

SZACOWANIE ZŁOŻ JAKO ZAGADNIENIE STATYSTYCZNE*

EKSPLLOATATORA ZŁOŻ INTERESUJĄ różne jego cechy, jak np. ilość rudy, procentowa zawartość cennego metalu, zawartość szkodliwych domieszek. Przed przystąpieniem do eksploatacji złoża bada się je dla ustalenia tych cech złoża. Od odkrycia złoża począwszy badanie to powtarza się kilkakrotnie, za każdym razem zagęszczając sieć robót poszukiwawczych, zaim przystąpi się do eksploatacji. Pierwsze systematyczne badanie złoża — to wiercenie bardzo rzadkich otworów poszukiwawczych, które ma na celu znalezienie złoża i określenie obszaru zalegania. Drugi etap to wiercenia rzadkie, które mają na celu wstępne zorientowanie w wielkości i jakości złoża. W trzecim etapie, gdy chodzi o bliższe zlokalizowanie zakładów wydobywczych, zagęszcza się jeszcze sieć wierceń, a w końcu bezpośrednio przed przystąpieniem do eksploatacji bada się złożo już po udośćpieniu do sieci chodników.

Ta wielostopniowość badań podyktowana jest z jednej strony ich kosztownością, a z drugiej potrzebą zbadania złoża na tyle dokładnie, aby móc zagwarantować rentowność zakładów wydobywczych. W szczególności wielkość zakładu wydobywczego musi być dostosowana do wielkości złoża.

W tym stanie rzeczy umiejętność maksymalnego wykorzystania informacji zawartej w próbkach jest niezmiernie ważna. O poszukiwaniu sposobów ocen najdokładniejszych świadczą wielka ilość metod szacowania opisanych w literaturze geologicznej. Np. Smirnow w swoim podręczniku (6, str. 70 i nast.) wylicza osiem różnych sposobów obliczania zasobów kopalni stałych, a Izakson (1, str. 5—37) wylicza ich aż siedemnaście.

W niniejszym artykule chcę dać krótki krytyczny przegląd tych sposobów, przy czym szczególną uwagę chcę skupić na metodologicznych podstawach metod statystycznych.

§ 1. *Metody geometryczne.* Przeważająca ilość metod szacowania opisanych przez zacytowanych autorów ma charakter geometryczny, a niemal z reguły używa się rozważań geometrycznych do uzasadnienia tych sposobów. Takimi są: metoda bloków geologicznych, metoda bloków eksploatacyjnych, metoda trójkątów, wieloboków, izolinii, przekrojów czy izohips. Uzasadnienie jest zawsze tego typu: zamiast prawdziwej bryły złoża, której kształt jest nieznany, buduje się konwencjonalnie pewne figury geometryczne, które w punktach, gdzie dokonano pomiarów, mają taką grubość, jak wynika z pomiarów; następnie zamiast objętości prawdziwej bryły złoża oblicza się objętość skonstruowanej figury geometrycznej i tę przyjmuje się jako oszacowaną objętość złoża. Jest to szczególnie widoczne w sposobie trójkątów i w sposobie wieloboków. W sposobie trójkątów sieć prób dzieli się na trójkąty o wierzchołkach w punktach pomiarów i za objętość złoża bierze się objętość bryły złożonej z graniastosłupów, których podstawami są owe trójkąty, a które w wierzchołkach swych podstaw mają wysokości takie jak wyniki pomiarów. Powstaje tu dodatkowa trudność wynikająca z niejednoznaczności podziału na trójkąty. W sposobie zaś wieloboków dzieli się badany obszar na wieloboki, przyłączając każdy punkt badanego obszaru do najbliższego miejsca pobrania próbki, a za objętość złoża bierze się

objętość bryły złożonej z graniastosłupów, których podstawami są owe wieloboki a wysokościami — wyniki pomiarów dokonanych na podstawie.

Myśl przyświecająca tym różnym sposobom — to zastąpienie rzeczywistej bryły złoża przez wyidealizowaną regularną bryłę geometryczną, dostosowaną w pewien sposób do wyników pomiarów, i obliczenie objętości tej bryły geometrycznej jako namiastki objętości złoża.

Zadaniem szacowania zasobów na podstawie danej sieci pomiarów jest nie tylko podanie ich przybliżonej wielkości, ale także ocena błędu podanego przybliżenia. Ocena błędu jest bodaj najsłabszym punktem metod geometrycznych. Różnicą między prawdziwą objętością bryły złoża a obliczoną objętością figury geometrycznej nazywa się tu błędem analogii. Podejmowano różne badania w celu zorientowania się w wielkości tych błędów i wyznaczenia ich zależności od gęstości sieci poszukiwawczej i od jej kształtu. Jak podaje Izakson (1, str. 72 oraz 77 i następne), w ZSRR były przeprowadzane badania na modelach polegające na tym, że porównywano prawdziwe objętości regularnych figur geometrycznych (np. elipsoidy obrotowej) oraz figur naśladujących prawdziwe ukształtowanie złoża z ocenami uzyskanymi metodami geometrycznymi na podstawie grubości tych figur w pewnej sieci punktów. Próbowano także empirycznego prześledzenia tego, jak zmniejsza się błąd oceny przy zagęszczaniu sieci pomiarów przez porównanie ocen otrzymanych z coraz gęstszej sieci pomiarów. Jak przyznaje Izakson, z gąszcza uzyskanych liczb nie zdołano wysnuć ogólnych wniosków. Wreszcie jedną z używanych tu metod oceny błędu jest porównywanie ocen obliczonych z tych samych pomiarów różnymi sposobami. Ze zgodności różnych ocen wnoszą się wówczas o dokładności oszacowania. Są to jednak tylko próby dotarcia do sedna zagadnienia. Najciekawsze są chyba badania na modelach naśladujących prawdziwe ukształtowanie złoża.

