GEOLOGIE EN MIJNBOUW VOLUME 55 (1-2), p. 105-109 1976

# SIMPLE USES OF COVARIOGRAMS IN GEOLOGY

S. DIJKSTRA[1)]


## ABSTRACT

The paper presents an elementary introduction into the theory of random functions and demonstrates the usefulness of covariograms in the analysis of geological data. Most examples refer to experiences in exploration geochemistry.


## INTRODUCTION

This paper provides a simple introduction to the of random functions is clearly recognized with regard to both mine of random functions is clearly recognized with regard to both mine evaluation and exploration geochemistry. In other applied earth sciences, however, the functions have, until now, received only limited attention.

Both the above mentioned disciplines employ random functions to describe the behaviour of continuous variables in terms of space coordinates. In mine evaluation the functions serve to describe the variation of metal values within the space of a mineralized body. Likewise, in exploration geochemistry the functions have been shown to be capable of characterizing the spatial properties of trace element distributions in various surface materials.

The author wishes to suggest that random functions may also prove useful to the solution of similar problems met in other geological disciplines; this includes the analysis of geological time series.

Let us begin our examination of the use of random functions by considering the data sequence given in Fig. 1. We shall assume this sequence to represent a series of $N$ measurements of values $v_i$ at equally spaced points $x_i$, along a profile $X$.

A visual inspection of the sequence reveals that the measurements fluctuate around a mean $\bar{v}$; it may also be observed that neighbouring measurements display some kind of relationship.

While it might be possible to describe the observed relationship with the aid of high order polynomial functions, fitting high order polynomials to geological data is rarely recommended.

1) Mining Exploration Dept., ITC, 3 Kanaalweg, Delft, The Netherlands

One of the reasons is that too great an accuracy may be accorded to function values lying between the measuring points. The sequence presented in Fig. 1 is an example of a sequence which may, in fact, be better described through the aid of the statistical concepts associated with the theory of random functions. These concepts are reasonably easy to handle and in many instances provide insight into the natural properties of geological data distributions.

## A PRACTICAL APPROACH TO THE CONCEPTS OF RANDOM FUNCTIONS

From a mathematical point of view an introduction into the theory of random functions requires a great deal of fundamental study. For practical mining engineers and geologists the most important aspect of the theory is that it provides a tool permitting characterization of data patterns by means of estimated statistical parameters:

$$\text{Mean} \quad \bar{v} = (1/N) \sum_{i=1}^{N} v_i \tag{1a}$$

$$\text{Variance} \quad S^2 = [1/(N-1)] \sum_{i=1}^{N} (v_i - \bar{v})^2 \tag{1b}$$

$$\text{Covariance} \quad C(v_i, v_{i+jk}) = [1/(P-1)] \sum_{i=1}^{N} (v_i - \bar{v})(v_{i+jk} - \bar{v}) \tag{1c}$$

Where
- N = number of measurement (samples)
- k = sampling interval
- j = 0, 1, 2 ..... N
- P = number of pairs of products in the summation.

The first two parameters need no explanation.

The final parameter refers to the spatial relationship

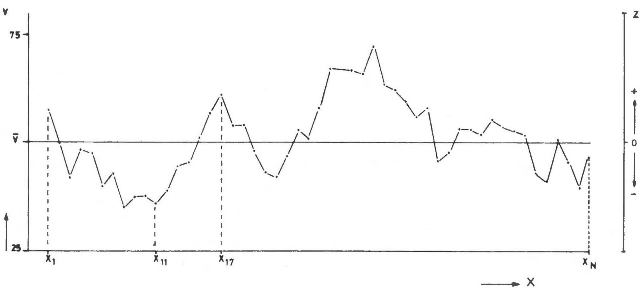


Fig. 1.
Sequence representing a hypothetical series of values $v_i$ along a profile $X$.

