DIMACS Technical Report 99-3
February 1999

Massive data set analysis in seismic explorations for oil and gas in crystalline basement interval

by

Ilya Muchnik[1]
DIMACS, Rutgers University
P.O. Box 8018, Piscataway,
NJ 08854-8018, USA
muchnik@dimacs.rutgers.edu

Vadim Mottl[2]
Tula State University
Lenin Ave. 92, 300600 Tula, Russia
mottl@atm.tsu.tula.ru

Vladimir Levyant
Central Geophysical Expedition
Narodnogo Opolcheniya St. 40/3, 123298 Moscow, Russia

[1]Assistant to the Director
[2]DIMACS Visitor

DIMACS is a partnership of Rutgers University, Princeton University, AT&T Labs-Research, Bell Labs, Bellcore and NEC Research Institute.

DIMACS is an NSF Science and Technology Center, funded under contract STC–91–19999; and also receives support from the New Jersey Commission on Science and Technology.

ABSTRACT

On the basis of the optimization-based approach to the analysis of massive ordered data sets, a new method is proposed for computer-aided interpretation of seismic exploratory data from the so-called crystalline basement of the Earth mantle, which underlies the relatively thin sedimentary cover having been, up to now, the almost exceptional object of seismic explorations. The seismic exploratory data sets, seismic sections and cubes, are a class of, respectively, two- and three-dimensional data arrays, which are analyzed in the course of gas and oil reserves prospecting with the purpose of studying the structure of the underground rock mass. The seismic data sets consist of synchronous records of reflected seismic signals registered by a large number of geophones (seismic sensors) placed along a straight line or in the nodes of a rectangular lattice on the earth surface. As the source of the initial seismic pulse, usually serves a series of explosions, responses to which are averaged in a special manner. The vertical time axis forming the resulting two- or three-dimensional picture is identified with depth, so that the peculiarities of the reflected signal under the respective sensor carry an information on the local properties of the rock mass at the respective point of the underground medium. In contrast to the above-lying sedimentary cover, the absence of pronounced reflecting surfaces in a crystalline body results in a great difficulty of inferring the geological information from the basement interval of the seismic picture. The new method of seismic data analysis proposed in this work is aimed at finding fractured zones of the basement rock mass capable to accumulate oil or gas. The essence of the method consists in numerical evaluating distinctions in the local spatial texture of the seismic picture that are caused by differences in physical properties of fractured and monolith rock. The problem of estimating the local texture over the whole data array at once is set as that of minimizing an objective function in that the texture model parameters at all the elements of the array occur as its arguments. A special separable structure of the objective function provides a high speed of the optimization procedure.

Contents

1 Introduction

This work continues the investigation started in the previous DIMACS Reports [1, 2] and aimed at creation of a unified mathematical framework for constructing processing algorithms for a wide class of typical problems arising in the practice of the analysis of massive ordered data sets. The original data set was considered as an array of experimentally acquired values of a number of variables each associated with the respective node of an undirected adjacency graph meant to preset the fixed structure of the array. In particular, it is natural to treat a signal on the axis of a discrete argument as an array formed by elementary variables bound up with each other by the chain-like neighborhood relation, and the structure of an image is defined by a lattice of pixel neighborhood, the rectangular or a more sophisticated one.

Among other kinds of ordered data sets, the two- and three-dimensional seismic data arrays were mentioned, so-called seismic sections and cubes. The arrays of this kind, that are analyzed in the course of gas and oil reserves exploration, consist of synchronous records of reflected seismic signals registered by a large number of geophones (seismic sensors) placed along a straight line or in the nodes of a rectangular lattice on the earth surface. As the source of the initial seismic pulse, usually serves a series of explosions, responses to which are averaged in a special manner. Although the initial argument of the seismic signals is time, a relevant preprocessing allows to identify the vertical axis of the two- or three-dimensional picture with depth, so that the peculiarities of the reflected signal under the respective sensor carry an information on the local properties of the rock mass at the respective point of the underground medium.

It is just this latter kind of ordered data sets which is scrutinized in this Report, the third one in the series started by [1, 2]. We pursue here not only the aim of illustrating the mathematical principles proposed for the analysis of massive ordered data sets. Our purpose will be to create, on the basis of those principles, a new method of interpretation of seismic data from the so-called crystalline basement of the Earth mantle, which underlies the relatively thin sedimentary cover having been, up to now, the almost only source of hydrocarbons and, hence, the almost exceptional object of seismic explorations.

What has drawn our attention to the problem of interpretation of seismic explorative data from the basement interval is the fact that the reserves of hydrocarbons in the upper sedimentary cover of the Earth mantle are quickly going exhausted. It has become obvious to date, that a significant increase of prospected oil and gas reserves in the coming XXI century may be provided through assimilating the basement rocks which bed under the sedimentary cover. There are already over 200 commercial oil and gas accumulations found in the basement. In the opinion of a number of leading geologists, it may give a second wind to the oil and gas industry.

However, seismic data have been used, up to now, solely for explorations in the sedimentary rock mass [4, 5, 6, 7]. A sedimentary massif is characterized by a system of quasi-parallel reflecting boundaries between single rock layers, and all the existing methods of seismic data analysis are aimed on the reconstruction of just this layered structure. The seismic data array registered on the daylight surface is formed primarily by the totality of seismic signals reflected by the boundaries. As a rule, the sedimentary structure is represented in the seismic picture so clearly, that in many cases it easily lends itself to an immediate visual analysis. Due to this fact, the computer-based methods of seismic data processing are aimed basically at a convenient representation of the seismic picture and fulfilling some routine operations.

As to the crystalline basement buried under the sedimentary cover and, respectively, represented in the seismic exploratory data in the interval of greater time values, the absence of pronounced reflecting surfaces makes the traditional visual analysis of the seismic picture ineffective. One of the main geological factors that determine the ability of a basement fragment to bear oil or gas is abundance of the so-called collector zones in it, i.e. zones characterized by an increased hollowness. An analysis of known occurrences in the basement rock mass we undertook, including such large and well-studied ones as La Paz and Mara in Venezuela, White Tiger and Dragon in Vietnam, Bombay High in India, showed that the prospected reserves of oil and gas primarily gravitate to buried basement areas, where forming of collector zones is provided by tectonic fracturedness and hydrothermal activity which leads to arising of caverns.

However, at the present time there are only rare publications on acoustic and seismic propagation in porous media [8], which do not make any commonly adopted theory of the seismic field in such bodies. As a consequence, the methods of interpretation of seismic exploratory data, which are the most preferable source of information on the structure of gas- and oil-promising objects, are extremely poorly developed for massive crystalline media.

In this work, we systematically exploit the analogy of seismic data arrays and visual images, but on the basis of a more detailed analysis of texture properties of the seismic picture. In contrast to [5, 6, 7], where coarse texture patterns of a layered seismic picture were analyzed, our aim will be to visualize subtle spatial texture distinctions, hardly discernible by the naked eye, which may be characteristic for local fracturedness of the basement rock. For estimating the local texture properties of seismic sections and cubes we use here the methods proposed in the foregoing Reports [1, 2] for the analysis of massive ordered data sets.

The class of data analysis problems under consideration was outlined there by the assumption that the ultimate aim of processing can be represented as a transformation of the original data array treated as an ordered set of variables into a secondary array of the same structure but formed by variables of, generally speaking, different nature, i.e. different ranges. The essence of data processing consists in coordinating two kinds of the available information on the sought-for values of the secondary variables, namely, the data-dependent information usually carried by respective fragments of the original data array, on the one hand, and some a priori neighborhood constraints upon the pairs of adjacent hidden values, on the other. The principal idea of algorithmical providing such a coordination was given more recently the name of the variational approach to the analysis of ordered data [3]. The idea consists in treating the generalized problem of data analysis as that of optimization (minimization or maximization) of a real-valued objective function of all the hidden variables. The objective function is assumed to be separable, i.e. to consist of additive constituents of one or two arguments associated, respectively, with the nodes and edges of the graph of the pair-wise adjacency of data set elements, which is meant to represent the structure of the data set.

Such a generalized statement covers many practical problems of signal and image analysis, for instance, those of segmentation, smoothing, matching of two or several signals or images, time frequency analysis of signals, inhomogeneous texture analysis of images. In the problems of seismic data analysis considered in this work, it is required to estimate some hidden characteristics of the rock mass at all the points of the underground medium. As those points are associated with the respective elements of the data array, the problems of such a kind also fall into the class outlined in the Reports.

As numerical local characteristics of the seismic picture, supposedly carrying some information on the fracturedness of the crystalline basement around the respective points of the underground medium, we use here parameters of an appropriate mathematical model of the spatial texture formed by ordered seismic signals. We obtain first fuzzy preliminary estimates on the parameters at all the elements of the array (nodes of the rectangular neighborhood graph) in the form of the set of local (node) elementary objective functions, and then find their final coordinated values with respect to a priori smoothness constraints imposed upon each pair of neighboring elements of the array (edges of the neighborhood graph) in the form of appropriate penalty (edge) functions. To find the coordinated estimates, it is sufficient to minimize the full objective function which is built as sum of all the node and edge functions and includes all the local texture parameters as its arguments. The necessity of coordinating (generalized smoothing) of preliminary local estimates stems from the fact that each of them is inferred from a very small elementary fragment of the array and, hence, is subject to inevitable interferences of various kinds.

Such a processing substitutes the original seismic data array for another one which consists of vector estimates on the local texture parameters. We don't consider here the problem of visual representation of the resulting vector field, in the simplest case it may be represented as several scalar ones. What is essential here is the fact that, when the texture model is chosen appropriately, the totality of the secondary vector fields will make discernible even subtle distinctions in the spatial mode of signal oscillations. In this Report, we use spectral properties of one-dimensional sections through the three-dimensional undersurface medium for a quantitative evaluation of the spatial texture.

Once the transformation of the original data array into the vector one of local texture parameters is done, it remains only to decide what changes are characteristic for the sought-for areas if increased hollowness. In this work, we content ourselves with a qualitative comparison of the estimated spectral parameters of the seismic picture with experts' assessment of the local oil-bearing ability of the basement rock. We present here the first positive results of application of the new method of seismic data processing to studying the crystalline basement of a concrete oil field, namely, Bombay High in the Arabian Sea on the coastal shelf of India.

As to establishing more justified relations between local texture properties of the seismic picture and the presence of collectors in the crystalline basement, this question falls, actually, into the competence field of pattern recognition. As the source of trainer's information, may serve investigations in exploratory wells, which are at present the main method of estimating collecting properties of the basement rock mass, though, an extremely expensive one. The inevitably sparse net of wells does not allow for extending, with a sufficient reliability, the results of such pointwise estimates onto the entire volume of the potential reservoir being investigated. The principles of pattern recognition can be used in the course of prospecting for oil and gas for combining indirect data of superficial seismic observations with direct information on the structure of the crystalline basement obtained from a relatively sparse set of exploratory wells . We hope that elaboration of a theoretical and algorithmical framework for recognition of collector areas in the crystalline basement will make the subject of a special investigation.

2 Seismic picture of the underground medium as a texture image

2.1 Registration and structure of a seismic exploratory data array

A two- or three-dimensional seismic data array, respectively, seismic section or cube, consists of synchronous records of reflected seismic signals registered as functions of time by a large number of geophones (seismic sensors) that are placed along a straight line or at the nodes of a rectangular lattice on the earth surface. The principal scheme of forming reflected seismic signals is illustrated by Fig. 1. As the source of the initial seismic energy, usually serves a series of explosions on the surface, in turn, at all the points where the sensors are located. Each pulse produces elastic oscillations of the rock mass that propagate through the underground medium, reverberate from sufficiently large heterogeneities, and return to the daylight surface, where they are synchronously registered, every time, by all the sensors at once. The method of seismic explorations was invented for reconstructing, primarily, the layered structures, therefore, it is meant that seismic waves are reflected, in the first line, by boundaries between geological layers with different mechanical properties.

So, in the course of a seismic experiment, every sensor registers several seismograms produced by pulses at all the points where the sensors are located. But in the final array subject to an analysis, each sensor is represented by only one signal obtained as result of averaging several seismograms from different sensors, including this sensor itself, and different pulses. One of the most spread method of averaging is the so-called common reflection point technique [9, 10]. The essence of the technique consists in that, first, those of seismograms are averaged which have been reflected from supposedly the same point on an underground boundary and, second, the scale of time is corrected in each seismogram so that the run time of the seismic wave from the source of the pulse to the respective sensor would be the same in all of them (Fig. 2). The application of the common reflection point technique presupposes that the sound speed is known for all the layers forming the underground structure, but there are methods of estimating it from the same seismic data, what is often possible by virtue of their redundancy.

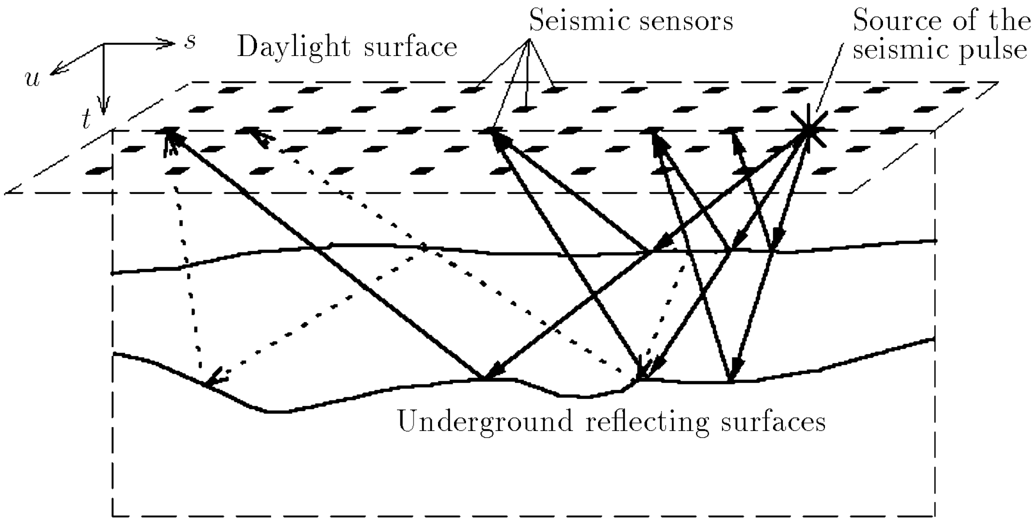


Figure 1: Simplified model of propagation and reflection of seismic waves in a layered rock mass. The solid lines designate ways of the first reflections from boundaries, and the dotted ones show an example of the stray (second and third) reflections to be suppressed.

