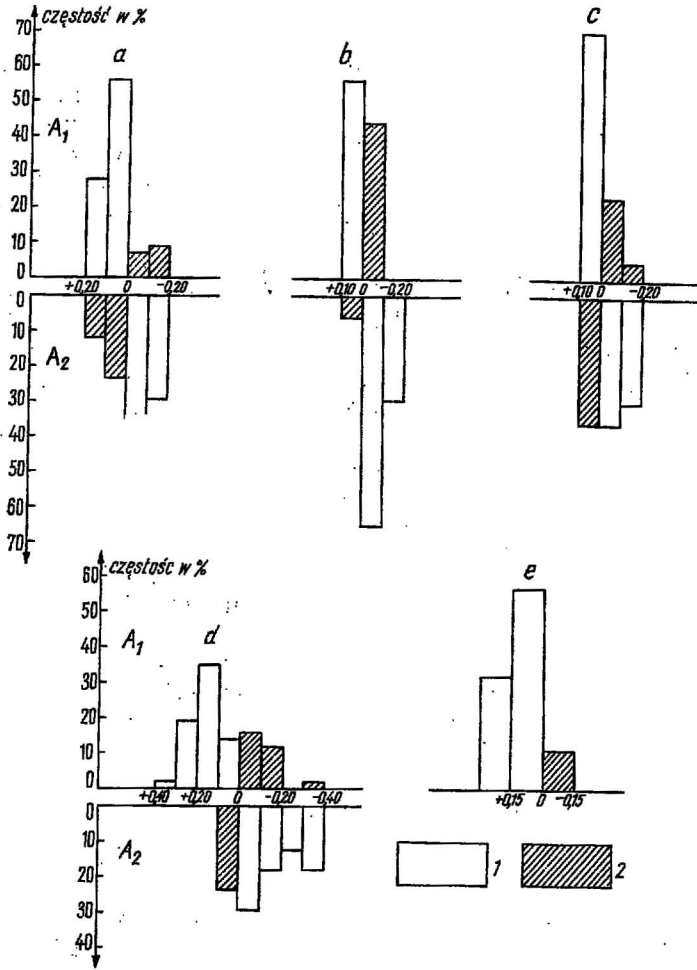


Fig. 3. Histogramy rozkładu informacji identyfikującej klasę A_1 dla obszarów pozytywnych (A_1) i negatywnych (A_2) computed from: a — według głębokości od stropu podłoża (H_n); b — według przewodnictwa podłużnego nadkładu (S); c — według oporności rzeczywistej warstwy leżącej na podłożu (ρ_{n-1}); d — według trzech parametrów łącznie; e — według wszystkich parametrów dla złóż udokumentowanych

Distribution histograms of information identifying the A_1 class for positive areas (A_1) and negative areas (A_2) computed from: a — depth to the high resistivity basement (H_n); b — longitudinal conductance of the cover (S); c — true resistivity of the layer directly overlying the basement, (ρ_{n-1}); d — all three parameters combined; e — all three parameters for known kaoline deposits

1 — informacje identyfikujące prawdziwe; 2 — informacje identyfikujące błędne; skala pozioma histogramów w bitach

1 — true identifying information; 2 — false identifying information; horizontal scale for histograms in bits

Stąd wniosek, że dzięki jednoczesnemu wykorzystaniu kilku parametrów z pseudoinformacji, czyli informacji dwuznacznych lub wieloznacznych otrzymuje się informacje jednoznaczne, czyli transinformacje.

Dlatego też z dużą dozą pewności można powiedzieć, że obszary o informacji identyfikującej powyżej $+0,15$ bita (przedstawione na fig. 2) są strefami kaolinizacji. Należy sobie jednak zdawać sprawę z tego, że rozważania powyższe oparte są na próbach stosunkowo mało licznych, w związku z czym wnioski nie mogą być zbyt kategoryczne.

Zastosowano jeszcze jeden sprawdzian. Wzięto pod uwagę znane i udokumentowane złoża kaolinu i obliczono dla nich rozkład sumarycznej ilości informacji identyfikującej. Tu trzeba dodać, że jako punkt wyjścia do wyboru 57 wierceń wyjściowych, czyli kryterium oceny czy dany otwór wiertniczy zaliczyć do pozytywnych, czy negatywnych, brano obecność kaolinu pierwotnego, tj. strefy skaolinizowanej bez względu na przydatność przemysłową tego kaolinu. Przy analizie informacji sumarycznej dla złóż kaolinu przydatnego dla ceramiki szlachetnej otrzymano histogram przedstawiony na fig. 3e. Częstość przedstawia tutaj procent powierzchni złoża objętej danym przedziałem wartości informacji sumarycznej. Jak widać, błąd wynosi tutaj tylko 11 procent.

Wynika stąd, że odkrycia nowych złóż można się spodziewać w pobliżu maksymalnych wartości informacji sumarycznej. Część tych obszarów została już przebadana. W szeregu przypadków napotkano na kaolin nieprzydatny dla celów przemysłowych, jednak niektóre partie czekają jeszcze na przebadanie. Obserwuje się wyraźny związek między obszarami perspektywicznymi a systemami uskoków przesłedzonymi przy pomocy profilowania elektrycznego.