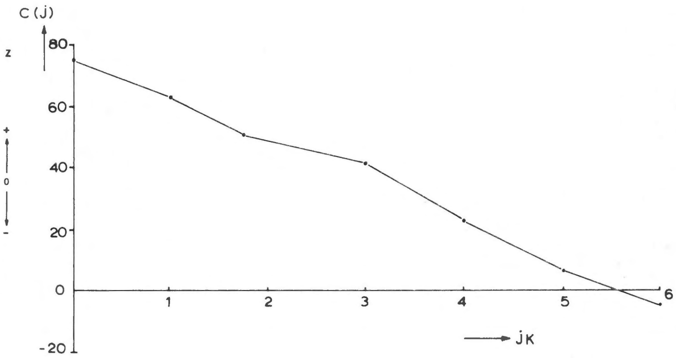


Fig. 2
Covariogram computed from data of Fig. 1.

between neighbouring data, thus adding an extra dimension to conventional statistics. The reader will be aware from his own experience that the relationship between geological measurements $v_i$ and $v_{i+d}$ generally decreases as the distance $d$ between measuring points increases. This relationship is formally described by covariance functions *Cov (d);* the estimation rule (1c) serves to estimate a discrete number of function values from actual data. Examination of this rule provides insight into the significance of covariance. To increase this insight let us replace all values $v_i$ in eq. (1c) by values $z_i$, using the transformation $z_i = v_i - \bar{v}$. The physical meaning of the tranformation is easily seen from Fig. 1. The mean of values $z_i$ being zero, eq. (1c) may be rewritten as follows:

$$C(z_i, z_{i+jk}) = [1/(P-1)] \sum_{i=1}^{N} z_i z_{i+jk} \tag{2}$$

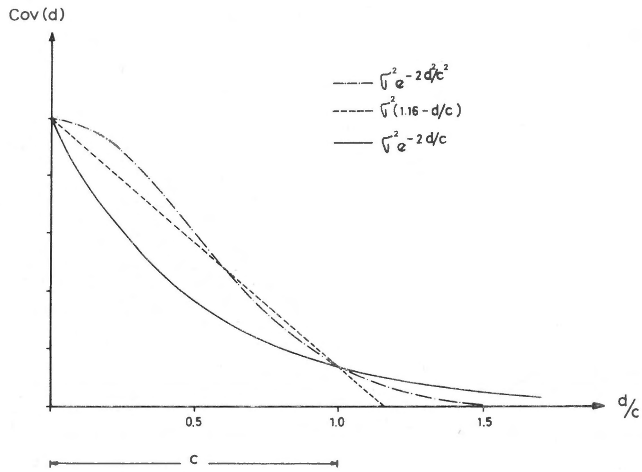


Fig. 3
Examples of covariance models: gaussian model (upper curve), linear model (in midst) and exponential model (lower curve); $\tau^2$ represents variance, $c$ is the effective correlation distance.

For $j = 0$ the estimation rule is made up exclusively of squares and, accordingly, all terms are positive. In this case the rule is the equivalent of the rule for the estimation of variances. Where $j = 1$, the rule begins to include some negative terms, thus decreasing the value of the covariance. The number of negative terms corresponds to the number of crossings of measurement interconnections and the reference line $\bar{v}$ in Fig. 1.

For $j = 2, 3...$ the number of negative terms increases in a stepwise fashion and the value of the covariance decreases accordingly until it reaches the limiting value zero for $j = 5$. At this distance, expressed in units of the measuring interval $k$, the values are no longer autocorrelated and the distance $5\,k$ is an estimate of the autocorrelation distance $c$.

Plots of estimated covariances against multiples $j$ of lags $k$ are referred to as covariograms. Fig. 2 shows the covariogram computed for the series of values $v_i$ found in Fig. 1.

The concepts "covariance" and "autocorrelation" are closely related. It holds for any value of $j$ that:

$$r_{jk} = C(v_i, v_{i+jk})/s^2 \tag{3}$$

where $r_{jk}$ is the estimated autocorrelation coefficient for values of the geological parameter at distance $jk$ apart and $s^2$ is the estimated variance.