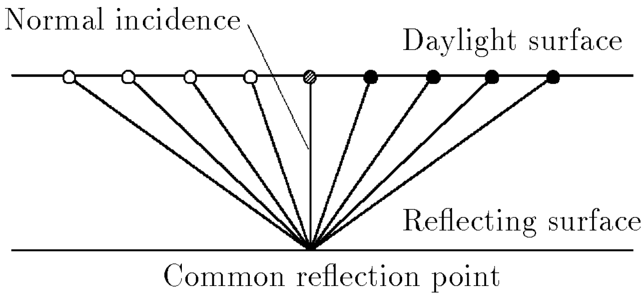


Figure 2: Averaging of seismograms by the common reflection point method.

After such a processing, every synthetic seismogram represents primarily the rock structure along the vertical line just under the seismic sensor which this seismogram is associated with, so, it becomes possible to identify the time axis of the synthetic seismogram with depth under the respective sensor. In addition, the stray multiple reflections of the seismic waves get suppressed to a great extent, because, due to the time correction, the first reflections are summed in phase whereas the phases of multiple ones are, as a rule, different at the point of summation.

The resulting data set consists of synchronous synthetic seismograms $\mathbf{y}_{su} = (y_{sut},\ t = 1, \ldots, N^v)$ registered as a real-valued function of time. Each seismogram is associated with a certain point on the daylight surface, respectively, $(s = 1, \ldots, N^l,\ u = \text{const})$, when geophones are located along a line, and $(s = 1, \ldots, N^l,\ u = 1, \ldots, N^f)$ in the case of lattice-like allocation. Here the upper indices v, l and f denote, respectively, the vertical, lateral

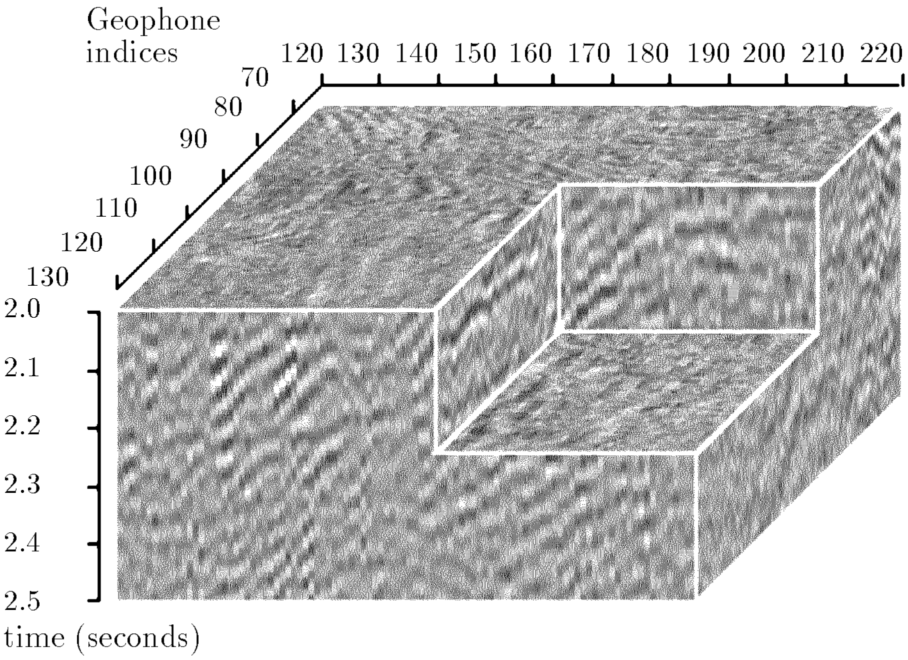


Figure 3: Fragment of a seismic cube.

and frontal axis of the array. Being placed vertically side-by-side, the seismograms form the vertical rows of a seismic section or cube $\mathbf{Y}_u = (\mathbf{y}_s,\ s = 1,\dots,N^l,\ u = \text{const}) = (y_{\mathbf{t}},\ \mathbf{t} = (s,u,t) \in T_u)$, $y_{\mathbf{t}} \in \mathsf{R}$, $T_u = \{\mathbf{t} = (s,u,t)\colon\ s = 1,\dots,N^l,\ t = 1,\dots,N^v,\ u = \text{const}\}$ or $T = \{\mathbf{t} = (s,u,t)\colon\ s = 1,\dots,N^l,\ u = 1,\dots,N^f,\ t = 1,\dots,N^v\}$. In spite of the non-visual origin of seismic data, it is the practice to represent them in a pictorial form, frequently, by displaying the signal value at each point of the array as darkness level of the respective point in the image (Fig. 3).

As we mentioned above, the vertical axis of time can be interpreted as that of depth under the assumption that the sound speed is known at all the points of the massif being examined. Each element of the array corresponds to the point under the earth surface whose horizontal coordinates coincide with those of the respective seismic sensor and the vertical one can be calculated from time. So, the seismic array offers, to a certain extent, an immediate model of the underground medium. However, it should be kept in mind that such a treatment of the seismic data array is based on the assumption that only one reflection from each inhomogeneity is represented in the signals. If multiple reflections are not completely suppressed, the peculiarities of upper layers will affect the seismic signals also after the time that corresponds to the depth of their bedding and hinder the interpretation of the seismic picture produced by the underlying rock.

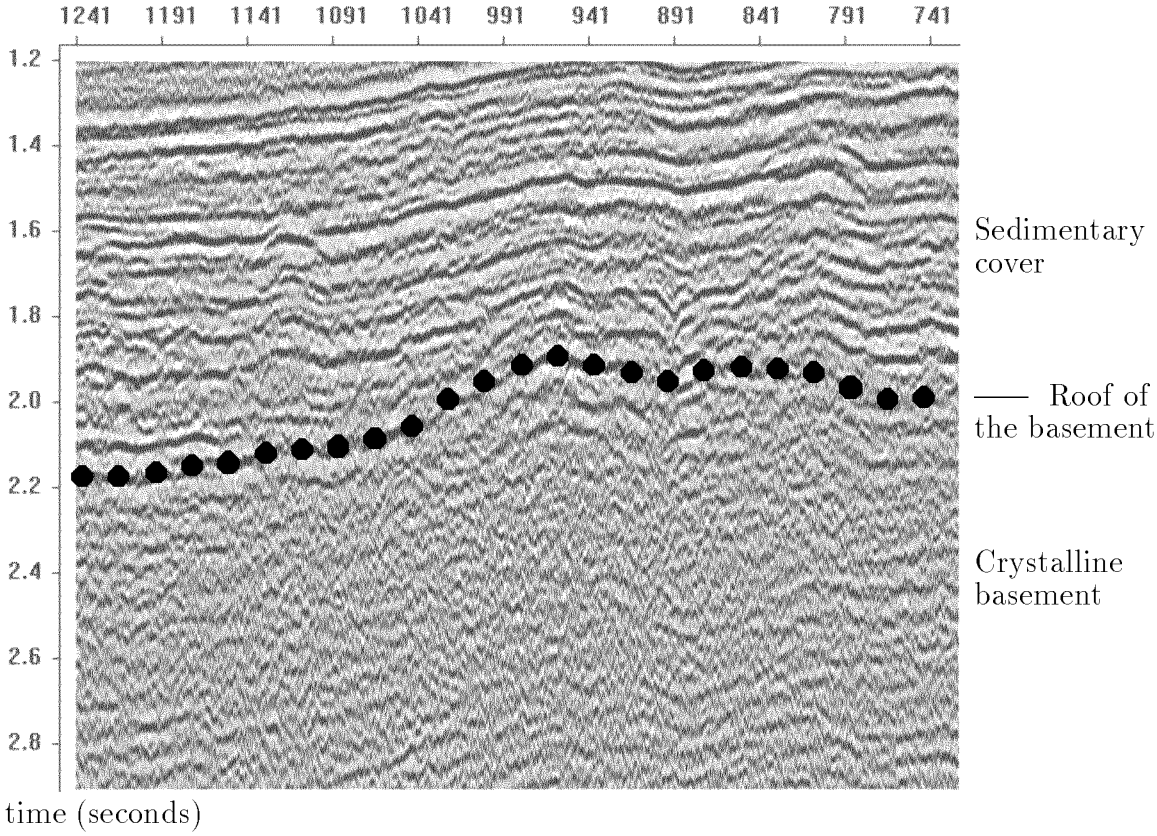


Figure 4: Seismic section through sedimentary cover and crystalline basement.

2.2 Seismic pictures of the sedimentary cover and crystalline basement of the Earth mantle

As the crystalline basement is buried under the sedimentary cover, it is represented in the seismic picture at greater time values. These two main parts of the upper Earth mantle are substantially different in the geological structure. Respectively, the mode of spatial signal oscillations that forms the pictorial view of the data array is markedly different above and under the time boundary corresponding to the roof of the basement. The difference between the picture patterns is well seen in Fig. 4 that shows a seismic section through both the sedimentary and basement time interval.

A sedimentary geological structure consists of several rock layers that bed upon each other and are much more stretched in the quasi-horizontal direction than in the vertical. As compared to the difference between mechanical properties of neighboring layers, each of them can be considered as an approximately homogeneous body. An incident seismic wave is reflected primarily by boundaries between the layers, whereas their interior is almost transparent for it. As a result, the idealized record of the reflected seismic signal registered

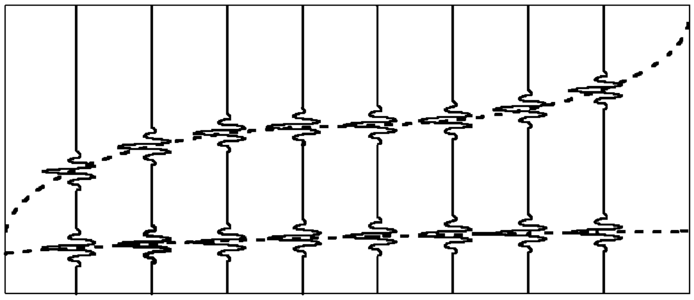


Figure 5: Blurred representation of boundaries between rock layers as a system of wavelet-like pulses in reflected signals forming the seismic picture.

by a seismic sensor will have the form of a sequence of pulses spaced just at the intervals proportional to the depth of the respective strata if measured along the vertical starting from this sensor.

In the two- or three-dimensional seismic data array, the pulses reflected by the same boundary in signals placed vertically side-by-side in accordance with the location of sensors on the earth surface will form a quasi-horizontal line or, respectively, surface playing the part of a spatial model of the respective reflection surface in the medium being examined. As a result, the seismic picture in the near (upper) interval of time values consists of well-ordered systems of quasi-horizontal and curvilinear patterns of considerable length formed by rock strata boundaries.

However, in consequence of damping higher frequencies by the medium on the way of seismic wave propagation, even the originally sharp pulse will be inevitably turned to a wavelet-like burst of low-frequency oscillations and, hence, get blurred in time, as it is shown in Fig. 5. In addition, it should be kept in mind that the local scale $\frac{dh}{dt}$ of mapping depth h onto the time axis t depends on the velocity of elastic waves propagation at the respective point of the underground medium.

As to the internal specificity of the rock within the strata, it also affects the seismic response registered on the daylight surface. But the respective fine variations of the signal oscillation mode practically sink in short intervals between quite long and intense bursts of main reflections.

Thus, the amount of information carried by the seismic picture of a sedimentary rock mass contains almost only that on the system of quasi-horizontal reflecting surfaces. Respectively, the analysis of seismic pictures in the sedimentary cover interval is directed primarily towards restoration of its stratigraphic structure. In explorations for oil and gas, the aim of the analysis is finding folds of a specific kind called collectors, in the first instance, convex structures and linear shifts, that might serve as traps for naturally ascending fluids.

The system of quasi-parallel reflecting boundaries that characterize a sedimentary structure is so clearly represented in the seismic picture that in many cases it easily lends itself to an immediate visual interpretation. The part of the computer analysis boils down to a visualization of the data array convenient to the researcher by way of a preprocessing aimed at demagnification of noise, suppression of multiple reflections, improvement of resolution and the like.

The situation is substantially different with the seismic picture in the far (lower) time interval that corresponds to the crystalline basement. There are no pronounced layers with sharp contours in the basement. The possible macro-inhomogeneities of the size comparable to the length of elastic waves, in other words, to the so-called Fresnel zone, are smoothly spread in space, and their reflection ability is much less than that of boundaries between sedimentary strata. The quasi-horizontally stretched patterns seen in the lower part of the seismic section in Fig. 4 are caused by multiple reflections from the sharply pronounced roof of the basement and do not carry any information on the internal structure of it. Except for these interferences that are to be removed by a suitable preprocessing, the ordered seismic signals form a three-dimensional field characterized by homogeneous spatial oscillations in the band of significantly higher frequencies than those characteristic for reflected signals.

The simultaneous presence of regular and chaotic elements and the generally adopted manner of seismic sections' visual representation in the form of a wave picture gives the spatial oscillations a deep analogy with proper images of the so-called texture type, like images of grass, rippled water surface, wooden board. Especially pertinent appears the comparison with the texture seen in the section of a stone, because, a seismic section does, in some more mediated manner, practically the same, namely, shows the hidden quite a complicated structure of a quasi-homogeneous underground medium.

The notion of texture stems from the visual image analysis, therefore, it implies only plane images. Nevertheless, it appears natural to extend it also onto the case of three-dimensional data arrays, despite the fact that only their sections have a visual representation.

However, the possibilities of the visual analysis of a seismic data array in the basement interval, even if it is a plane seismic section considered as a texture image, are quite limited all the same, because the objectively existing fine differences in the wave picture may be hardly discernible by the naked eye. In addition, because of the deep individuality of visual perception, different interpreters may involuntarily pay attention to different texture peculiarities and, thus, differently mark fragments of the picture.

