AD-A089 154 PRINCETON UNIV NJ DEPT OF STATISTICS F/G 8/7

STATISTICS IN GEOLOGY.(U)

JUL 80 G S WATSON N00014-79-C-0322

UNCLASSIFIED TR-168-SER-2 NL

1 OF 1

AD A089154

END

DATE FILMED

10-80

DTIC

STATISTICS IN GEOLOGY

by

G. S. Watson
Princeton University

Technical Report No. 168, Series 2
Department of Statistics
Princeton University
June 1980

Research supported in part by a contract with the Office of Naval Research, No. N00014-79-C-0322, awarded to the Department of Statistics, Princeton University, Princeton, N. J.

ABSTRACT

This article, prepared for the *Encyclopaedia of Statistical Sciences*, gives many references to geological papers according to the statistical methods used. It should be useful to anyone preparing a course for geologists. It may also aid statisticians looking for applications of techniques.

Key words: Geology, Erosion, Sedimentation, Earthquakes, Volcanoes, Size Distributions.

1. INTRODUCTION

Geology seeks to describe and understand the processes which have acted in the past, and are acting now, to form the continents and oceans with their mountains and valley and which have led to the varied sequences of rocks of differing compositions and structures. At the time of Darwin's 1832-1836 voyage on the *Beagle*, geology was closely linked with biology as the study of "natural history," and both then made great leaps forward. In fact, Lyell's book (1830-1832) on stratigraphy (epochs were defined statistically by fossil contents) was Darwin's inspiration. It is a curious historical fact that, while the intense application of the physical sciences led the subjects to diverge, both had their next revolution at about the same time -- molecular biology in the 1950's and plate tectonics in the 1960's.

Mathematics entered geology when the physics of the earth was studied -- gravity and the figure of the earth, tides in the oceans, air and solid earth, the cooling of the earth, earthquakes and the propagation of waves around the earth. Most study of the earth must be a matter of inference, because it is necessarily indirect -- only Jules Verne could imagine a "Voyage to the Center of the Earth." Also, geological field measurements are subject to greater errors than laboratory work in chemistry and physics, and it is often not possible to take "random samples." It is something of a coincidence that a little after Sir Ronald Fisher's (q.v.) development of statistics largely for biologists, an eminent geophysicist, Sir Harold Jeffreys, should develop his own theory of statistics (1931-1939). Jeffreys' logical predisposition led him to a mathematical rule for deriving priors rather than to use a purely subjective origin. For his geological achievements and classical mathematical geophysics, readers should also consult his famous text, *The Earth* (1924-1961).

The earth sciences may also claim to have initiated several areas of statistical theory and practice. There are so many periodic or pseudo-periodic earth phenomena that Sir George Stokes' (1879) introduction of the Fourier transform of data and its development by Schuster -- see, e.g., Brillinger (1975) -- was natural. The most advanced applications of Time Series Analysis (q.v.) are still to be found in geophysics -- a comprehensive bibliography has been given by Tukey (1965). The orientation of pebbles (Krumbein, 1939) and the direction of magnetism of rocks (Fisher, 1953) led to the development of methods for Directional Data Analysis (q.v.); a survey with many references to papers in this area was given by Watson (1970). More recently, economic geology and efficient mining have led to (Geostatistics (q.v.)) an extensive application of random function theory by Matheron (1965). Chemical petrologists study the proportions of substances making up rocks, so their data add to unity. The study of the correlation of proportions raises special problems that have occupied geologists more than other scientists -- see, e.g., Chayes (1971). The study of their sections (e.g., Chayes, 1956) has led to stereological and geometrical probability problems -- as has exploration geology. Geologists have always needed maps and photographs of sections. Now the computer is being used heavily to produce and process such information; Matheron (1967) provides a theoretical background -- see also Mathematical Morphology (q.v.).