Zbadane za pomocą wierceń grubości złoża w niektórych punktach można potraktować tak, jak się to robi w statystyce reprezentacyjnej: jako próbki pobraną z „populacji” grubości złoża we wszystkich punktach badanego obszaru. Dlaczego by więc nie skorzystać z bogatego arsenału metod statystycznych przy określaniu dokładności oszacowania? I rzeczywiście podjęto próby zastosowania tych metod. Natknięto się przy tym jednak na pewne naradoksy, które u pewnych badaczy wzbudziły wątpliwości, czy metody statystyczne w ogóle nadają się do zagadnień związanych z szacowaniem złóż. Przechodzę do omówienia tych spraw.

§ 2. *Pierwsze próby zastosowania metod statystycznych* podjęte np. przez Ryżowa (5) a opisane w podręczniku Kreitera (4), były bezpośrednim przeniesieniem rozwiniętych już wówczas metod statystycznych opartych na koncepcji statystycznej niezależności obserwacji albo, innymi słowy, na koncepcji zmiennych losowych niezależnych. Prześledźmy pokrótce podstawy metodologiczne tych metod.

Klasyczne zadanie można sformułować tak. Dana jest populacja złożona z wielkiej liczby przedmiotów. Chodzi o szacowanie na podstawie próbki złożonej z niewielkiej ilości przedmiotów średniej wartości jakiejś cechy tych przedmiotów, np. ich masy. Dla każdej liczby x jest w tej populacji określony (choć dla nas nieznan) stosunek ilości tych przedmiotów, które mają masę mniejszą od x , do ogólnej ilości

* Referat wygłoszony na konferencji naukowej, zorganizowanej przez Katedrę Geologii AGH w Krakowie, na temat zastosowania rachunku statystycznego przy ustalaniu zasobów.

przedmiotów w populacji. Oznaczmy go przez $F(x)$. Wobec tego masę przedmiotu wylosowanego na chybił trafił z naszej populacji traktuje się jako realizację zmiennej losowej o dystrybucji $F(x)$, to znaczy realizację zmiennej losowej, która przyjmuje wartości mniejsze od x z prawdopodobieństwem $F(x)$. Powiada się dalej, że przedmioty mające wejść do próbki należy losować niezależnie, a więc gdy x_1, x_2, \dots, x_n są zaobserwowanymi masami przedmiotów, które weszły do próbki, można traktować te masy jako obserwacje niezależne statystycznie albo, innymi słowy, jako realizacje niezależnych zmiennych losowych o dystrybucji $F(x)$. Zakładamy tu milcząco, że dystrybuanta badanej cechy przedmiotów nie zmienia się wskutek wylosowania kilku przedmiotów do próbki; jest to praktycznie usprawiedliwione przy dostatecznie dużej liczności populacji. Skoro się na to zgodzimy, mamy otwartą drogę do stosowania tych wszystkich klasycznych metod statystycznych, które są oparte na koncepcji zmiennych losowych niezależnych. Można więc odpowiadać na pytanie, z jaką dokładnością średnia z próbki przybliży średnią z populacji, ile trzeba wziąć dodatkowych przedmiotów do próbki, aby oszacować średnią z populacji z zadaną dokładnością itp. Podkreślamy, że stosowanie wspomnianych metod uzasadnione jest wtedy, gdy można uważać za spełnione istotne dla nich założenie niezależności. Dla praktycznej realizacji tego założenia zaleca się różne środki ostrożności przy losowaniu przedmiotów do próbki, jak np. używanie tablic liczb losowych.

Przeniesienie powyższego na przypadek szacowania złóż jest następujące. Przypuśćmy, że chodzi o oszacowanie średniej miąższości złoża na danym obszarze D . Wówczas dla każdej liczby x jest określony choć nam nieznan, stosunek pola tej części obszaru D , na której miąższość złoża jest mniejsza od x , do pola całego obszaru D . Oznaczamy ten stosunek jak poprzednio przez $F(x)$. Powiada się więc, że miąższość złoża w wylosowanym punkcie obszaru D jest realizacją zmiennej losowej o dystrybucji $F(x)$. Dalej żąda się, aby punkty, w których mają być zrobione pomiary miąższości, były losowane z obszaru D niezależnie. Gdy te warunki są spełnione, traktuje się miąższości zaobserwowane w wylosowanych punktach jako realizację niezależnych zmiennych losowych o dystrybucji $F(x)$. To uzasadnia korzystanie z metod statystycznych wymagających niezależności obserwacji. Z wymienionych założeń wynika np. twierdzenie, że jeśli s jest odchyleniem standardowym (inaczej: dyspersją) miąższości na obszarze D , to s/\sqrt{n} jest błędem szacowania średniej miąższości na obszarze D przez średnią miąższości zaobserwowanych w n wylosowanych punktach. Tego twierdzenia m. in. używa Krajewski (2) przy wyznaczaniu gęstości robót poszukiwawczych.

