STATISTICS APPLIED TO FACTS AND CONCEPTS IN GEOSCIENCE

F. P. AGTERBERG [1]

ABSTRACT

Agterberg, F. P. 1979 Statistics applied to facts and concepts in geoscience. *In*: W. J. M. van der Linden (ed.): Fixism, mobilism or relativism: Van Bemmelen's search for harmony — Geol. Mijnbouw 58: 201-208.

This is a brief review of alternative methods of problem-solving in geoscience. Special attention is given to applications of the theory of probability, mathematical statistics, computers and artificial intelligence. It is desirable to maintain a clear-cut distinction between reliable facts which can be stored in data banks and concepts which should be incorporated in the specifications of statistical models designed for specific purposes. Two illustrative examples deal with the probability of occurrence of mineral deposits. This probability is conditional upon the occurrences of geological features systematically quantified and processed for large regions.

INTRODUCTION

This paper is concerned with the 'computerization' of geoscience, a topic discussed by VAN BEMMELEN (1972) in the first chapter of his book on geodynamic models.

According to VAN BEMMELEN (1972, p. 8), the computerization and adoption of mathematics in geoscience presents a wide ranging and challenging field of future developments but is fraught with organizational, educational and technical difficulties. The usage of mathematics has progressed further in geophysics and geochemistry than in geology because of differences in the nature of the data and other sampling methods. The relative weights to be assigned to data collected from geological observations at the surface of the earth are often unknown. A fact can be important and applicable at other places, frequently at great distances from the original observation point, or it may be unique and unimportant. Most geological facts are not unique but greatly influenced by other observations and by deductions from specific concepts on geological processes. Basic measurements, such as the determination of the strike and dip of a structural plane, might represent local exceptions and their importance can be evaluated only against a background of regional data. Satisfactory statistical averaging may not be possible due to lack of exposures.

Van Bemmelen continues his argument by stating that the shortcomings of the classical methods of geological observations constrain the quantification of geological data. Much is left to the 'feeling' and experience of the individual geologist. The results of his work, presented in the form of maps, sections and narratives with hypothetical reconstructions of the geological evolution of a region, do not have the same exactitude as the records and accounts of geophysical and geochemical surveys which are more readily computerized even though the results may be equally accurate in an interpretive sense. Geophysical and geochemical variables generally are determined by the characteristics of the bedrock geology which, in a given area, is likely to be nonuniform because of the presence of different rock types. Thus the heterogeneous nature of the geological framework will be reflected in these other variables.

Not a single geoscientist will be able to contain in his memory and to have at his immediate disposal all significant

[1] Geological Survey of Canada, 601 Booth Street, OTTAWA KlA OE8, Canada.

data concerning his field of research. For this reason, Van Bemmelen suggested that a worldwide network of cells should be devised and that as many relevant geoscience data as possible should be collected for these cells and stored in computer-based data banks with preservation of the geographical locations of the features and their depths below the topographic surface. Because of the great fluctuations in relative importance and reliability of the data, the evaluation of features in data banks by the individual geologist may in several respects remain superior to the indiscriminate manipulation by a machine.

Some of the ideas developed in this paper have previously been expressed elsewhere (AGTERBERG & ROBINSON, 1972; AGTERBERG, 1974-a, chapter 1; AGTERBERG, 1978; AGTERBERG & DAVID, 1979). Geology is essentially a probabilistic science in which all conclusions and predictions should be qualified because they are subject to a variable amount of uncertainty. It is desirable to make a clear-cut distinction between facts and concepts. Facts should become part of a data base and concepts should guide the choice of variables and their interrelationships in statistical models which, in addition to the variables, contain (constant) parameters that are estimated by systematic comparison with the facts.

Most other sciences having to deal with uncertainty have been subjected to two important changes during the past fifty years. The first change consisted of a widespread acceptance of probabilistic concepts and statistics. For example, agriculture was quick in adopting the inductive statistical approach to sampling and data analysis developed by R. A. Fisher and his colleagues during the 1920's and 1930's. The second important change occurred during the 1960's and 1970's when digital computers were widely adopted as tools for performing basic calculations and for the manipulation of facts stored in data banks.

The impact of these changes has remained relatively minor in geoscience where most important progress was triggered by new methods of transportation and the great refinement of instruments which led to new types of observations. These new observations, however, did not necessarily mean important progress until a suitable conceptual framework became available to interpret them.

As new facts continue to accumulate in geoscience it will become increasingly necessary to integrate the results obtained by the various subdisciplines. It is likely that the implications of the fact that the amounts of mineral resources hidden in the earth are limited will be even more seriously regarded in the future. The relatively new field of resource analysis provides one good reason for the integration of geoscience data of different types. The examples given in this paper for illustration deal with the probabilistic estimation of occurrence of undiscovered mineral deposits from facts stored in special-purpose, experimental data banks. The methods used for data processing are based on the theory of mathematical statistics and on the theory of regionalized random variables developed by MATHERON (1965).