On the force of this circumstance, the problem of computer-based objectification of similarities and differences in local texture properties within the bounds of the basement interval of a seismic data array, as well as in different arrays, appears to be quite actual, maybe, more actual than in the upper sedimentary interval. It is just this problem which is partially solved in this work.

2.3 Fluid collector zones in the crystalline basement and their expected features in the seismic picture

Collecting properties of crystalline massifs consisting of granites, gneisses and other rocks are formed by the growth of fissures caused by tectonic stresses and deformations in the course of geological evolution. The hollowness of the massif (volume of the hollow space) may increase through the circulation of mineralized solutions which gives rise to pores and caverns by hollowing-out some minerals.

The principal idea of our heuristic approach to the analysis of seismic exploratory data aimed at finding collector zones in the crystalline basement is based on the following reasoning. Fractures, pores and caverns create in a crystalline massif a great number of microinhomogeneities, hence, an incident seismic wave will generate in such a medium a great number of micro-diffracted waves or scatter essentially. On the contrary, a monolith homogeneous crystalline body must be almost transparent in the seismic sense. It appears that the high-frequency component of the seismic field, or, in other words, its texture, must be different in monolith and substantially fractured zones. The little size of inhomogeneities in comparison with the seismic wave length does not exclude the possibility to observe the micro-diffracted waves or zones of scattering, like a fish shoal is well seen on the screen of a sonar whereas a single fish is practically not able to reflect the too long acoustic waves. Thus, it must be possible, in analyzing the field of micro-diffracted waves in the respective time interval, to obtain some information on the fine structure of the crystalline massif.

It should to be expected that the fractured areas of an increased hollowness would be characterized by:

- *an increased irregularity of seismic records* as a result of interference of diffracted waves;

- a greater attenuation of high-frequency components in the spectrum of the reflected seismic energy as a result of its scattering, i.e. *an increased relative level of the low-frequency part of the seismic signal spectrum*; as an accompanying feature, *changes in the spectral properties of spatial signals* must occur also in other directions, in particular, *in paleohorizontal sections through the seismic array's volume*, i.e. surfaces quasi-parallel to the upper boundary (roof) of the basement (Fig. 4);

- *an increased level of the seismic energy* reflected by the points of diffraction relative to the homogeneous monolith areas of the basement, because the latter ones must be more transparent.

In this work, the above mentioned effects are quantitatively evaluated, respectively, by the following characteristics of local texture at each point of a seismic array:

- *estimates on the local parameters of a spatial linear autoregression model* of the seismic section considered as a two-dimensional picture;

- *local spectra of both the reflected seismic signals and paleohorizontal sections* through the two- or three-dimensional seismic picture with respect to the roof of the basement being examined;

- *estimates on the local intensity of the reflected seismic signals* forming the seismic section.

So, it is proposed to seek features of the basement rock's local hollowness, i.e. its potential ability to bear oil or gas, among appropriate texture characteristics of the two- or three-dimensional seismic image of the medium being examined. We do this here by way of local identification of a suitable texture model whose parameters $x_{\mathbf{t}}$ are assumed to smoothly change over the basement volume $\mathbf{t} = (s, u, t) \in T$. The result of such a processing will be the array of estimated parameter values $\mathbf{X} = (x_{\mathbf{t}}, \, \mathbf{t} \in T)$ that has the same structure as the original seismic data array $\mathbf{Y} = (y_{\mathbf{t}}, \, \mathbf{t} \in T)$.

Before estimating the basement texture, the seismic picture is to be cleared of the trails of multiple reflections produced by the upper sedimentary layers. These high-magnitude interferences basically repeat the form of the basement roof. We delete or, at least, weaken them by a light smoothing the picture along the paleohorizontal surfaces with subsequent subtraction of the result from the original signals.

3 Data models and algorithms of data processing

3.1 Paleohorizontal alignment of the seismic data array

In the original form, a seismic data array $\mathbf{Y} = (y_{\mathbf{t}}, \, \mathbf{t} \in T)$ is defined on the set of discrete points $T = \{\mathbf{t} = (s, u, t) : \; s = 1, \dots, N^l, \, u = 1, \dots, N^f, \, t = 1, \dots, N^v\}$, where vector $\mathbf{t} = (s, u, t)$ consists of the index of seismic record (s, u) and that of discrete time moment t. When studying the crystalline basement, only part of the time interval is to be processed, namely, starting from the time moment associated with the upper boundary of the basement.

An estimate on the upper boundary of the monolith basement to be examined is inputted manually by an expert as the roof surface t^*_{su}, $s = 1, \dots, N^l$, $u = 1, \dots, N^f$ or, in the case of seismic sections, roof lines, as it is shown in Fig. 4. In what follows, the points of the array are characterized not by the original coordinates $\mathbf{t} = (s, u, t)$ but by those warped in the vertical direction $\mathbf{t}' = (s, u, t')$, $t' = t - t^*_{su}$, so that time is counted in each signal from the preset roof surface or line. So, the totality of all the points at a fixed moment of the resulting relative time through all the signals in the original array $T_{t'} = \{\mathbf{t} = (s, u, t) : \; s = 1, \dots, N^l, \, u = 1, \dots, N^f, \, t - t^*_{su} = t' = \text{const}\}$ will make a surface (line) of the same form as the roof. In the original seismic array, the subsets of such point are called paleohorizons or paleoisochrones. In the array warped in accordance with the new zero value of time, in other words, aligned to the roof, the elements with the fixed value of time will be located at the same horizontal level. It is just this transformed version of the seismic array which will be subjected to all the kinds of processing. For the sake of writing simplicity, we will omit the prime-sign at t' and keep the original denotations, namely, t for the relative time, $T = \{\mathbf{t} = (s, u, t)\}$ for the set of discrete elements after their alignment to the roof, and $\mathbf{Y} = (y_{\mathbf{t}}, \, \mathbf{t} \in T)$ for the aligned array as a whole.

3.2 General outlines of the variational approach to the local analysis of massive ordered data sets

A seismic explorative data array $\mathbf{Y} = (y_{\mathbf{t}}, \, \mathbf{t} \in T)$ is a typical massive ordered data set constituted, in this case, by seismic signal values $y_{\mathbf{t}}$ considered as real numbers $y \in \mathcal{Y} = \mathsf{R}$.

The elements of a seismic data array are ordered along two $\mathbf{t} = (s, t)$ or three discrete axes $\mathbf{t} = (s, u, t)$ and form a function of, respectively, two or three arguments.

What is especially important here is the fact that each element of such an array is naturally associated with a number of its spatial neighbors. It is convenient to express the neighborhood relation on the set of data elements $\mathbf{t} \in T$ in the form of an undirected graph G which is understood as set of neighboring element pairs $G \subset T \times T$. The natural neighborhood graphs for elements of a seismic data array will be two- or three-dimensional rectangular lattice (Fig. 6).

The problem of evaluating the local spatial texture at every point of the seismic picture can be represented as that of transforming the original array $\mathbf{Y} = (y_{\mathbf{t}}, \mathbf{t} \in T)$, $y \in \mathcal{Y}$, into another function $\mathbf{X} = (x_{\mathbf{t}}, \mathbf{t} \in T)$, which would be defined on the same set of the array elements $\mathbf{t} \in T$ and take values $x \in \mathcal{X}$ from the set of all the possible model parameter values.

The generalized approach to such a class of data analysis problems proposed in [1, 2] leans upon the natural and commonly adopted way to create algorithms of data analysis of practically any kind as those of optimization of a suitable objective function defined on a set which is meant to represent the whole variety of feasible results. For the problem under consideration, the result of seismic data analysis is the sought-for secondary array of texture model parameters $\mathbf{X} = (x_{\mathbf{t}},\ \mathbf{t} \in T)$, so, it must occur as argument in an objective function $J(\mathbf{X}|\mathbf{Y})$ that should assess the discrepancy between each admissible version of the result $\mathbf{X}$ and the given data array $\mathbf{Y}$. The range of definition of the objective function is completely determined once the range of the elementary variables $x_{\mathbf{t}} \in \mathcal{X}$ is chosen. Thus, the result of data analysis $\hat{\mathbf{X}}(\mathbf{Y})$ is defined by the condition

$$\hat{\mathbf{X}}(\mathbf{Y}) = \arg \min_{x_{\mathbf{t}} \in \mathcal{X}} J(\mathbf{X}|\mathbf{Y}).$$

Let the local parametric model of the seismic image characteristic under study be chosen. Then, to estimate the vector of its parameters $x_{\mathbf{t}}$ at a certain point $\mathbf{t}$, we have to process the image within a vicinity of this point $\mathbf{Y}_{\mathbf{t}}$ in the original array $\mathbf{Y}$. It is convenient to express a preliminary judgment on the value of a single hidden variable $x_{\mathbf{t}}$ in the fuzzy form of a data-dependent function $\psi_{\mathbf{t}}(x) = \psi_{\mathbf{t}}(x|\mathbf{Y}_{\mathbf{t}})$ which must be defined over the set $\mathcal{X}$ and take the greater value the more evident is the contradiction between the hypothesis that $x \in \mathcal{X}$ is just the "correct" local value we are seeking and the current fragment $\mathbf{Y}_{\mathbf{t}}$ of the data array. Let such functions be called node functions, because each of them is associated with a node of the neighborhood graph on the set of data array elements.

It is clear that the size of the elementary vicinities should be taken sufficiently small for not to come into conflict with the initial assumption that the image characteristic to be evaluated does not remain the same in any region. But in this case practically inevitable interferences of various kinds will prevent from taking the totality of locally optimal values

$$\hat{x}_{\mathbf{t}}(\mathbf{Y}_{\mathbf{t}}) = \arg \min_{x_{\mathbf{t}} \in \mathcal{X}} \psi_{\mathbf{t}}(x_{\mathbf{t}}|\mathbf{Y}_{\mathbf{t}}), \quad \mathbf{t} \in T,$$

as the final decision on the sought-for secondary array $\hat{\mathbf{X}}(\mathbf{Y}) = (\hat{x}_{\mathbf{t}}(\mathbf{Y}_{\mathbf{t}}),\ \mathbf{t} \in T)$.

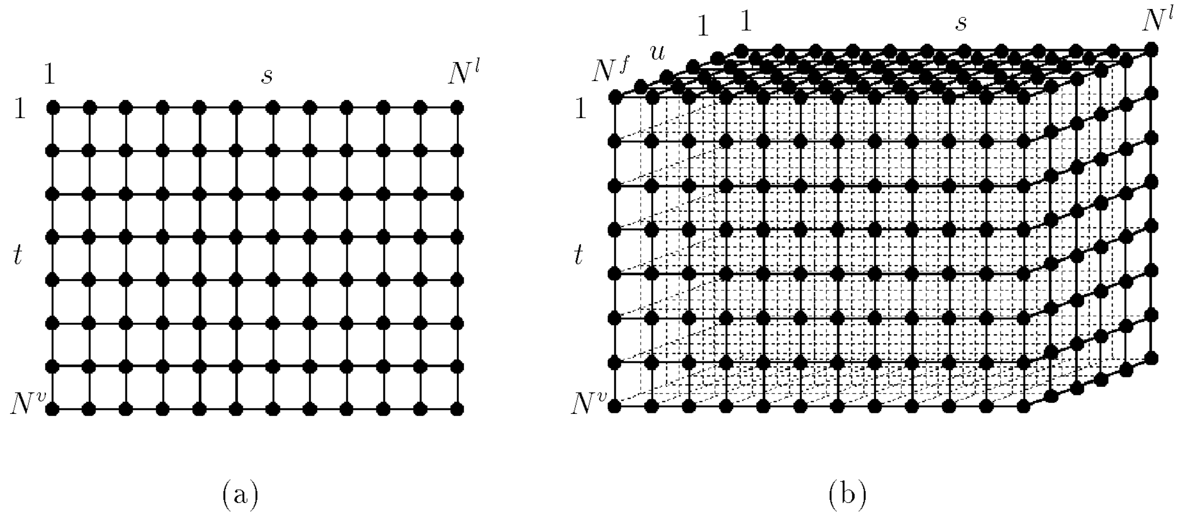


Figure 6: Graphs of the spatial neighborhood between elements of a seismic data array that serve as variable adjacency graphs of the separable objective function in the problems of seismic data analysis: two-dimensional (a) and three-dimensional (b) lattice.

To overcome this obstacle, we have to mutually coordinate the preliminary fuzzy estimates with respect to the assumption that the sough-for local characteristic of the image is sufficiently smooth in the space, what means that the final estimates at immediately neighboring discrete elements of the array should not differ too much from each other. This a priori knowledge can be expressed in local terms as a model-based penalty function $\gamma_{\mathbf{t}'\mathbf{t}''}(x_{\mathbf{t}'},\, x_{\mathbf{t}''})$ which is defined over $\mathcal{X} \times \mathcal{X}$ at each pair of neighboring elements of the data array and takes the greater value the greater is the discrepancy, in accordance with the accepted model, between the respective neighboring hidden values. These functions will be referred to as edge functions, since each of them is associated with an edge of the neighborhood graph G.

Given a data array $\mathbf{Y}$, the decision on the secondary array $\mathbf{X} = (x_{\mathbf{t}},\, \mathbf{t} \in T)$ as a whole should be sought for as a compromise between the mutually contradictory requirements to minimize both node and edge functions that are, in the general case, different at different nodes and, respectively, edges. It is natural to take the sum

$$J(\mathbf{X}|\mathbf{Y}) = \sum_{\mathbf{t}\in T} \psi_{\mathbf{t}}(x_{\mathbf{t}}|\mathbf{Y}_{\mathbf{t}}) + \sum_{(\mathbf{t}',\mathbf{t}'')\in G} \gamma_{\mathbf{t}'\mathbf{t}''}(x_{\mathbf{t}'}, x_{\mathbf{t}''}) \tag{1}$$

as a combined criterion of decision making:

$$\hat{\mathbf{X}}(\mathbf{Y}) = \arg\min_{x_t\in\mathcal{X}} J(\mathbf{X}|\mathbf{Y}). \tag{2}$$

As we see, the criterion to be minimized is a sum of items each of which is function of only one or two variables. The arguments of the criterion along with data-dependent functions $\psi_{\mathbf{t}}(x_{\mathbf{t}}|\mathbf{Y}_{\mathbf{t}})$ are associated with elements of the data array, i.e. with nodes of the neighborhood graph G (node functions), whereas the functions of two arguments $\gamma_{\mathbf{t}'\mathbf{t}''}(x_{\mathbf{t}'}, x_{\mathbf{t}''})$ are assigned to edges of the graph (edge functions). We will call an objective function of such a kind to be separable, pairs of its variables occurring in summands $\gamma(x_{\mathbf{t}'}, x_{\mathbf{t}''})$ to be adjacent, and the graph G, that determines the structure of the function, its adjacency or supporting graph. We will also say that an objective function of kind (1) is a separable G-supported function. The number of its variables is equal to the number $|T|$ of nodes in G, i.e. the number of elements in the data array.

