GEOPHYSICAL TRANSACTIONS 1984
Vol. 30. No. 2. pp. 141—155

APPLICATIONS OF STATISTICAL THEORY TO IMPROVEMENT OF TRACING OF REFLECTORS AND IDENTIFICATION OF ANOMALIES

F. M. HOLTZMAN**, E. A. KOZLOV*, O. A. POTAPOV*,
O. G. KUT'INA*, V. N. TROYAN**

The paper deals with identification of anomalies of parameters related to structural and lithological properties. Methods of tracing (classification) reflectors on seismic section are proposed. Some examples of utilizing algorithms to search for inhomogeneities on a seismic section and to specify the location of an oil reservoir based on the methods of multidimensional discriminant analysis are considered.

d: statistical analysis, inhomogeneity, wave attenuation, interval energy, resistivity, gravity anomaly

1. Introduction

The main task of modern geophysical prospecting for oil-promising zones for gas and other minerals is the location of areas with anomalous parameters in the geologic environment under study. The anomalies in question can be associated with structural or lithological peculiarities of the medium.

The search for structural anomalies is usually based on the methods of wave correlation: the detection of correlation breakpoints corresponding to an abrupt change in the reflecting properties of a boundary or tectonic faults. However, the problem of horizon tracing does not reduce to a simple determination of the structural features of a boundary; we should also consider the problem of tracing along the boundary those parameters which are correlatable with variations in the physical composition of rocks, e.g. layer velocity, amplitudes, wave attenuation, etc.

When searching for non-structural anomalies, the study of the distribution of a set of features and the search for areas where the statistical properties of the features essentially differ from each other becomes the main tool for dividing the section into blocks characterized by common properties.

* NPO Neftegeofizika, 103062 Moscow, Chernyshevskogo 22, USSR
** NIIF LGU 198904 Leningrad, St. Petergof, 1 Maya ul. 100, USSR

Paper presented at the 28th International Geophysical Symposium, Balatonszemes, Hungary, 28 September—1 October, 1983

The theory and algorithms for solving the mentioned tracing and grouping problems are thoroughly treated on the basis of mathematical statistics [Ed. HOLTZMAN 1981]. A short summary of the topic is to be found in Holtzman [1981], Practical applications are discussed in [KUT'INA 1982] snf [TROYAN 1982].

A set of arbitrary observational data can be regarded as a collection of realizations of random variables. Obviously conclusions drawn from observational data have also probabilistic character. The statistical interpretation theory postulates, as the most important requirement for evaluation algorithms, the possibility of checking the quality of their conclusions. As a measure characterizing the quality of conclusions may serve on of the statistical estimations on their information content, reliability, efficiency.

The present paper shows some application examples of tracing and grouping algorithms constructed on the above mentioned principles.

2. Defining the location of lithological anomalies

The problem of locating lithological anomalies can be solved together with the problem of horizon tracing or separately. In both cases, statistical analysis is based upon the values of the features (obtained during seismogram and time-section processing), that characterize a certain portion of a record covering usually 2—3 wave periods along T and 2—3 wavelengths along X. The main rule in selecting events is based on the signal-to-noise ratio in the analysis window; according to the assumed properties of the investigated material, the threshold value is selected from tables for distribution of random variables: normal, χ-square, Student's or Fischer's [KUT'INA 1982].

Figure 1 shows some examples of selecting events on a time section measured along one of the lines in the Aktyubinsk area. It can be seen that the widening of the time window for analysis (Fig. 1c) leads to better elimination of noise, improves the conditions of horizon traceablility and is useful for eveluating the integral characteristics of a record, for example, velocities. Determination of noise variance within the same time window as that of the signal energy (Fig. 1d) provides the possibility to emphasize wave behaviour in complicated wave zones (see parts of the section at $X = 4000$ m, $T = 2.1$ s).

Estimation of the parameters of selected events and subsequent tracing of the latter in time and along the line results in the creation of a field of features. Accumulation of features is performed at several successive stages of data processing: time, angle of wave arrival, signal-to-noise ratio, average amplitude and duration of basic peaks are determined from the time section while the velocity V_{CDP} from seismograms; tracing along the line permits one to determine the degree of wave regularity along X tracing in time allows one to determine the frequency, attenuation, number of phases, intensity ratio of various phases, ratio of the times of events and degree of traceability along T.

The next stages of data processing and interpretation are the study of regularities of feature variation in various trends, determination of peculiarities of their behaviour, localization of anomalous portions of the section and interpretation of the obtained results, i.e. the main task is to decide on the physical nature of objects.