INTEGRATION OF DIVERSE DATA; PROGNOSIS-DIAGNOSIS METHOD

MARSCHAK (1964) has made a sharp distinction between 'normative' and 'descriptive' sciences. The normative sciences such as engineering, medicine and agriculture address themselves to the question of what is optimal behaviour in the face of a given task. The descriptive sciences such as physics, biology and psychology are concerned with the behaviour of nature. DE FINETTI (1972) has pointed out that the theory of probability is a normative theory. Uncertainty is subjective and caused by lack of knowledge. Hence it is desirable to know how right or wrong any predictions are in the light of all available evidence. De Finetti's 'good probability appraisers' consider a diversity of relevant facts and reach conclusions primarily by inductive reasoning.

Decision-making in the mineral industries and broad regional resource appraisal are obviously normative topics in geoscience. However, probabilistic arguments are used in the formulation of most hypotheses. VAN BEMMELEN (1961) has defined the prognosis-diagnosis method of research as follows. The arrangement of observations into a pattern of relationships is a mental process of induction. To reason by induction is nothing other than to learn from experience (DE FINETTI, 1972, p. 147). A working hypothesis obtained by intuition has to be in agreement with the facts and gains in functional validity if it leads to deductions that can be verified. New facts can lead to modification of the hypothesis or theory. The result is a recurrent cycle of inductions and deductions.

De Finetti proposed an intuitive maxim according to which a hypothesis is considered as practically certain when it is supported by numerous highly diverse facts in its favour, even if, taken separately, none of these facts would satisfy our requirements for proof. He used Wegener's theory of continental drift for example. Wegener used facts from many geoscience fields because he asked himself the question: 'How would we judge the judge who based his sentence on one part only of the available evidence'. It is well known that Wegener was ridiculed especially on account of his poorly conceived speculations concerning the forces capable of transporting continents. Nevertheless his prognosis turned out to be correct.

DE FINETTI (1972) has pointed out that in order to appraise correctly any hypothesis the weight of all arguments pro and con known at the time must be considered. He deplores the prejudiced attitude of many scientists against broadly based hypotheses which indicates 'a dangerous unilateralness and a deplorable preconception against taking into account evidence from fields extraneous to one's own bailiwick'. Most geoscientists subscribe to Chamberlin's method of 'multiple working hypotheses'. In practice, however, a single person may not be able to form more than one preferred opinion which is in agreement with the rules in the school of thought to which he belongs.

PREDICTION IN GEOSCIENCE

The goal of the earth scientist is to create theoretical models from his observations on a great variety of features visible in exposures of rocks. These models, whether they are qualitative or statistical, should provide a description of the facts but, in addition to this, they should possess predictive potency. For verification the features predicted by a concept to exist at a given location can be compared with the features that actually exist at that place. In practical applications of geologic theory, the capability of a concept to predict may provide the only direct measure of its usefulness. This prediction primarily takes place in three-dimensional space. For example, when a new tunnel is constructed the geoscientist is asked to tell in advance which rock types and structures will be encountered. The observable features represent the end products of chiefly slow processes which took place in the course of geologic time. For the construction of most types of models such as the geological map of a region, an understanding of the processes that have led to the origin of the phenomena may be essential. Geologic time is an important variable in most prognostic models although the diagnosis takes place in three-dimensional space.

A simple example may illustrate the importance of concepts in the prediction of unobserved facts. NIEUWENKAMP (1968) pointed out that by his adherence to Werner's neptunist point of view L. von Buch in 1842 was led to assume that the interior of the hills of Kinnehulle in Sweden consisted of basaltic rocks. The observations in this region consist of granitic rocks in the valleys, subhorizontal Palaeozoic sedimentary strata on the slopes and subhorizontal basaltic rocks on the tops of the hills. At the core of Von Buch's prediction was a stratigraphic column in which granite was overlain by basalt, and basalt by sedimentary rock. Because of his better understanding of processes such as erosion, sedimentation, and the movement of basaltic flows, to-day's geologist would rapidly conclude that the interior of these hills consists of Palaeozoic sedimentary rocks. The challenge of the geomathematician is not only to store the relevant facts in a data base but also to develop statistical models which incorporate conceptual knowledge. Obviously both Von Buch's and the modern interpretation can be simulated on a computer depending on whose concepts are programmed.