The covariogram is an estimate of the covariance function. Various mathematical models of covariance functions are available, including the gaussian model, the linear model and the exponential model. The cited models are presented in Fig. 3. The experimental covariogram in Fig. 2 approximates a linear model; $5\,k$ and $s^2$ are estimates of the parameters $c$ and $\tau^2$ respectively.

It has been tacitly assumed in the above discussion that value distributions of the geological variable under consideration are homogeneous. This implies that mean value, variance and covariance properties remain constant all along the profile; the relationship between any pair of values $v_i$ and $v_{i+d}$ must be independant of the absolute location of the measured points and only depend on their relative position. If the condition of homogeneity is not reasonably satisfied, profiles may be split into separate sections. In the presence of low order deterministic trends, data sequences can be described by the summation of a deterministic function plus a random function. However, it is certainly possible for a deterministic trend in a small interval to be revealed to be part of a local undulation when the interval is enlarged. Referring to Fig. 1, closely spaced measurements within the limits of the interval $x_{11}$ to $x_{17}$ would reveal the presence of a first order polynomial trend. Enlargement of the interval, however, would show this trend to be a local undulation. Such reasoning may sound rather speculative, nevertheless it is not without merit. Based on plausible assumptions, the accuracies of apparent deterministic trends can be delineated with the aid of the theory of random functions. It may therefore be suggested that random functions have a wider scope than deterministic functions.

In the above introduction we have restricted ourselves to data sequences along profiles. However, the use of random functions can easily be expanded to two dimensional space. The covariance function then serves to measure the natural relationship of values $v_i(x_i, y_i)$ and $v_j(x_j, y_j)$ where $x$ and $y$ are plane coordinates. As it is not necessary for covariance properties to be isotropic, covariograms must be computed for several directions.

## APPLICATIONS

A few simple uses of cavariograms are outlined in the following sections. In all instances it is assumed that changes in measured values obey homogeneous random functions.

### *Comparison of data patterns*

The applicability of the concept of covariance to the comparison of value patterns in different areas is apparent. Care must be exercised that covariograms are computed on the basis of adequate numbers of equally spaced measurements; the number 100 may serve as a guide. Covariograms of parallel profiles may be combined to form a composite covariogram. The replacement of conventional statistics by covariance studies is not advocated; the study of frequency distributions and covariance properties are best done in combination.

### *Selection of sampling interval and map accuracies*

Systematic measurements employing a sampling grid demand the selection of an appropriate sampling interval. These intervals can be efficiently selected with the aid of covariograms computed from the data obtained from orientation surveys. The production of isograd maps is often the ultimate goal of measuring; these maps represent estimates of a geological surface and are commonly constructed by means of linear interpolation between measurements. The mean accuracy of the surface depends on both the character of the prevailing covariance function and on the sampling interval. K u b i k et al (1976) have shown that in most instances fairly good accuracies are obtainable when sampling intervals are less than one third the correlation distance. An example may be illustrative. For two dimensional value patterns with isotropic covariance properties approximating the gaussian model in Fig. 3, mean map accuracies for spacings *1/6c, 1/3c, 1/2c, 2/3c* and *1c* are *0.04 s, 0.16 s, 0.38 s, 0.62 s* and *0.84 s* respectively; $c$ is the correlation distance and $s$ is the standard deviation of the entire data distribution.

In the presence of point deviations the mean map accuracy will be inferior. The term point deviation may need some explanation: it refers to the effect of human measuring errors combined with the effect of rapid fluctuations in the value of the geological variable at the very local scale. The adverse influence of point deviations on map accuracies may be reduced if raw data are smoothed employing moving average techniques prior to contouring. The definition of point deviation and the application of moving average techniques will be further elaborated in subsequent sections of this paper.

### *Smoothing*

Since the object of smoothing is to eliminate the effect of point deviations, the relative magnitude of this effect should be known before moving average rules are designed. This is another situation in which covariograms are very helpful. The covariogram shown in Fig. 4 illustrates how the magnitude of the effect of point deviations is estimated by hand extrapolation (see dotted line). The effect appears to account for almost half the total variance.