Opisany sposób stawiania zagadnienia ma jednak pewne mankamenty. Przede wszystkim wymaga losowania punktów, w których mają być zrobione pomiary, i to losowania niezależnego poszczególnych punktów. To wymaganie losowania punktów, w których mają być zrobione pomiary, budzi szczególnie sprzeciw geologów-praktyków. Powiadają oni, że przecież w wyniku losowania może się zdarzyć, że wszystkie wylosowane punkty wypadną w jednym zakątku badanego obszaru D , a jest niezmiernie trudno uwierzyć, żeby szacowanie miąższości na całym obszarze D na podstawie pomiarów w jednym tylko zakątku mogło być rozsądne. Intuicja stanowczo temu przeciwi. Ma ona mocne podstawy w doświadczeniu, które poucza, że miąższość złoża nie zmienia się na ogół zbyt nagle, wobec czego liczne próby w jednym miejscu po pierwsze powtarzają poniekąd jedna drugą, a po drugie słabo informują o miąższości w części obszaru D nie pokrytej pomiarami. W tym stanie rzeczy postępuje się praktycznie tak, że pomiary robi się w mniej lub bardziej regularnej sieci punktów, a błąd oszacowania oblicza się tak, jak gdyby pomiary można uważać za statystycznie niezależne. Grozi to przecenieniem wielkości błędu oszacowania.

Nie jest to jedyna trudność. Jednym z zadań przykuwających uwagę geologów jest klasyfikacja złóż ze względu na trudności związane z szacowaniem ich parametrów. Na takiej klasyfikacji oparte są normy gęstości sieci poszukiwawczej (2). Otóż próbowano oprzeć taką klasyfikację na tak zwanym wskaźniku zmienności, to jest na stosunku standardowego odchylenia danej cechy do jej średniej. Ta średnia i to odchylenie standardowe obliczane są z reguły z obserwacji rozciągających się na dużych obszarach, gdy tymczasem zagęszczenie sieci poszukiwawczej ma między innymi na celu uzyskanie możliwości porównywania ze sobą mniejszych kawałków złoża. Otóż złoża o takim samym współczynniku zmienności mogą się znacznie różnić pod względem ciągłości. Grozi to tym, że zaprojektowana na podstawie samego współczynnika zmienności sieć poszukiwawcza będzie dla złóż ciągłych aż za gęsta, a dla innych jeszcze ciągle zbyt rzadka.

Traktowanie poszczególnych pomiarów parametru złoża jako obserwacji niezależnych jest niewątpliwie pociągające z tego powodu, że pozwala na mechaniczne przeniesienie znanych metod statystycznych. Jednak wskutek leżącego u jego podstaw wymagania niezależnego losowania punktów pomiarowych z badanego obszaru takie ujęcie zagadnienia uniemożliwia uchwycenie i uwzględnienie w rachunku różnic w ciągłości poszczególnych złóż, jakie zachodzą np. między pokładowymi złożami węgla a impregnacyjnymi złożami cynku.

W dalszej części artykułu chcę wskazać takie probabilistyczne postawienie zagadnienia szacowania parametrów złoża, przy którym można nie tylko uzasadnić przewagę systematycznego rozmieszczania punktów pomiarowych, ale także obliczyć numerycznie, jak różne sposoby rozmieszczania punktów pomiarowych wpływają na dokładność szacowania.

§ 3. Jak uwzględnić dotychczasowe obserwacje przy szacowaniu? Metody statystyczne, o których dotychczas mówiłem, wykorzystywały przy szacowaniu tylko aktualne obserwacje. W archiwach biur geologicznych są jednak nagromadzone wielkie ilości obserwacji i warto by się pokusić o wykorzystanie doświadczenia tam zawartego do ulepszenia aktualnych ocen.