Fig. 1. Comparison of different processing schemes

1. ábra. A feldolgozási eljárások összehasonlítása

Рис. 1. Сопоставление различных режимов обработки

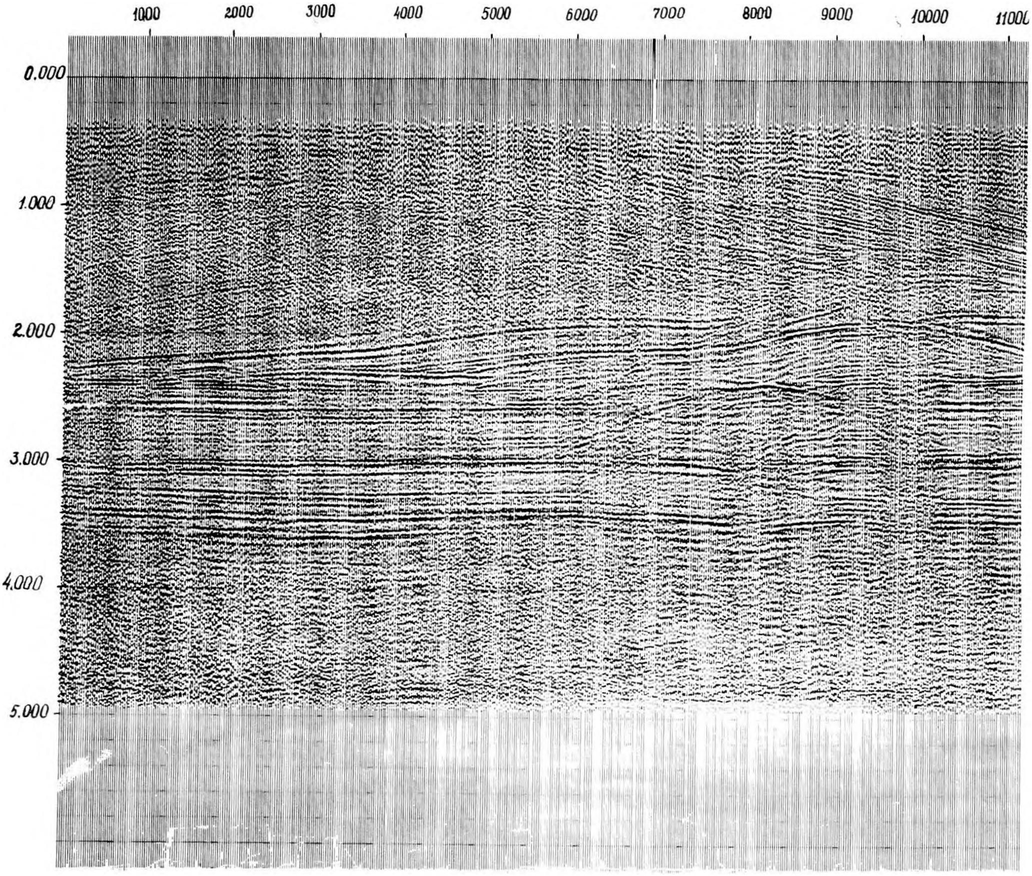

Fig. 1a. Standard processing

1a. ábra. Az időszelvény rutinfeldolgozása

Рис. 1a. Стандартная обработка

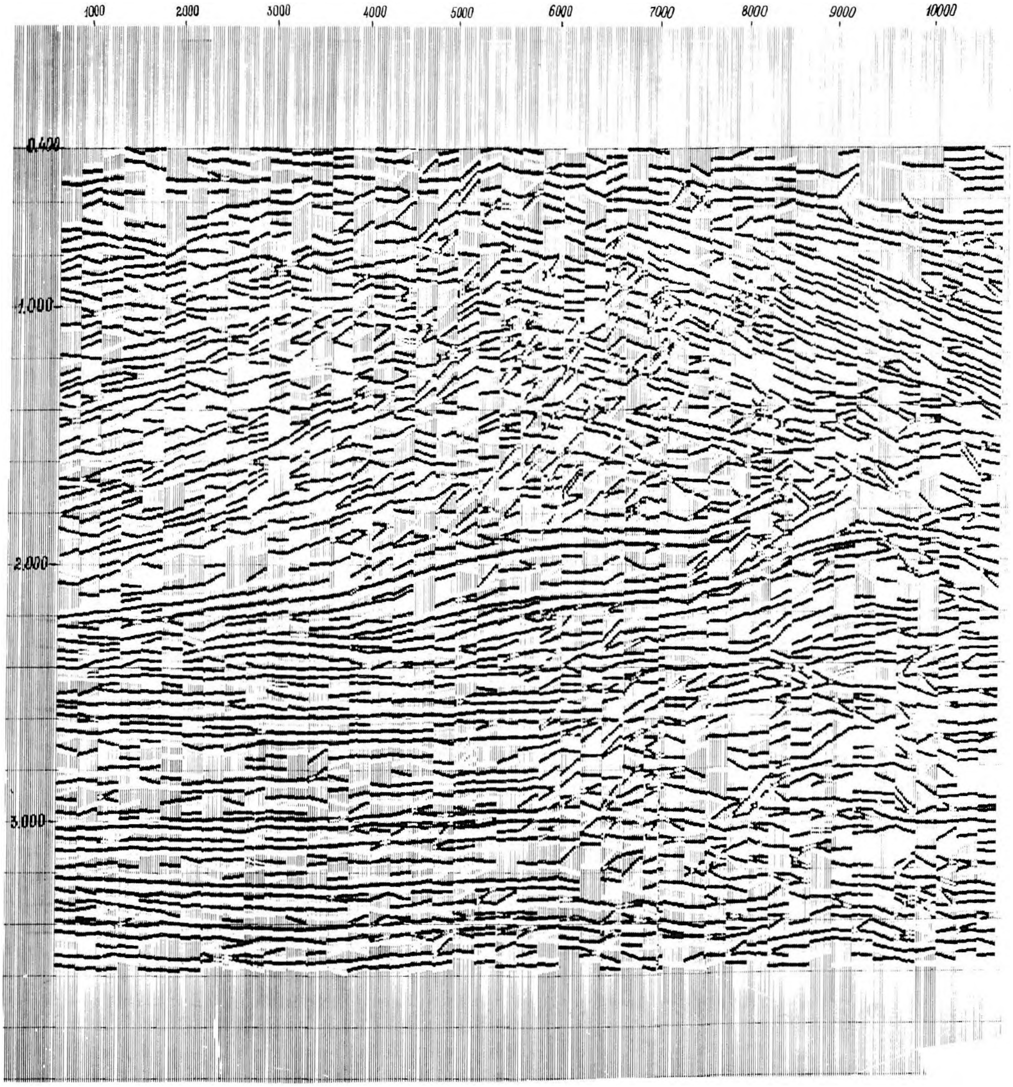

Fig. 1b. Results of signal enhancement in time sections (window of estimating noise dispersion is 1600 ms)

