10

Statistical Analysis of Particle Sizes and Sediments

Nick Fieller
Department of Probability and Statistics, University of Sheffield

Eleanor Flenley
Department of Probability and Statistics, University of Sheffield

10.1 Introduction

Particle size data are routinely collected in palaeoecological and other archaeological investigations. Usually material is passed through a sequence of sieves of decreasing mesh sizes, perhaps supplemented by some technique based on hydrometer measurements to size the very fine fraction. Recently, more automatic and faster techniques, such as laser particle sizers, have become available and so the opportunities for obtaining such data are increasing.

The traditional statistical analysis of particle size data rests on approximate calculation of sample moments (sorting, skewness etc) by graphical techniques, the so-called Folk and Ward estimates (Folk & Ward 1957 and Inman 1952). The geological usefulness of such analyses has often been questioned, (Ehrlich 1983), and there are some doubts as to the statistical validity of the method (Bagnold 1979, Bagnold & Barndorff-Nielsen 1980, Christiansen *et al.* 1984 and Fieller *et al.* 1984).

This paper describes the use of statistical models for particle size distributions. The class of models proposed, the log skew Laplace family, has geological interpretation and appears to be of wide applicability. The technique has been applied to a variety of problems, including the resolution of the question of the siting of Mesolithic middens in Oronsay, the investigation of quaternary sediments from the caves in Creswell Crags and an examination of the depositional processes of aeolian sands on a partially vegetated climbing dune in Libya. The purpose of this paper is twofold. The main objective is to describe effective and informative ways of analysing particle size data. A secondary objective, but one perhaps of much more general importance, is to illustrate that investment in statistical modelling and analytic techniques can yield dividends.

In the present context, it is quite possible to analyse simple particle size data without any regard to the statistical models underlying the data. Calculation of Folk and Ward estimates may be all that is required to discriminate between two distinct types of sand, beach and dune sand for example. Performing the analysis by estimating the parameters of a proposed statistical model for the particle size distribution may appear to be an over-sophisticated approach to a simple problem. However, the advantage of using a model for the distribution is that it can be modified and adapted to new and more complex situations whereas mere calculation of descriptive Folk and Ward statistics can not. In particular, the examples illustrate the adaptation and extension of a simple basic model to handle firstly data obtained by two different measuring techniques (sieves and hydrometer) and secondly data that are mixtures of two different size distributions. Similarly, analysis of particle size data obtained by direct measurement of grains in thin sections (*e.g.* of inclusions in pottery fabric) would need to take account of potential stereological bias induced by the sectioning. This could only be tackled by incorporation of the mathematical effects of sectioning in the underlying model.

10.2 Measurement of particle size data

The most widely used technique for measuring the relative proportions of particles of different sizes, at least in geological work, is to pass the sample through a sequence of sieves of decreasing mesh sizes. The material entrapped in each sieve is weighed and so the 'size-distribution' is recorded as the relative proportions by weight of particles within each size range determined by the mesh sizes of the preceding and entrapping sieve. The fact that this produces technically a 'mass-size distribution' rather than a 'frequency-size distribution' as would be obtained by counting rather than weighing the particles has some statistical implications but these will not be detailed here.

The technique is adequate for only a limited range of sizes. The lower limit is determined by the finest sieve and is typically 63 microns, though finer sieves do exist. The upper limit is determined by the dynamics of shaking particles through sieves (the heavier the particle the more vigorously must the sieve be shaken) and is effectively a few millimetres. To measure the proportions of particles of sizes finer than 63 microns, a common technique is to record the rate of change of density with time of a mixture of a known weight of the material in a known volume of water. Application of Stokes' Law permits the calculation of the empirical mass-size distribution for this fine fraction. This 'hydrometer method' gives the relative proportions of particles in size ranges determined largely by the times at which the readings are taken. Details of the method are given in Kaddah 1974. The method is applicable again only to a limited range of sizes, perhaps a few microns to about 70 microns.' Particles larger than that settle out too quickly (*i.e.* within a few seconds) to be measured accurately, finer particles settle too slowly (*i.e.* many hours). In many cases, the material of interest will be composed of particles whose sizes lie entirely within one or other of these size ranges. In such cases the appropriate measuring method can be selected and the 'sizes' measured without much consideration being given to precisely what aspect of size of the particle is being recorded. The size of an object is a property which is intuitively obvious to appreciate but is difficult to define mathematically. Particles possess length, breadth, thickness, mass, surface area, maximal diameter and many other properties each of which contribute to the overall picture of size. The relationship between them is determined by the shape of the particle. Different measuring techniques measure different size properties. When a particle passes through a sieve with known square mesh size, this provides information on the magnitude not of its length but of its second principal diameter. More exactly, the situation is very much more complicated by the geometry of aligning an irregular shape in a square mesh, but the second principal diameter (*i.e.* 'breadth') is a reasonable approximation.