Omówię tu jeden sposób pozwalający wykorzystać przeszłe obserwacje przy aktualnych oszacowaniach. Wyobraźmy sobie w tym celu rodzinę bloków eksploatacyjnych o ustalonym kształcie i wielkości wyciętych w złożu o ustalonej strukturze geologicznej. Wyobraźmy sobie dalej, że na pewnej ilości tych bloków, które można uważać za reprezentacyjną próbkę z naszej populacji bloków, w punktach wylosowanych niezależnie z tych bloków zostały dokonane pomiary miąższości złoża. Te dane chcemy wykorzystać przy szacowaniu objętości złoża w blokach dotychczas nie badanych. Jako podstawę do szacowania objętości złoża na nowym bloku mamy przeto pomiary miąższości w punktach wylosowanych niezależnie w tym nowym bloku i dane archiwalne dotyczące bloków pomierzonych dawniej. Otóż z dawnych pomiarów możemy obliczyć rozkład objętości złoża na blokach dotychczas badanych. Ten rozkład możemy przyjąć za rozkład objętości wśród bloków jeszcze nie zbadanych jako tak zwany rozkład a priori. Jeżeli jeszcze uda się nam ustalić, jak zależy rozkład miąższości w punktach wylosowanych z bloku od objętości złoża na tym bloku, to będziemy mieli otwartą drogę do zastosowania schematu Bayesa, który uwzględni dotychczasowe doświadczenie w postaci rozkładu a priori i opiera oszacowanie objętości w nowym bloku na rozkładzie a posteriori, ten zaś oblicza się łącznie z rozkładem a priori i aktualnych pomiarów. Sens orzeczenia, że według rozkładu a posteriori objętość złoża na badanym bloku jest mniejsza od ustalonej uprzednio liczby x z jakimś prawdopodobieństwem, jest wówczas taki, że gdybyśmy ponawiali pomiary na coraz nowych blokach i za każdym razem notowali prawdziwą objętość złoża na tych blokach, na których wynik pomiarów byłby taki

sam jak na bloku aktualnie badanym, to stosunek ilości bloków, na których objętość byłaby mniejsza od x , do ogólnej ilości tych bloków byłby równy właśnie owemu prawdopodobieństwu a posteriori. O takim postawieniu zagadnienia pisze np. Trembecki (7). Nie rozumiem jednak, jak doszedł on do wniosku, że gęstość sieci poszukiwawczej ma być odwrotnie proporcjonalna do zmienności złoża?

Opisane tu bayesowskie potraktowanie zagadnienia jest krokiem naprzód: pozwala wykorzystać doświadczenie zawarte w dawniej wykonanych pomiarach. Jesteśmy jednak dalej skrepowani tym, że rozważa się tu jedynie losowanie niezależne punktów pomiarowych. Nie ma więc nadal mowy o uwzględnieniu wpływu sposobu rozmieszczania punktów pomiarowych na dokładność oszacowania. Powiedzieliśmy również, że potrzebna jest przy tej metodzie znajomość zależności rozkładu miąższości w wylosowanych punktach od prawdziwej objętości złoża na bloku. Otóż zależność ta w znacznym stopniu zmienia się wraz z wielkością i kształtem bloku, brak zaś w niej wskazówek co do sposobu wyeliminowania tej zależności.

§ 4. Ujęcie zagadnienia na gruncie teorii procesów stochastycznych. Przystępujemy do opisu takiego probabilistycznego ujęcia zagadnienia, przy którym jest możliwe badanie wpływu sposobów rozmieszczania punktów na dokładność oszacowania jak również odpada wspomniany dopiero co kłopot z zależnością rozkładu miąższości od wielkości i kształtu bloku. Jest to sposób zaproponowany przeze mnie w pracy (9).

Dla wskazania istoty tego ujęcia zaczniemy od bardzo uproszczonego przykładu. Wyobraźmy więc sobie rząd beczek nie różniących się wyglądem zewnętrznym, z których jedne są puste, a inne pełne i zawierają po 2 hl ropy. Ustawione są tak, że następuje po sobie kolejno po osiem beczek pustych i po osiem beczek pełnych (to jest nasze złożo). Nam wolno wybrać dowolnie pięć kolejnych beczek (to nasz blok eksploatacyjny), zmierzyć ilość ropy w dwu z nich, które wolno nam dowolnie wskazać (to nasze wiercenie próbne) i na tej podstawie oszacować ogólną ilość ropy we wszystkich pięciu wybranych beczkach (to nasze oszacowanie objętości złoża na bloku eksploatacyjnym). Owo fikcyjne beczkowe złożo pokazuje schematycznie ryc. 1; czarne kółka przedstawiają beczki z ropą, białe kółka — beczki puste.



Ryc. 1

Wybierając na chybił trafił początek naszej piątki beczek, natrafimy z jednakowym prawdopodobieństwem na jeden z szesnastu przypadków przedstawionych schematycznie na ryc. 2, na którym kółka ujęte w ramkę pokazują rozmieszczenie pustych i pełnych beczek w wybranej piątce.