HOW TO EXPRESS UNCERTAINTY

The question of how good is your prediction is continuously asked in economic geology. For example, in the oil industry it is known from experience that most holes, especially the wildcats, will remain dry. Nevertheless, the geologist is asked to provide an opinion as to whether it is worthwhile drilling a hole at a particular site. This problem has been discussed in detail by GRAYSON (1960) and DE FINETTI (1974). The geologist himself does not have any say in the final decision of whether or not to drill. This is the responsibility of the decision-maker who will reach his conclusion after considering all different pieces of information available to him of which the geologist's report is just one piece. The geologist cannot state categorically that oil is present or absent. Neither can he restrict himself to a mere listing of the reliable facts. His conclusion about the probable outcome of the drilling is precisely what the geologist is called upon to provide. GRAYSON (1960) found that most geologic reports contain probabilistic answers which are disguised in vague adjectives ('fairly good prospect', 'favourable', 'permissive', 'it's difficult to say', etc.). He has proposed methods for translating the geologist's opinion into subjective probabilities. In recent years methods of this type have been used extensively for the regional appraisal of undiscovered oil and gas resources (see e.g. MILLER, 1977).

GRAYSON's (1960) book deals primarily with drilling decisions by oil and gas operators. A sign hanging in the office of one of the operators interviewed by him states 'Holes that are going to be dry shouldn't be drilled', although some years previously this particular operator had drilled thirty consecutive dry holes. DE FINETTI (1974) used this paradox to urge the geologist to express his prediction in a probabilistic manner rather than translating it into the inadequate logic of absolute certainty.

One method of avoiding the logic of certainty would consist of obtaining the predictions from reliable information by means of methods akin to multivariate analysis on the one hand and based on geostatistical reasoning on the other (AGTERBERG, 1978). Suppose that all relevant facts for a region have been stored in a data bank and that the prediction is to be carried out by statistical treatment of these facts. Then the relevant knowledge of the geological evolution of the region should be incorporated in the specifications of the statistical model. In this approach the uncertainty might be estimated at the same time as the expected values.

WHICH LOGIC TO USE?

In the computerization of geoscience it can be attempted to simulate the prognosis-diagnosis method practiced by most individual geologists. In Novosibirsk, at the Institute of Geology and Geophysics of the Siberian Branch of the Soviet Academy of Sciences, a team consisting of eight geologists and eight mathematicians use formal logic and methods of computer programming in the attempt to mimic the mind of the experienced geologist (DMITRIEV, 1976). At another institute, also in Novosibirsk, VORONIN (1976) and his group are developing novel semantically-based schemes for finding undiscovered mineral deposits by computer. Voronin also uses formal logic but he tries to steer clear of any subjective formulation of concepts as takes place in the mind of the geologist. Most other workers in this field use geological concepts and try to use as much mathematical statistics as is possible.

The human mind allows the formulation of hypotheses which are flexible to the extent that they may immediately incorporate all new facts before the hypotheses could be properly tested. On the other hand, the advantage of using the logic of mathematics is that it is indisputable and, when random variables are used, it is possible to check the deductions against reality. P.A.M. Dirac (in MARLOW, 1978) has advocated the use of mathematics in physics as follows: 'One should keep the need for a sound mathematical basis dominating one's search for a new theory. Any physical or philosophical ideas that one has must be adjusted to fit the mathematics. Not the other way around. Too many physicists are inclined to start from preconceived physical ideas and then try to develop them and find a mathematical scheme that incorporates them. Such a line of attack is unlikely to lead to success'.

It is likely that most geologists agree with T. C. Chamberlin's famous dictum: 'There is perhaps no beguilement more insiduous and dangerous than an elaborate and elegant mathematical process built upon unfortified premises'. It should be kept in mind, however, that Chamberlin's criticism was directed against Lord Kelvin who did not use random variables. Some resistance by geoscientists to the more widespread use of mathematics may be caused by their early exposure to calculus only and their lack of training in the theory of probability.

COMPUTER HARDWARE AND SOFTWARE

Today one can buy in the store a machine with as much speed and precision as the first experimental computers of the 1940's. Large computers are millions of times more powerful and are still becoming faster, more accurate, more reliable, and about 25% less expensive per year (RAPHAEL, 1976). On the other hand, the development of artificial intelligence (A.I.) has been slow. According to WINSTON (1977), A.I. is the study of ideas which enable computers to accomplish the things that make people seem intelligent. A.I. will be reviewed in more detail later in this paper.

The hardware of a computer consists of (a) memory, (b) central processing unit (CPU), and (c) peripheral equipment. The term 'software' refers to the instructions and computer programs. The memory can remember a large quantity, frequently millions, of numbers and words which can be manipulated in the CPU. Magnetic tapes, disks and drums in the peripheral equipment are used to extend the memory of a computer to billions of facts to which there is almost instantaneous access.

Special peripheral equipment can increase the input capabilities by giving the digital computer a numeric representation of the physical phenomenon it is experiencing. For example, a geologic map could be digitized automatically as about a million grey-level values measured for separate picture points whose locations are memorized. Other peripheral devices such as plotters and TV-like cathode-ray tubes (CRT's) facilitate graphical display. These technological advances have been accompanied by advances in computer-based statistical analysis, improved data management techniques, and the development of new methods of pattern analysis (cf. WATSON, 1975).