There has been a very rapid growth in the use of computers, mathematics and statistics in geology in the 1970's. This literature is fairly easy to enter. The American Geological Institute publishes a Bibliography and Index in which most of the relevant articles appear under the main heading of "Automatic Data Processing," though some appear under "Mathematical Geology." Two journals, *Mathematical Geology* and *Computers and Geosciences*, specializing in these topics, began in the 70's. There are a number of general texts (e.g., Agterberg, 1974; Davis, 1973) and a number devoted to specific topics to be mentioned below. D. F. Merriam has edited many symposia. As these quantitative methods become a recognized part of all subdivisions of geology, the specialized journals (*Sedimentology*, to give just one example) all carry articles of statistical interest.

The following sections are chosen to show the methods and problems of special interest which are to be found currently in geology. The references given will lead the reader further. Exploration and resource estimation and exploitation is ignored here but partly covered in geostatistics (q.v.).

2. DATA BANKS

Efforts are being made to computerize data so it can be accessed easily -- see, e.g., Chayes (1979) for igneous petrology. Much interesting data is unavailable because of its economic value to those who possess it.

3. STOCHASTIC MODELS

This field is very wide indeed. Earth movements lead to an interest in the growth of cracks or fractures -- see, e.g., Vere-Jones (1977). The occurrence and strength of earthquakes and volcanic eruptions have been the subject of much point processes modelling -- see, e.g., Adamopoulos (1976). Erosion and sedimentation require a knowledge of particle size distributions (q.v.). Models for forming sands and powders often lead to the lognormal (e.g., Kolmogoroff, 1941) and Weibull (see, e.g., Kittleman, 1964) distributions. Considerations of the transport and deposition of sand (see, e.g., Bagnold, 1954; Sen Gupta, 1975) lead to other distributions and stochastic processes. Kolmogoroff (1949) first modelled the deposition and subsequent erosion of sediments. His model was studied further by Hattori (1973). This theory is distinct from the literature that tries to fit Markoff chains (q.v.; see below) to the succession of beds according to their composition rather than thicknesses, though

Hattori deals with both approaches. The present writer regards the "Markoff approach" as more data analysis than modelling. Grenander (1975) has provided a stationary stochastic model (q.v.) on the circle (which is easily generalized to the sphere) for the height of the land surface. Erosion is modelled by diffusion which always smooths, and inequalities are maintained by uplifts at random time points described by random independent functions (random function, q.v.).

The study of streams in drainage basins, their lengths and topology is fascinating -- see, e.g., Dacey and Krumbein (1976).

The distribution of elements has a large literature, but statistical models to explain them are few. Kawabe (1977) gives a model and a literature list, including references to papers by Ahrens (1963), who felt lognormal (q.v.) distributions of elements were a law of nature.

4. DATA ANALYSIS

Nowadays, all the common statistical procedures are used widely. Most data is observational. One collects rocks where they happen to be exposed and accessible, so the problems of "non-random samples" are very serious. The earth is a sample of one. The list below gives leads to areas of particular interest.

Clustering Methods (q.v.)

Dendrograms (q.v.) and other methods are often used to relate fossils, rocks, etc., to help explain their evolution. Petrofabric and other studies yield orientations plotted on a sphere. Deciding whether the points fall in groups or clusters is a common problem but may be attacked differently.

Factor Analysis

Factor analysis (q.v.) is widely used in palaeontology and elsewhere -- see, e.g., Jöreskog, Klovan and Reyment (1976). Temple (1978) gives a very critical review. In the analysis of data on the sedimentary composition of a closed basin the factors might be the few inputs to the basin. In the fossil content of oceanic cores, they might be the depositorial climates -- tropical, polar, etc. This latter problem has recently been studied differently by Sachs, Siegel and Goldburg (unpublished Princeton report).

Markov Chains

In studying the succession of different lithologies, often a small set (e.g., sand, silt and clay) recurs in a partially cyclic way. It might be that the failure to be strictly cyclic is due to the complete erosion of some parts of the record. See Casshyap (1975), Miall (1973).