When calculations are performed to convert a hydrometer reading made at a known time after settling began, the size property measured is a rather unusual one. It is the diameter of the 'equivalent sphere', that is the diameter of a spherical particle whose rate of settling is the same as that of the particle. In effect it is essentially the viscous drag coefficient of the particle. Particles of different sizes (in the intuitive sense, or in the sieve sense) can have identical rates of settling in water and so would be recorded by the hydrometer method as being of the same 'size'. For this to be possible the particles would have to be of different shapes, that is have different ratios of first, second and third principal diameters.

If the material to be measured consists of particles whose sizes span the two ranges appropriate for sieve and hydrometer measurement, then the two techniques are used in combination. However, the properties measured by the two techniques are quite different. It would be unjustified to combine the two sets of measurements without making some adjustment for this.

Section 10.6 below presents an example of how such an adjustment can be made in a much simplified case. The two measuring techniques described above do not exhaust the possibilities for measuring particle sizes; they merely typify the problems that can arise. Other devices for sizing particles which are becoming increasingly widely used are Coulter counters (which measure electrical resistance), laser-sizers (which measure the light scattering properties of particles, *i.e.* essentially their surface curvature) and image analyzers. Each of these devices measures distinct size properties and their use, particularly in combination, will raise complex statistical problems.

10.3 Statistical modelling and analysis of particle size data

Analysis of particle size data can proceed by either of two distinct routes. The first attempts only to obtain a simple numeric summary and description of the observed data; the second postulates a statistical model for the distribution of the sizes and proceeds to estimate the parameters of the model by statistical techniques. The first approach describes only the actual data obtained, the second attempts to investigate the process underlying the data.

Simple data description, using perhaps sample moments (mean, sorting, skewness and kurtosis), can be very effective in simple situations, distinguishing between beach and dune sand for example. The approach is exemplified in the work of Friedman (Friedman 1961, Friedman 1967, Friedman 1979a, Friedman 1979b). Typically, these estimates are calculated semi-graphically from a log Normal probability plot of the data, following suggestions of Folk & Ward 1957 and Inman 1952. However, the technique does suffer from numerous deficiencies which severely restrict its use in more complex situations. In particular, calculation of such descriptive statistics for mixed grain-size distributions can be misleading (Flenley *et al.* 1987), especially when calculated from log Normal probability plots of the data. Section 10.5 below returns to this problem. The second approach is to estimate the parameters of some proposed model for the particle size distribution. Models that have been considered are the log Normal (e.g. Wyrwoll & Smyth 1985) with two parameters, the log hyperbolic (Barndorff-Nielsen 1977, Bagnold & Barndorff-Nielsen 1980) with four parameters and the log skew Laplace (Olbricht 1982, Fieller *et al.* 1984, Fieller & Gilbertson 1985, Flenley 1985, Flenley *et al.* 1987) with three parameters. The particular model considered here, which will be extended and adapted to encompass more complex situations, is the last of these, the log skew Laplace. The description of this model is best achieved by discussion of a specific example.

Fig. 10.1 presents four graphical representations of one particular sample of sand ('Libyan sample 276') sieved at quarter-phi intervals taken from a barchan dune near Gasr Banat in the Libyan Pre-Desert. The diagram in the top left is a histogram of the actual sizes of the particles. The diagram at the top right is the same histogram displayed on a logarithmic horizontal scale, *i.e.* it is a histogram of the log sizes. The diagram at the bottom left is a 'log histogram' of the log sizes, that is the histogram is displayed with the vertical scale also logarithmic. This follows a tradition established by Bagnold (Bagnold 1937) who discovered empirically that