TWO EXAMPLES OF THE USE OF COMPUTERS AND STATISTICS

Figure 1A shows the occurrence of acidic volcanic rocks as represented on geological maps with a scale of about 1:125,000 of the Bathurst-Newcastle area in the Canadian Appalachian Region. The C-shaped area of rhyolitic rock represented in black is surrounded by metabasalt and sedimentary rocks with which it belongs to the Tetagouche Group of Middle-Late Ordovician age. The rhyolitic core has been referred to by some geologists as a basin structure and by others as a dome. SKINNER (1974) has argued convincingly that the C-shape is the result of two periods of folding. According to Skinner, the Tetagouche Group was folded into northwesterly trending recumbent folds overturned toward the southwest during the late Ordovician Taconic Orogeny, then refolded with an about northeasterly trending axis during the Devonian Acadian Orogeny. The rhyolitic core would be the youngest part of the Tetagouche Group and stratigraphically underlain by the metabasalt and sedimentary rock. Skinner suggested an ignimbritic (pyroclastic flow) origin for most of the acidic volcanics. During their deposition, these ignimbrites covered a region several times larger than the part of the Bathurst-Newcastle area now underlain by this type of rock. SKINNER (1974) has subdivided the rhyolitic unit into a rhyolite tuff subunit and a rhyolite crystal tuff subunit which differ only in the relative proportions of nonporphyritic and porphyritic rhyolitic rocks they contain. The rhyolitic crystal tuff subunit usually underlies the rhyolite tuff subunit.

Figure 1F which is for the same area as figure 1A shows the occurrences of forty volcanogenic massive sulphide deposits located at the centres of the black squares. Most of these mineral deposits occur within or adjacent to rhyolite crystal tuff although the host rock is generally a chlorite schist associated with magnetic iron formation. On the basis of the regional geological model it is to be expected that most massive sulphide deposits occur within the pattern shown as acidic volcanic rocks in figure 1A but relatively close to contacts between acidic volcanics and the other rock units which would be older.

AGTERBERG & FABBRI (1978) have studied the relationship between the patterns of figures 1A and 1F after digitizing them on a flying spot scanner and by using Minkowski operations (cf. WATSON, 1975) such as dilatation (Figs. 1B and 1C) and erosion (Figs. 1D and 1E). Two dilatations of the forty mineral deposits are shown in figure 1F and 1G. Figure