Bounded Sum (or Closed) Data

Bounded sum (or closed) data, e.g., the proportions $p_1, p_2, \ldots, p_k$ of the k constituents of a rock. It is natural to study such data to see if the relative amounts of substances 1 and 2 are associated. The facts that $p_1 + p_2 < 1$, $\sum_{1}^{k} p_i = 1$ make the usual methods invalid. See Chayes (1971) and Darroch and Ratcliff (1978) for later work.

Orientation Data

Normals to bedding planes, the directions of cracks, and joints provide examples of axial data and the flow of glaciers, directions of magnetization, examples of Direction Data Analysis (q.v.). See Watson (1970), McElhinny (1973).

Time Series Analysis

Times Series Analysis (q.v.) is basic to seismological data processing. See Brillinger (1975), Tukey (1965).

REFERENCES

ADAMOPOULOS, L. (1976). Cluster models for earthquakes,regional comparisons. *Int. Assoc. Math. Geol.*, *8*, 463-475.

AGTERBERG, F.P. (1974). *Geomathematics*. Elsevier Publishing Co., Amsterdam, The Netherlands.

(A course on calculus, matrix algebra, probability statistics, sound but classical followed by two chapters on spatial statistics, time series, Markov chains, stochastic petrology, directional data, multivariate analysis of spatial data.)

AHRENS, L.H. (1963). Lognormal-type distributions in Igneous Rocks IV, *Geochim. et Cosmochim., Acta 27*, 333-343.

(Gives references to three earlier papers and many other leads to this literature.)

BAGNOLD, R.A. (1954). *The Physics of Blown Sands and Desert Dunes*. Chapman and Hall, London, England.

BRILLINGER, D.R. (1975). *Time Series, Data Analysis and Theory*. Holt, Rinehart and Winston, Inc., New York, New York.

(Best modern account of spectrum analysis, combining numerical insight and examples with the requisite formulae and theorems.)

CASSHYAP, S.M. (1975). Cyclic characteristics of coal-bearing sediments in the Bochumer Formation (Westphal A2) Ruhrgebiet, Germany, *Sedimentology*, 237-255.

CHAYES, F. (1956). *Petrofabric Modal Analysis*. John Wiley and Sons, New York.

(Now out of date, but shows the origins of the topic.)

__________ (1971). *Ratio Correlation*. University of Chicago Press, Chicago, Ill.

(For recent advances, see Darrock and Ratcliff.)

CUBITT, J.M., HENLEY, S. (1978). Statistical Analysis in Geology. *Benchmark Papers in Geology/37*, Dowden, Hutchinson & Ross, Inc., Stroudsberg, Pennsylvania.

(A collection of 23 papers on statistical ideas important in geology rather than techniques.)

DACEY, M.F., KRUMBEIN, W.C. (1976). Topological properties of disjoint channel networks with enclosed regions. *Int. Assoc. Math. Geol.*, *8*, 429-461.

DARROCK, J.H., RATCLIFF, D. (1978). No-association of proportions. *Int. Assoc. Math. Geol.*, *10*, 361-368.

DAVIS, J.C. (1973). *Statistics and Data Analysis in Geology*. John Wiley and Sons, New York, New York.

GRIFFITHS, J.C. (1967). *Scientific Method in the Analysis of Sediments.* McGraw-Hill, New York, New York.

(Rather old-fashioned statistically -- largely sedimentology.)

JEFFREYS, SIR HAROLD (1961). *The Earth, 4th Edition.* Cambridge University Press, London, England.

(Classic text on geophysics.)

__________________ (1939). *Theory of Probability.* Clarendon Press, Oxford, England.

(Author's version of Bayesian statistics with many novel geophysical applications.)

JÖRESKOG, K.G., KLOVAN, J.E., REYMENT, R.A. (1976). *Geological Factor Analysis.* Elsevier Scientific Publishing Co., New York, New York.

KAWAKE, I. (1977). A stochastic model of fractional crystallization and its application to the Japanese Geosynclinical Basalt, *Int. Assoc. Math. Geol.*, 9, 39-54.