For vertical seismic sections, as being two-dimensional data arrays, the natural neighborhood relation is rectangular lattice (Fig. 6), and, so, the summands in the objective function (1) will be double sums over its nodes

$$J(\mathbf{X}|\mathbf{Y}) = \sum_{s=1}^{N^l}\sum_{t=1}^{N^v}\psi(x_{st}|\mathbf{Y}_{st}) + \sum_{s=1}^{N^l}\sum_{t=2}^{N^v}\gamma^v(x_{s,t-1}, x_{st}) + \sum_{s=2}^{N^l}\sum_{t=1}^{N^v}\gamma^h(x_{s-1,t}, x_{st}). \tag{3}$$

The edge functions at horizontal and vertical edges are taken different, what is necessary for vertical seismic sections, in which the respective axes have different physical meanings.

Seismic cubes differ from vertical sections only by the presence of an additional horizontal coordinate, and the neighborhood relation and adjacency graph for them will be three-dimensional rectangular lattice (Fig. 6,b). The structure of the objective function remains the same, just becomes a little more bulky:

$$J(\mathbf{X}|\mathbf{Y}) = \sum_{s=1}^{N^l}\sum_{u=1}^{N^f}\sum_{t=1}^{N^v}\psi(x_{sut}|\mathbf{Y}_{sut}) + \sum_{s=1}^{N^l}\sum_{u=1}^{N^f}\sum_{t=2}^{N^v}\gamma^v(x_{sut-1}, x_{sut}) + \sum_{s=1}^{N^l}\sum_{u=2}^{N^f}\sum_{t=1}^{N^v}\gamma^h(x_{su-1,t}, x_{sut}) + \sum_{s=2}^{N^l}\sum_{u=1}^{N^f}\sum_{t=1}^{N^v}\gamma^h(x_{s-1,ut}, x_{sut}). \tag{4}$$

We use here the same edge function in both horizontal directions s and u because of their identical physical meaning.

However, the intent to create a non-iterative procedure of finding the minimum point $\hat{\mathbf{X}}$ of a separable function (1) supported by a graph of general kind comes up against a fundamental barrier. Unfortunately, rectangular lattice does not possess any specific properties from this viewpoint in comparison with the general case. Hence, the algorithm for finding the minimum point of the criterion (3) or (4) can be only of iterative type what is utterly undesirable for massive data sets.

One of the main results of the previous study [1, 2, 3, 7] is a noniterative global optimization procedure for the particular case when the graph of the pair-wise adjacency of the data set elements has the form of a tree, i.e. has no cycles. The procedure is based on a recurrent decomposition of the initial optimization problem over all the node variables into a succession of partial problems each of which consists in optimization of a function of only one variable like Bellman functions in the classical dynamic programming.

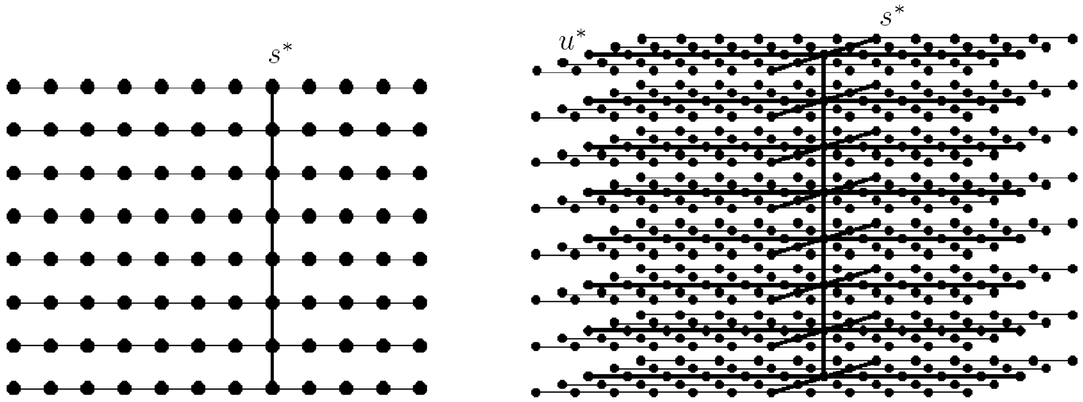


Figure 7: Adjacency graphs of the partial objective function with tree-like separability for evaluating the goal variables in a single vertical row of a two-dimensional (a) and three-dimensional (b) rectangular array.

However, a tree is not yet a lattice, therefore, the basic optimization procedure cannot be directly used for the processing of seismic data sets. It is clear that there is no way to replace the lattice-like neighborhood graph by a tree-like one without loss in the ability to ensure the required smoothness of the secondary data field in all directions from each element of a spatially ordered data array.

To avoid this obstacle, for finding the values of the hidden field at each vertical row of the picture, it was proposed in [1, 2] to use a separate pixel neighborhood tree which is defined, nevertheless, on the whole pixel set and has the same horizontal branches as the others (Fig. 7). The resulting generalized data processing procedure is aimed at finding the optimal values only for the hidden variables at the stem nodes in each tree, and boils down to a combination of two usual dynamic programming procedures, each dealing with single image rows, respectively, the horizontal and vertical ones.

We don't go here into details of the generalized data processing procedure that remains the same for many kinds of applications. The specificity of each particular data processing problem falls only on the way of extracting single units of the local information on the hidden variables immediately from the original data set and on the formalization of neighborhood constraints, i.e. on forming the data-dependent node functions and model-based edge ones. In the next two Sections, we shall do this for clearing the seismic data array from trails of multiple reflections and estimating parameters of the local texture model, the two stages constituting the technique being proposed for processing seismic explorative data in the crystalline basement interval.

3.3 Suppression of multiple reflections

As we pointed it out at the end of Section 2.3, the stray patterns caused by multiple reverberation of seismic waves from the roof of the basement and the above-lying sedimentary strata

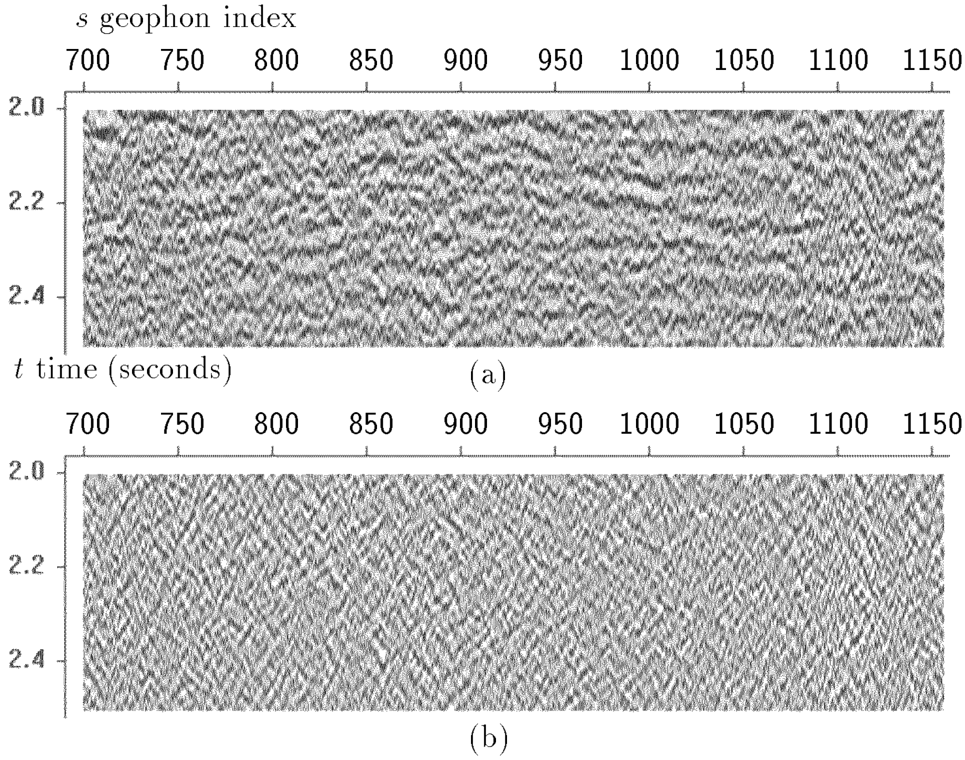


Figure 8: A seismic section interval aligned to the paleohorizon line (a) and result of preprocessing aimed at suppression of multiple reflections (b).

basically repeat the form of the roof. Once the basement interval of the array is aligned to it, the trails of multiple reflections become practically horizontal interferences, as it is well seen in Fig. 8. Thus, in the horizontal sections through the aligned array, one-dimensional or two-dimensional ones, respectively, in the case of two-dimensional and three-dimensional data, these will look as quite low-frequency patterns, in contrast to high-frequency oscillations which are referred to micro-inhomogeneities of the basement rock being the object of interest in this study. If we slightly smooth the array in the horizontal direction and subtract the smoothed picture from the original, the low-frequency component will be essentially attenuated. It is just this technique that we use to suppress the multiple reflections.

As the elements of the original seismic data array are real numbers y_{st} or $y_{sut} \in \mathcal{Y} = \mathsf{R}$, the result of smoothing will consist of real numbers, too, x_{st} or $x_{sut} \in \mathcal{X} = \mathsf{R}$. Smoothing of an ordered data array boils down to estimating the average value of its elements in a vicinity of each point. So, the smoothed seismic picture must be as close to the original as possible and, simultaneously, as smooth as possible. In the version of smoothing technique proposed here, we allocate the responsibility for these two mutually contradictory requirements, respectively, to the node and the edge functions in the generalized separable objective function $J(\mathbf{X}|\mathbf{Y})$ (3) or (4) that is to be minimized 2. For definiteness sake, we consider here the three-dimensional version of the problem.

From the viewpoint of the intent to retain the original values of the seismic signals, the node function $\psi_{sut}(x)$ in (4) should reach its minimum value when x coincides with the actual value at the current element y_{sut}. Because no further preferences are meant to be acquirable from the original data, the node function must be symmetric relative to this point. The simplest quadratic function will be the appropriate choice:

$$\psi_{sut}(x) = \psi(x|y_{sut}) = (x - y_{sut})^2, \quad x \in \mathcal{X} = \mathsf{R}. \tag{5}$$

So, the elementary vicinity of the original array consists, in this case, of only one element $\mathbf{Y}_{sut} = y_{sut}$.

It is just the purpose of the edge function in (4) to provide the desired smoothing effect by imposing a penalty upon the eventual difference of neighboring goal variables in the vertical direction $\gamma^v(x_{sut-1}, x_{sut})$ and two mutually orthogonal and physically identical horizontal directions $\gamma^h(x_{su-1,t}, x_{sut})$ and $\gamma^h(x_{s-1,ut}, x_{sut})$. The edge functions can be taken in the simplest quadratic form with coefficients specific for each kind of them:

$$\begin{array}{l} \gamma^v(x_{sut-1}, x_{sut}) = \alpha^v\,(x_{sut-1} - x_{sut})^2, \ \alpha^v \geq 0, \\ \gamma^h(x_{su-1,t}, x_{sut}) = \alpha^h\,(x_{su-1,t} - x_{sut})^2, \\ \gamma^h(x_{s-1,ut}, x_{sut}) = \alpha^h\,(x_{s-1,ut} - x_{sut})^2, \ \alpha^h \geq 0. \end{array} \tag{6}$$

The greater the coefficient α^v or α^h, the deeper will be the resulting smoothing in the respective direction.

Thus, to provide the designed smoothing only in the horizontal sections of the array, the vertical coefficients must taken equal to zero $\alpha^v = 0$. In deciding on the form of the edge functions in the case of two-dimensional vertical seismic sections, the reasons remain the same with the only difference that there will be one horizontal direction instead of two.

Once we have found the solution of the minimization problem 2, i.e. obtained the smoothed array $\hat{\mathbf{X}}(\mathbf{Y})$, it remains only to subtract it from the original array for the operation of multiple reflections suppression would be completed $\bar{\mathbf{Y}} = \mathbf{Y} - \hat{\mathbf{X}}(\mathbf{Y})$. Fig. 8 shows the result of such a preprocessing of a vertical seismic section through basement rock.

3.4 Parametric autoregressive model of seismic texture and estimation of texture parameters

The aim of seismic data processing is to make visible fine distinctions in the spatial oscillation mode of the reflected seismic signals forming the array to be analyzed, i.e. local texture properties of the array considered as a two- or three-dimensional texture image. The final result of processing should be represented as an array $\mathbf{X} = (\mathbf{x_t}, \ \mathbf{t} \in T)$ of estimated vector parameters $\mathbf{x_t} \in \mathsf{R}^n$ of an appropriate local oscillation model. Just as it is accepted throughout this work, we do this by way of minimization of a separable objective function $J(\mathbf{X}|\mathbf{Y})$ and taking the minimum point $\hat{\mathbf{X}}(\mathbf{Y})$ as the field of local texture parameters. So, it remains only to specify the node and edge functions constituting the criterion (3) or (4) and apply the respective version of the basic optimization procedure described in [1, 2].

As a mathematical basis for evaluating local texture properties of spatial signal oscillations, we use, in this work, the autoregression model of the seismic image. Such a model presupposes treating the data array subject to an analysis as a realization of a random field, two- or three-dimensional one depending on the kind of the data array.