1b. ábra. Az időszelvény feldolgozása 1600 ms-os időablakban végzett zaj-szórás meghatározással

Рис. 1b. Результаты выделения волн по временному разрезу (оценка дисперсии помехи в окне 1600 мсек, сигнал во времени не суммируется)

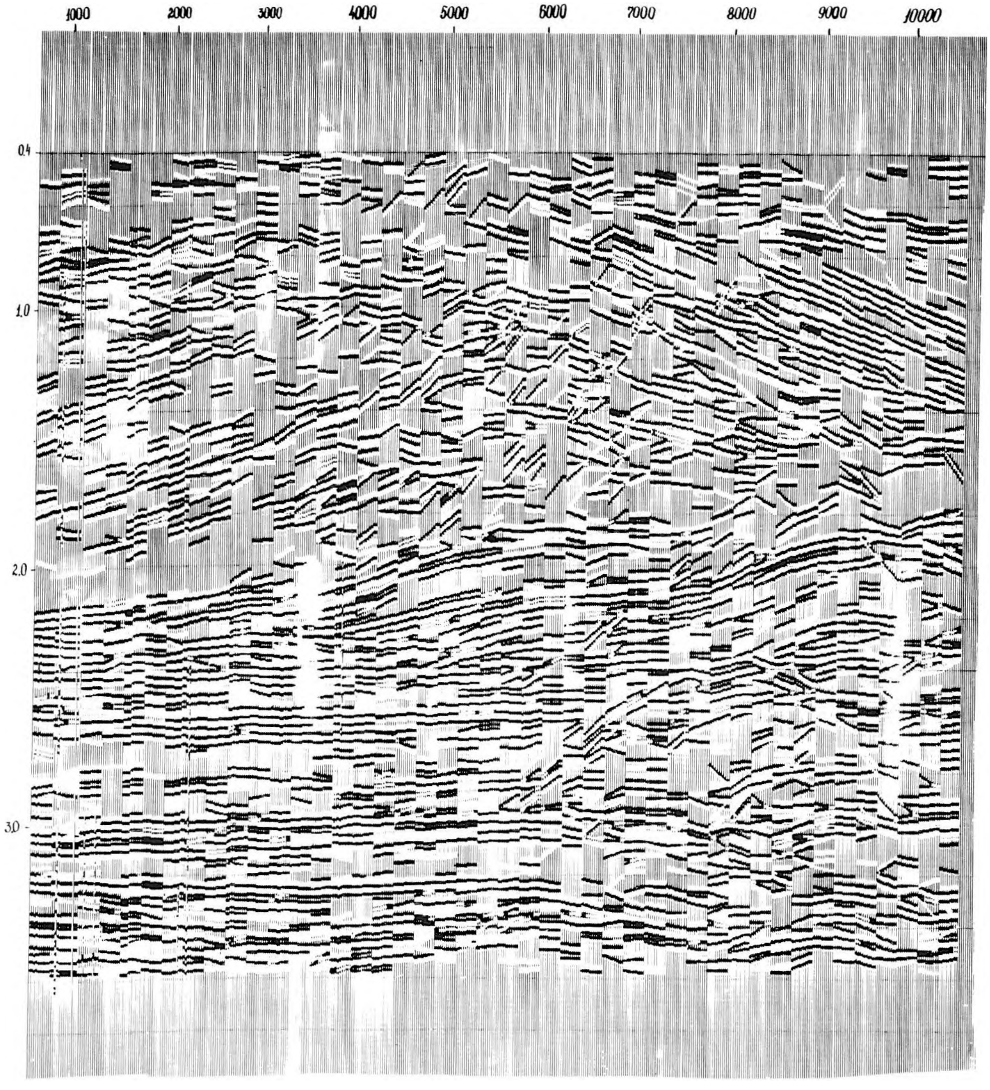

Fig. 1c. Results of signal enhancement in time sections (window of estimating noise dispersion is 1600 ms, window of signal summing is 40 ms)