Indiscriminate use of arbitrary moving average rules may lead to illusionary results. Optimum moving average rules are designed with the aid of a technique known as Kriging. The application of the technique requires advanced knowledge of geostatistics. Lacking such knowledge the following formula may be used for smoothing data sequences along profiles:

$$h_x = (1/n) \sum_{i=1}^{i=n} v_{x_i} \tag{4}$$

where $h_x$ is the smoothed value at coordinate $x$ *in is the number of sampling points within the interval* $x - ½c$ to $x +$

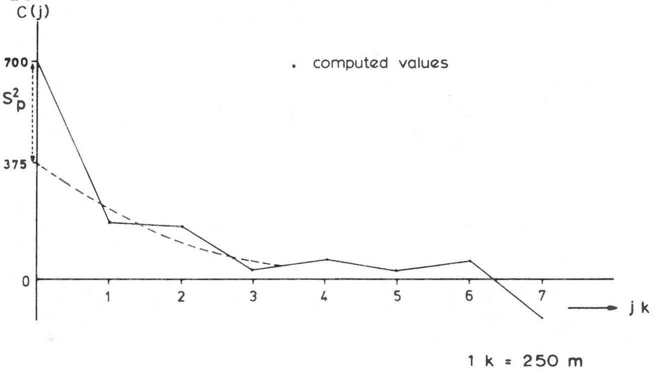


Fig. 4
Correlogram of copper values (in ppm) from a stream sediment survey in the Ingessana Hills area (Sudan). The correlogram is estimated from values of 100 samples, collected at 250 *m* intervals along four different profiles, all parallel to the general direction of the drainage system. $3k = 750\ m$ is a conservative estimate of the correlation distance.

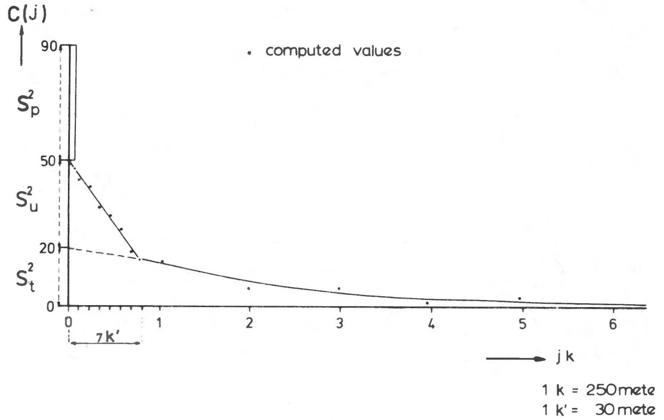


Fig. 5
Covariogram of lead values (in ppm) from stream sediment samples of an orientation survey in the Belgian Ardennes. The covariogram is estimated from over 100 pairs of values of three different sets of samples, collected at intervals of 250 *m*, 30 *m* and 3 *m*, respectively, along a single stream profile; $S_p^2$, $S_u^2$ and $S_t^2$ are estimates of variances, relating to point deviations, undulations and trend, respectively.

½ *c*, with *c* being the estimated correlation distance
$v_{x_i}$ raw values within the above interval.

With reference to the data of Fig. 4, the formula leads to the following moving average rule:

$$h_x = \tfrac{1}{5}\,(v_{x-2k} + v_{x-k} + v_x + v_{x+k} + v_{x+2k}) \qquad (5)$$

The above formula can be expanded for use with two dimensional patterns. Studies at the author's institute proved that it yields accuracies that differ up to a factor 2 from the optimum achievable accuracy. Such results are better than the results to be expected employing arbitrary rules.