Rozważmy szacowanie zawartości ropy w pięciu beczkach na podstawie jej zawartości w skrajnych beczkach piątki (to jakby jeden z możliwych sposobów systematycznego rozmieszczania wierceń próbnych). Wówczas w przypadkach 1 do 4 i od 9 do 12 nie popełnimy błędów przy szacowaniu: w pierwszych czterech przypadkach średnia zawartość ropy we wskazanych dwu beczkach wynosi 0 hl i na tej podstawie zawartość we wszystkich pięciu beczkach szacujemy na 0 hl, ta ocena jest zgodna ze stanem faktycznym. W przypadkach od 9 do 12 średnia zawartość we wskazanych dwu beczkach jest 2 hl, a więc zawartość we wszystkich pięciu beczkach szacujemy na 10 hl, co jest również zgodne ze stanem faktycznym. W pozostałych ośmiu przypadkach popełniamy błędy, gdyż średnia zawartość w skrajnych beczkach jest wówczas stale równa 1 hl i na tej podstawie szacujemy ogólną zawartość w pięciu beczkach na

5 hl, gdy prawdziwa zawartość ropy we wszystkich pięciu beczkach razem wynosi kolejno 2, 4, 6, 8, 8, 6, 4 i 2 hektolitry. Na szesnaście równoprawdopodobnych przypadków 8 razy nie omylimy się, 2 razy oszacowanie będzie o 3 hl za mało, 2 razy o 1 hl za mało, 2 razy o 1 hl za dużo i 2 razy o 3 hl za dużo. Średni kwadrat błędów jest równy

$$\frac{1}{16} [8 \cdot 0^2 + 2 \cdot (-3)^2 + 2 \cdot (-1)^2 + 2 \cdot 1^2 + 2 \cdot 3^2] = \frac{40}{16} = 2 \frac{1}{2}$$

Gdybyśmy szacowali nie na podstawie zawartości w beczkach skrajnych, lecz np. na podstawie zawartości w beczce drugiej i czwartej, to oszacowanie polepszyłoby się w przypadkach 5, 8, 13 i 16 (ryc. 2). Np. w przypadku 5 oszacowalibyśmy ogólną zawartość na 0 hl, a nie jak przedtem na 5 hl, i oszacowanie byłoby o 2 hl za mało zamiast jak przedtem o 3 hl za dużo. Średni kwadrat błędów zmalałby dzięki temu dwukrotnie, z $2\frac{1}{2}$ na $1\frac{1}{4}$. Tym razem beczki próbne mają względem badanej piątki inne, bardziej centralne położenie. Widzimy, że wpływa to wyraźnie na polepszenie dokładności oszacowania.

Porównaliśmy przed chwilą dwa systematyczne sposoby pobierania próbek. Zapytajmy teraz o dokładność oszacowania przy losowym pobieraniu dwu beczek do próbki. Otóż dwie beczki z pięciu można wylosować na 10 różnych sposobów. W przypadkach od 1 do 4 i od 9 do 12 (ryc. 2) przy każdym wyniku tego losowania doszlibyśmy do trafnej oceny zawartości ropy we wszystkich pięciu beczkach, ale np. w przypadku 5 myliliibyśmy się zawsze. Mianowicie na 10 możliwości 4 razy otrzymalibyśmy ocenę 5 hl, a 6 razy otrzymalibyśmy ocenę 0 hl, gdy prawdziwa zawartość ropy wynosi w tym przypadku 2 hl. Tak więc 4 razy podalibyśmy ocenę o 3 hl za dużą i 6 razy o 2 hl za małą. Średni kwadrat błędów w tym przypadku równy

$$(6 \cdot 2^2 + 4 \cdot 3^2) \cdot 10 = 6.$$

Gdy obliczymy średnią analogicznych liczb dla wszystkich szesnastu przypadków przedstawionych na ryc. 2, otrzymamy jako średni kwadrat błędów oszacowania przy losowym wybieraniu beczek próbkowych liczbę 4; widzimy, o ile gorsze jest losowe wybieranie beczek próbkowych w porównaniu z rozpatrzonym uprzednio systematycznym ich rozmieszczeniem w badanej piątce.

Zależność dokładności oszacowania od sposobu rozmieszczenia beczek próbkowych w badanej piątce wystąpiła tu dlatego, że w naszym „złożu” puste i pełne beczki stoją seriami. Skutek tego jest taki, że gdybyśmy na podstawie zawartości w beczce wybranej na chybił trafił chcieli odgadnąć, czy jej sąsiadka jest pełna czy pusta, to potrafilibyśmy odgadnąć trafnie średnio 7 razy na 8 prób. Wystarczyłoby orzekać, że sąsiadka jest taka sama jak beczka, której zawartość sprawdziliśmy. Natomiast, jak łatwo sprawdzić, znajomość zawartości wylosowanej na chybił trafił beczki nie nam nie mówi o tym, czy beczka czwarta licząc od niej jest pełna czy pusta.