1c. ábra. Az időszelvény feldolgozása 1600 ms-os időablakban végzett zaj-szórás meghatározással, valamint 40 ms-os ablakban végzett jelösszegezéssel

Рис. 1c. Резултаты выделения волн по временному разрезу (оценка дисперсии помехи в окне 1600 мсек., окно суммирования сигнала: 40 мсек)

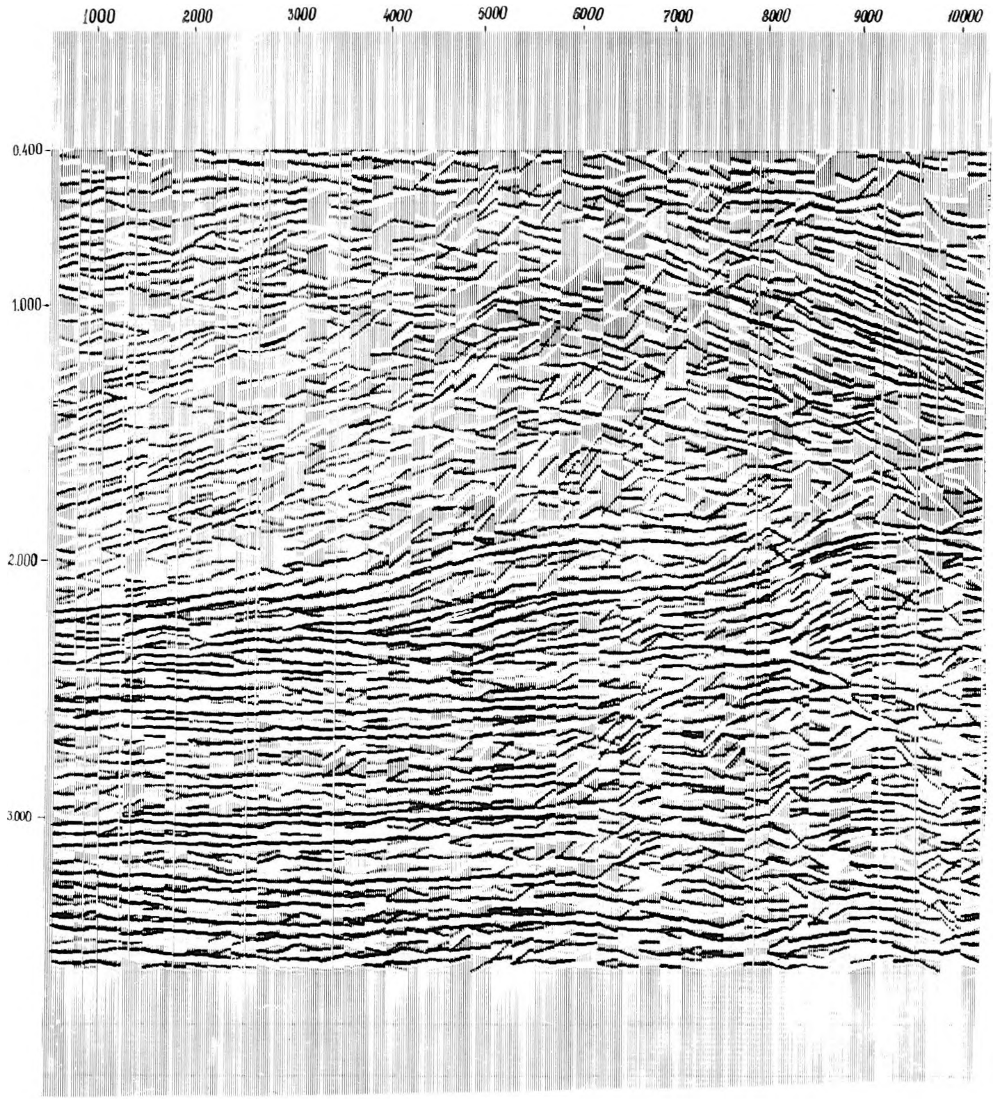

Fig. 1d. Results of signal enhancement in time sections (noise dispersion and signal energy are estimated in the same window 1600 ms)

1d. ábra. Az időszelvény feldolgozása 1600 ms-os időablakban végzett zaj-szórás és jelenergia meghatározással

Рис. 1d. Резултаты выделения волн по временному разрезу (оценка дисперсии помехи и енергии сигнала в одном и том же временном окне; суммирование во времени не производится)

