A SEDIMENTOLOGICAL PATTERN RECOGNITION PROBLEM

Malcolm W. Clark and Isobel Clark
University of London

ABSTRACT

Analysis of grain-size distributions of coastal sands reveals that the distributions may be considered as composed of two (or more) lognormal components. It is tempting to infer that these components are derived from different depositional mechanisms. More faith could be placed in this inference if other characteristics of the deposit were describable in terms of the mixing of a similar number of components. Attention has been directed to the shape characteristics of the deposits. Feature extraction was achieved by digitizing the perimeter of the silhouettes of about 700 grains and fitting a truncated Fourier series to the outlines. The first eight harmonic amplitudes of these series were analyzed to detect naturally occurring clusters. Nonlinear mapping, fuzzy-set analysis and multivariate-mixing analysis were employed to determine clusters. KEY WORDS: *data display, mapping, cluster analysis, discriminant analysis, Fourier analysis, fuzzy-set analysis, multivariate mixing, statistics, sedimentology.*

INTRODUCTION

It has long been realized by sedimentologists that there is a correspondence between the size-frequency distributions of many sediments and the lognormal distribution. Some arguments have been advanced, notably by Middleton (1970) and Mahmood (1973), to provide some theoretical justification for this observation. However, despite these arguments, observed distributions continue to be wayward, and the correspondence between observed and theoretical distributions is convincing in only a few of the reported situations. In order to account for these discrepancies suggestions have been made to employ probability densities other than the lognormal. Tanner (1958) examined the Pearson Type I and IV distributions, Krumbein and Jones (1970) used a Gamma distribution, and Bagnold (1941) suggested the use of a function akin to the lognormal; Kittleman (1964) applied the Rosin-Rammler distribution.

Although these alternative distributions seem to provide a closer fit, they lack the general applicability of the lognormal, and, with the exception of the Rosin-Rammler, seem to have no theoretical justification. The Rosin-Rammler can be derived for crushed materials, and thus should apply to broken, unsorted rock material.

An alternative, which retains the generality of the lognormal, but introduces more flexibility, is to regard the frequency distribution as a sum of several lognormals, or, because the problem may be specified in terms of logarithms, a sum of normals

$$q(x;\underline{\theta}) = \sum_{j=1}^{i=m} \alpha_i \varphi[(x - \mu_i) \div \sigma_i] \quad (1)$$

where m is the total number of components;

μ_i is the mean of the i th component distribution;

σ_i is the standard deviation of the i th component distribution;

α_i is the proportion of the overall population deriving from component distribution i ;

$\varphi(z)$ is the probability density function of the standard normal distribution; and

$\underline{\theta}$ is a vector of parameters $(\mu_1 \ \sigma_1 \ \alpha_1 \ \mu_2 \ \sigma_2 \ \alpha_2 \ \dots \ \mu_m \ \sigma_m \ \alpha_m)$

This model has been discussed in sedimentological terms by Tanner (1964), Spencer (1963), and Folk (1971), among others. A similar model has been employed by Visser (1969), where he considers a sediment-size frequency curve to be composed of a sequence of truncated lognormal components. A similar model has been proposed by McKinney and Friedman (1970), where they extract a major lognormal component, and two smaller subsidiary components which need not be lognormal. Doeglas (1946) also suggested an additive scheme, but maintained that the underlying components were normal, not lognormal. This conflict can be resolved to some extent, because one of his reasons for choosing normal components was that coarse deposits commonly have symmetric size-frequency distributions. It can be demonstrated that as the mean of a lognormal distribution increases, its skewness decreases, until it becomes an essentially symmetric distribution.

There are some problems associated in this type of size-component analysis, the major problem lies in determining how many components are needed to approximate the observed distribution adequately. Whereas Walger (1961) suggested that no deposit is composed of more than three lognormal components, both van Andel (1973) and Curran (1960) published accounts where more than three components are present. One reason for this contradiction is that no obvious method exists of deriving sample size from weight-frequency data, so that "goodness of fit" tests, such as χ^2 , can be applied. Jones (1969) discussed this problem, and suggested that a value for the sample size may be determined considering the "smallest reproducible weight" of the weighing procedure, and using this as the basis for the total number of units in the frequency distribution. Using this convention it is possible to fit a model, using some objective criterion to decide if sufficient components have been fitted. The multitude of components observed by van Andel and Curran also may be attributed to the fact that they were taking offshore samples, where control to sample only a single sedimentation unit would have been impossible; thus their deposits were likely to be mixtures of several layers.