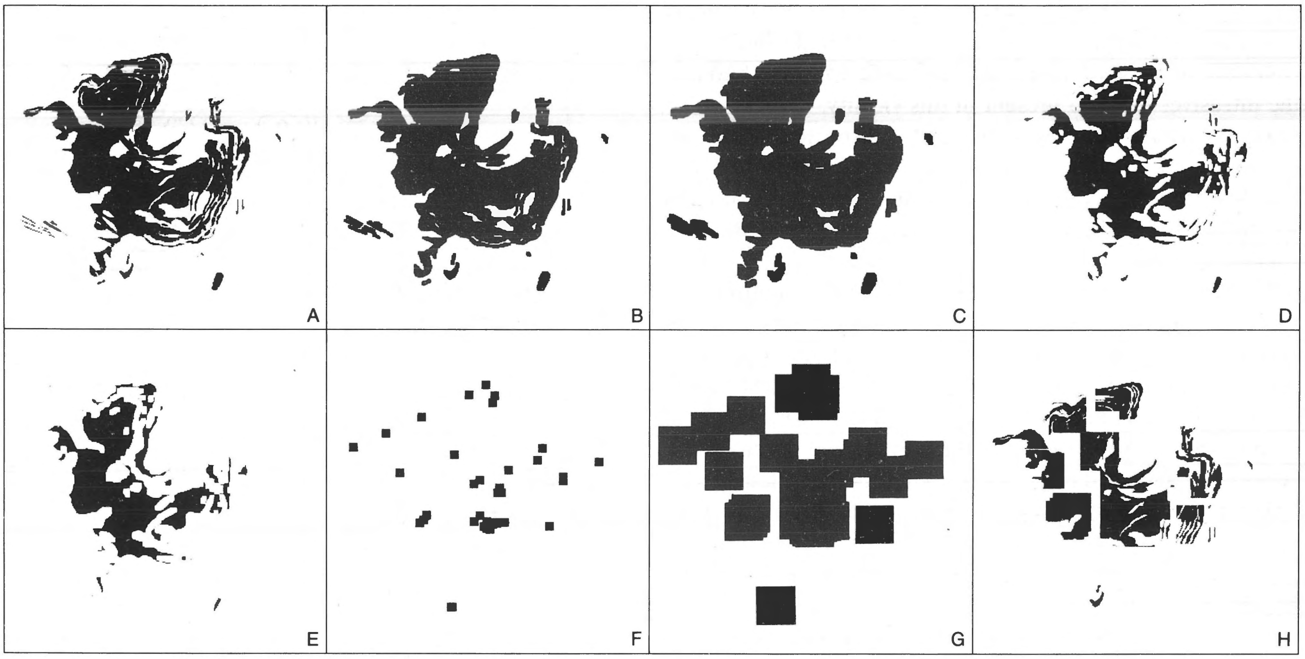

Fig. 1
Image analysis of acidic volcanic rocks (A) and volcanogenic massive sulphide deposits (F) in the Bathurst-Newcastle area, New Brunswick (from Agterberg & Fabbri, 1978). B and C represent first two dilatations of pattern of acidic volcanics (A); D and E are first two erosions of A; F. and G are 9-point and 19-point dilatations of the original pattern of 40 points for mineral deposits; H represents area of coincidence of A and G. A single dilatation (or erosion) consisted of adding (or subtracting) adjoining picture points along all boundaries. Distance between original picture points was 259 m in the north-south and east-west directions. North points upwards. Frame measures approximately 80 km on a side.

1H is the intersection between the patterns of figure 1A and 1G. The area of each black pattern can be precisely measured by counting the picture points which are black. These measurements could be used to answer, for example, the following five questions: (1) What is the probability that a very small area underlain by acidic volcanic rocks (as shown on a 1:125,000 geological map) contains a sulphide deposit? (2) What is this probability if the very small area is not more than a given distance removed from at least one contact between acidic volcanics and other rock units? (3) What is the probability that a square cell of a given size placed at random on the study area contains at least one sulphide deposit? (4) What is the latter probability given the proportion of the cell underlain by acidic volcanics? (5) What is the proportion of acidic volcanics in a cell of given size for which the probability that this cell contains at least one mineral deposit is a maximum? For answers of these questions, the reader is referred to AGTERBERG & FABBRI (1978).

A second example (from AGTERBERG, 1974-b) is illustrated in figure 2. Cells maeasuring 10 km on a side are shown in the bottom diagram for an area underlain by Archean rocks within the Abitibi Volcanic Belt on the Canadian Shield. The percentage of rock types shown on geological maps with a scale of about 1:250,000 was determined for each of the cells (by point-counting). The objective of this project was to determine for each cell the conditional probability that it contains one or more deposits of a given type, this probability being conditional upon the rock type percentage values. Conditional probabilities of this type can be estimated by a variety of multivariate statistical methods. The model used for figure 2 is named the logistic model. It is noted that from the measured percentage values new variables had been defined which accounted for the possible coexistence of pairs of different rock types in the same cell.

Each probability estimated for a single cell can also be interpreted as the expected number of 'events' for that cell. An 'event' then consists of the occurrence of one or more mineral deposits in that cell. The advantage of this interpretation is that expected values can be added for groups of adjacent cells and this gives expected numbers of events in larger unit areas which can be contoured. Figure 2A shows contoured expected values for volcanogenic massive sulphide deposits with the unit area measuring 40 km on a side. The contours were scaled with respect to a relatively well explored region for which it was assumed that all 10 km cells with and without deposits had been correctly identified. The values in figure 2A are relatively high when there are many contacts between acidic volcanics and other rocks in the surrounding 40 km unit area and when mafic volcanics and acidic intrusives are also present in this vicinity.

Figure 2B shows a similar pattern for expected number of events per 40 km unit area for nickel-copper sulphide deposits. Now the relatively high values are primarily correlated to the presence of ultramafic bodies and mafic intrusive rocks.

The example of figure 2 indicates that the multivariate statistical approach may agree with concepts developed by economic geologists regarding the genesis of ore deposits. A disadvantage of the statistical approach is that the information which can be systematically quantified for a region at a given time is much less than the information which can be considered and used by geologists with first-hand knowledge of the region. An advantage of the statistical approach, however, is that it may be relatively easy to integrate different types of geoscience information including geophysical and geochemical measurements.

Contours such as those of figure 2 can have prognostic value. AGTERBERG & DAVID (1979) have discussed the location of seven newly discovered large copper deposits with respect to a set of prognostic contours similar to those shown in figure 2, but originally constructed in 1971 for large copper deposits in the vicinities of Timmins, Ontario, and Noranda-Val d'Or, Quebec.

ARTIFICIAL INTELLIGENCE

According to WINSTON (1977) the central goal of artificial intelligence is to make computers more useful and to understand the principles which make human intelligence possible. SLAGLE (1971) distinguished between three different approaches which are being taken to A.I.: (a) artificial networks, (b) artificial evolution, and (c) heuristic programming. An artificial network (a) consists of a large number of single elements and their interconnections. This approach tries to simulate the human brain which consists of approximately 10^{10} inter-connected neurons. Artificial evolution (b) attempts to simulate the process of evolution by mutation and natural selection. Both (a) and (b) are hampered by the fact that little is known about the natural processes that are being simulated.