3. Determination of inhomogeneities in the section

In complicated geoseismic conditions, in the presence of noise the detection of inhomogeneity boundaries seems to be difficult. In this case, in order to determine the boundaries where the lithological composition changes, one needs, to search for the boundaries corresponding to the changes of integral characteristics of a record in various areas of the sequence under study. Accordingly, it is assumed that information on the properties of rocks is contained in seismograms not only from the boundaries where lithological composition changes but from the regions between the boundaries as well. The parameters essentially changing their values when crossing an interface are as follows: regular wave amplitude attenuation, ratio between time and amplitude of separate phases, duration of periods, velocities, etc. Variation in the values of these parameters may be obscured by noise and are shown only as various distributions of parameter values on both sides of the interface. The search for an interface can be accomplished by sorting out possible interface positions in a sequence, calculation of values of parameter distribution at each step on both sides of the interface and determination of the position corresponding to the greatest difference in distribution. Assuming equal covariance matrices for the distributions, the so called Machalonobis' distance can be used to express their difference. However, data analysis on the areas of inhomogeneities displays an essential difference in parameter variances and in correlation between separate parameters on both sides of the interface. Therefore, in order to estimate the degree of distribution variance when searching for inhomogeneities, it is advisable to apply probability measures regarding possible differences in covariance matrices. The methods of multivariate statistical analysis are often used in cases when learning samples with known classification are available. However, it seems impossible to create required learning samples (training sets) for solving the most important problems of subdividing objects into classes. Due to costly drilling and stripping work, objective identification of geophysical fields with mineral resources is possible only at some points of the territory; thus, the required typical learning samples cannot be ensured.

The frequency of correct attribution of an arbitrary spatial point to the areas being investigated can be considered as a measure allowing differences in covariance properties of distributions and not requiring samples with known classification (so called learning procedure without teacher). Establishment of the classes is performed using the rule elaborated with the help of the values of sample moments of distribution.

Let us consider an example of running the searching programme for inhomogeneities in some portions of the time section. The areas with relatively simple elliptical or partly hyperbolic boundaries served as models of anomalies. The search for an anomalous portion of the section was performed by means of the sorting out of all possible boundaries of the 2nd order. For each subdivision a decision rule was elaborated that effectively recognized in the parameter space the alternative sets obtained at the current step of selecting and achieved degree of differentiation was calculated. The rule being elaborated in the course of learning is the optimum one among all possible types of decision rules for the case of arbitrary distributions with the two lowest moments coincident with their typical sample estimates from learning material.

The algorithm for the search for inhomogeneities includes:

- setting up the type of subdivision of observated space into alternative parts;
- calculation of statistical characteristics within these parts: sample mean and covariance matrices;
- calculation of coefficients of a linear discriminant function taking into account possible distinguishing of covariance matrices in alternative classes;
- calculation of frequencies of recognition errors of the 1st and 2nd types.

An example for one of the lines on the Jetybai oil field is shown in *Fig. 2.* The zone of seismic section, 10,000 m long (along the line) and 1000 m deep, included 882 observations of parameter vectors containing the following components: mean-interval coefficient of seismic wave attenuation, mean-interval energy, apparent electric resistivity, estimate of mean cross correlation between trace pairs of seismograms, gravity anomaly. Sample positions of boundaries (horizontal, inclined and curved) are shown in Fig. 2. Among the versions of the boundaries separating the section into left and right parts, two patterns having poor correlation with the geological data characterized by error probabilities of 0.43—0.49 are approaching an upper limit of 0.5. It means that the statistical distributions of features in such patterns hardly differ from one another and, consequently, cannot correspond to any inhomogeneities in the section. The boundary with a frequency estimate of 0.18 corresponds to one of the most abrupt changes in the set of parameters under consideration.

The obtained results are confirmed even in the case of a smaller number of samples as well as in the case of a reduced number of features. In the latter case, the efficiency of recognizing the alternative classes and, accordingly, the reliability of distinguishing the boundary corresponding to inhomogeneity decreases. However, there are some combinations of a small number of features for which the efficiency of recognition only slightly differs from the limiting one. For instance, the combinations of such parameters as mean energy and the estimate of mean cross correlation or apparent electrical resistivity and the estimate of mean cross correlation differ in efficiency from a complete set of features by 0.02 and less. It is clear that the majority of boundaries corresponding to increased values of differentiation efficiency are characterized by a certain persistence and repeated shape. It should be noted that the discriminant method of recognizing the inhomogeneities does not require a strict correlation between the parameters and the points on a section. Some shift of the values of a separate element in the distribution space (on a section plane) relative to the true distribution will lead to decreased correlation between a given feature and the features identified by other geophysical methods; as a consequence, the maximum difference between the features which is the basis for distinguishing the physical object in question will be less pronounced. If the inhomogeneity is determined in spite of a shift, it can be said with confidence that it would also be determined using features better correlated with the particular points of the section. The above-mentioned facts supports the application of approximation methods when identifying features. Once an inhomogeneity is identified, specification of its position can be performed by means of the more precise processing of each geophysical method.