1	○ ○ ○ ○ ○	○ ○ ○ ○ ○
2	○ ○ ○ ○ ○	○ ○ ○ ○ ○
3	○ ○ ○ ○ ○	○ ○ ○ ○ ○
4	○ ○ ○ ○ ○	● ● ● ● ●
5	○ ○ ○ ○ ○	● ● ● ● ●
6	○ ○ ○ ○ ○	● ● ● ● ●
7	○ ○ ○ ○ ○	● ● ● ● ●
8	○ ○ ○ ○ ○	● ● ● ● ●
9	● ● ● ● ●	● ● ● ● ●
10	● ● ● ● ●	○ ○ ○ ○ ○
11	● ● ● ● ●	○ ○ ○ ○ ○
12	● ● ● ● ●	○ ○ ○ ○ ○
13	● ● ● ● ●	○ ○ ○ ○ ○
14	● ● ● ● ●	○ ○ ○ ○ ○
15	● ● ● ● ●	○ ○ ○ ○ ○
16	● ● ● ● ●	○ ○ ○ ○ ○

Ryc. 2

Rozpatrzone zadanie można sformułować w terminach pięciu zmiennych losowych przedstawiających zawartość ropy w pięciu beczkach stanowiących nasz „blok eksploatacyjny”. Ich łączny rozkład prawdopodobieństwa łatwo odczytać z ryc. 2, który można potraktować jako listę szesnastu równoprawdopodobnych układów wartości tych pięciu zmiennych losowych. Odczytujemy stamtąd, że każda z tych zmiennych losowych przyjmuje z prawdopodobieństwem 1:2 wartość 0 i z takim samym prawdopodobieństwem 1:2 wartość 2 (ilość hektolitrów ropy), że więc wszystkie one mają wartość oczekiwaną 1 i dyspersję 1. Współczynniki korelacji zależą tylko od różnicy numerów naszych zmiennych losowych i dla zmiennych losowych o numerach różniących się o 1 wynoszą 3:4, przy różnicy numerów 2 są równe 2:4, przy różnicy numerów 3 są równe 1:4, a skrajnie zmienne losowe są nie skorelowane. Zadanie zaś formuluje się tak: oszacować sumę naszych pięciu zmiennych losowych znając średnią wartość dwu z nich, których numery wolno nam wybrać. Podkreślimy na koniec, że probabilistyczny lub jak kto woli, statystyczny sens naszego zadania jest w gruncie rzeczy zawarty w ryc. 2, która określa nam „populację” „bloków eksploatacyjnych”, z jakimi możemy się spotkać.

Sądźmy, że po tym przykładzie będą zrozumiałe następujące wyjaśnienia ukazujące sens zastosowania procesów stochastycznych do opisu struktury złóż. Mianowicie rozważania nasze odnosimy do rodziny bloków eksploatacyjnych, jakie można wyciąć z wielkiego złoża o ustalonej strukturze geologicznej. Bardziej obrazowo można to powiedzieć tak, że dywan takiej wielkości i kształtu jak blok eksploatacyjny kładziemy kolejno w różnych miejscach wielkiego złoża i za każdym razem notujemy pod każdym punktem dywanu aktualną miąższość złoża. Dla każdego punktu dywanu z osobna możemy się więc zapytać o to, jak często pod tym punktem dywanu pojawia się miąższość mniejsza od ustalonej uprzednio liczby x ? Innymi słowy, pytamy o rozkład prawdopodobieństwa miąższości złoża pod ustalonym punktem dywanu. Jest naturalne przyjąć, że przy złożu dostatecznie wielkim w porównaniu z wymiarami bloku eksploatacyjnego ten rozkład miąższości będzie praktycznie taki sam dla wszystkich punktów dywanu i równy rozkładowi miąższości w całym złożu. Podobnie dla ustalonego układu punktów na dywanie można zapytać, jak często pod tymi punktami pojawiają się miąższości mniejsze odpowiednio od ustalonych liczb? Innymi słowy, pytamy o rozkład prawdopodobieństwa układu miąższości pod wskazanymi punktami dywanu. Wydaje się rozsądne przyjąć, że dla złoża izotropowego rozkład ten będzie zależny tylko od wzajemnych odległości punktów wybranych na dywanie.

W tym właśnie sensie, to znaczy przez odwołanie się do rodziny możliwych bloków eksploatacyjnych, będziemy strukturę złoża opisywać matematycznie przez rodzinę zmiennych losowych $y(p)$ przypisanych poszczególnym punktom p bloku eksploatacyjnego i reprezentujących miąższość złoża w punkcie p . Matematycznie mówiąc, strukturę złoża będziemy więc opisywać za pomocą rodziny zmiennych losowych przypisanych punktom płaskiej powierzchni — a to znaczy za pomocą płaskiego procesu stochastycznego.

Z powyższego wynika, że strukturę złoża izotropowego można opisywać za pomocą płaskiego procesu stochastycznego, w którym

(a) wszystkie zmienne losowe $y(p)$ mają jednakowy rozkład prawdopodobieństwa;

(b) łączny rozkład prawdopodobieństwa zmiennych losowych $y(p), \dots, y(q)$, przyporządkowanych punktom p, \dots, q , zależy tylko od wzajemnych odległości tych punktów, a nie od ich absolutnych współrzędnych.

Założenie wymienione pod (a) i (b) charakteryzują rodzinę zmiennych losowych $y(p)$ jako płaski proces stochastyczny stacjonarny i izotropowy. Istotny postęp w stosunku do sytuacji opisanej w § 3 polega

na tym, że obecnie strukturę złoża opisujemy przez łączne rozkłady prawdopodobieństwa zmiennych losowych $y(p), \dots, y(q)$. Oto dlaczego w tym ujęciu otwarta jest droga do uwzględnienia wpływu rozmieszczenia punktów pomiarowych na dokładność oszacowania. Te łączne rozkłady prawdopodobieństwa, gdy się je pozna ze starych pomiarów, grać będą taką rolę jak rozkłady a priori w § 3.