Let $\mathbf{t} = (s, t)$ or $\mathbf{t} = (s, u, t)$ be a point of the array $\mathbf{Y} = (y_{\mathbf{t}},\ \mathbf{t} \in T)$ to be analyzed, and $\mathbf{Y}_{(\mathbf{t})} = (y_{\mathbf{t}+\mathbf{i}},\ \mathbf{i} \in \mathcal{F})$, where $\mathbf{i} = (i^l, i^v)$ or $\mathbf{i} = (i^l, i^f, i^v)$, be a relatively small fragment of the image around point $\mathbf{t}$ excluding this point itself. The form of the fragment is defined by the so-called autoregression mask $\mathcal{F}$ which is nothing else than the set of admissible non-zero spatial shifts from $\mathbf{t}$ that is taken, as a rule, symmetric with respect to the pairs of opposite directions from this point. The number of points in the autoregression mask will be denoted by n.

Mark, that $\mathbf{0} = (0, 0, 0) \notin \mathcal{F}$ and, so, $\mathbf{Y}_{(\mathbf{t})}$ does not contain the signal value $y_{\mathbf{t}}$ at the central point. The autoregression model is based on the idea of an approximate representation of each of these values as a linear combination of the surrounding values within the mask, so that the remainders

$$y_{\mathbf{t}} - \sum_{\mathbf{i} \in \mathcal{F}} a_{\mathbf{t}}^{(\mathbf{i})} y_{\mathbf{t}+\mathbf{i}} = \sigma_{\mathbf{t}} \xi_{\mathbf{t}} \tag{7}$$

be independent random variables formed by discrete white noise $\xi_{\mathbf{t}}$ with zero mathematical expectation and variance equal to unity. The coefficients of the linear combination $\left(a_{\mathbf{t}}^{(\mathbf{i})},\ \mathbf{i} \in \mathcal{F}\right)$ and the mean root square deviation $\sigma_{\mathbf{t}}$, both assumed to change not too quickly over the data array space T, are considered here as local texture parameters. We shall use a several autoregression masks $\mathcal{F}$ to evaluate varied aspects of the spatial texture.

The equality (7) is usually referred to as autoregression equation with changing parameters. In these terms, to evaluate local texture of the seismic data array means to estimate the autoregression parameters at each point $\mathbf{t} \in T$ under some preset smoothness constraints. We show now that such a problem can be solved as that of finding the minimum point of a separable objective function of type (3) or (4).

It will be easier to estimate at each point not the white noise intensity $\sigma_{\mathbf{t}}$ but the full variance of the random field or, in other words, the full local intensity of seismic signals oscillations, let it be denoted by $d_{\mathbf{t}}$. The latter parameter is proportional to $\sigma_{\mathbf{t}}^2$, but the proportionality coefficient is quite a complicated function of autoregression coefficients $\left(a_{\mathbf{t}}^{(\mathbf{i})},\ \mathbf{i} \in \mathcal{F}\right)$.

Let $\mathbf{Y}_{\mathbf{t}}$ be the elementary fragment cut out of the seismic section by the autoregression mask $\mathcal{F}$ including $y_{\mathbf{t}}$, and $\mathbf{y}_{\mathbf{t}} = \left(y_{\mathbf{t}+\mathbf{i}},\ (\mathbf{i}) \in \mathcal{F}\right)$ be the same fragment without the current point. We will assume that $\mathcal{F}$ consists of n elements ordered in an appropriate manner and consider $\mathbf{y}_{\mathbf{t}} \in \mathsf{R}^n$ as a real-valued column vector. Respectively, the autoregression coefficients ordered in the same way will be considered as column vector $\mathbf{a} = \left(a^{(\mathbf{i})},\ (\mathbf{i}) \in \mathcal{F}\right) \in \mathsf{R}^n$ forming, along with the full intensity of signal oscillations, the full vector of texture parameters at each point $\mathbf{x}_{\mathbf{t}} = (\mathbf{a}_{\mathbf{t}},\ d_{\mathbf{t}}) \in \mathsf{R}^{n+1}$.

The node function associated with point $\mathbf{t} \in T$ is intended to be an appropriate measure of mismatch between a hypothesis $\mathbf{x} = (\mathbf{a},\ d)$ on the hidden value of the autoregression parameter and the respective elementary fragment $\mathbf{Y}_{\mathbf{t}} = (y_{\mathbf{t}}, \mathbf{y}_{\mathbf{t}})$. It is convenient to consider two different measures of such a kind, separately for autoregression coefficient vector $\mathbf{a}$ and signal oscillation intensity d.

The squared left part of the autoregression equation (7) is a function which shows to what extent the hypothesis that $\mathbf{a}$ is just the current autoregression coefficient vector $\mathbf{a_t}$ contradicts the elementary fragment $\mathbf{Y_t}$:

$$\psi^{(\mathbf{a})}(\mathbf{a}|\mathbf{Y_t}) = \left(y_\mathbf{t} - \sum_{(\mathbf{i})\in\mathcal{F}} a^{(\mathbf{i})} y_{\mathbf{t}+\mathbf{i}} \right)^2 = \left(y_\mathbf{t} - \mathbf{y}_\mathbf{t}^T \mathbf{a} \right)^2. \tag{8}$$

It is easy to see that the degenerate matrix $\mathbf{B}_\mathbf{t}^0 = \mathbf{y_t}\mathbf{y}_\mathbf{t}^T$ is Hessian matrix of the degenerate quadratic function $\psi^{(\mathbf{a})}(\mathbf{a}|\mathbf{Y_t})$, and each vector satisfying the condition $\mathbf{x}_\mathbf{t}^T\mathbf{a} = y_\mathbf{t}$ is its minimum point. The shortest of them in the sense of Euclidean norm $\|\mathbf{a}\|$ is vector $\mathbf{a}_\mathbf{t}^0 = \left(y_\mathbf{t}/\mathbf{y}_\mathbf{t}^T\mathbf{y_t} \right) \mathbf{y_t}$. So, function (8) can be represented in the form

$$\begin{aligned} &\psi^{(\mathbf{a})}(\mathbf{a}|\mathbf{Y_t}) = (\mathbf{a} - \mathbf{a}_\mathbf{t}^0)^T \mathbf{B}_\mathbf{t}^0 (\mathbf{a} - \mathbf{a}_\mathbf{t}^0), \quad \mathbf{a} \in \mathsf{R}^n, \\ &\mathbf{a}_\mathbf{t}^0 = \frac{y_\mathbf{t}}{\mathbf{y}_\mathbf{t}^T\mathbf{y_t}} \mathbf{y_t} \in \mathsf{R}^n, \quad \mathbf{B}_\mathbf{t}^0 = \mathbf{y_t}\mathbf{y}_\mathbf{t}^T, \ (n \times n). \end{aligned} \tag{9}$$

Analogously, quadratic function

$$\psi^{(d)}(d|\mathbf{Y_t}) = \left(d - y_\mathbf{t}^2 \right)^2, \quad d \in \mathsf{R},$$

shows the extent to which real number d is contradictory, when judging from the single observation $y_\mathbf{t}$, to the hypothesis that it coincides with the current variance of the random field $d_\mathbf{t}$.

Thus, all the requirements placed upon a data-dependent node function in the generalized optimization-based statement of the data analysis problem are met by the sum of $\psi^{(\mathbf{a})}(\mathbf{a}|\mathbf{Y_t})$ and $\psi^{(d)}(d|\mathbf{Y_t})$. If we consider the full parameter vector $\mathbf{x} = (\mathbf{a}, d)$ as column vector in which d is the last element, the desired quadratic data-dependent node function will obtain the form

$$\begin{aligned} &\psi(\mathbf{x}|\mathbf{Y_t}) = (\mathbf{x} - \mathbf{x}_\mathbf{t}^0)^T \mathbf{Q}_\mathbf{t}^0 (\mathbf{x} - \mathbf{x}_\mathbf{t}^0), \quad \mathbf{x} = \begin{pmatrix} \mathbf{a} \\ d \end{pmatrix} \in \mathsf{R}^{n+1}, \\ &\mathbf{x}_\mathbf{t}^0 = \begin{pmatrix} \dfrac{y_\mathbf{t}}{\mathbf{y}_\mathbf{t}^T\mathbf{y_t}} \mathbf{y_t} \\ y_\mathbf{t}^2 \end{pmatrix} \in \mathsf{R}^{n+1}, \quad \mathbf{Q}_\mathbf{t}^0 = \begin{pmatrix} \mathbf{y_t}\mathbf{y}_\mathbf{t}^T & \mathbf{0} \\ \mathbf{0}^T & y_\mathbf{t}^2 \end{pmatrix}, \ [(n+1)\times(n+1)], \end{aligned} \tag{10}$$

where $\mathbf{0}$ is column vector of n zero elements.

If it were allowed to choose the autoregression coefficient vectors $\mathbf{a_t}$ independently of each other, there would exist at each point $\mathbf{t} \in T$, in consequence of the degeneracy of the quadratic functions (9) occurring in the node functions (10), an infinite set of vectors providing an absolutely exact representation of the central value $y_\mathbf{t}$. As to the estimates $d_\mathbf{t}$, these would repeat the squared instantaneous values of the seismic signals $y_{\mathbf{t}}^2$. It is just the purpose of edge functions to constrain the freedom of choosing both of them $\mathbf{x_t} = (\mathbf{a_t}, d_{\mathbf{t}})$ by imposing upon the combined vectors at the neighboring points the requirement to be close to each other in all the directions in the space of the array. The constraints of such a kind are meant to provide a coordination or, in other terms, smoothing of the fuzzy and even ambiguous local decisions, but the degree of smoothing should be taken different in the vertical direction and in the horizontal ones, because there is a fundamental distinction in a seismic data array between the nature of the vertical axis interpreted as time and that of the horizontal ones formed by the increasing lateral and frontal geophone indices. In addition, the autoregression coefficient vector $\mathbf{a}$ and oscillation intensity d occur in the node functions (10) in different manners, therefore, it may turn out to be needed to set a specific smoothing degree for each of these variables. Two identical quadratic functions taken with different coefficients would be just the appropriate choice:

$$\begin{aligned}
\gamma^v(\mathbf{x}_{su,t-1}, \mathbf{x}_{sut}) &= \beta^v\,(\mathbf{x}_{su,t-1} - \mathbf{x}_{sut})^T \mathbf{A}(\mathbf{x}_{su,t-1} - \mathbf{x}_{sut}),\\
\gamma^h(\mathbf{x}_{s-1,ut}, \mathbf{x}_{sut}) &= \beta^h\,(\mathbf{x}_{s-1,ut} - \mathbf{x}_{sut})^T \mathbf{A}(\mathbf{x}_{s-1,ut} - \mathbf{x}_{sut}),\\
\gamma^h(\mathbf{x}_{su-1,t}, \mathbf{x}_{sut}) &= \beta^h\,(\mathbf{x}_{su-1,t} - \mathbf{x}_{sut})^T \mathbf{A}(\mathbf{x}_{su-1,t} - \mathbf{x}_{sut}),
\end{aligned}
\qquad \mathbf{A} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0}^T & \alpha \end{pmatrix}, \qquad (11)$$

where $\mathbf{I}$ is identity matrix $(n \times n)$. Coefficient $\alpha > 0$ is intended to align the smoothing degree of d relative to that of $\mathbf{a}$, and coefficients $\beta^v > 0$ and $\beta^h > 0$ determine two different smoothing degrees in, respectively, the vertical direction and the horizontal ones.

So, the estimation of the local autoregression parameters for a preset mask $\mathcal{F}$ boils down to the minimization of the separable objective function (3) or (4) with node and edge functions defined by (10) and (11). The result will have the form of an array assembled from vector elements $\hat{\mathbf{X}} = \big(\hat{\mathbf{x}}_{\mathbf{t}} = (\hat{\mathbf{a}}_{\mathbf{t}}, \hat{d}_{\mathbf{t}}), \mathbf{t} \in T\big)$, but it is convenient to consider it as two separate arrays, namely, the vector array of autoregression coefficients $\hat{\mathbf{A}} = (\hat{\mathbf{a}}_{\mathbf{t}}, \mathbf{t} \in T)$ and the scalar one of instantaneous signal intensities $\hat{\mathbf{D}} = (\hat{d}_{\mathbf{t}}, \mathbf{t} \in T)$.

4 Evaluation of local texture features of fluid collectors

4.1 Seismic image irregularity

As we supposed in Section 2.3, local fracturedness of the monolith basement at a point in the three-dimensional undersurface medium must reveal itself, in particular, by a higher irregularity of spatial signal oscillations around the respective point in the seismic data array than that around the points corresponding to areas with unbroken crystalline structure. We evaluate this characteristic of the array, as well as all other texture properties of it, via estimating parameters of the spatial autoregression model (7) with an appropriate form of the mask $\mathcal{F}$.