Fitting the model also may be a problem, but Clark (1976) suggested that a variety of numerical techniques may be employed. In analyzing about 120 beach and dune samples, it was determined that a two-component lognormal model provides a satisfactory fit in almost all

situations. This approach seems to provide a reasonably objective methodology for the analysis of sediment-size frequency distributions. It is necessary to emphasize that although the frequency distributions may be composed of more than one component, they need not have several distinct modes. If the component means are close enough together, a distribution will be observed which has only one mode, but it may be skewed. Despite this, we will use the loose terminology of "bimodal" to convey the concept of a single frequency distribution comprised of two or more components, but not necessarily with two or more distinct modes.

Having suggested that the size distributions may be thought of as consisting of more than one component, it becomes interesting to speculate whether this seeming structure is merely a fortuitous artefact, or whether it represents a real aspect of the deposit. If two size components are present, it seems reasonable to expect that these components also may be reflected in some other characteristics of the deposit. Following Griffiths (1967), a rock specimen may be defined uniquely as

$$P = f(m, s, sh, o, p) \quad (2)$$

where the properties (P) are a function of mineralogy (m), size (s), shape (sh), orientation (o) and packing (p). Where there is evidence that the size shows bimodal characteristics, are any of the other properties bimodal?

EXAMPLE

To test this concept, six samples were taken from the swash-backwash zone of the beach at Dungeness, Kent. The samples were taken from the west side of this cusate foreland, near the middle of a nearly straight section, about 6 km long. The six samples were taken at 25-m intervals, parallel to the shore line. About 40 gm of sand were collected from the top layer of the sand on the beach, over an area of about 1 sq m. The size frequency of the samples were determined by sieving at 3 ϕ intervals. The results of the sieving were modified by eliminating the contribution made by shell fragments. (These fragments made up less than 3 percent of the total distribution.)

The size-frequency distribution was analyzed by the method of nonlinear least squares (Clark and Garnett, 1974), and gave the results shown in Table 1. The analysis indicates the presence of two lognormal components in each sample, with the proportions of the components remaining reasonably constant between the samples.

Of the possibilities, the shape characteristics were chosen to examine more closely. Moss (1962, 1963, 1972) considered this in some detail, and was able to identify components on the basis of size and shape characteristics, but he did not relate them to the underlying lognormality of the size components.

A size grade was chosen which was well represented in the samples (2.75 to 3.00 ϕ); this was done to try to eliminate the confounding effect of size. A proportion of the grains were mounted in Canada balsam, on a glass slide. The expected proportion of the hypothetical shape components are

given under the column headed α^* in Table 1. Measurement of grain shape is a fairly routine procedure in sediment analysis. It was felt however that any shape differences which might occur were likely to be rather subtle, and that relatively simplistic measures of "a" and "b" axes, roundness and sphericity, were unlikely to yield the fine detail which might be required.

We have assumed, in common with many others, that shape information of a three-dimensional grain may be derived adequately from the two-dimensional outline of that grain. Some method is required here which will permit the grain periphery to be represented in a manner which is unique, and also tractable for some type of numerical analysis.

The analysis of closed curves like these is not restricted to sedimentary studies. Freeman (1961) suggested methods in which any arbitrary geometric shape may be encoded for further analysis, where a continuous figure can be represented in a discrete form. This is clearly of great merit in reducing the problem to manageable proportions.

TECHNIQUES AND DATA

There seem to be at least three categories of techniques which have utility:

(1) Fourier techniques as suggested by Brill (1968), Schwarcz and Shane (1969), Ehrlich and Weinberg (1970), Graniund (1972), and Zahn and Roskies (1972). The use of Fourier models for image encoding suggests a kinship with optical methods, which in fact turns out to be a close relationship (Pincus and Dobrin, 1966; Kaye and Naylor, 1972).

(2) Slope density, introduced by Nahin (1972), (also Sklansky and Nahin, 1972; Nahin, 1974).