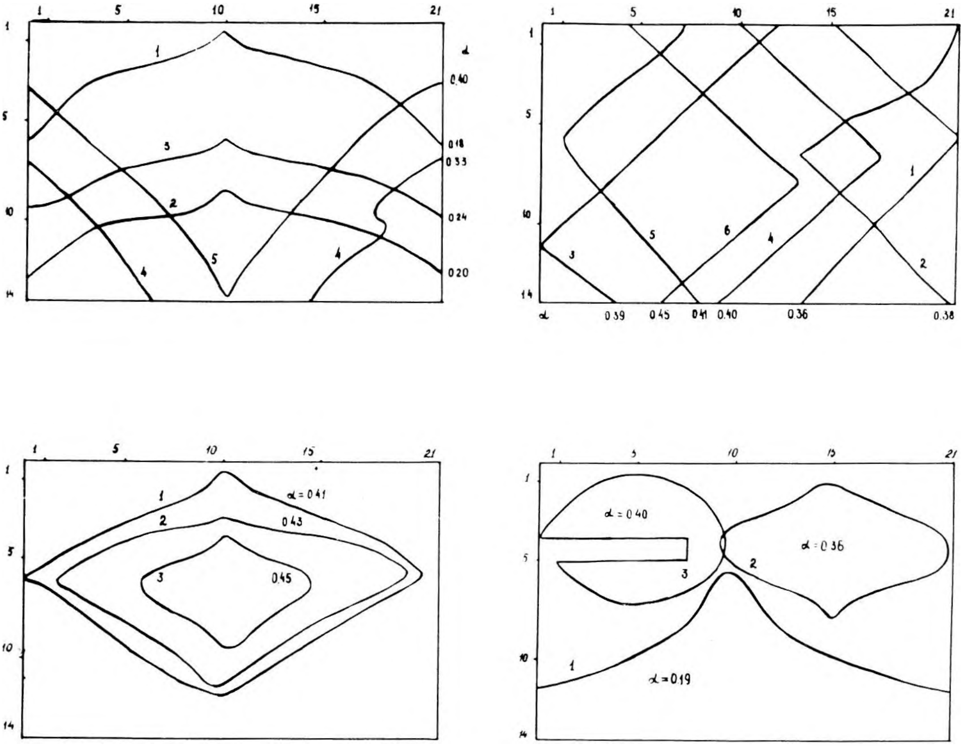

Fig. 2. Variants of dividing a cross-section into homogeneous zones
α — efficiency of dividing the cross-section into two zones (cummulative error probability)

2. ábra A heterogén szelvény zónákra osztása
α — a két zónára történő osztás hatékonysága (halmozott hibavalószínűség)

Рис. 2. Варианты разбиения участка разреза на зоны
α — эффективность разбиения (суммарная вероятность ошибки)

The absence of *a priori* information on the location of intervals possibly associated with lithological inhomogeneities necessitates subdivision of the whole medium to statistically homogeneous zones. It is difficult to carry out statistical analysis because neither the number of objects, nor their boundaries and statistical characteristics are known beforehand. Sometimes, a preliminary analysis of the obtained field of features can be performed by plotting a bar chart of their distribution on some selected intervals in a few portions of the line. If we do not know which elements form a class of objects and at the same time the average values are also unknown and can vary within the limits of a particular population of a single class, the procedure of gradual selection of elements assigned to the same class is required; this is performed by tracing [HOLTZMAN 1981, KUT'INA 1982]. Tracing reduces the problem of partitioning the space of features with unknown number of classes and unknown statistical characteristics to the problem of alternate scanning of possible combinations of feature values beginning with various reference points, calculation of the current average value and deciding about attributing alternative values of the features to one or another set of reference values. In more simple cases, when the characteristics of the objects are considered to be invariant within the limits of the classes and there is no restriction to attributing features to some local area of the space, only the problem of classification with an unknown number of classes remains. The most simple problem involves a known number of classes, namely two classes. The other problems are usually reduced to the above mentioned one, i.e. they are solved by means of multiple checking of two alternatives.

4. Localization of spatial anomalies

Localization of anomalies on the area is connected with the same problems but they are solved in the space of greater dimension. When tracing, the problems of approximation designed for calculating variable mean values, detection of a regular background, etc. are of particular importance [TROYAN 1982]. Tracing makes it possible to locate approximately the boundaries of spatial blocks having similar statistical properties. Then the problem concerning the specifying of object boundaries is to be solved: whether two objects under consideration belong to one or two classes (tracing through break, location of an inhomogeneous zone between the boundaries being identified, assignment of two separately traced phases to the same boundary etc.), whether a certain object could be subdivided into two (checking the uniformity of distribution of feature values), assignment of the remainder of single elements to one of the distinguished classes or neglecting these elements.

To a greater degree than the problem of distinguishing inhomogeneities on a section, the problem of localization of spatial anomalies is connected with the necessity to perform simultaneous analysis of the distribution of a great number

