INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION

INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
34100 TRIESTE (ITALY) - P.O.B. 586 - MIRAMARE - STRADA COSTIERA 11 - TELEPHONES: 224281/2/3/4/5/6
CABLE: CENTRATOM - TELEX 460392 - I

SMR/107 - 7

WORKSHOP ON PATTERN RECOGNITION AND ANALYSIS OF SEISMICITY

(5 - 16 December 1983)

SOURCE MECHANISM AND EARTHQUAKE PREDICTION

V. KEILIS-BOROK
F. PRESS

ON SEISMOLOGICAL APPLICATIONS OF PATTERN RECOGNITION

by Volodia KEILIS-BOROK

Institute of Physics of the Earth USSR Academy of Sciences Moscow

and Frank PRESS

Washington D.C. USA

ABSTRACT

In this paper we give a brief review of pattern recognition methods and applications in solid Earth physics. These include the determination of areas where strong earthquakes may occur, and we describe relevant major control experiments. Another application concerns the times at which strong earthquakes may occur and in particular their connection with other geophysical phenomena such as changes in the earth's rotation. It seems to us that the formal, logical approach of pattern recognition is well suited at the present time for research in solid earth physics and would deserve greater attention.

RÉSUMÉ

Dans cet article, nous donnons un bref aperçu des méthodes et applications de la reconnaissance des formes à la physique de l'intérieur de la Terre. Il s'agit en particulier de la détermination des régions où peuvent se produire des forts séismes, et nous décrivons les principaux tests de contrôle correspondants. L'autre sujet abordé est celui des époques auxquelles les forts séismes peuvent avoir lieu et en particulier, leur relation avec d'autres phénomènes géophysiques tels que les irrégularités de la rotation de la Terre. Il nous semble que le point de vue formel et logique de la reconnaissance des formes est bien adapté à l'heure actuelle, à la recherche en physique de l'intérieur de la Terre et mériterait qu'on y attache plus d'attention.

INTRODUCTION

Research in solid Earth physics, and especially in major problems such as the mechanism of geodynamics, earthquakes prediction and control, deep lateral inhomogeneities in the composition and state of the Earth's interior, etc., requires joint analysis of large volumes of very diverse quantitative and qualitative data.

Traditional methods of statistics and mathematical physics are inapplicable in many situations : first, because the available data base is too small for information of such multidimensional nature : and second, because of the absence of mathematical models and corresponding equations.

In such situations, pattern recognition becomes a powerful tool. In geological exploration, for example, this technique often increases at least by a factor of 3 the reliability of prediction of mineral deposits. Pattern recognition applications in solid Earth's physics are reviewed here.

The problem of pattern recognition can be formulated in the following manner : A set of vectors is given. Each vector belongs to one and only one of several (usually two) classes. The goal is to find the distinctive features of each class, i.e., a rule of recognition. Using this rule, we may further decide which class each vector belongs to.

In applications, each vector describes some real object, for example, a structure in studying earthquake-prone areas, a time-interval in earthquake prediction, a structure of another kind in geological prospecting, etc. Components of each vector are the answers to a questionnaire, i.e., the parameters of the object.

The existing algorithms of pattern recognition are of heuristic nature : no theory exists so far and the problem belongs to the still unexplored fields between statistics of small samples and logics. Usually, recognition is divided into the following three stages :

Learning : Determination of the rule of separation. Examples of the vectors of each class (" learning material ") are usually (but not necessarily) used at this stage.

Discrimination : Application of this rule to actual separation of vectors into classes.

Control experiments : Logical tests of the reliability of results obtained at previous stages. In the absence of a theory, this stage is decisive and absorbs most of the efforts.

Physical interpretation of results, especially of the rule of separation, often presents independent difficulties since [illegible] may be formulated in rather peculiar ways.