§ 5. **Uproszczenie zadania.** Mówiliśmy dotychczas o łącznym rozkładzie prawdopodobieństwa zmiennych losowych $y(p)$. Rozkłady takie zadane są za pomocą dystrybuant, które w przypadku łącznego rozkładu prawdopodobieństwa n zmiennych losowych są funkcjami n zmiennych rzeczywistych. Nic dziwnego, że estymowanie tych funkcji, acz teoretycznie możliwe, jest praktycznie bardzo trudnym zadaniem. Dlatego w przypadku jednej zmiennej losowej zadowolamy się zwykle znajomością tylko dwu charakterystyk liczbowych jej rozkładu: średniej m i dyspersji s . Okazuje się, że do wyznaczenia średniego kwadratu błędu, który charakteryzuje dokładność oszacowania, wystarczy ponadto znajomość współczynników korelacji między parami zmiennych losowych $y(p)$ i $y(q)$. Łączny rozkład prawdopodobieństwa i współczynnik korelacji zależy jedynie od odległości punktów p i q . Funkcja odległości d tych punktów podająca tę zależność zwie się funkcją korelacyjną. Oznaczmy ją przez $f(d)$. Tak więc do scharakteryzowania rodziny zmiennych losowych $y(p)$ ze względu na dokładność oszacowania mierzoną za pomocą średniego kwadratu błędu wystarczy znajomość dwu stałych liczbowych, m i s , oraz jednej funkcji korelacyjnej $f(d)$.

W pracy (9) starałem się wyznaczyć funkcję korelacyjną dla złóż cynkowych. W świetle tych wyników oraz tego, co udało się znaleźć w literaturze, wydaje się, że można pójść jeszcze dalej i ograniczyć się do zastąpienia funkcji korelacyjnej parametrem liczbowym. Okazało się bowiem, że we wszystkich przypadkach obliczone z obserwacji empiryczne funkcje korelacyjne dały się zaskakująco dobrze wyrównywać funkcją wykładniczą $f(d) = \exp(-Cd)$, gdzie C jest stałą dodatnią. Wobec tego funkcja korelacyjna mogłaby być scharakteryzowana stałą C i tylko tą stałą trzeba by wyestymować. Tak więc zadanie upraszcza się o tyle, że charakterystykę struktury złoża daje się zamknąć w trójce liczb, średniej m , dyspersji s i stałej dodatniej C , charakteryzującej funkcję korelacyjną. Do tego dochodzić może jeszcze liczba δ — to jest dyspersja błędu losowego przy pomiarach (9).

§ 6. **Zależność dokładności oszacowania od sposobu rozmieszczenia punktów próbných.** W pracy (9) podałem wzory wyrażające błędy oszacowania przy ustalonym rozmieszczeniu punktów na bloku w zależności od charakteryzującej złożę funkcji korelacyjnej i stałych m, s . Rozpatrzyłem też, jak od tych wielkości zależą najlepsze liniowe estymatory średniej wartości parametru na bloku. Są to wzory dość trudne z punktu widzenia rachunków. Zależnością dokładności oszacowania średniej wartości parametru złoża na bloku, np. średniej miąższości, zająłem się z nieco innego punktu widzenia w pracy (10). Rozważyłem tam szacowanie średniej parametru na bloku jedynie za pomocą średniej arytmetycznej z wyników obserwacji i studiowałem zależność średniego kwadratu błędu oszacowania od sposobu rozmieszczenia punktów pomiarowych.

Rozważyłem trzy sposoby rozmieszczenia punktów. Jeden — losowy — polegający na niezależnym losowaniu punktów pomiarowych z całego bloku eksploatacyjnego, na którym ma się oszacować średnią wartość parametru. Drugi — warstwowy — polega na tym, że blok dzieli się na tyle przystających i rozłączonych części, ile punktów pomiarowych ma się wybrać, i losuje się niezależnie po jednym punkcie z każdej części. Trzeci — systematyczny — polega na tym, że blok dzieli się na tyle części, ile punktów pomiarowych trzeba wybrać, z tym że teraz żąda się, aby części były przystające przez przesunięcie. Następnie losuje się jeden punkt w części pierwszej,

a w pozostałych częściach bierze się punkty odpowiadające mu przez przesunięcie ustalające przystawanie odpowiednich części. Okazało się, że niezależnie od kształtu części i od tego, jaka jest funkcja korelacyjna, sposób warstwowy prowadzi zawsze do mniejszych błędów niż losowy. Potwierdza to intuicję, w myśl której ograniczenie elementu losowego przy rozmieszczaniu punktów jest korzystne.