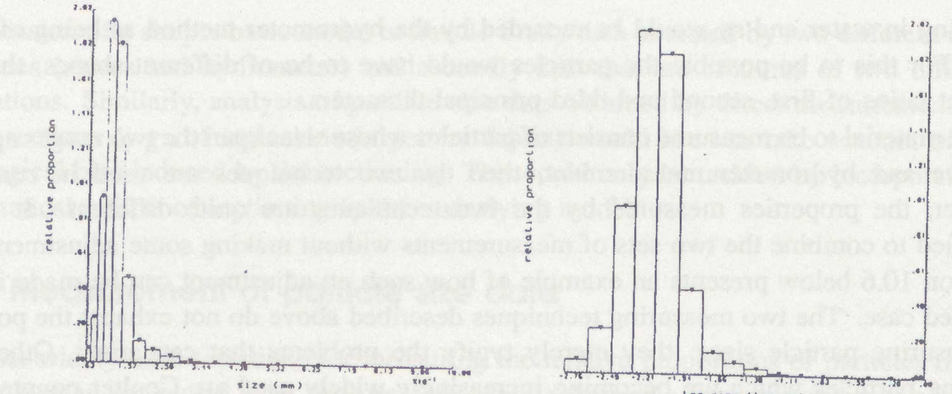

Fig. 10.1: Libyan sample 276: four representations

many examples of particle size data exhibited a particularly simple form when displayed in this way. It is clear, on this choice of scales, that the log relative proportions increase linearly with log size to a maximum and then decrease linearly. This feature is the basis of the use of the log skew Laplace distribution as a model for particle size data. Additionally, there are mathematical theories based on random breakage which would also tend to suggest that this form of model would be appropriate.

The fitted (*i.e.* estimated) log skew Laplace density is displayed in the final diagram of Fig. 10.1, at the bottom right, again on log scales for both axes. On this logarithmic scaling it is represented by a pair of straight lines, one 'fitted' to the log relative proportions of the finer sized grains, the other to those of the coarser particles. The 'fitting' is actually performed by the method of maximum likelihood, and necessarily gives more importance to the size ranges with most particles. The apparent discrepancies in the fit in the right hand tail are exaggerated by the vertical logarithmic scaling. Fig. 10.2 displays the fitted density and natural histogram of log sizes without the vertical logarithmic transformation. It illustrates an alternative description of the density as two 'back-to-back' exponential densities or an asymmetric double exponential