The results discussed in this paper were obtained mainly by algorithm " Cora-3 " [1] and its modification, " Clusters " [2]. Their brief description follows. The vectors are presented in binary form, with numerical parameters, discretized in few (usually two to four) intervals,

$$A = A_1 A_2 \ldots\ldots\ldots A_L$$

All binary digits A_i, $i = 1 \ldots . L$, and all combinations of two or three digits (A_i, A_j) and (A_i, A_j, A_k) are considered as the *traits* of the vector so that each vector has $L + C_L^2 + C_L^3$ traits, and the number of different possible traits is

$$2L + 4C_L^2 + 8C_L^3$$

The *features* of each class are looked for among these traits in the learning stage. The trait is a feature of some class if it is present relatively frequently in this class and relatively infrequently in other classes.

For each feature the set of vectors of the same class which have this feature is considered. The feature is not used for the recognition (second stage) if the corresponding set is part of some other set. If several of these sets coincide, then only one of the corresponding features is used for recognition.

In the algorithm " Clusters " groups of vectors are considered instead of separate vectors ; by definition, a group has some trait if at least one vector in the group has it. More detailed descriptions of both algorithms can be found in [2]. Many other algorithms of pattern recognition have been developed. Presently, there is no objective way to choose among them, and the success seems to depend mainly on the choice of questionnaire and on the thoroughness of the control experiments. The latter may partly compensate for the absence of strict estimations of the reliability of the results.

The rule of recognition in both algorithms is the following : the numbers of the features of the first and second class are determined for each vector. If their difference is large enough, the vector is identified as belonging to the first class and vice versa.

We see from this description that application of pattern recognition includes many arbitrary decisions – on the numerical parameters, the choice of questionnaire, etc. This creates the danger of self-deception which could be avoided, by control experiments.

We shall now describe the applications of pattern recognition in the Physics of Solid Earth.

EARTHQUAKE-PRONE AREAS

The determination of areas where strong earthquakes may occur can be reduced to pattern recognition in the following way. We shall call " strong " the earthquakes with magnitude M not lower than some threshhold M_o.

Vectors represent the structures which can be earthquake prone : segments of active faults, zones of intersection of lineaments, etc. In the studies which are reviewed here, the possible position of epicenters, i.e., the places where an earthquake can be initiated, were examined; the conclusion that an earthquake can be initiated only at some specific part of the entire faultbreak may be of independent interest.

We shall call such vectors D (Dangerous), and the rest of them N. We have *a priori* division of learning material into two classes : (D_1, N_1) and (D_2, N_2) – correspondingly close to and far from the known epicenters of strong earthquakes. The first class may contain some vectors N due to the errors in location or magnitude; the second class may contain some vectors D because strong earthquakes may occur where they are still unknown. The vectors of the first class are subdivided into clusters; each cluster corresponds to the places near a known epicenter of a strong earthquake.

The most detailed and reliable results so far were obtained for strike-slip earthquakes of California, $M \geqslant 6.5$ [2]. The objects of recognition were the segments of the major strike-slip faults of the San Andreas system. The finally selected questionnaire is given in Table 1. Due to the limited volume of learning material only a small part of possibly relevant parameters could be used for learning. The features of D and N, determined by the algorithm " Clusters ", are shown in Table 2. Each line in this table represents one feature – the conjunction of two or three inequalities. It was found that a significant part of the major faults of the San Andreas system is of the N type.

We shall now describe the major control experiments.

(i) Some strong earthquakes were considered as unknown. The objects around their epicenters were reassigned in learning to the second class. Then we applied the same algorithm, from the very beginning, in order to check whether these objects which are actually D would be recognized as D.

In one version of this experiment, six strong earthquakes which occurred after 1911 were considered as unknown, by an admittedly theatrical reason : to simulate the situation in which we conducted such an analysis some years ago and are now testing the results on subsequent earthquakes. At least one segment of the faults near epicenter of each of these six earthquakes was recognized as D.