(3) Moments, presented by Hu (1962) and Alt (1962).

A useful review of descriptions of line and shape is given by Duda and Hart (1973). Each of these approaches has merits, but, with the exception of the "radial" Fourier method, used by Schwarcz and Shane (1969) and Ehrlich and Weinberg (1970), none of them have been used in a geological context. The method used here was the radial Fourier method, not for any known superiority, but simply because we were not aware then of the work which had been done in other fields. In fact the radial method has one major disadvantage compared with the other Fourier methods, because it cannot handle curves with substantial reentrants. However, published accounts suggest that the method preserves useful information, and has the advantage (Tilman, 1973) that it was not critical that the maximum projective area be considered.

The mounted grains were magnified 250 times, and their outlines drawn, for about 100 to 120 grains at each site. These outlines then were digitized (Piper, 1970). The grains were digitized into between 36 and 60 points, depending on the size and complexity of the outline. The Cartesian coordinates of each of these outlines were converted into polar coordinates by first determining the grain "center of gravity", and using this point as the origin for the polar coordinates. The Fourier descriptor of the outline was determined in terms of the polar coordinates, but in a manner which differed slightly to that of Ehrlich and Weinberg. It is easier

to solve a Fourier series if the data points are equally spaced (in this example, at equal angular separation). Ehrlich and Weinberg use a linear interpolation scheme to provide the equal spacing, but this could introduce unwanted bias. Here equal spacing was achieved by first fitting a bicubic spline (Ahlberg, Nilson, and Walsh, 1967) to the grain periphery. A spline has the property of passing through all the data points, as a smooth curve. New data points at equal angular increments were calculated on the basis of the spline (Fig. 1). A Fourier series then was fitted to the new points.

The Fourier series may be expressed as

$$r(\theta) = a_0/2 + \sum_{i=1}^{\infty} a_i \cos(i \theta - \phi_i)$$

where r is the radius at any given angle θ ,
 a_i represents the contribution of the i th harmonic, and
 ϕ_i represents the phase angle (offset) of that harmonic.

This form of the expression would describe a continuous periphery. Because the periphery is not continuous in this example, but quantized, the series is truncated to $n/2$ terms, where n is the number of data points. In fact, in this, application, the series was truncated further, to only eight terms. The Fourier equation therefore becomes

$$r(\theta) = a_0/2 + \sum_{i=1}^8 a_i \cos(i \theta - \phi_i) \quad (3)$$

In order to standardize the a_i terms to a size-independent form, they were each divided by the average radius term ($a_0/2$). The eight terms of the truncated Fourier series retain about 85 to 90 percent of the information contained in the quantized curve (Fig. 2). A typical line spectrum is given in Figure 3. The a_i terms (the harmonic amplitudes) have the convenient property of being origin independent (or rotation invariant), which allows the amplitudes from one grain to be compared with those from another. The phase angles are clearly not rotation invariant, and therefore were dropped from the subsequent analysis.

The procedure adopted is a fairly standard one in pattern recognition. Meisel (1972) outlined the methodology as one which proceeds from the physical system (the sand grains), to the measurement space (the shape descriptors), into pattern space and "reduced" pattern space (the truncated Fourier series), and from there into some type of clustering procedure, from which a decision rule may be constructed in order to classify other data points (Fig. 4). The groupings themselves also may be used to summarize or exhibit the data.

Attention must be given to the clustering or grouping techniques, whereby naturally occurring homogeneous groups are determined in the data, remembering that it is anticipated that these groups may be present in the proportions as given in Table 1. Many of the classical clustering techniques suffer from the drawback that they are not able to handle large data sets. A total of 713 samples, with eight variables, is not an intolerably large data set, but simplistic number

crunching perhaps is not the most subtle or rewarding technique to employ. With this in mind, each of the six sites was analyzed individually. Site one (DW1) was used as a type of training set, where some conclusion was drawn about the nature of the samples. These conclusions then were tested on the other sites. This permits the consistency of the conclusions to be evaluated.