O potrzebie takich badań niech jednak poświadczy fakt, że w pewnych okolicznościach sprecyzowanych w (10), kiedy części, na jakie jest podzielony blok, są jak najbardziej zwarte, i przy „złożu” izotropowym z wykładniczą funkcją korelacyjną warstwowe rozmieszczenie punktów jest korzystniejsze niż systematyczne. Jest to sprzeczne z wymienioną intuicją, gdyż przy systematycznym rozmieszczeniu punktów element losowy jest jeszcze mocniej ograniczony niż przy warstwowym. Ideą przyswiewającą tym badaniom jest znalezienie takiego sposobu rozmieszczania punktów pomiarowych, który by się wyróżniał pewnymi optymalnymi własnościami w stosunku do innych sposobów rozmieszczania punktów i żeby te optymalne własności były jak najbardziej niezależne od aktualnej funkcji korelacyjnej złoża.

§ 7. Program dalszych badań. Na zakończenie kilka uwag o badaniach, jakie wydaje się można i warto przeprowadzić. Należą do nich przede wszystkim dalsze badania funkcji korelacyjnych różnych złóż. Za ciekawe i potrzebne uważam porównanie pod tym względem złóż o różnych strukturach geologicznych, np. złóż cynku ze złożami węgla. Do tego rodzaju badań najlepiej nadają się materiały składające się z pomiarów wzdłuż ustalonej prostej poczynionych w równych odstępach (porównaj „stumetrówkę” w pracy 9, str. 122). Należałoby zebrać więcej takich materiałów. Mniej się natomiast nadają do wyznaczenia funkcji korelacyjnej materiały złożone z punktów wyrzucanych nieregularnie. Można je jednak łatwo wykorzystać do badania struktury złoża w nieco inny sposób. Warto mianowicie przestudiować, jak zwiększa się dyspersja parametru wraz ze zwiększaniem obszaru, z którego pochodzą obserwacje; jest to ściśle związane z funkcją korelacyjną. Dalej, w pracy (9) okazało się, że niemają rolę w ogólnym błędzie oszacowania odgrywają błędy pomiarów. Należy je przestudiować. Za ciekawe i potrzebne uważam skompletowanie oszacowania błędu pomiarowego wynikającego z ekstrapolacji w kierunku zera funkcji korelacyjnej (9) z oszacowaniami tego błędu uzyskanymi na innej drodze. Za potrzebne uważam także badania teoretyczne w rodzaju przedstawionych w (10). Mają one zresztą znaczenie nie tylko dla badania złóż, ale także np. w polowych doświadczeniach rolniczych.

LITERATURA

1. Izakson S. S. — Kontrolnyje wycislenija pri podsczotie zapasow poleznych iskopaemych i opriedielenije pogrieszności podsczota. Moskwa 1953.
2. Krajewski R. — Ustalenie gęstości sieci rozpoznawczej złóż na podstawie stopnia wiaro-

godności zasobów. „Przegląd Geologiczny” 1956, nr 1.

3. Krajewski R. — Z badań nad wskaźnikami zmienności polskich złóż kruszcowych. „Zeszyty Naukowe AGH w Krakowie” Geologia, 1956, z. 1.
4. Krieffier W. M. — Poiski i razwiedki poleznych iskopaemych. Moskwa 1940.
5. Ryżow P. A. — Matematyczeskoe opriedelenije „geologiczeskoi osyżki” pri podsczotie zapasow poleznych iskopaemych. „Trudy kazach. gorńo-mietal. inst.” 1938, nr 1.
6. Smirnow W. I. — Ustalenie zasobów surowców mineralnych. Warszawa 1954.
7. Trembecki A. — Metodyka ustalania zasobów w świetle analizy klasycznej oraz w świetle analizy statystycznej. „Zeszyty Naukowe AGH w Krakowie” 1954, nr 1.
8. Trembecki A. — Ustalanie gęstości robót poszukiwawczych. „Zeszyty Naukowe AGH w Krakowie” 1954, nr 2.
9. Zubrzycki S. — O szacowaniu parametrów złóż geologicznych. „Zastosowanie Matematyki” 3 (1957), nr 2.
10. Zubrzycki S. — Remarks on random, stratified and systematic sampling in a plane. „Colloquium Mathematicum” 6 (1958).

STEFAN ZUBRZYCKI

ESTIMATING ORE-DEPOSITS AS A STATISTICAL PROBLEM

Summary

The author, beyond discussing in short the methodological basis of geometrical methods used in estimating parameters of ore-deposits, concerns himself with statistical methods. The very first applications of such methods consisted in using fairly developed statistical methods based on a concept of statistical independence of observations. However, in order to justify the assumption of independence underlying these methods the measurement points must necessarily be distributed at random on the investigated domain, s. g. exploitation block. It is why the very important question how the way of distributing the measurement points influences the exactness of estimation can not be answered on the basis of these methods. The author proposes and discusses another method of dealing with a problem in question which is based on theory of plane stationary stochastic processes. Here the measurements are not more regarded as independent and the observed mildness in changes of parameter values in an ore-deposit can be described in terms of correlation function. This enables us to give answers to questions concerning the dependence of exactness of estimation on the distribution of measurement points. The results obtained up to now are reviewed and suggestions of further researches are given.