In another version of this experiment, the earthquakes of some area were considered as unknown. We successfully recognized as D the faults' segments near eliminated epicenters of Northern and Southern California : Mendocino zone, or San Jacinto mountains, or Garlock fault. However, the experiment failed for Central California (the area between San Francisco and Parkfield).

(ii) The distinctive features, obtained for California, were applied to Anatolia, and vice versa. The experiment was surprisingly successful : with learning of one region we recognized the places of occurence of all strong earthquakes in another region (with only one exception for Anatolia). In other control experiments we tested the stability of recognition against variation of

Table 1

Parameters used for recognition (after [2])

N	Name of parameter	Symbol	Assigned threshold
1	maximal elevation	h_{max}	≤ 500 m $\leq 1{,}250$
2	minimal elevation	h_{min}	≤ 50 m ≤ 500
3	elevation difference	Δh	≤ 500 m $\leq 1{,}000$
4	"gradient"	$\Delta h/l_1$	≤ 30 m/km ≤ 70
5	relative area of soft sediments	q	≤ 10 ≤ 50
6	type of rocks	–	presence of I
7	distance to closest fault	r_1	≤ 12.5 km ≤ 37.5
8	distance to closest end or intersection of faults	r_2	≤ 12.5 km ≤ 37.5
9	distance to geothermal zone	r_3	≤ 25 km ≤ 75
10	distance to the closest zone of divergence	r_4	≤ 100 km ≤ 200
11	distance to the reference point (intersection of San Andreas and Big Pine faults)	r_5	≤ 125 km ≤ 375
12	distance to closest water reservoir	r_6	≤ 0 km ≤ 25
13	the number of unnamed faults on *Tectonic Map of U.S. (USGS, 1962)*	n_1	≤ 3 ≤ 5
14	the number of changes in types of relief	n_2	$= 0$
15	maximal elevation	H_{max}	≤ 500 m $\leq 1{,}500$
16	minimal elevation	H_{min}	≤ 0 m ≤ 200
17	the number of contacts between rocks of different age on *Geological Map of North America* (USGS, 1965)	n_3	≤ 8
18	the number of parallel faults	n_4	≤ 2 ≤ 3
19	the number of faults	n_5	≤ 1 ≤ 3
20	the number of ends and intersections	n_6	≤ 1 ≤ 2
21	the angle between the fault and the dominant structural trend in the region (°)		≤ 10 ≤ 20
22*	distance to a region of large precipitation	R_3	≤ 62.5 km
23*	elevation differences	ΔH	$\leq 1{,}500$ m

The rule for coding the description of a point is the following: "1" means that the inequality in the last column is fulfilled.
* These parameters were used in control experiments only with learning based on earthquakes prior to 1934.

Table 2

Distinctive D and N features (after [2])

Feature number	Maximum elevation, h_{max} (meters)	Gradient $\Delta h/l_1$	Relative area of soft sediments, q %	Type of rocks	Distance to closest end or intersection of faults, r_2, km	Distance to nearest geothermal zone, r_3, km	Distance to nearest body of water, r_6, km	Minimum elevation, H_{min} (meters)	Number of contacts between rocks of different age on geologic map	Number of faults n_5	Angle between fault and dominant regional strike (deg.)
	1	4	5	6	8	9	12	16	17	19	21
					D						
1					≤ 37.5	≤ 75					≤ 10
2				no I	≤ 37.5						≤ 10
3				no I	≤ 37.5					≤ 3	
4			> 10		≤ 12.5					> 1	
5					≤ 37.5			≤ 200	≤ 8		
6			> 10		≤ 37.5			≤ 0			
7				no I	≤ 37.5	> 25					
8			≤ 50		≤ 12.5	> 25					
					N						
1							> 20		> 8	≤ 3	
2					> 37.5			≤ 0		≤ 3	
3				I		> 25				≤ 3	
4	> 500			I						≤ 3	
5					> 12.5		> 20		> 8		
6					> 12.5	> 75			> 8		
7				I		> 75			> 8		
8					> 37.5	> 25		≤ 200			
9			> 10		> 37.5			≤ 200			
10	> 500				> 37.5			≤ 200			
11	> 500		> 10		> 12.5						

the data, the questionnaire, the criteria for selection of the features, the criteria for recognition, etc. We also changed the objects of recognition, from all segments of major faults to the intersection of major lineaments.