WNIOSKI

1. Przedstawiona metoda wieloparametrowej analizy wyników badań geofizycznych, sprawdzona na konkretnym przykładzie, okazała się przydatna dla oceny perspektywiczności obszarów przy poszukiwaniu surowców mineralnych. Jej stosowanie może prowadzić do wyższej efektywności prac poszukiwawczych.
2. Przy wykorzystaniu większej ilości parametrów uzyskuje się wyższą wiarygodność wyników.
3. Istnieje potrzeba ulepszenia metody, zwłaszcza jeżeli chodzi o uwzględnienie korelacji parametrów.
4. Przytoczony przykład udowadnia dużą przydatność stosowania metod geofizycznych przy poszukiwaniu surowców mineralnych.

PIŚMIENNICTWO

- BATES R. C. (1959) — An application of statistical analysis to exploration for uranium on the Colorado Plateau. *Econ. Geol.*, 54, nr 3, p. 449—466. Urbana.
- FISZ M. (1969) — Rachunek prawdopodobieństwa i statystyka matematyczna. PWN. Warszawa.
- MAZUR M. (1970) — Jakościowa teoria informacji, WTN, Warszawa.
- АЙЗЕНБЕРГ М. А., ЗАХАРОВА Т. Л., КАВИНА В. О., КОЛМОГОРОВА Н. А. (1972) — Опыт прогнозирования по геофизическим данным нефтеносных площадей с помощью ЭВМ. *Геология и геофизика*, No 2, стр. 71—85. Новосибирск.
- ВЫСОКООСТРОВСКАЯ Е. Б., ЗЕЛЕНЦКИЙ Д. С. (1968) — О количественной оценке перспектив территории при поисках месторождений рудных полезных ископаемых. *Сов. геол.*, No 8, стр. 58—71. Москва.
- НИКИТИН А. А. (1970) — Применение теории статистических решений для классификации геологических объектов по данным комплекса геофизических методов. *Изв. ВУЗ, Геология и разведка*, No 10, стр. 135—144. Москва.

Станислав ВЫБРАНЕЦ

**СТАТИСТИЧЕСКО-ИНФОРМАЦИОННЫЙ АНАЛИЗ РЕЗУЛЬТАТОВ
ГЕОФИЗИЧЕСКИХ ИССЛЕДОВАНИЙ ДЛЯ ОПРЕДЕЛЕНИЯ ПЕРСПЕКТИВ
ПОИСКОВЫХ ТЕРРИТОРИЙ**

Резюме

В статье представлен способ определения перспектив территории с точки зрения залежей минерального сырья. Этот способ заключается в статистическо-информационном анализе результатов геофизических и геологических исследований. В нем использовано понятие числа отождествляющей информации (количества информации). Для подсчета числа отождествляющей информации применяется известная формула Шаннона $H = -\sum P(X_i) \log_2 P(X_i)$. Вводится понятие информации, отождествляющей избранный класс по отношению ко всей популяции или к другому классу. Она подсчитывается в виде разницы между количеством отождествляющей информации выбранного класса и всей популяции или выбранного класса и класса относимости. Отождествленным классом является площадь месторождения, а классом относимости — вся территория или территория без залежей. Отождествляющие информации, подсчитанные для нескольких параметров, могут суммироваться, если эти параметры независимы друг от друга.

Этот метод использовался для оценки перспектив залегания каолина в районе Стшегом-Собутка. Использовано 3 геоэлектрических параметра, полученных при интерпретации вертикального электрического зондирования: глубина основания с высоким сопротивлением, действительное сопротивление пласта, залегающего на основании и продольная проводимость покрова. На фиг. 2 представлено поверхностное распределение подсчитанной информации отождествляющей зоны каолинизации. Проведенный анализ результатов указывает на выгоды одновременного использования нескольких параметров.

Stanisław WYBRANIEC

**STATISTICAL-INFORMATIONAL ANALYSIS OF THE GEOPHYSICAL
DATA FOR EVALUATING THE PROSPECTS OF MINERAL OCCURRENCE**

Summary

A method based on statistical-informational analysis of geophysical and geological data to evaluate the prospects of mineral occurrence is presented. The concept of the number of identifying informations (amount of information) is applied. The known Shannon's formula $H = -\sum P(X_i) \log_2 P(X_i)$ is used for computing the number of identifying informations. The concept of the information identifying a given class with reference to the whole population or another class is introduced. It is computed as a difference between the number of identifying informations of one given class and that of the whole population or that of another class. The mineralized area is the class to be identified and the whole area or barren area are the reference classes. The identifying informations computed for a number of parameters can be added when the latter are independent.

The presented method was tested in the Strzegom-Sobótka area to evaluate the prospects of kaoline occurrence. Following the interpretation of vertical electrical soundings three geoelectrical parameters have been used: depth to the high resistivity basement, true resistivity of the layer directly overlying the basement and longitudinal conductance of the cover. The surface distribution of the information identifying the kaolinization zones is presented on fig. 2. It has been recognized that the combined analysis of a number of parameters, proposed in this paper provides favourable results.