The heuristic approach (c) has been more successful. Heuristics are strategies or tricks used to improve the efficiency of a system which tries to discover the solutions to complex problems (SLAGLE, 1971). In computer science there are many striking examples of inductive heuristics which drastically reduce the amount of computer time required for solving a specific problem.

According to BERLINER (1978) there are three types of computer programs in A.I. Type 1 relies on computational power. Billions of logical operations are performed within a

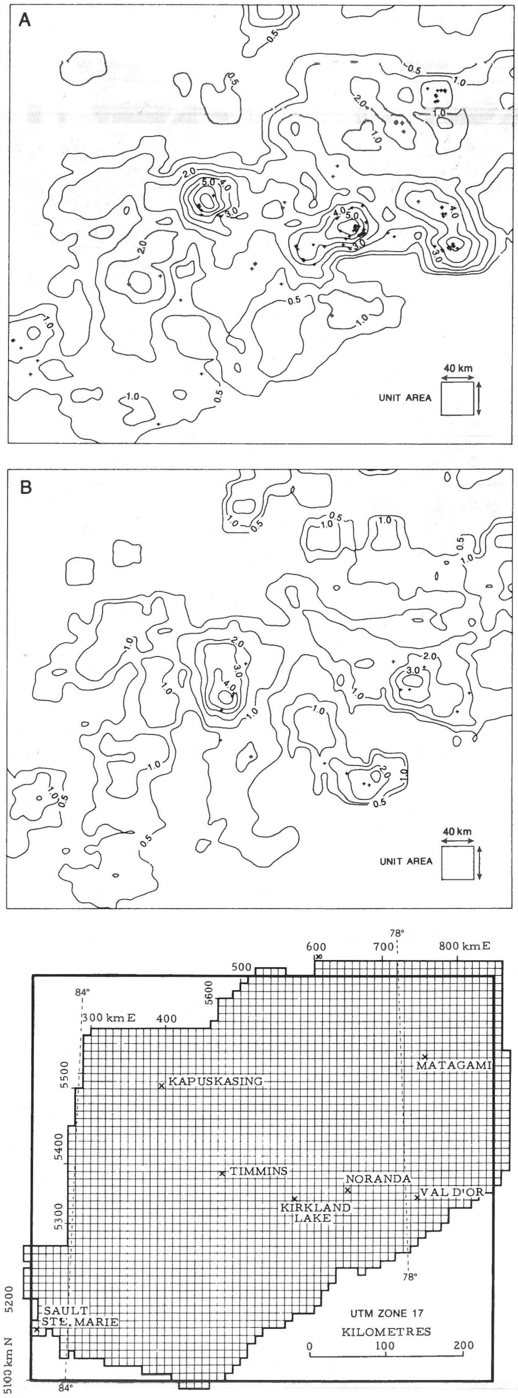

Fig. 2
Contour map for expected number per (40km × 40 km) unit area of (10 km × 10 km) cells that contain one or more deposits (from Agterberg, 1974-b). The contoured value defines a Poisson random variable for the probabilities of having 0, 1, 2, ... events in the surrounding unit area. The multivariate logistic response model was applied to 36 lithological variables defined for the 10 km cells shown on the location map. 4A: Volcanogenic massive sulphide deposits; 4B: Nickel-copper sulphide deposits. Known deposits are shown as crosses.

short time. In the search for an optimum solution, each possible move may open up several new possibilities. The result can be an exponential explosion of branching possibilities which severely limits the depth of the search. Type 2 programs embody the notion of directing the activity of the machine into more promising areas by suppressing those possibilities which seem less promising. Contrary to Type 3, both Type 1 and Type 2 programs are nonhuman. Type 3 attempts to simulate human strategies by formulating ideas and investigating the merits of these ideas before the next move is made. Type 3 programs use a method of analogies and catalogue the salient features of specific situations. If similar situations are encountered later, the result is assumed to be known.

BERLINER (1978) has pointed out that the nonhuman types of program are superior to Type 3 at this point in time. He illustrated this point by a review of the history of computer chess. Many computer scientists regard computer chess as the guinea pig of A.I. Until about 1967 all programs played very badly according to human tournament standards. At present, the best program which is of a type somewhere between Types 1 and 2 has about Expert strength on the human scale for chess players. It has beaten a Grandmaster in 'blitz' chess.