The stability is by and large quite sufficient, though in some cases it is marginal.

The distinctive features of *D* areas can be qualitatively summarized as follows: *epicenters of strong strike-slip earthquakes can occur only in areas characterized by proximity to intersections or ends of major faults and by neotectonic subsidence against a background of weaker uplifts.* The study of Bouguer anomalies shows that these areas have deep roots; namely, fractured density inhomogeneities are typical for the Earth's crust below these areas. Pattern recognition allowed us to find similar regularities in the location and distinctive features of earthquake-prone areas for other regions as well: for Pamir, Tien-Shan, Anatolia, Little Caucasus, Southern Balkans, Eugian basin, Sierra-Nevada, Italy and for the strongest ($M \geq 8$) earthquakes of the Circumpacific seismic belt.

In this example the emphasis was placed on the study of the physical mechanism of phenomena, using distinctive features as a guide. The connection between the Chandler wobble, strong earthquakes, rotation of the Earth and geomagnetic changes was studied in [3].

Objects of recognition were the consecutive years. The years of the build-up and decay of the amplitude of the Chandler wobble were assigned to different classes, G_1 and G_2, in learning.

The questionnaire is shown in table 3. The distinctive features in table 4 are formed from the first fourteen questions relevant to seismicity. This striking pattern emerges from these features : most of the seismic belts (with only two exceptions) either show major earthquake

TABLE 3

Questions posed for pattern recognition (after [3])

1. In the preceding 5th, 4th, or 3rd year WTAEO* ⩾ 1.0 EM-8† in the New Guinea-New Hebrides region ?
2. In the following 0th, 1st, 2nd year WTAEO ⩾ 1.0 EM-8 in the New Guinea-New Hebrides region ?
3. In the following 0th, 1st, 2nd, 3rd or 4th year WTAEO ⩾ 0.5 EM-8 in the Tonga-Kermadec region ?
4. In the preceding 2nd, 1st, or 0th year WTAEO > 1.5 EM-8 in the Kurile-Japan-Mariana region ?
5. In the preceding 1st, 0th or following 1st year WTAEO > 2.0 EM-8 in the Peru-Chile region ?
6. In the preceding 3rd, 2nd, 1st, 0th or following 1st year WTAEO > 1.0 EM-8 in central America ?
7. In the preceding 2nd, 1st or 0th year WTAEO > 0.8 EM-8 in Java ?
8. In the following 0th, 1st, 2nd, 3rd, or 4th year WTAEO > 1.5 EM-8 in the Ryukyu-Philippines region ?
9. In the following 0th, 1st, 2nd, 3rd, or 4th year WTAEO > 0.5 EM-8 in the San Andreas-Gorda Ridge region ?
10. In the preceding 2nd, 1st, 0th, or following 1st or 2nd year WTAEO > 0.5 EM-8 in China ?
11. In the preceding 2nd, 1st, or 0th yar WTAEO > 0.5 EM-8 in the Aleutians ?
12. In the preceding 1st, 0th, or following 1st year WTAEO > 0.5 EM-8 along a continent-continent collision zone ?
13. In the preceding 5th, 4th, 3rd, 2nd, 1st, 0th, or following 1st year WTAEO > 0.5 EM-8 in the Caribbean ?
14. In the preceding 1st, 0th, or following 1st year were there > 15.0 EM-8 integrated over the whole Earth ?
15. In the preceding 7th, 6th, 5th, 4th, 3rd, 2nd, or 1st year was there a change in the absolute value of Earth's rotational acceleration $(|\dot{\omega}/\omega|)$ of $> 2 \times 10^{-9}$ in units $\Delta\omega/\omega$ yr^{-1} ?
16. In the present year was the absolute value of the magnitude of Earth's rotational acceleration $(|\dot{\omega}/\omega|) < 1.0 \times 10^{-9}$ in units $\Delta\omega/\omega$ yr^{-1} ?
17. In the present year was the Earth's rotational acceleration negative ?
18. In the present year was the length of day (l.o.d.) more than 0.5 ms longer than the mean value (corrected for secular slowdown) ?
19. In the following 2nd year was the speed (v) of the westward drift of the eccentric geomagnetic dipole > 0.25 degrees yr^{-1} ?
20. In the following 3rd or 4th year was the acceleration ($\dot{v}$) of the eccentric geomagnetic dipole motion negative ?
21. In the preceding 1st year was the second derivative of the squared amplitude of the Chandler Woble positive ?
22. In the preceding 5th, 4th, or 3rd year was there a major volcanic eruption on Earth ?