Some limitations of the classical clustering techniques are summarized by Howarth (1973). To avoid many of the usual drawbacks of clustering three techniques were employed. Nonlinear mapping (Sammon, 1969, 1970) was introduced into geology by Howarth (1973). Nonlinear mapping (NLM) is a method in which a multidimensional situation is represented in fewer dimension:: (commonly, but not necessarily, two), with a minimum amount of induced distortion. The rationale of the approach is that the human eye (together with the human brain) is better able to distinguish groups than any inflexible algorithm.

The nonlinear maps, however, proved to be of limited value in this instance. The map of the first site is given in Figure 5. No groups are readily apparent. Table 2 gives the error present in the mapping, together with the probable dimensionality. The maps suggest one of two things; either there are no groups, or they are overlapping to a fairly high degree. Given the fairly high error present, it is perhaps not surprising that clusters are not observed. The mapping of all 713 individuals indicated no grouping either. This was somewhat encouraging, because it suggests that there was no "drift" or change in the shape characteristics between the six sites (Fig. 6).

Fuzzy-set analysis also was used to seek out the groups (Zadeh, 1965). An example of a fuzzy set (Gitman and Levine, 1970) is the set defined as "all the very tall buildings", thus it is possible to see that it is a class of objects with a continuum of grades of membership. The algorithm of Gitman and Levine (1970) will detect unimodal fuzzy sets, and as such, will detect concentrations of points which may have irregular shapes (Fig. 7). A threshold parameter is used, whose value is somewhat arbitrary; different values of the threshold parameter can give different numbers of groups (and perhaps different groups). An example of how the number of groups may differ with the threshold is given in Table 3. Site one was analyzed extensively, with the object of determining those threshold values which gave two main clusters with approximately the expected proportion of members. Eight such values were determined, and are given in Table 4, together with the group sizes. These eight groupings were examined closely to derive consistently appearing groups. This provided three groups, one of 46 members which made up group 1, one of 55 members making up group 2, and a further 20 members which were unclassified, because they did not occur in the two core groups with regularity.

The other five sites were analyzed with the eight thresholds, and provide the results in Table 4. Although the results are not as decisive as might have been hoped, they do indicate the possible

presence of two major clusters at most of the sites. In interpreting the results, it is probably wise to regard groups of ten or fewer members as spurious, resulting from the fact that we are dealing with a finite (sampling) situation. Gitman and Levine (1970) note that a finite sample from a Gaussian distribution can be composed of several modes.

A decision rule also was constructed, based on the two "core groups of site one. Because these groups are likely to be rather irregular an empirical discriminant method (Howarth, 1971; Specht, 1967a, 1967b) was employed. This has the virtue of embodying no assumptions about the nature of the underlying distributions. The classification provided by this polynomial discriminant function corresponds to a fair degree with the groupings provided by the fuzzy-set analysis. The proportions of the two groups present at the six sites is given in Table 6. The twenty unclassified individuals of site one were classified by the polynomial discriminant function and added into the cores for the table. The relative consistency of the proportions again confirms the absence of drift in the shape characteristics.

The techniques used do not allow for any great amount of overlap of the components, but it seems reasonable to suggest that a high degree of overlap is present (assuming the components themselves exist). Multivariate-mixture analysis (Wolfe, 1970) permits clustering of overlapping groups. In providing this highly sophisticated analysis, the method is highly parametric. It can be seen as a multivariate extension of the methods used in analyzing the size-frequency distributions. It is assumed that the observed distribution comprises a mixture of several multivariate normal distributions. The distribution therefore is characterized by the vector of means, the covariance matrix, and the proportion, for each of the components. This requires the estimation of a large number of parameters. In an effort to reduce the computational effort, an alternative is given by Wolfe. Instead of allowing the covariance matrices to be unconstrained, the alternative requires that the covariance matrix for each of the components is equal. This reduces the number of parameters considerably. Wolfe terms the unconstrained solution NORMIX, and the constrained NORMAP. Both forms of the analysis were used. Clearly, there is no a priori reason to suppose that the shape characteristics should correspond to a multivariate normal distribution of the form required by the analysis; there is no reason to suggest that it should not. The marginal distributions for the amplitude of harmonic 2 and 3 are given in Figure 8. These two variables are the two with maximum variance in every situation. There is some evidence on the basis of these marginal distributions for suspecting bimodality. A peculiarity of the variables is that they are bounded; the lowest value possible for an amplitude is zero, and the highest (by the definition used here) is unity.