The following example from computer-based multivariate statistics further illustrates the performance of programs of Types 1 and 2. Suppose that a dependent variable is to be correlated with p independent variables. Then there are (2^p-1) subgroups of the p variables with which the dependent variable can be correlated. All these possibilities are indeed considered in the 'method of all possible regressions' which is of Type 1. However, when p=100, the number of possibilities exceeds 10^{30} and that many correlation coefficients could never be computed. On the other hand, a solution is always provided by the 'method of forward selection' which is of Type 2. Here the individual independent variables that show maximum partial correlation coefficients are successively included in the subgroup being sought. It is possible that the method of forward selection will not yield the same solution as the method of all possible regressions but the combined strength of the variables eventually selected will be approximately equal for both procedures.

A recent example of A.I. techniques applied in geoscience is 'Prospector', a computer-based consultant for mineral exploration developed by DUDA ET AL. (1977) at the Artificial Intelligence Center of the Stanford Research Institute in California. It consists of semantic networks encoding models developed by economic geologists, e.g. C.F. Park's model for Kuroko-type massive sulphide deposits. The purpose of Prospector is to enable the field geologist to compare via remote terminal his observed data with the information incorporated in the model. The geologist is asked for his degree of belief in the certainty of each of his observations on a scale ranging from -5 to 5. Prospector can compute the degree of certainty of a hypothesis on the same scale. For example, Prospector might conclude that it is 93% certain that there is a Kuroko type massive sulphide deposit present in the vicinity. This conclusion is reached automatically by means of a flexible system of rules for subjective probabilities. Prospector is a knowledge-intensive program. The use of many facts is considered a necessary part of the advance of A.I. into more complex domains.

SUBJECTIVE AND OBJECTIVE PROBABILITIES

The history of the concept of probability has been outlined by HACKING (1975). It came into being during the 16th century as an essentially dual concept, on the one hand having to do with degrees of belief and, on the other, with devices tending to produce stable long-run frequencies. The forerunner of probability was the concept of sign in the 'low' sciences such as medicine which had to deal in opinion. For example, Agricola in 1556 taught miners how to read the signs on the earth's surface that say what minerals are down below. In 1640 Thomas Hobbes wrote: 'If the signs hit twenty times for one missing, a man may lay a wager of twenty to one of the event, but may not conclude it for a truth'.

Probabilities dealing in degrees of belief are named 'subjective'. They cannot be tested by experiment. On the other hand, an objective probability does not refer to a judgment but to the possible outcome of a real or conceptual experiment involving stable long-run frequencies. Statistical techniques are based on objective probabilities. The majority of geologists is concerned primarily with subjective probabilities. Mathematical geologists have to formulate their problems in such a manner that there is a statistical 'population' with a sampling space from which a 'sample' is drawn. Then the probabilities of the events of the sampling space are objective in that they satisfy a set of formal rules, for example those formulated by A.N. Kolmogorov in his axiomatic theory of 1933.

A combination of subjective probabilities and objective use of observed data is found in Bayesian methods of statistics. These techniques use subjective probabilities measuring degrees of belief about the values of unknown parameters in a statistical model. This determines what is named the prior distribution for each parameter. Consequently, in Bayesian methods an unknown parameter is a random variable with a known prior distribution which summarizes the subjective degree of belief about the unknown value of this parameter. For example, the prior distribution is given a large variance if the uncertainty is believed to be great. The main difference between Bayesian methods and classical non-Bayesian statistical methods is that in the latter a statistical parameter such as the population mean is a constant without a frequency distribution.

After the subjective specification of the prior distribution of a parameter in Bayesian statistics, a set of observed values is used to compute what is named the posterior distribution of the parameter. The posterior distribution is used to construct an estimator of the unknown parameter. The final estimates of the parameters will to some extent reflect the initial subjective probabilities. Although mathematical statisticians agree that the specification of a statistical model is essentially inductive, they disagree about the use of subjective probabilities in their models. For example, in their book on probability methods in oil exploration, HARBAUGH ET AL. (1977) have warned that subjective estimates 'are fraught with uncertainty and may lend a false sense of security because estimates are represented by numbers which in turn are subjected to involved mathematical manipulations'.

CONCLUDING REMARKS

Statistics applied to facts in geoscience is subject to many difficulties as outlined in the introduction. The great fluctuations in the weights to be assigned to separate observations at the surface of the earth present significant problems which cannot be solved without the incorporation of concepts involving geological time into the statistical models. Most geologists tend to be sceptical about any statistical manipulations of field data whereas statistically-inclined geologists are not satisfied with the lack of formal proof associated with most geoscience concepts.

Statistical models in geoscience should be essentially spatial in that specific features are to be correlated with features both in their vicinity and occurring at greater distances. For example, on a local scale the mineral deposits of figure 1F are most strongly correlated with sedimentary rocks, whereas, on a 1:125,000 geological map, most of them fall on acidic volcanic rocks. On an even more regional basis, their location is correlated with the contacts between acidic volcanics and other rocks. This is illustrated by the fact that the probability that a cell measuring 10 km on a side contains one or more mineral deposits is almost zero if it is fully underlain by acidic volcanics.