* WTAEO means 'was there an event of'.

† EM-8 signifies a measure of seismicity in units of equivalent magnitude 8 earthquakes, obtained by dividing the cumulative, annual seismic moment released in a region by the seismic moment for a magnitude 8 earthquake.

TABLE 4 a

Characteristic traits for G_1 years (after [3])

	Question													
Trait	1	2	3	4	5	6	7	8	9	10	11	12	13	14*
1	1	1	.	0	.	.	.	.	.	.	.	.	.	.
2	1	1	.	.	.	1	.	.	.	.	.	.	.	.
3	1	.	.	.	.	.	.	.	.	.	1	.	.	.
4	.	1	.	.	.	.	.	.	.	.	1	.	.	.
5	.	.	1	0	.	.	.	.	.	0	.	.	.	.
6	.	.	1	.	.	.	.	1	.	.	.	.	.	0
7	.	.	.	0	1	.	.	.	.	.	.	.	.	.
8	.	.	.	0	.	1	.	.	.	.	.	.	.	.
9	.	.	.	.	1	1	.	.	.	.	.	.	.	.
10	.	.	.	.	1	.	1	.	.	.	.	.	.	.
11	.	.	.	.	1	.	.	.	.	.	1	.	.	.
12	.	.	.	.	1	.	.	.	.	.	.	.	.	1
13	.	.	.	.	.	1	1	.	.	.	.	.	.	.
14	.	.	.	.	.	1	.	1	.	.	.	.	.	.
15	.	.	.	.	.	1	.	.	.	.	1	.	.	.
16	.	.	.	.	.	.	1	.	1	.	.	.	.	.
17	.	.	.	.	.	.	1	.	.	.	1	.	.	.
18	.	.	.	.	.	.	.	1	.	.	1	.	.	.
19	.	.	.	.	.	.	.	.	1	.	.	.	1	.
20	.	.	.	.	.	.	.	.	.	.	1	.	.	1

* Columns describe activity in seismic belts (1 = yes, 0 = no); rows are combinations of answers which form individual traits.

TABLE 4 b

Characteristic traits for G_2 years (after [3])

	Question													
Trait	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	.	.	1	.	.	0	.	.	.	.	.	.	0	.
2	.	.	.	1	0	.	.	.	.	.	.	.	.	.
3	.	.	.	1	.	0	.	.	.	.	.	.	.	.
4	.	.	.	1	.	.	.	.	.	.	0	.	.	.
5	.	.	.	.	0	.	.	.	.	.	.	0	.	.
6	.	.	.	.	0	.	.	.	.	.	.	.	0	.
7	.	.	.	.	.	0	0	.	.	.	.	.	0	.
8	.	.	.	.	.	0	.	.	.	.	0	0	.	.
9	.	.	.	.	.	0	.	.	.	.	.	0	.	0
10	.	.	.	.	.	0	.	.	.	.	.	.	0	0
11	.	.	.	.	.	.	.	0	.	.	.	.	.	0
12	.	.	.	.	.	.	.	.	.	.	0	.	0	.