### *Analysis of surface components*

Geological surfaces often represent a summation of a number of components, each differing in regional significance. These components are referred to by such terms as trend, undulations and point deviations. With regard to the separation of surface components, covariograms of raw data may provide some insight into data structure. Fig. 5 provides an illustration relating to a geochemical orientation survey. The covariogram refers to lead values in sediments of a stream profile in the Barvaux area (Belgium). It indicates the presence of point deviations *p*, local undulations *u* and an areal trend *t*.

The contribution of each one of the components to the total variance is estimated by extrapolation.

On the basis of this covariogram it may be suggested that sampling intervals of *400 m* being one third the correlation distance, will quantify an areal trend and that delineation of local undulations demands sampling intervals of less than *100 m*. Smoothed profiles can be computed using suitable moving average rules (see above). As far as the computation of the areal trend is concerned, the operation should aim at the elimination of the combined effect of $sp^2$ and $su^2$. The earlier definition of point deviation is thus extended. In view of the high ratio $(s_p^2 + s_u^2)/s_t^2$, the trend surface is bound to be of low accuracy. This accuracy may be improved, either by revising sampling techniques, increasing sample volume, and/or employing a closer sample interval than was initially suggested.

### *Accuracy of volume estimates*

The accuracy of volume estimates has been investigated by specialists in the field of ore reserve estimation. Some of the results may be of use to other disciplines as well. Two types of sampling arrangements have been examined in detail, namely arrangements with regular grid spacings and arrangements employing perimeter control. The first arrangement refers to surface drilling, the second to underground sampling in drives and raises.

In all instances where variations in metal values can be described by homogeneous random functions with given covariance properties, the accuracy of mean blockvalues $h_{mean}$ depends on sampling arrangement, sampling spacing and block size. Fig. 6 and Fig. 7 are examples of accuracy results where covariance functions are of the exponential type and measurements contain no point deviations (B o t m a n et al, 1975). Fig. 6 represents an isometric grid; sample spacing and block size are measured in units of the correlation distance *c*. The accuracy of the estimated mean value of the central block is *0,20 s*, where *s* is the standard deviation of the entire raw data distribution. Fig. 7 represents a sampling arrangement employing perimeter control. The error in the mean value of the central block is *0.09 s*. In the actual practice of underground sampling measurements are not without point deviations, however, the actual number of measurements far exceeds the number in the example

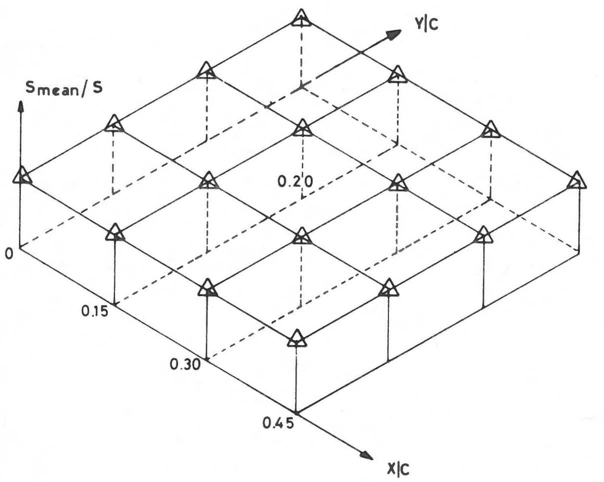


Fig. 6
Example of accuracy of mean block value for a specific block size with regular sampling control.

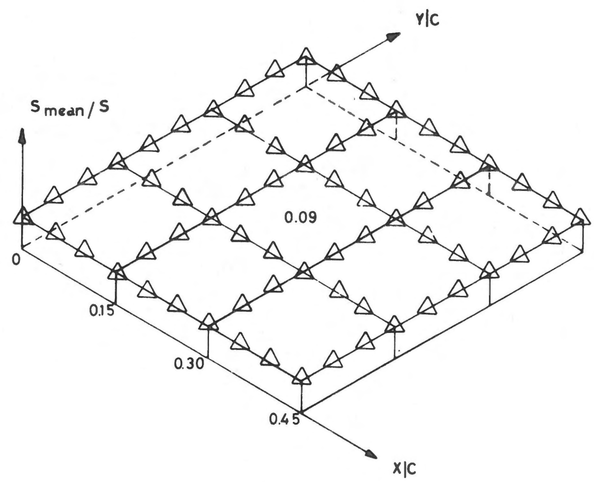


Fig. 7
Example of accuracy of mean block value for a specific block size with perimeter control.

shown in Fig. 7. Theory and practice can be brought together by means of (moving) averaging techniques. When such techniques are used, the effect of point deviations is largely reduced.