KITTLEMAN, L.R. (1964). Application of Rosin's distribution in size frequency analysis of clastic rocks, *J. Sed. Pet.*, *34*, 483-502.

KOCH, G.S., LINK, R.F. (1971). *Statistical Analysis of Geological Data, Vols. 1 and* John Wiley and Sons, New York, New York.

(More of interest to economic geologists than statisticians.)

LYELL, C. (1800, 1832, 1833). *Principles of Geology, Vols. I, II, III.* (Republished in 1969 by the Johnson Reprint Corp., New York.)

(Darwin took Vol. I with him and received Vol. II in Montevideo in 1832.)

MARDIA, K.V. (1972). *Statistics of Directional Data.* Academic Press, London and New York.

(The only text on the topic -- for statisticians mainly.)

MATHERON, G. (1967). *Elements pour un theorie des milieux poreux.* Masson et Cie., Paris, France.

(In the first part, random sets are used to model sediments and tie descriptive theory to observations made by an Image Analyser. In part two, the flow through porous media is studied -- outdated.)

____________ (1965). *Les variables regionalisees et une estimation.* Masson et Cie., Paris, France.

(Difficult but now outmoded by Huisbreghts and Journel's book.)

McELHINNY, M.W. (1973). *Palaeomagnetism and Plate Tectonics.* Cambridge University Press, Cambridge, England.

(Most up-to-date with palaeomagnetism and its implications with many illustrations of statistical procedures.)

MIALL, A.D. (1973). Markov chain analysis applied to an ancient alluvial plain succession, *Sedimentology, 20*, 347-364.

RAMSEY, J.G. (1967). *Folding and Fracturing of Rocks.* McGraw-Hill, New York, New York.

REYMENT, R.A. (1971). *Introduction to Quantitative Paleoecology.* Elsevier Publishing Co., Amsterdam, The Netherlands.

SENGUPTA, S. (1975). Size-sorting driving suspension transportation-lognormality and other characteristis, *Sedimentology, 22,* 257-273.

TUKEY, J.W. (1965). Use of numerical spectrum analysis in geophysics. *Proc. 35th Session, Int. Stat. Inst. (Belgrade), Book 1*, 267-305.

(Much scientific detail, large bibliography.)

VERE-JONES, D. (1977). Statistical theories of crack propagation, *Int. Assoc. Math. Geol.*, *9*, 455-482.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(14) TR-168-SER-2

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: TECHNICAL REPORT NO. 168, SERIES 2

2. GOVT ACCESSION NO.: AD-A089 154

3. RECIPIENT'S CATALOG NUMBER:

4. TITLE (and Subtitle): Statistics in Geology

5. TYPE OF REPORT & PERIOD COVERED: Technical

6. PERFORMING ORG. REPORT NUMBER:

7. AUTHOR(s): Geoffrey S. Watson

8. CONTRACT OR GRANT NUMBER(s): N00014-79-C-0322

9. PERFORMING ORGANIZATION NAME AND ADDRESS:
Department of Statistics
Princeton University
Princeton, NJ 08544

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:

11. CONTROLLING OFFICE NAME AND ADDRESS:
Office of Naval Research (Code 436)
Arlington, VA 22217

12. REPORT DATE: July 1980

13. NUMBER OF PAGES: 8

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office):

15. SECURITY CLASS. (of this report):

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report):
Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report):

18. SUPPLEMENTARY NOTES:

19. KEY WORDS (Continue on reverse side if necessary and identify by block number):
Geology, Erosion, Sedimentation, Earthquakes, Volcanoes, Size Distributions

20. ABSTRACT (Continue on reverse side if necessary and identify by block number):

This article, prepared for the Encyclopaedia of Statistical Sciences, gives many references to geological papers according to the statistical methods used. It should be useful to anyone preparing a course for geologists. It may also aid statisticians looking for applications of techniques.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

406873