TABLE 5 a

Display of 8 G_1 traits associated with years in which amplitudes of the Chandler Wobble are increasing. The component features of each trait are shown in boxes in the proper time sequence with respect to a G_1 year. The number of the question (Table 3) from which the feature is derived is given in each box as an aid to the abbreviations. Eq, Earthquake; l.o.d., length of day (after [3])

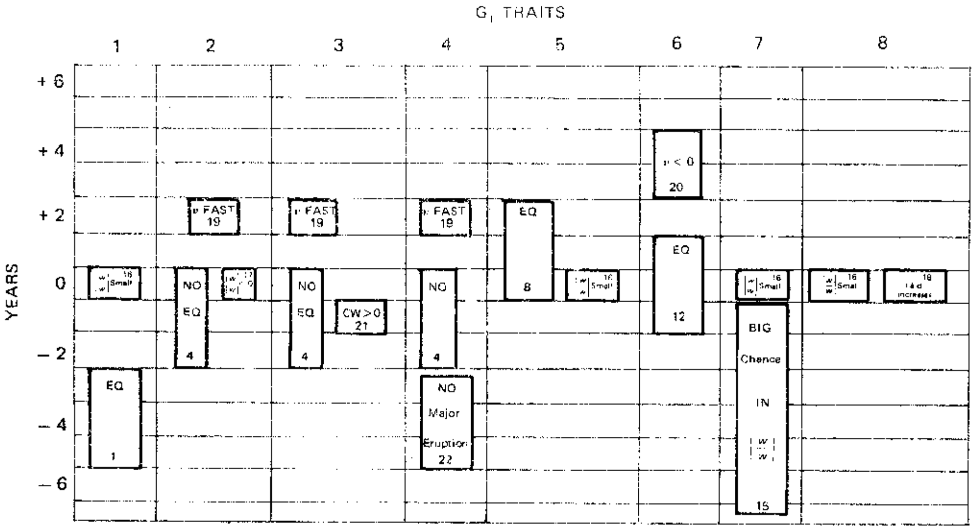
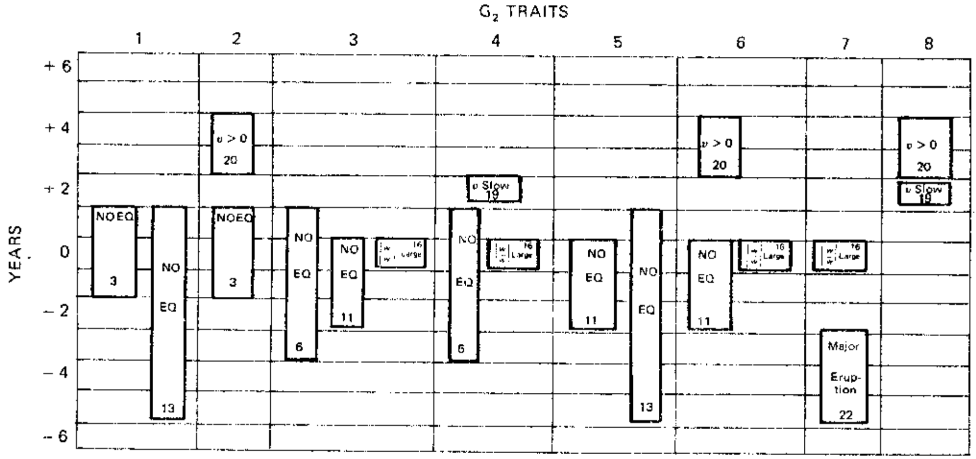


TABLE 5 b

Display of eight traits associated with years in which amplitudes of the Chandler Wobble are decreasing. Abbreviations as in Table 5 a (after [3])



activity during G_1 years or no such activity during G_2 years. This implies a preferential excitation of the wobble by strong earthquakes in certain areas.