It is possible to test the results of the multivariate-mixture analysis to some extent. The optimum number of components can be established by testing the hypothesis that there is one type, two types, three types, etc. Given the fairly low number of individuals present, and the large number of parameters to be estimated, it was indeed unwise to proceed to more than three types, or components. The hypothesis testing is given in Table 5. It can be seen that for the NORMAP analysis (with equal covariance matrices) the favored solution tends to be a three-component solution, whereas for the NORMIX analysis, it is a two-component solution. This may be explained with reference to Figure 9, where it is suggested that two of the equal covariance matrices are attempting to approximate the single larger covariance matrix in the NORMIX analysis. Testing the NORMAP 3-type results against NORMIX 2-type results tends to confirm this view. In Table 6, where the proportions expected on the basis of the size analysis are compared, the proportions are those derived from the NORMIX analysis, except in the situation of DW3 and DW4, where the NORMAP results were used, because the tests suggest that these were the better choice of solution.

RESULTS

The results (Table 6) are in fair agreement although derived by three different methods. The poorer performance of the multivariate-mixture analysis probably can be attributed to the problems associated in estimating the covariance matrices from a data set which was on the small side (Ball, 1965). These results would seem to confirm the view that there exist, in the sediments analyzed, two lognormal size components which are reflected in the shape characteristics of the sediment. We intend extending the analysis both to consider the other size ranges within the same deposits, and to consider other sites.

How may these shape components be explained? Two possibilities seem attractive. The two-shape components may be either the result of different transport mechanisms, or be inherited characteristics.

The results obtained by Kolmer (1973) tend to support the concept of two transport mechanisms. He suggested the presence of two saltation populations, one associated with the swash and the other with the backwash. The results in Table 1 may be interpreted in this light easily. The coarser component may be deposited in the swash. The backwash may be lower in transporting power, due to the return of some of the water as percolation. This may account for the finer component. The slightly less well sorted nature of the coarser component may be related to the higher degree of turbulence in the swash. Waddell (1973) indicated that the sand arrived on the beach "as suspended load entrained in the uprush. The subsequent downslope movement of this material occurred as bed load in the backwash". This again may relate to the two size components. The suspended load may tend to favor the more angular grains, whereas in the backwash the more rollable grains may move more easily. Morris (1957) indicated that the roundness of grains is related in a rather complex manner to fluid velocity.

As an alternative explanation, component one may be derived from one environment (e.g. a river), whereas the other may be derived from another (e.g. offshore). This could account similarly for the two shape and size components.

The work presented here suggests that the lognormal components determined in size analysis are real features, and are reflected in the shape characteristics. The actual mechanisms giving rise to these shape and size components are not clear.

ACKNOWLEDGMENTS

This analysis presented here represents part of the work being carried out by M.W. Clark for the degree of Ph.D. in the University of London. Financial support for the presentation of the paper was granted by the Royal Society, the Hilary Bauerman Trust, the Department of Mining and Mineral Technology (Imperial College), and the Department of Geography (London School of Economics). Thanks also must be expressed to the Computer Units of Imperial College, Kings College and the London School of Economics. The nonlinear mapping algorithm (NLM)

was developed at the Rome Air Development Center, New York. The NORMIX/NORMAP and unimodal fuzzy-set analysis programs were made available through the University of London Computer Centre.