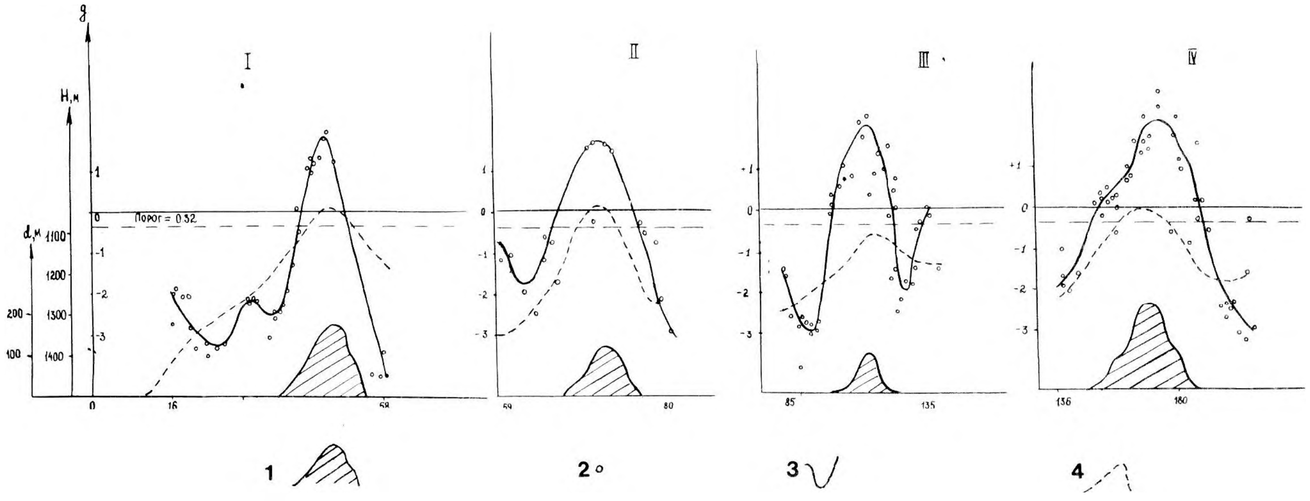

Fig. 3. Results of two-class grouping of data by discriminator function method
1 — reservoir thickness (d); 2 — values of discriminator function at separate points;
3 — smoothed values of discriminator function;
4 — depths of structures

3. ábra. A diszkriminátor-függvény módszerrel végzett kétosztályos felbontás eredményei
1 — tároló vastagság; 2 — a diszkriminátor függvény értékei; 3 — a diszkriminátor függvény simított értékei; 4 — a szerkezetek mélysége

Рис. 3. Результаты классификации по минимаксному решающему правилу
1 — мощность залежи (d); 2 — значения дискриминантной функции в отдельных точках;
3 — сглаженные зхачения дискриминантной функции; 4 — глубина залегания структуры

of feature values; this is too great a task for an interpreter. Application of discriminant functions for this problem allows one to reduce the whole set of multidimensional data into a one-dimensional file of discriminant function values. As in the previous case, the problem can be reduced to two alternatives assuming that the field under study could be subdivided into areas characterized by different statistical properties, discriminant function coefficients can be calculated and a chart of their values plotted. However, even the solution of this simple problem with known but different covariance matrices leads to quadratic rules which require a great number of operations for recognizing procedure, and they give unstable classification. A linear decision rule can be constructed only on the basis of an interation procedure [HOLTZMAN 1981].

Let us consider an example of formalized subdivision of territories into areas associated with the presence or absence of raw minerals. Complex studies with application of the methods of seismic, electric and gravity surveys as well as geochemical methods were carried out along the lines on the Jetybai field. The following elements were subjected to processing: mean value of an absorption coefficient of waves reflected from a reference boundary, number of reflections in a time interval, time of last reflection, density of events in time, value of apparent resistivity and its derivative, gravity anomaly, recalculated vertical gravity gradient, beta-activity from geochemical data.

Subdivision into alternative classes was performed according to preliminary geological interpretation on the basis of well data, and processed reflection data on structures resulted in the determination of the total thickness of the reservoir along individual lines. The total number of observation points was 175; for the learning process 12 points belonging to the class of reservoir and 27 points attributed to the area outside the reservoir were used. Using the values of sample moments, the coefficients of the linear minimax rule were calculated.

Figure 3 shows the results of recognizing all 174 points along individual lines. The resulting classification corresponds to the initial geological classification in 87% cases. The decision rule coefficients were varied in different combinations in the vicinity of obtained values with increments of ± 10%. Accordingly, empirical frequencies of correct recognition changed by not more than 2—3% thus pointing to the stable nature of the procedure for creating the decision rule.

Analysis of plots enables one to draw the following conclusions: 1) there are practically no errors when using the learning material; 2) recognition errors are generally present in a transition area between the reservoir and the outer part, in relation to the preliminary partitioning; 3) the decision rule values along the lines follow the behaviour of curves of the reservoir thickness and the depth of the structure and, hence, they could be used for approximate prediction of these values.

In *Figure 4* smoothed values of the decision rule are shown plotted as levels with similar values. It can be seen that recognition makes it possible to correct the prediction maps as well as to elaborate a quantitative measure of the degree of formal separability of the classes under consideration.