The features in table 5 are formed from all questions listed in table 3. The position of the feature on the time-axis shows the optimal time shift between the G_1 or G_2 year and the year of event, described by the questionnaire (for details see [3]). These features show that G_1 years are preceded by a decrease of the Earth's rotational acceleration, which remains low during G_1 years.

During G_2 years, the rotational acceleration is large and westward geomagnetic drift is slow. These years are preceded by a drop in volcanic activity and are followed by a speed-up in the westward drift three to five years later. Thus, the algorithm recognized the well known lag between rotational acceleration and the increase of the westward drift.

The described patterns led to the development of a hypothetical qualitative mechanism that the Chandler wobble is excited by tectonic deformations preceding and accompanying the earthquakes. (Deformations, realized by the earthquakes alone, would not be sufficient).

In the subsequent period the angular momentum is transferred to the Earth fluid core. This transfer causes attenuation of the Chandler Wobble and geomagnetic perturbations.

PERSPECTIVES

There is at present no physical model of the major features of the constitution and development of the Earth. In the absence of such a model, one must resort to intuition or to formal logic to comprehend all relevant data. This alternative can be illustrated by table 6.

The left column, representing analysis " by eye ", is taken from L. Tolstoy: Pierre Besuhoff is trying to recognize the potential victor over Napoleon (*War and Peace*, vol. 3, pt. 1, ch. XIX); the name of this victor, according to belief, should be equivalent to the apocalyptic number 666. The right column is a summary of logical rules formulated by R. Descartes in " Treatise on methodology ". It is by no means our intention to underestimate the intuitive approach: it seems indispensible at the beginning of research, for discovery of a new hypothesis, which could be lost due to premature formalization. An example is the discovery of the low-velocity zone in the upper mantle. However, the intuition may fail at later stages, if the volume of accumulated data is too large. This is apparently the present state of art in many major problems of solid Earth science.

Major discoveries over the past two decades, for example plate tectonics, are due to the intuitive approach, with physical and mathematical methods used mainly for preliminary data processing.

TABLE 6

P. BESOUHOFF		R. DESCARTES
L'Empereur Napoléon	*666*	1. Make only reliable a priori assumptions.
L'Empereur Alexandre	691	2. Divide the problem into provable parts.
La Nation Russe	640	3. Organize the analysis into logical tree.
Comte Pierre Besouhoff	754	4. List all assumptions and allow for all facts.
Comte Pierre Bezouhoff	824	
Comte Pierre de Besouhoff	763	
Le comte Pierre de Besouhoff	813	
Le Russe Besuhof	671	
L'Russe Besuhof	*666 !*	

It seems, however, that at the present time, the more formal, logical approach deserves greater attention, as illustrated by the experience described above. It allows extraction of more definite conclusions from a given set of data, which is a useful supplement and sometimes possibly an alternative to the growing data accumulation. It is also a necessary preliminary stage to development of a physical foundation for Earth science.

We have chosen this review for contribution to the volume in honor of Professor Coulomb, since he belongs to the grand school, which represents the physical and mathematical strain of Earth' science.

REFERENCES

[1] **Bongard, M.**, 1968. – Raspoznavanie obrazov (Pattern recognition), Moscow *Nauka* (in Russian).

[2] **Gelfand, I. M., Sh. A. Guberman, V. I. Keilis-Borok, L. Knopoff, F. Press, E. Ya. Ranzman, I. M. Rotwain and A. M. Sadovsky**, 1976. – Pattern recognition applied to Earthquake epicenters in California, *Physics of the Earth and Planetary Interior* **11**: 227-283.

[3] **Press, Frank and P. Briggs**, 1975. – Chandler Wobble, earthquakes, rotation, and geomagnetic changes, *Nature*, **256**: 270-273, July 24.