REFERENCES

- Ahlberg, J.H., Nilson, E.N., and Walsh, J.L., 1967, The theory of splines and their application: Academic Press, New York, 284 p-
- Alt, F.L., 1962, Digital pattern recognition by moments: Jour. Assoc. Comput. Mach., v. 9, no. 2, P. 240-258.
- Bagnold, R.A., 1941, The physics of blown sand and desert dunes: Chapman L Hall, London, 265 p.
- Ball, G.H., 1965, Data analysis in the social sciences: What about the details?: AFIPS Conf. Proc. v. 27, no. 1, p. 533-559.
- Brill, E.L., 1968, Character recognition via Fourier descriptors: WESCON Tech. Papers, Ses. 25 (Qualitative pattern recognition through image shaping), p. 1-10.
- Clark, I., and Garnett, R., 1974, The detection of multiple mineralisation phases by statistical methods: Trans. Inst. Min. Metall., v. 83, no. 809, p. A43-A52.
- Clark, M.W., 1976, A discussion of some methods for the statistical analysis of multimodal distributions, and their application to grain size data: Jour. Math. Geology, v. 8, in press.
- Curry, J.R., 1960, Tracing sediment masses by grain size modes: Rept. 21st Sess. Intern. Geol. Congress (Norden), pt. 23, p. 119-130.
- Doeglas, D.J., 1946, Interpretation of the results of mechanical analyses: Jour. Sed. Pet., v. 16, no. 1, p. 19-40.
- Duda, R.O., and Hart, P.E., 1973, Pattern classification and scene analysis: Wiley Interscience, New York, 502 p.
- Ehrlich, R., Orzeck, J.J., and Weinberg, B., 1974, Detrital quartz as a natural tracer - Fourier grain shape analysis: Jour. Sed. Pet., v. 44, no. 1, p. 145-150.
- Ehrlich, R., and Weinberg, B., 1970, An exact method for characterization of grain shape: Jour. Sed. Pet., v. 40, no. 1, p. 205-212.
- Folk, R.L., 1971, Longitudinal dunes of the northwestern edge of the Simpson Desert, Northwestern Territory, Australia, 1. Geomorphology and grain size relationships: Sedimentology, v. 16, no. 1, p. 5-54.

- Freeman, H., 1961, on the encoding of arbitrary geometric configurations: IRE Trans., Elec. Comp., EC-10, no. 2, p. 260-268.
- Gitman, I., and Levine, M.D., 1970, An algorithm for detecting unimodal fuzzy sets and its application as a clustering technique: IEEE Trans. Comp., C-19, no. 7, p. 583-593.
- Graniund, G.H., 1972, Fourier preprocessing for hand print character recognition: IEEE Trans. Comp., C-21, no. 2, P. 195-201.
- Griffiths, J.C., 1967, Scientific method in analysis of sediments: McGraw-Hill Book Co., New York, 508 P.
- Howarth, R.J., 1971, An empirical discriminant method applied to sedimentary rock classification from major element geochemistry: Jour. Math. Geology, v. 3, no. 1, P. 51-60.
- Howarth, R.J., 1973, Preliminary assessment of a nonlinear mapping algorithm in geological context: Jour. Math. Geology, v. 5, no. 1, p. 39-57.
- Hu, M-K., 1962, Visual pattern recognition by moment invariants: IRE Trans., Inf. Theory, IT-8, no. 2, P. 179-187.
- Jones, T.A., 1969, Determination of 'n' in weight frequency data: Jour. Sed. Pet., v. 39, no. 4, P. 1473-1476.
- Kaye, B.H., and Naylor, A.G., 1972, An optical information procedure for characterizing the shape of fine particle images: Pattern Recognition, v. 4, no. 2, p. 195-199.
- Kittleman, L.R., 1964, Application of Rosin's distribution in size-frequency analysis of clastic rocks: Jour. Sed. Pet.-i v. 34, no. 3, p. 483-502.
- Kolmer, J.R., 1973, A wave tank analysis of beach foreshore grain size distribution: Jour. Sed. Pet., v. 43, no. 1, P. 200-204.
- Krumbein, W.C., and Jones T.A., 1970, The influence of areal trends on correlations between sedimentary parameters: Jour. Sed. Pet., v. 40, no. 2, P. 656-685.
- McKinney, T.F., and Friedman, G.M., 1970, Continental shelf sediments of Long Island, New York: Jour. Sed. Pet., v. 40, no. 1, P. 213-248.
- Mahmood, K., 1973, Lognormal distribution of particulate matter: Jour. Sed. Pet., v. 43, no. 4, P. 1161-1166.
- Meisel, W.S., 1972, Computer-oriented approaches to pattern recognition: Academic Press, New York, 250 P.

Middleton, G.V., 1970, Generation of the log-normal frequency distribution in sediments, in Topics in mathematical geology: Consultants Bur., New York, P. 34-42.