The local estimates on the autoregression parameters are sought for as two arrays $\hat{\mathbf{X}} = (\hat{\mathbf{A}}, \hat{\mathbf{D}})$ that deliver the minimum value to the separable objective function. In particular, in

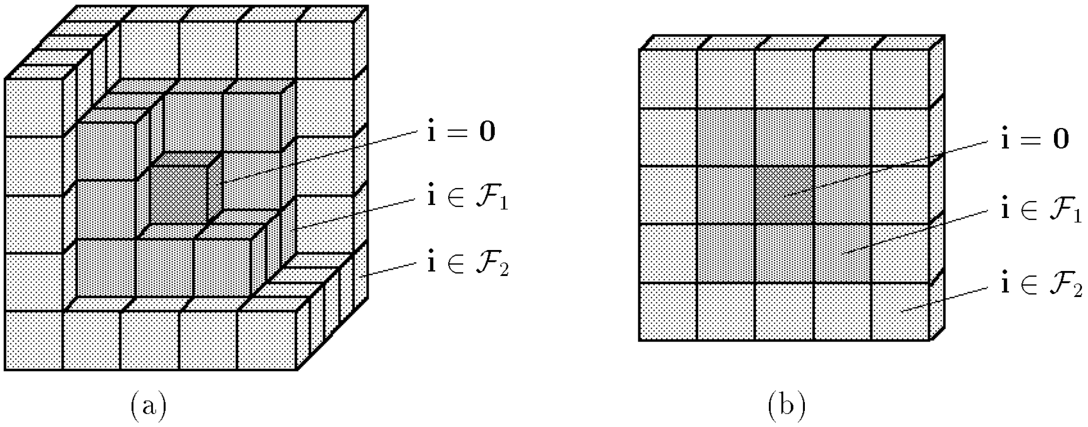


Figure 9: Autoregression masks for evaluating the local irregularity of spatial signal oscillations in a three-dimensional (a) and two-dimensional (b) seismic data array.

the search for the assembly of autoregression coefficients $\hat{\mathbf{A}} = (\hat{\mathbf{a}}_{\mathbf{t}}, \mathbf{t} \in T)$, the optimization procedure will try to minimize each of the functions (9) with respect to the smoothness constraints, i.e. to approximate all the signal values $y_{\mathbf{t}'}$ at the points $\mathbf{t}'$ in a vicinity of each point $\mathbf{t}$ by linear combinations of their immediate neighbors $\mathbf{Y}_{\mathbf{t}'} = (y_{\mathbf{t}'+\mathbf{i}}, \mathbf{i} \in \mathcal{F})$ using the same or, at least, almost the same vector of coefficients $\mathbf{a}_{\mathbf{t}'} \cong \mathbf{a}_{\mathbf{t}}$. It appears to be natural to consider the spatial oscillations around a particular point $\mathbf{t}$ to be the more regular or irregular, the more successful or unsuccessful such an attempt is. It follows from this reason that, for the degree of local irregularity of spatial oscillations would be revealed, quite a high level of smoothing is required when estimating the autoregression coefficients.

Let the compromise assembly $\hat{\mathbf{A}} = (\hat{\mathbf{a}}_{\mathbf{t}}, \mathbf{t} \in T)$ be found. The squared residual error $\sigma^2_{\mathbf{t}} = \left(y_{\mathbf{t}} - \sum \hat{\mathbf{a}}^{\mathbf{i}}_{\mathbf{t}} y_{\mathbf{t}+\mathbf{i}}\right)^2$ which remains at a particular point $\mathbf{t}$ after the optimal coefficients are found seems to be an appropriate quantitative characteristic of the local irregularity of the seismic image at this point, but the array of such values $\mathbf{S} = (\sigma^2_{\mathbf{t}}, \mathbf{t} \in T)$ will inherit the spatial chaotic component of the original oscillations. Therefore, as the result of processing, we take the spatially smoothed array $\hat{\mathbf{S}} = (\hat{\sigma}^2_{\mathbf{t}}, \mathbf{t} \in T)$ that is obtained from $\mathbf{S} = (\sigma^2_{\mathbf{t}}, \mathbf{t} \in T)$ by the same smoothing operation we used in Section 8, i.e. by application the basic optimization procedure once again, but in this case with non-zero coefficients $\alpha^v 0$ and $\alpha^h 0$ of the quadratic edge penalties (6) in all the directions. The greater the local value $\hat{\sigma}^2_{\mathbf{t}}$, the more irregular should be considered the seismic image at the respective point $\mathbf{t}$.

To evaluate the local irregularity of a seismic image, we used quite a simple autoregression mask $\mathcal{F}$ shown in Fig. 9. It consists of two areas $\mathcal{F}_1$ and $\mathcal{F}_2$ that encircle the current point $\mathbf{t}$ in the space of the array. The vector of autoregression coefficients is defined by two values $\mathbf{a} = (a^{(1)}, a^{(2)})$ each of which remains the same for all the elements within the respective mask area $\mathcal{F}_1$ and $\mathcal{F}_2$:

$$y_{\mathbf{t}} \cong a^{(1)} \frac{1}{|\mathcal{F}_1|} \sum_{\mathbf{i} \in \mathcal{F}_1} y_{\mathbf{t}+\mathbf{i}} + a^{(2)} \frac{1}{|\mathcal{F}_2|} \sum_{\mathbf{i} \in \mathcal{F}_2} y_{\mathbf{t}+\mathbf{i}}, \tag{12}$$

where $|\mathcal{F}_1|$ and $|\mathcal{F}_1|$ are, respectively, the number of elements in the inner and the outer area of the mask. These numbers are $|\mathcal{F}_1| = 26$, $|\mathcal{F}_2| = 98$ for seismic cubes, and $|\mathcal{F}_1| = 8$, $|\mathcal{F}_2| = 16$ for plane sections.

4.2 Local seismic signal intensity and one-dimensional local spectra

A seismic data array is formed by a multitude of seismic records (y_{sut}, $s =$ const, $u =$ const) each of which is function of one variable, namely, time t. From the "orthogonal" point of view, the same array can be considered as consisting of its horizontal, or, to be exact, paleohorizontal rows, the lateral (y_{sut}, $u =$ const, $t =$ const) and, in the case of three-dimensional arrays, also frontal ones (y_{sut}, $s =$ const, $t =$ const), that are functions of one variable as well, namely, seismic sensor index s or u. Distinctions of the local oscillation mode in different regions of a seismic image must reveal themselves, in particular, as differences in the local oscillation mode of such one-dimensional signals generated by the two- or three-dimensional picture.

It is just this kind of the analysis that is provided by autoregression masks $\mathcal{F}$ completely concentrated in one of the image lines, namely, in the vertical $(\dots, t-1, t, t+1, \dots)$ or one of the horizontal directions $(\dots, s-1, s, s+1, \dots)$, $(\dots, u-1, u, u+1, \dots)$, as it is shown in Fig. 10. When we use a one-dimensional mask, each vector of autoregression parameters $\mathbf{x}_{sut} = (\mathbf{a}_{sut}, d_{sut})$ will actually represent spatially smoothed local properties of the respective one-dimensional random process formed by a single seismic signal or a single horizontal one-dimensional section through the data array. In the case of one-dimensional models, it is sufficient to use one-sided masks, i.e. masks concentrated before the current point according to the direction in which the respective index increases. Hence, the local autoregression model (7) becomes more simple:

$$\begin{gathered} y_{sut} - \sum_{i=1}^{n} a_{sut}^{(i)} y_{sut-i} = \sigma_{sut} \xi_{sut}, \\ y_{sut} - \sum_{i=1}^{n} a_{sut}^{(i)} y_{s-i,ut} = \sigma_{sut} \xi_{sut}, \quad y_{sut} - \sum_{i=1}^{n} a_{sut}^{(i)} y_{su-i,t} = \sigma_{sut} \xi_{sut}. \end{gathered} \tag{13}$$

Models of such a kind describe the texture through describing its one-dimensional horizontal or vertical sections, so, these are actually models of the respective one-dimensional random processes.

The geophysical reasons mentioned in Section 2.3 suggest the existence of a relation between local fracturedness of the crystalline basement and local spectral properties of both the vertical and paleo-horizontal sections through the respective regions of the seismic picture. Thus, the local peculiarities of these kinds of one-dimensional oscillations should be studied just in the spectral form.

Variance (dispersion) of a stationary random process on the discrete argument axis $y_t,\ t = \ldots,1,2,3,\ldots$ is mathematical expectation of its squared random values $d = \mathrm{E}(y_t^2)$, assumed that the mean value of the process equals zero. Spectral density of variance is a function of frequency $S(f) \geq 0$ which is defined in the range $0 \leq f \leq 1/2$, because frequencies higher than $1/2$ periods per sample are senseless in the case of discrete argument. Integral of the spectral density over the whole range of definition gives the variance of the random process, and integral over a frequency interval $f_{\min} \leq f \leq f_{\max}$, where $f_{\min} \geq 0$ and $f_{\max} \leq 1/2$, shows what partial variance is carried by harmonic constituents within this frequency band:

$$d = \int_0^{1/2} S(f)df, \qquad d(f_{\min}, f_{\max}) = \int_{f_{\min}}^{f_{\max}} S(f)df.$$

Spectral density of an autoregressive random process

$$y_t = \sum_{k=1}^{n} a_k y_{t-k} + \xi_t$$

is a function $S(f;\mathbf{x})$ of the vector of autoregression parameters $\mathbf{x} = (\mathbf{a}, d)$, namely, of the autoregression coefficients $\mathbf{a} = (a_1, \ldots, a_n)$ and full variance of the process d. This function is especially simple in the case of second-order autoregression $n = 2$:

$$S(f) = \frac{(1+a_2)\left((1-a_2)^2 - a_1^2\right)}{(1-a_2)} \frac{2d}{1 + a_1^2 + a_2^2 - 2a_1(1-a_2)\cos 2\pi f - 2a_2 \cos 4\pi f}\,.$$

One-dimensional local texture models (13) are based on just the opposite assumption than that of stationarity of the random process being analyzed, namely, it is assumed that

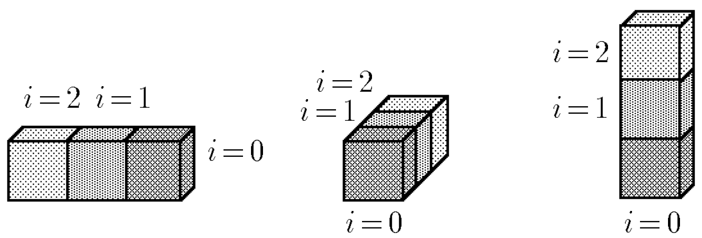


Figure 10: Autoregression masks for estimating one-dimensional local spectra of the seismic image.

the estimated autoregression parameters $\hat{\mathbf{x}}_{sut}(\mathbf{Y})$ of a horizontal or vertical process cut out of a seismic data array $\mathbf{Y}$ are different at different points (tut) as along the axis of this process as well as across it. A series of spectra representing gradually changing oscillation mode of a signal on the time axis is usually referred to as result of its time-frequency analysis. In the case of seismic data arrays, we will obtain a two- or three-dimensional array of horizontal or vertical spectra $S\big(f;\hat{\mathbf{x}}_{sut}(\mathbf{Y})\big)$, $0 \le f \le 1/2$, but the term "time-frequency analysis" is applicable in full sense only to the vertical direction, because it is just the vertical axis of a seismic data array which is associated with time.

Of course, a series of spectral densities is a too immense representation of the local oscillation modes in a nonstationary signal, the more so, in a two- or three-dimensional image. In this work, we represent local spectral properties of a seismic data array at each point $\mathbf{t}$, separately in each of the two or three directions, namely, in the vertical ver, lateral lat and frontal fr one, by the full variance of the respective vertical or horizontal random process and its fractions in the low, medium and high frequency band considered as the first $0 \le f \le 1/6$, the second $1/6 \le f \le 2/6$ and the third $2/6 \le f \le 3/6$ of three equal non-overlapping intervals making the full range of frequency values $0 \le f \le 1/2$.

For instance, for the vertical direction the instantaneous partial constituents of the full signal variance are integrals of the vertical instantaneous spectral density

$$\hat{d}_{\mathbf{t}}^{ver}(0,1/6) = \int\limits_0^{1/6} S\big(f;\hat{\mathbf{x}}_{\mathbf{t}}^{ver}(\mathbf{Y})\big)df, \quad \hat{d}_{\mathbf{t}}^{ver}(1/6,2/6) = \int\limits_{1/6}^{2/6} S\big(f;\hat{\mathbf{x}}_{\mathbf{t}}^{ver}(\mathbf{Y})\big)df,$$
$$\hat{d}_{\mathbf{t}}^{ver}(2/6,3/6) = \int\limits_{2/6}^{3/6} S\big(f;\hat{\mathbf{x}}_{\mathbf{t}}^{ver}(\mathbf{Y})\big)df,$$

the instantaneous full variance itself is the sum

$$\hat{d}_{\mathbf{t}}^{ver}(0,1/2) = \int\limits_0^{1/2} S\big(f;\hat{\mathbf{x}}_{\mathbf{t}}^{ver}(\mathbf{Y})\big)df = \hat{d}_{\mathbf{t}}^{ver}(0,1/6) + \hat{d}_{\mathbf{t}}^{ver}(1/6,2/6) + \hat{d}_{\mathbf{t}}^{ver}(2/6,3/6), \quad (14)$$

and fractions of the full variance are ratios

$$\hat{\delta}_{\mathbf{t}}^{ver}(0,1/6) = \frac{\hat{d}_{\mathbf{t}}^{ver}(0,1/6)}{\hat{d}_{\mathbf{t}}^{ver}(0,1/2)}, \quad \hat{\delta}_{\mathbf{t}}^{ver}(1/6,2/6) = \frac{\hat{d}_{\mathbf{t}}^{ver}(1/6,2/6)}{\hat{d}_{\mathbf{t}}^{ver}(0,1/2)},$$
$$\hat{\delta}_{\mathbf{t}}^{ver}(2/6,3/6) = \frac{\hat{d}_{\mathbf{t}}^{ver}(2/6,3/6)}{\hat{d}_{\mathbf{t}}^{ver}(0,1/2)}, \quad (15)$$

that meet at each $\mathbf{t}$ the obvious condition

$$\hat{\delta}_{\mathbf{t}}^{ver}(0,1/6) + \hat{\delta}_{\mathbf{t}}^{ver}(1/6,2/6) + \hat{\delta}_{\mathbf{t}}^{ver}(2/6,3/6) = 1. \quad (16)$$

The formulas for the lateral and frontal direction are absolutely analogous.

There is no need to calculate the full signal variance in each of the directions at a point $\mathbf{t}$, because these will coincide. It is sufficient to find one of them and take it as estimate on the local seismic signal intensity at the current point of the array:

$$\hat{d}_{\mathbf{t}} = \hat{d}^{ver}_{\mathbf{t}}(0,1/2) = \hat{d}^{lat}_{\mathbf{t}}(0,1/2) = \hat{d}^{fr}_{\mathbf{t}}(0,1/2). \tag{17}$$

So, we characterize spectral properties of the spatial texture at a point of the data array by ten real numbers for three-dimensional arrays and seven for two-dimensional ones: the seismic signal intensity estimated, for instance by (14), (17), and three spectral fractions (15) in each of the three directions. Actually, on the force of condition (16), it is sufficient to consider, in each direction, only two fractions of three, so, the number of spectral characteristics reduces, respectively, to seven and five for the two versions of the data dimensionality.