## FURTHER READING

For detailed discussion regarding the topics raised in this paper reference is made to Dijkstra and Kubik (1975) and to Botman et al (1975).

Blais et al (1968) provide a clear and informative introduction to Matheron's theory of regionalised variables. The latter theory employs so-called (semi-) variograms instead of covariograms. Examples of practical applications in ore evaluation are included. An application of a differing nature is described by Boyle and Hosgit (1974).

While the above cited authors approach the subject matter from a practical point of view, authors such as Matheron (1963) and Agterberg (1970) provide mathematical background.

Davis (1973) has compiled an elementary textbook on geostatistics. It includes an introduction to various techniques of geological data analysis, as well as many examples and extensive lists of references.

Kendall (1966) and Wiener (1949) have published accounts of the statistical properties and uses of random functions. Although they refer exclusively to time series, and no reference is made to problems in geology, many earth scientists will find certain sections of their books to be highly instructive.

Agterberg's textbook on geomathematics (1974) is the most comprehensive textbook of its kind available. It is directed to an advanced audience, however, and makes no easy reading. Matheron's formal derivation of the theory of regionalised variables (1965) is exclusively recommended to skilled mathematicians.

## REFERENCES


Agterberg, F.P. (1970) – Autocorrelation functions in geology. Proceedings of a Colloquium on Geostatistics, University of Kansas, Lawrence, June 1970. In: D.F.Ed. Merriam (Editor), Series Computer Applications in the Earth Sciences, Geostatistics. Plenum Press, N.Y., p. 113-141.

, (1974) – Geomathematics. Elsevier Scientific Publishing Company, Amsterdam, pp. 596.

Blais, R.A. and P.A. Carlier (1968) – Applications of geostatistics in ore evaluation. In: J.E. Gill at al (Editors), Ore Reserve Estimation and Grade Control, Special Vol. 9, 1968, The Canadian Institute of Mining and Metallurgy, p. 41-68.

Botman, A.G., S. Dijkstra and K. Kubik (1975) – Theoretical accuracy of volume determination for topographic and geological surfaces, ITC Journal 1975-3, p. 331-340.

Davis, J.C. (1973) – Statistics and Data Analysis in Geology. John Wiley and Sons, New York, N.Y. pp. 550.

Dijkstra, S. and K. Kubik (1975) – Autocorrelation studies in the analysis of stream sediment data. In: I.L. Elliott and W.K. Fletcher (Editors), Developments in Economic Geology 1, Geochemical Exploration 1974, Elsevier Scientific Publishing Company 1975, p. 141-161.

Kendall, M.G. (1966) – The Advanced Theory of Statistics, 3. Griffin, London, pp. 552.

Kubik, K. and A.G. Botman (1976) – On the interpolation accuracy for topographic and geological surfaces. ITC Journal 1976-2 (in print), pp. 20

Matheron, G. (1963) – Principles of geostatistics, Economic Geology Vol. 58, p. 1246-1266.

Matheron, G. (1965) – Les Variables régionalisées et leur Estimation. Masson, Paris, pp. 306.

Royle, A.G. and E. Hosgit (1974) – Local estimation of sand and gravel reserves by geostatistical methods. Transactions, Vol. 83 1974, Institution of Mining & Metallurgy, p. A53-A62.

Wiener, N. (1949) – Extrapolation, Interpolation and Smoothing of Stationary Time Series. John Wiley and Sons, New York, N.Y., pp. 163.