The spatial characteristics of geoscience facts and the importance of the nonobservable and frequently unknown processes which led to their origin dictate that interpretations of statistics in geoscience differ significantly from those of other types of statistics. As pointed out by Van Bemmelen, the new problems that arise offer a wide ranging and challenging field of research.

REFERENCES

Agterberg, F. P. 1974-a Geomathematics (Mathematical background and geoscience applications) – Dev. Geomathematics 1, Elsevier (Amsterdam): 596 pp.

—— 1974-b Automatic contouring of geological maps to detect target areas for mineral exploration – J. Math. Geol. 6: 373-394.

—— 1978 Analysis of spatial patterns in the earth sciences. In: D. F. Merriam (ed.): Geomathematics, past, present, and prospects – Syracuse University Geology Contribution 5:7-18.

Agterberg, F. P. & M. David 1979 Statistical exploration. In: A. Weiss & T. J. O. O'Neil (eds.): Computer methods for the '80s – Amer. Inst. Mining Eng. (AIME) (New York) (to appear in October 1979).

Agterberg, F. P. & A. G. Fabbri 1978 Spatial correlation of stratigraphic units quantified from geological maps – Computers & Geosciences 4: 285-294.

Agterberg, F. P. & S. C. Robinson 1972 Mathematical problems of geology – Bull. Intern. Statist. Inst. 44: 567-596.

Berliner, H. J. 1978 Computer chess – Nature 274: 745-748.

De Finetti, B. 1972 Probability, induction and statistics – Wiley (New York): 266 pp.

—— 1974 Theory of probability (a critical introductory treatment), Vol. 1–Wiley (New York): 300 pp.

Dmitriev, A. N. (ed.) 1976 Logico-mathematical analysis of geological information (Theory and mathematical apparatus) – Inst. Geol. Geoph. Novosibirsk: 152 pp. (in Russian).

Duda, R. O., P. E. Hart, N. J. Nilsson, R. Reboh, J. Slocum & G. L. Sutherland 1977 Development of a computer-based consultant for mineral exploration – Stanford Res. Inst. (Menlo Park, Calif.): 202 pp.

Grayson, Jr. C. J. 1960 Decisions under uncertainty (drilling decisions by oil and gas operators) – Harvard Business School (Boston, Mass.) 14: 402 pp.

Hacking, I. 1975 The emergence of probability (a philosophical study of early ideas about probability, induction and statistical inference) – Cambridge Univ. Press: 209 pp.

Harbaugh, J. W., J. H. Doveton & J. C. Davis 1977 Probability methods in oil exploration – (New York) 9: 269 pp.

Marlow, A. R. (ed.) 1978 Mathematical foundations of quantum theory – Academic Press, (New York): 372 pp.

Marschak, J. 1964 Actual versus consistent decision behaviour – Behavioural Sci. 9: 103-110.

Matheron, G. 1965 Les variables régionalisées et leur estimation – Massen (Paris): 306 pp.

Miller, B. M. 1977 Probabilistic and computer methodologies used by the U.S. Geological Survey for geological estimates of undiscovered oil and gas resources in the United States. In: R. V. Ramani (ed.): Application of computer methods in the mineral industry – Amer. Inst. Mining Metall. Petrol. Eng. (New York): 419-430.

Nieuwenkamp, W. 1968 Natuurfilosofie en de geologie van Leopold von Buch – K. Ned. Akad. Wet., Proc., ser. B 71 (4): 267-278.

Raphael, B. 1976 The thinking computer (Mind inside matter) – Freeman (San Francisco, Calif.) 14: 322 pp.

Skinner, R. 1974 Geology of Tetagouche Lakes, Bathurst, and Nepisiguit Falls Map-areas, New Brunswick – Geol. Surv. Canada Mem. 371: 133 pp.

Slagle, J. R. 1971 Artificial intelligence (the heuristic programming approach) – Mc Graw-Hill (New York) 10: 196 pp.

Van Bemmelen, R. W. 1961 The scientific character of geology – J. Geol. 69: 453 463.

—— 1972 Geodynamic models (an evaluation and synthesis) – Dev. Geotectonics 2, Elsevier (Amsterdam) 11: 267 pp.

Voronin, Yu. A. 1976 Improvement of the methodological, theoretical and organizational basis of mineral prospecting in connection with the application of mathematics and computers – Computing Centre, Siberian Branch, Acad. Sci. USSR (Novosibirsk): 129 pp. (in Russian).

Watson, G. S. 1975 Texture analysis – Geol. Soc. Amer. Mem. 142: 367-391.

Winston, P. H. 1977 Artificial intelligence – Addison-Wesley (Reading, Mass.) 16: 444 pp.