Finally, with regard for the seismic image irregularity considered in the previous Section, we obtain eight numerical texture parameters for seismic cubes and six for seismic sections. The expected association between the local texture parameters and the fractured or monolith basement areas are summarized in the Table 1.

4.3 Identification of collector zones in the crystalline basement of the Bombay High oil field

On the basis of such an approach, we attempted to forecast the location of fractured and, maybe, cavernous zones in the crystalline basement of the oil field Bombay High on the western coastal shelf of India. The topicality of this problem is referred to the following circumstances:

- the minable and signs of nonminable oil and gas yields were obtained in 23 of about 30 examined wells that open the crystalline basement;
- because of the huge space of the crystalline block Bombay High, the hydrocarbon resources in the basement of this field may be quite large, if the location of collector zones in the massif is known.

We took for the analysis quite an old data set obtained in 1972 as a system of profiles covering a part of the field Bombay High in the form of a grid 5×7 km. More recent data were available only in the form after a deep preprocessing undertaken, in due time, with the purpose to emphasize the information on the layered structure of the sedimentary interval, whereas the high-frequency component, especially important when studying the basement interval, turned out to be suppressed to a great extent. The estimation of the above mentioned parameters was fulfilled from fragments of 13 two-dimensional seismic sections cut out from the profiles usable for the texture analysis. The map of the original profiles, their fragments taken for the experiment, and the results obtained is shown in Fig. 12 [11].

At every point of each seismic section, we estimated six texture parameters introduced in the two previous Sections, namely, the irregularity of the seismic image $\sigma^2_{\mathbf{t}}$, local seismic

Table 1: Expected distinctions between values of local texture parameters of the seismic picture in fractured and monolith zones of the crystalline basement.

	Local parameters of spatial texture	**Zones of supposed fracturedness**	**Zones of supposed monolith basement**
1	Irregularity of seismic picture	High values of noise coefficient $\sigma^2_{\mathbf{t}}$	Low values of noise coefficient $\sigma^2_{\mathbf{t}}$
2	Spectrum of vertical frequencies	Low intensity of high frequencies $\delta^{ver}_{\mathbf{t}}(2/6, 3/6)$ and high intensity of low frequencies $\delta^{ver}_{\mathbf{t}}(0, 1/6)$	High intensity of high frequencies $\delta^{ver}_{\mathbf{t}}(2/6, 3/6)$ and low intensity of low frequencies $\delta^{ver}_{\mathbf{t}}(0, 1/6)$
3	Spectrum of horizontal frequencies	High intensity of high frequencies $\delta^{lat,fr}_{\mathbf{t}}(2/6, 3/6)$ and low intensity of low frequencies $\delta^{lat,fr}_{\mathbf{t}}(0, 1/6)$	Low intensity of high frequencies $\delta^{lat,fr}_{\mathbf{t}}(2/6, 3/6)$ and high intensity of low frequencies $\delta^{lat,fr}_{\mathbf{t}}(0, 1/6)$
4	Intensity of signal oscillations	High values of $d_{\mathbf{t}}$	Low values of $d_{\mathbf{t}}$

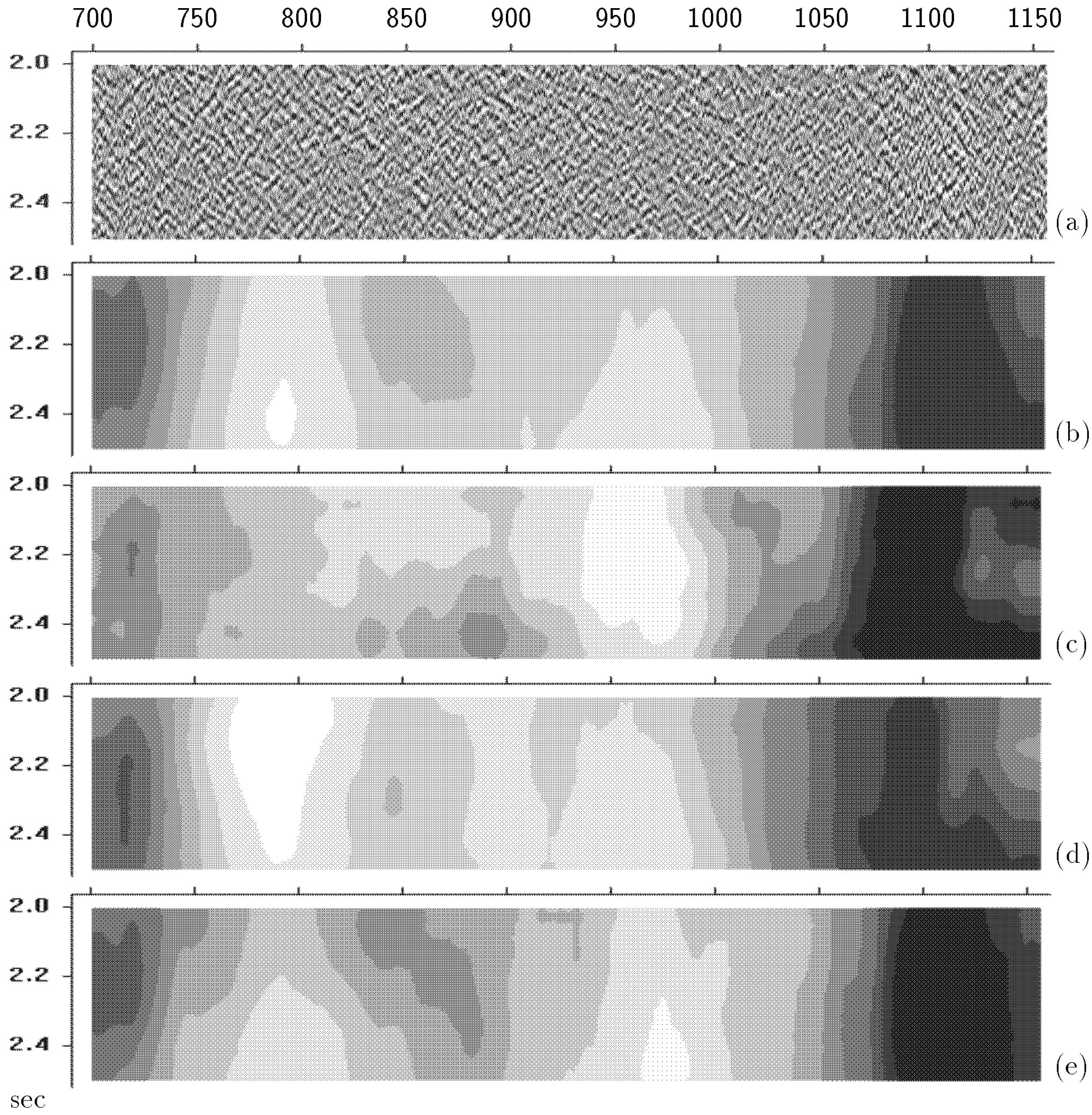


Figure 11: Fragment of an original seismic section after alignment to the upper basement boundary and suppression of multiple reflections (a), and estimates on the texture parameters: irregularity of seismic picture $\sigma_{\mathbf{t}}^2$ (b), local intensity of vertical one-dimensional sections through the two-dimensional array, i.e. of seismic signals, in the low-frequency band $\delta_{\mathbf{t}}^{ver}(0, 1/6)$ (c), local intensity of paleohorizontal cross-sections in high-frequency band $\delta_{\mathbf{t}}^{lat}(2/6, 3/6)$ (d), total local intensity of signal oscillations $d_{\mathbf{t}}$ (e). All of these local parameters, shown in accordance with the rule the darker the greater, serve as indicators of local fracturedness of the crystalline rock.

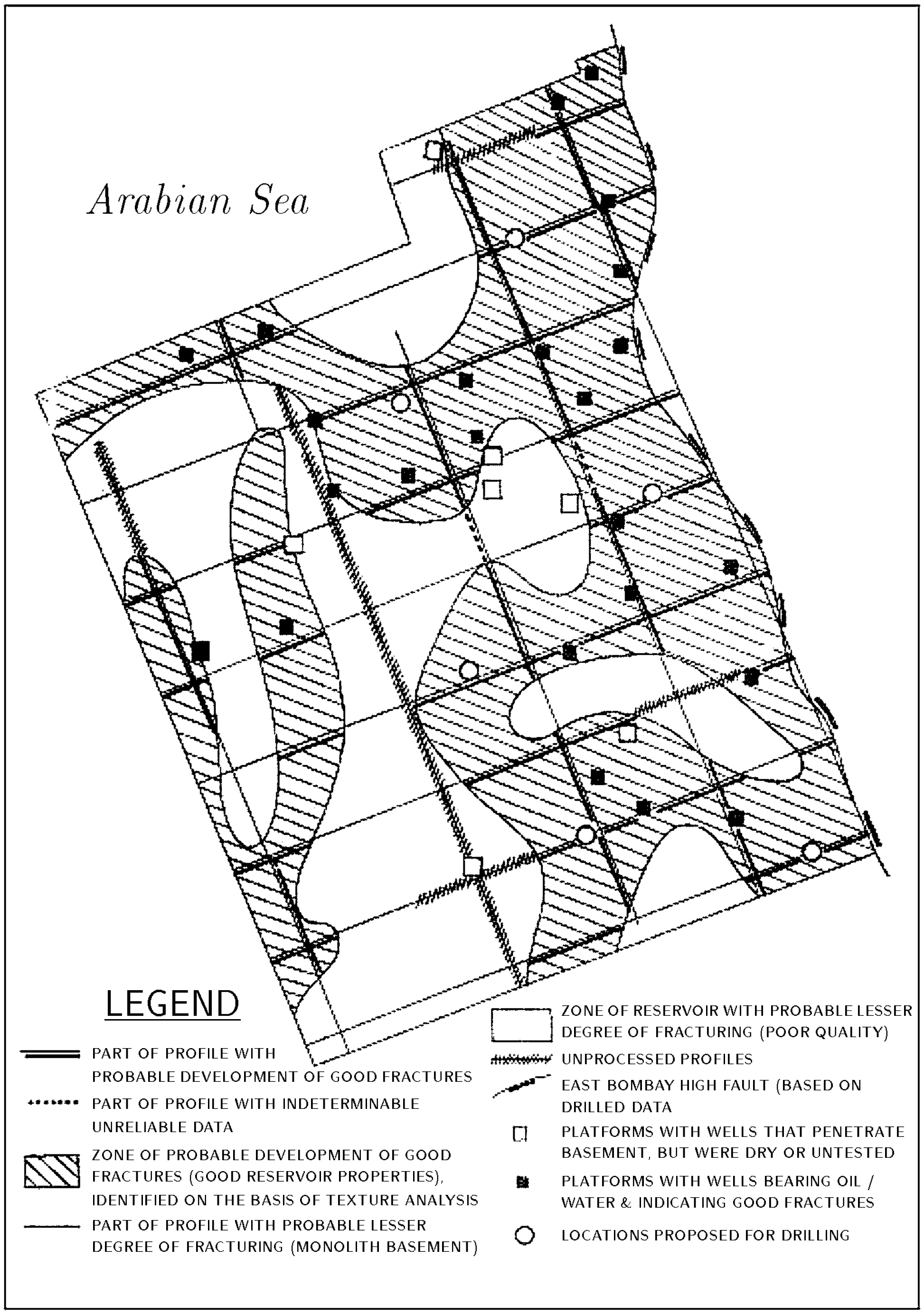


Figure 12: Map of sesmic profiles and estimated fractured zones in the oil field Bombay High.

signal intensity $d_{\mathbf{t}}$, partial intensities in the low-frequency and high-frequency band in both vertical $\delta_{\mathbf{t}}^{ver}(0, 1/6)$, $\delta_{\mathbf{t}}^{ver}(2/6, 3/6)$ and horizontal direction $\delta_{\mathbf{t}}^{lat}(0, 1/6)$, $\delta_{\mathbf{t}}^{lat}(2/6, 3/6)$.

Fig. 11 shows the same example of an original seismic section and result of its preprocessing that were given in Fig 8, along with the respective sections of four estimated parameters of the six, excluding the vertical partial intensity in the high-frequency band and the horizontal one in the low-frequency band because these are practically complementary to their partners. As we see, the results of processing show a good consistency of all the parameters with each other. This phenomenon is observed in 90% of the profile length. The nonrandom behavior of the six parameters examined is confirmed also by a good coincidence of their values in the majority of profile crossings, namely, in 21 of 23.

To summarize data obtained from all the 13 profiles, we plotted the zones of supposed fracturedness in the map of the field fragment under investigation (Fig. 12). In accordance with qualitative reasons summarized in Table 1 in Section 4.2, we considered large values of parameters $\sigma_{\mathbf{t}}^2$, $\delta_{\mathbf{t}}^{ver}(0, 1/6)$, $\delta_{\mathbf{t}}^{lat}(2/6, 3/6)$, $d_{\mathbf{t}}$, and small values of $\delta_{\mathbf{t}}^{ver}(2/6, 3/6)$, $\delta_{\mathbf{t}}^{lat}(0, 1/6)$ as indications in favour of fracturedness. The critical intervals were taken over 70% of the maximum value of the parameter over the respective seismic section for the former four of them and below 30% of the maximum value for the latter two. We assumed a point of the field to possess good reservoir properties if one of the estimated parameters fell in the critical interval at least once in the vertical line under this point. In Fig. 12, the zones of parameter values referred to fractured collectors are shaded.

The map of estimated fractured zones Fig. 12 allows to compare the results of seismic data analysis in accordance with the proposed technique, on the one hand, and those of hydrodynamic experiments in previously bored wells, on the other. The wells that open the basement and bear fluids, namely, oil, gas, or water, are marked in Fig. 12 by black squares. As we see, the shaded zones of estimated fracturedness cover or almost cover the locations of all of them. On the contrary, the locations of dry or untested wells marked by light squares fall outside the estimated zones or are located on their skirts.

We call the results obtained positive because almost all of the wells having had given appreciable yields of fluids fall inside the shaded zones, i.e. those of collectors if judging from the seismic data. Nevertheless, we call the results conditionally positive for the following reasons.