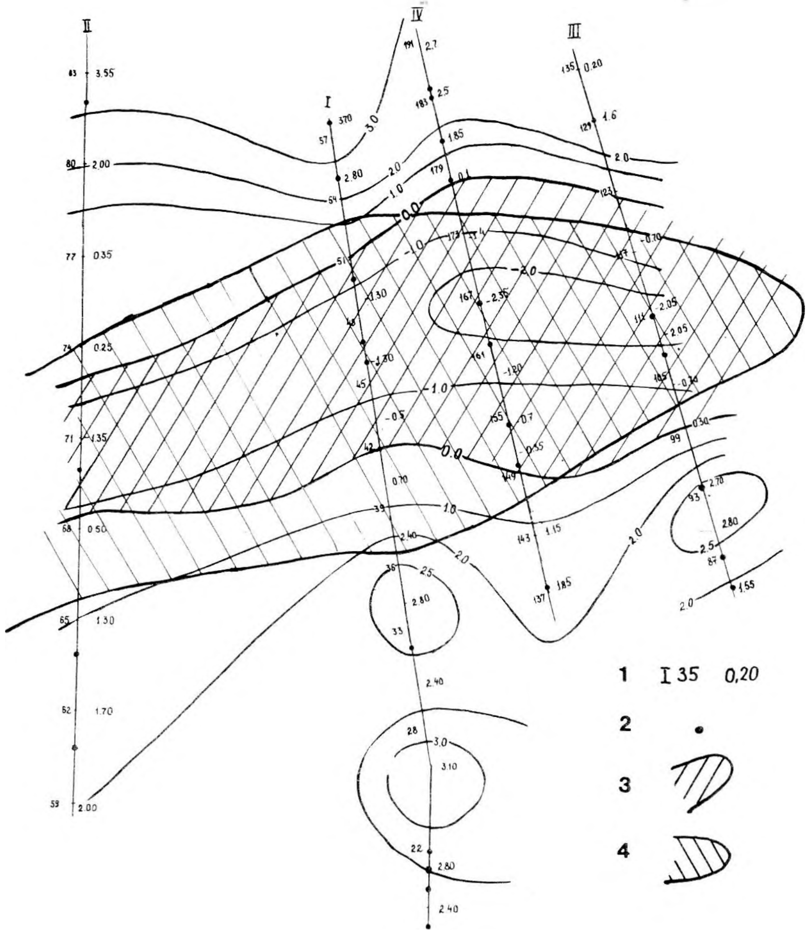

Fig. 4. Improvement of reservoir positioning
1 — pegmarks and values of discriminator function (g); 2 — points used for learning; 3 — reservoir outline from geological data; 4 — computed reservoir outline (zero level of discriminator function)

4. ábra. A tároló helyzetmeghatározásának javítása
1 — karószám, illetve a g diszkriminátor függvény értéke; 2 — a tanuláshoz használt pontok; 3 — a tároló határvonala fúrási adatok alapján; 4 — számított tároló határvonal (a diszkriminátor függvény zérus nívója)

Рис. 4. Уточнение положения залежи
1 — №№ пикетов и зхачения дискриминантной функции; 2 — точки, использовавшиеся для обучения; 3 — контур залежи по геологическому прогнозу; 4 — рассчитанное положение контура залежи (нулевой уровень значений дискриминантной функции)

5. Conclusions

A study of the information capacity of various groups of features for separation of alternative sets was carried out. The values associated with the probability of correct separation of two sets with partly known distributions was considered as a measure of information capacity. Possible combinations of features for small extension of groups were studied. Groups of features with high efficiency of separation have been distinguished. The best combinations containing two or three features are based on all the types of prospecting. It can be seen that the high efficiency of integration is not connected with the predominant application of any single feature. When simultaneously using the data of electrical and seismic prospecting, both the number of reflections in a time interval and the time of the last reflection have the same efficiency. When combining seismic, gravity and geochemical methods, the above mentioned elements from seismic data or the density of events in time could be applied. In all cases, any combination of features which is sufficiently effective uses the data obtained from several geophysical methods.

REFERENCES

HOLTZMAN F. M. (Ed.) 1981: Statistical interpretation of geophysical data. L. LGU, 256 p.

HOLTZMAN F. M. 1981: Statistical algorithms for tracing and classification. Izv. AN SSSR, ser. Fizika Zemli, **4**, pp. 58—72

KUT'INA O. G. 1982: Elaboration of statistical algrithms for processing and interpretation of seismic data, 165 p., Nedra, Moscow

TROYAN V. N. 1982: Statistical methods for processing seismic information in interpretation of complicated areas. 202 p., Nedra, Moscow

REFLEKTOR-KIMUTATÁS ÉS ANOMÁLIA-AZONOSÍTÁS MINŐSÉGÉNEK JAVÍTÁSA STATISZTIKUS MÓDSZEREKKEL

F. M. Holtzman, E. A. Kozlov, O. A. Potapov, O. G. Kut'ina, V. N. Troyan

A szerzők szerkezeti és litológiai paraméterekben jelentkező anomáliák azonosításával foglalkoznak. Egy új módszert javasolnak az anomáliák helyének az időszelvényen történő meghatározására. Egy olajtároló meghatározásának példáján mutatják be a sokváltozós diszkriminátor analízis módszerével végzett időszelvény-feldolgozás eredményét.

ПРИМЕНЕНИЕ СТАТИСТИЧЕСКОЙ ТЕОРИИ ДЛЯ УЛУЧШЕНИЯ ПРОЦЛЕЖИВАНИЯ ОТРАЖАЮЩИХ ПОВЕРХНОСТЕЙ И ИДЕНТИФИКАЦИИ АНОМАЛИЙ

Ф. М. ГОЛЬЦМАН, Е. А. КОЗЛОВ, О. А. ПОТАПОВ, О. Г. КУТЬИНА, В. Н. ТРОЯН

В работе обсуждаются вопросы идентификации аномалий параметров, связанных со структурными и литологическими свойствами. Предлагаются способы прослеживания (классификации) отражающих поверхностей по сейсмическому разрезу. Приведены некоторые примеры применения алгоритмов для отыскивания неоднородностей по сейсмическому разрезу и для выделения местонахождений залежи нефти ха основе методов многомерного дускриминасионного анализа.