Morris, W.J., 1957, Effects of sphericity, roundness and velocity on traction transportation of sand grains: Jour. Sed. Pet., v. 27, no. 1, P. 27-31.

Moss, A.J., 1962, The physical nature of common sandy and pebbly deposits. part I: Am. Jour. Sci., v. 260, no. 5, P. 337-373.

Moss, A.J., 1963, The physical nature of common sandy and pebbly deposits. part II: Am. Jour. Sci., v. 261, no. 4, P. 297-343.

Idoas, A.J., 1972, Bed-load sediments: Sedimentology, v. 18, nos. '3/4, P. 159-219.

Nahin, P.J., 1972, A parallel machine for describing and classifying silhouettes: unpubl. doctoral dissertation, Univ. California , Irvine, 184 p.

Nahin. P.J., 1974, The theory and measurement of a silhoij(--tte de-scriptor for image pre-processing and recognition: l'attc-rn Recognit Zion. v. 6. no. 2, P. 85-95.

Pincus,,H.J., and Dobrin, M.B., 1966, Geological applications of optical data processing: Jour. Geophysical Res., v. 71, no. 20, p. 4'861-4869.

Piper, D.J.W., 1970, The use of the D-Mac pencil follower in routine determinations of sedimentary parameters, *in* Data processing in biology and geology: Academic Press, New York, p. 97103.

Sammon, J.W., 1969, A non-linear mapping for data structure analysis: IEEE Trans. Comp., C-18 no. 5, p. 401-409.

Sammon, J.W., 1970, Interactive pattern analysis and classification: IEEE Trans. Comp., C-19, no. 7, p. 594-616.

Schwarcz, M.P., and Shane, X.C., 1969, Measurement of particle shape by Fourier analysis: Sedimentology, v. 13, nos. 3/4, p. 213-231.

Sklansky, J., and Nahin, P.J., 1972, A parallel mechanism for describing silhouettes: IEEE Trans. Comp., C-21, no. 11, P. 1233-1239.

Specht, D.F., 1967a, Generation of polynomial discriminant functions for pattern recognition: IEEE Trans. Elec. Comp., EC-16, no. 3, p@ 308-319.

Specht, D.F., 1967b, Vectorcardiographic diagnosis using the polynomial discriminant method of pattern recognition: IEEE Trans. Bio-med. Eng., BME-14, no. 2. p. 90-95.

Spencer, D.W., 1963, The interpretation of grain size distribution curves of clastic sediments: Jour. Sed. Pet., v. 33, no.1, P. 180-190.

Tanner, W.F., 1958, The zig-zag nature of type I and type IV curves: Jour. Sed. Pet., v. 28, no. 3, p. 372-375.

Tanner, W.F., 1964, Modification of sediment size distribution: Jour. Sed. Pet., v. 34, no. 1, p. 156-164.

Tilman, S.E., 1973, The effect of grain orientation on Fourier shape analysis: Jour. Sed. Pet., v. 43, no. 3, p. 867-869.

van Andel, T.H., 1973, Texture and dispersal of sediments in the Panama Basin: Geol. Soc. America Bull., v. 81, no. 4, p. 434-457.

Visher, G.S., 1969, Grain size distributions and depositional processes: Jour. Sed. Pet., v. 39, no. 3, p. 1074-1106.

Waddell, E., 1973, Dynamics of swash and implication to beach response: Coastal Studies Inst., Tech. Rept. 139, Louisiana State Univ., Baton Rouge, 49 P.

Walger, E., 1961, Grain size distribution within single arenaceous beds and their genetic meaning (in German): Geologische Rundschau, v. 51, no. 2, P. 494-507.

Wolfe, J.H., 1970, Pattern clustering by multivariate mixture analysis: Multivariate Behav. Res., v. 5, P. 329-350.

Zadeh, C.T., 1965, Fuzzy sets: Information and Control, v. 8, no. 3, P. 336-353.

Zahn, C.T., and Roskies, R.Z., 1972, Fourier descriptors for plane closed curves: IEEE Trans. Comp., C-21, no. 3, P. 269-281.

Table 1. Analysis of size frequency distribution into two lognormal components.