First, the strategy of hydrodynamic experiments absolutely did not take into account the specificity of fractured collectors in the crystalline basement, so, the well data cannot be considered as completely reliable. Second, the depth of the wells was in most cases substantially less than that of the seismic analysis. And third, the seismic data had a low density of observations forming a grid 5×7 km, what is far of being sufficient if one takes into account the modern potentials of seismic data processing, in particular, those promising by the approach depicted in this Report.

For the further prospecting of oil and gas reserves in the basement of the field Bombay High, it appears to be expedient to drill the wells in the areas of most firm manifestation of texture parameter values associated with fractured zones. The locations of the recommended wells are marked in Fig. 12 by light circles.

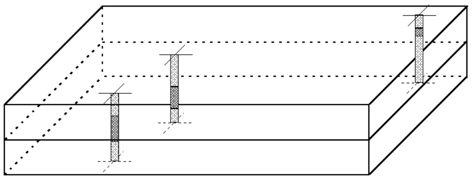
5 Outlines of further research: Recognition of collector zones

We have depicted here only the main idea of the new approach to examining the Earth mantle basement from the daylight surface, and this idea consists in evaluating some local texture characteristics of seismic exploratory data sets. Actually, we have only elaborated an algorithm of estimating local parameters of an appropriate spatial texture model in the two- or three-dimensional basement interval of the array and proposed a heuristic version of such a model borrowed from the analysis of visual images. But we did not give the model any serious justification from the viewpoint of the theory of seismic wave propagation, did not show how to calibrate the model parameters on the basis of reliable data from a sparse set of exploratory wells, and, so, did not propose any regular way of a geological interpretation of estimated values.

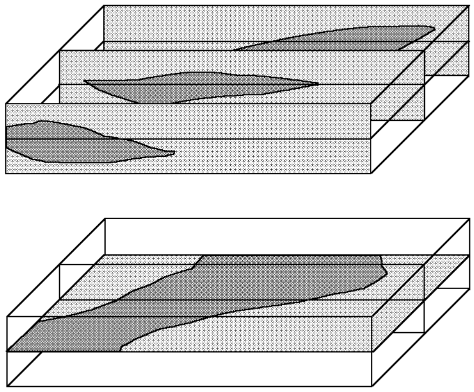
Creation of a sound theoretical basis for choosing local seismic texture parameters in the basement interval presupposes elaboration of mathematical models of the wave field in a three-dimensional heterogeneous medium with known quantitative characteristics of fracturedness and cavern porosity. But it is doubtful that such models can be precise to the extent that would be sufficient for a final reconstruction of collector zones in the crystalline medium from its seismic image. More likely, a mathematical model will only qualitatively sketch the picture of the wave field and show what aspects of it are informative for judging about the fine structure of a crystalline body. To all appearance, these aspects will lead just to the local seismic texture parameters we are looking for.

From the geological point of view, the ultimate aim of seismic data analysis is classifying all the elements of the array $\mathbf{t} \in T$, i.e. the points of the underground medium, with one of the types of basement rock $k \in \mathcal{K}$, for instance, $\mathcal{K} = \{\text{monolith, fractured}\}$. It is natural to assume the values of this index variable to be objectively defined at each position as an array $\mathbf{K} = (k_{\mathbf{t}}, t \in T)$ of the same structure as the data array $\mathbf{Y} = (y_{\mathbf{t}}, \mathbf{t} \in T)$ to be analyzed. Let us consider such an array as a first approximation of the hypothetical "full knowledge" on the sought-for hidden local properties of the basement rock: if the geophysicist new it, the ultimate aim would be achieved from the outset, and no data analysis would be required at all. In accordance with this assumption, the final result of processing should be understood as an array $\hat{\mathbf{K}}(\mathbf{Y}) = (\hat{k}_{\mathbf{t}}(\mathbf{Y}), t \in T)$ of estimates on the unknown values $k_{\mathbf{t}}$ inferred from the original seismic data.

Under the conditions when there is no complete mathematical model that would explain how the local properties of the basement rock exhibit themselves in the seismic picture registered on the daylight surface, the initial information on this association can be taken only from an independent source. The role of such a source may be played by information mined from the inevitably sparse net of exploratory wells drilled through the basement rock. As the mathematical expression of this knowledge, it appears pertinent to assume the actual rock-type indices $k_{\mathbf{t}}$ to be partially known in the form of an array $\mathbf{K}^0 = (k_{\mathbf{t}}, \mathbf{t} \in T^0)$ defined within a subset $T^0 \in T$ vertically piercing the rock body from a few points on the upper surface. Then, the aim of data analysis will consist in an extrapolation of the partially



(a)



(b)

Figure 13: Trainer's information on the "true" local geological structure of the basement rock acquired from exploratory wells (a), and its extrapolation onto the whole cube (b) to be obtained by a trainable pattern recognition algorithm as result of an analysis of both trainer's and seismic data.

known indices onto the entire data array space with respect to the seismic image as a whole, like it is shown in Fig. 13. To solve the problem, it is required to construct an appropriate extrapolation operator $\hat{\mathbf{K}} = \hat{\mathbf{K}}(\mathbf{Y}, \mathbf{K}^0)$.

By its spirit, such a problem falls into the competence area of the theory of supervised pattern recognition, where the part of trainer's data is played, in this case, by the information on the class membership acquired from exploratory wells only for a part of the array elements. But even if we disengage ourselves from the problem of training, the essence of the recognition problem differs from the traditional statement in a series of important details.

The first distinction stems from the fact that the entities to be allocated over a finite number of classes, namely, elements of the array $\mathbf{t} \in T$, are associated with respective points of the underground medium and, consequently, ordered in the three-dimensional space. This means that the decision on the class membership of a point cannot be made independently of the analogous decisions for its natural neighbors. Namely, there is an insistent a priori preference for the neighboring points to be marked by the index of the same class, because the sought-for areas of the massif are expected to be essentially stretched in space.

The second distinction resides in the practical impossibility to infer the decision on the class of the current point only from the single value of the reflected seismic signal at the respective position in the array $\hat{k}_\mathbf{t} = \hat{k}_\mathbf{t}(y_\mathbf{t})$. Each local decision must take into account also the signal values at neighboring points, at least, within the bounds of an immediate vicinity $\hat{k}_\mathbf{t} = \hat{k}_\mathbf{t}(\mathbf{Y}_\mathbf{t})$, $\mathbf{Y}_\mathbf{t} = (y_{\mathbf{t}+\mathbf{i}}, \, \mathbf{i} \in \mathcal{E})$.

Some versions of such a generalization of the classical pattern recognition problem for the case of signals, i.e. one-dimensional data arrays, are proposed in [12] and [13]. The latter version is especially close to the needs of seismic data analysis, because it considers the result of recognition as a classified segmentation of the signal, whereas the former one is primarily aimed at the recognition of a succession of isolated events. To be completely accurate, it would be best to generalize the theoretical framework of supervised segmentation still further onto the case of multi-dimensional arrays, and then construct algorithms of finding collectors in the basement rock just on that theoretical basis.

But if we content ourselves with some loss in the accuracy of contouring-out the areas with different properties of the rock, we can go a simpler way. This way is based on the traditional version of the pattern recognition theory, namely, on treating each element of the data array as a separate entity to be associated with one of the classes independently of other elements, even its immediate neighbors. The principal soundness of such a technique is anticipated by the approach to seismic data analysis we have proposed in this Report.

The central hypothesis set up in this work is that the local seismic image texture characterized by a suitable parameter vector $\mathbf{x_t} = (x_{\mathbf{t}}^{(1)}, \dots, x_{\mathbf{t}}^{(n)})$ may carry a sufficient information on the hidden rock-type index $k_{\mathbf{t}}$. If it is really so, the array of estimated texture parameters $\hat{\mathbf{X}}(\mathbf{Y}) = \big(\hat{\mathbf{x}}_{\mathbf{t}}(\mathbf{Y}), \mathbf{t} \in T\big)$ will be a sufficient representation of the original local properties of the basement rock $\mathbf{K} = (k_{\mathbf{t}}, \mathbf{t} \in T)$. Thus, there must exist a decision rule $\hat{k}(\mathbf{x})$ that, if applied to the pre-estimated texture parameter vector $\hat{\mathbf{x}}_{\mathbf{t}}(\mathbf{Y})$ individually at the respective element of the array, would allow to recognize the actual unknown class of the current point $\hat{k}_{\mathbf{t}}(\mathbf{Y}) = \hat{k}\big(\hat{\mathbf{x}}_{\mathbf{t}}(\mathbf{Y})\big)$.

The possibility of independent recognition of the hidden rock type at each element of the array from the sole feature vector is provided by the fact that the estimates $\hat{\mathbf{x}}_{\mathbf{t}}(\mathbf{Y})$ are already coordinated with each other in the course of minimizing the separable objective function (3) or (4) and smoothly change over the space of the array $\mathbf{t} \in T$. As a result, the local decisions on the rock type $\hat{k}\big(\hat{\mathbf{x}}_{\mathbf{t}}(\mathbf{Y})\big)$ will be implicitly coordinated with each other to the same extent. But the coordination is undertaken too early, at the previous stage of processing before the final decision making, and the abrupt changes in texture parameters at the boundaries between the areas with different physical properties will be inevitably blurred. It is just this why the contours of the sought-for areas may turn out to be displaced in the array of estimated class indices $\hat{\mathbf{K}}\big(\hat{\mathbf{X}}(\mathbf{Y})\big)$ relative to their actual positions in the hypothetical hidden array $\mathbf{K}$.

However, there exists a way to hinder blurring interfaces of regions with different texture properties when estimating local texture parameters of the data array and overcome, to some extent, the inaccuracy in finding contours of collectors. This way consists in using a special technique of preserving essential abrupt changes of the spatial signal oscillation mode in the course of coordinating preliminary local estimates expressed by node functions $\psi(\mathbf{x_t}|\mathbf{Y_t})$ that occur in the separable objective function (3), (4) to be minimized. The technique has procured the name of generalized edge preserving smoothing, because it was proposed first as a means of usual smoothing of signals and visual images [14]. Its essence consists in individual adaptive adjusting each of the edge functions $\gamma_{\mathbf{t}'\mathbf{t}''}(\mathbf{x}_{\mathbf{t}'}, \mathbf{x}_{\mathbf{t}''})$ responsible for the degree of smoothing at the respective pair of neighboring elements of the data array $(\mathbf{t}', \mathbf{t}'')$.

So, the approach to the analysis of seismic exploratory data in the basement interval we have outlined in this work is far of being completely elaborated. The remaining questions should make the subject of a special comprehensive investigation.

References

[1] Muchnik, I., and Mottl, V. *Optimization Algorithms for Separable Functions with Tree-Like Adjacency of Variables and their Application to the Analysis of Massive Data Sets.* DIMACS Technical Report 97-16, April 1997. Center for Discrete Mathematics and Theoretical Computer Science. Rutgers University, the State University of New Jersey.

[2] Muchnik, I., and Mottl, V. *Bellman Functions on Trees for Segmentation, Generalized Smoothing, Matching and Multi-Alignment in Massive Data Sets* . DIMACS Technical Report 98-15, April 1998. Center for Discrete Mathematics and Theoretical Computer Science. Rutgers University, the State University of New Jersey.

[3] Mottl, V., Blinov, A., Kopylov, A., Kostin, A., and Muchnik, I. Variational methods in signal and image analysis. *Proceedings of the 14th International Conference on Pattern Recognition.* Brisbane, Australia, August 16-20, 1998. Volume I, pp. 525-527.

[4] Pitas, I., and Venetsanopoulos, A.N. Towards a knowledge based system for automated geophysical interpretation. *Signal Processing*, Vol. 13, 1987, pp. 229-253.

[5] Pitas, I., and Kotropoulos, C. Texture analysis and segmentation of seismic images. *Proc. IEEE Int. Conf. on ASSP.* Glasgow, U.K., 1989, pp. 1437-1440.

[6] Pitas, I., and Kotropoulos, C. A texture-based approach to the segmentation of seismic images. *Pattern Recognition*, Vol. 25, 1992, No. 9, pp. 929-945.

[7] Mottl, V.V., Blinov, A.B., Kopylov, A.V., and Kostin, A.A. Optimization techniques on pixel neighborhood graphs for image processing. In: *Graph Based Representations in Pattern Recognition* (J.-M. Jolion and W.G. Kropatsch, Eds.) Springer Wien New York, 1997, pp. 135-145.

[8] Biot, M.A. Mechanics of deformation and acoustic propagation in porous media. *Journal of Applied Physics*, Vol. 33, 1962, No. 4, pp. 240-253.

[9] McQuillin, R., Bacon, M., and Barclay, W. *An Introduction to Seismic Interpretation.* Graham & Trotman Limited, 1979.

[10] Wood, L.C., and Treitel, S. Seismic signal processing. *Proceedings IEEE*, Vol. 63, 1975, No. 4, pp. 649-661.

[11] Leviant, V. , Mottl, V., Lal, N.K., Dirghangi, R.S., and Varun, T.R. Experience of reservoir zone identification in crystalline basement of Bombay High field from seismic data. *Proceedings of the 3rd International Petroleum Conference and Exhibition PETROTECH-99*, New Delhi, India, January 9-12, 1999.

[12] Dvoyenko, S.D., Mottl, V.V., and Muchnik, I.B. Pattern recognition in experimental waveforms. *Pattern Recognition and Image Analysis. Advances in Mathematical Theory and Applications.* 1991, Vol. 1, No. 1, pp. 87-98.

[13] Mottl, V.V., and Dvoyenko, S.D. Supervised recognition of events in signals with changing probabilistic properties. *1995 IEEE Workshop on Nonlinear Signal and Image Processing*. Neos Marmaras, Greece, June 20-22. Vol. 1, pp. 238-241.

[14] Mottl, V., Kopylov, A., and Kostin, A. Edge-preserving in generalized smoothing of signals and images. *Proceedings of the 14th International Conference on Pattern Recognition*. Brisbane, Australia, August 16-20, 1998. Volume II, pp. 1579-1581.