Site	Component 1			Component 2			df	α^*
	mean	s.d.	prop.	mean	s.d.	χ^2		
DW1	2.4745	0.3165	0.5713	2.7209	0.1637	0.32	1	0.3318
DW2	2.4672	0.3406	0.6447	2.7603	0.1005	0.08	2	0.3297
DW3	2.5380	0.3539	0.5788	2.7635	0.1232	0.19	2	0.3949
DW4	2.5463	0.3533	0.6627	2.7737	0.0976	0.11	2	0.3798
DW5	2.5526	0.3386	0.6555	2.7662	0.1000	0.65	2	0.3905
D116	2.6636	0.2846	0.5750	2.7697	0.0925	1.11	1	0.3803

Table 2. Error on nonlinear mapping.

Site	mapping error	probable dimensionality
DW1	15.845	2
DW2	16.604	2
DW3	17.243	2
DW4	19.939	2
DW5	23.128	2
DW6	15.710	2

Table 3. Results with different thresholds for fuzzy-set analysis at site DW1.

Threshold value	no. of groups	size of each group
0.0069	4	8 71 32 10
0.0088	3	12 105 4
0.0107	2	109 12
0.0126	3	104 12 5
0.0145	3	102 7 12
0.0164	4	102 7 11 1
0.0183	3	107 11 3
0.0202	4	113 1 6 1
0.0221	9	1 7 85 7 6 1 10 1 3
0.024	5	91 12 7 10 1
0.0259	4	111 6 1 3
0.0278	6	7 49 52 9 1 3

**Table 4. Selected thresholds, with group sizes from fuzzy-set analysis, for all sites.
Groups with 10 or fewer members have been omitted.**

threshold	DW1		DW2		DW3		DW4			DW5				DW6		
0.0069	71	32	106	12	117		86	35		73	30			86	16	
0.0278	49	52	104	12	81	33	96	15		53	15	28	21	58	14	40
0.0354	72	28	99	12	117		56	31		65	49			107		
0.0373	69	31	102	12	117		67	22	21	91	17			99		
0.0411	40	57	105	12				17	49	33	77	39		13	62	22
0.0525	54	49	115		117		103			79	38			104		
0.0544	67	40	75	48	105	11	78	27	117					109		
0.0563	50	66	65	58			82	26						86	11	11

**Table 5. Results from multivariate-mixture analysis,
giving selected number of components.**

Site	Chosen number of components		
	NORMAP	NORMIX	NORMAP3/NORMIX2
DW1	3	2	2
DW2	3	2	2
DW3	2	*	3
DW4	2	*	*
DW5	3	2	2
DW6	3	2	2

*no solution possible

**Table 6. Larger proportions of shape components derived from size analysis,
polynomial discriminant function based on cores from fuzzy-set analysis (PDF), and
multivariate-mixture analysis (MM).**

Site	1- α^*	PDF	MM
DW1	0.6682	0.5702	0.6502
DW2	0.6703	0.6239	0.5642
DW3	0.6051	0.6410	0.8000
DW4	0.6202	0.6160	0.9600
DW5	0.6095	0.6016	0.7925
DW6	0.6197	0.6429	0.6338

Figure 1. A, original data points outlining grain periphery; and B, equally spaced points calculated from bicubic spline fit.

Figure 2. Typical relationship of variance explained (in percent) by Fourier series fit (harmonic number). Almost 90 percent is explained by truncated series at eight terms.

Fig. 3

Figure 3. Typical line spectrum from Fourier fit.

Figure 4. Standard procedure in pattern recognition (after Meisel, 1972).

Figure 5. Nonlinear map of site one (DW1).

Figure 6. Nonlinear map of 713 individuals at all six sites.

Figure 7. Shapes of point concentrations, a=circular, b=elongated, c=irregular.

NORMAP
3 components

NORMIX
2 components

Figure 9. Graphic representation of NORMAP analysis (with equal covariance matrices) which result in three-component solution, and NORMIX analysis which tends toward two-component solution.

Figure 8A. Distribution of amplitudes (from 0 to 1) of harmonic 2 and harmonic 3 for all sites (DW1 through DW6). These are maximum variance directions.

Figure 8B. Distribution of natural logarithms of amplitudes of harmonic 2 and harmonic 3 for all sites. Because this transformation does not seem to normalize distributions, it is not used in analysis.