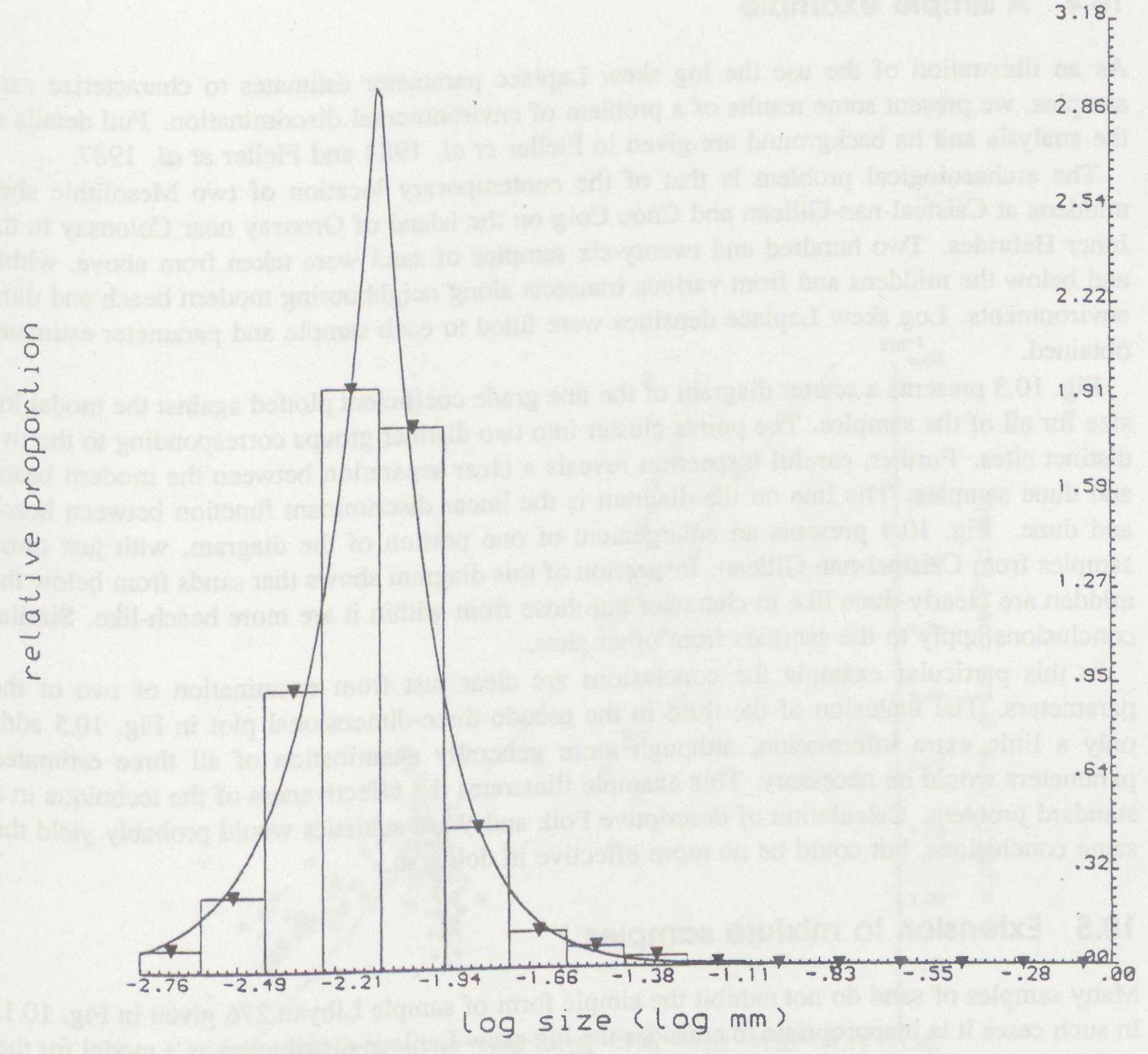

Fig. 10.2: Libyan sample 276: natural/log scales, fitted Laplace density.

The three parameters required to specify this distribution are

1. the slope of the left hand line on the log-log diagram,
2. the slope of the right hand line and
3. the abscissa of their point of intersection.

The first of these reflects the relative proportion of finer material in the sample, the second that of the coarser and the third gives a 'typical' or most frequently occurring log size.

Thus the three parameters have meaningful interpretations and will be referred to as (1) the fine grade coefficient, (2) the coarse grade coefficient and (3) the modal log size.

The estimates of these three parameters can be used to characterize the sample in much the same way as can Folk and Ward estimates of sorting and skewness etc. The advantage of using the former method rather than the latter, even in simple situations, is that the parameters have geological interpretability.

10.4 A simple example

As an illustration of the use the log skew Laplace parameter estimates to characterize sand samples, we present some results of a problem of environmental discrimination. Full details of the analysis and its background are given in Fieller *et al.* 1984 and Fieller *et al.* 1987.

The archaeological problem is that of the contemporary location of two Mesolithic shell middens at Caisteal-nan-Gillean and Cnoc Coig on the island of Oronsay near Colonsay in the Inner Hebrides. Two hundred and twenty-six samples of sand were taken from above, within and below the middens and from various transects along neighbouring modern beach and dune environments. Log skew Laplace densities were fitted to each sample and parameter estimates obtained.

Fig. 10.3 presents a scatter diagram of the fine grade coefficient plotted against the modal log size for all of the samples. The points cluster into two distinct groups corresponding to the two distinct sites. Further, careful inspection reveals a clear separation between the modern beach and dune samples. The line on the diagram is the linear discriminant function between beach and dune. Fig. 10.4 presents an enlargement of one portion of the diagram, with just those samples from Caisteal-nan-Gillean. Inspection of this diagram shows that sands from below the midden are clearly-dune like in character but those from within it are more beach-like. Similar conclusions apply to the samples from other sites.

In this particular example the conclusions are clear just from examination of two of the parameters. The inclusion of the third in the pseudo-three-dimensional plot in Fig. 10.5 adds only a little extra information, although more generally examination of all three estimated parameters would be necessary. This example illustrates the effectiveness of the technique in a standard problem. Calculation of descriptive Folk and Ward statistics would probably yield the same conclusions, but could be no more effective in doing so.

10.5 Extension to mixture samples

Many samples of sand do not exhibit the simple form of sample Libyan 276 given in Fig. 10.1. In such cases it is inappropriate to consider the log skew Laplace distribution as a model for the underlying particle size distribution. Some samples exhibit a clear bimodality and the obvious interpretation is that they are composed of a mixture of two underlying simple particle size distributions. Fig. 10.6 shows 'Libyan sample 550' taken from a seif dune in Wadi el Amud in Libya. The diagram in the upper left of the figure is a histogram of the log sizes. Also shown is a fitted mixture of two log skew Laplace densities. This will be described in more detail below. The diagram in the upper right of the figure is a log Normal probability plot of the data. This would be used for calculation of Folk and Ward estimates. The clear bimodality apparent in the histogram is not revealed in the probability plot. This illustrates the danger of using such probability plots and the associated simple summary statistics which fail to capture the most obvious feature of the sample distribution.

The diagram in the lower part of Fig. 10.6 is the log histogram of the sample with the fitted Laplace mixture. This mixture distribution is essentially a weighted average of two separate log skew Laplace densities. It is characterized by seven parameters, three from each of the two components (carrying corresponding interpretations for each component to those above) and the seventh giving the relative proportions of the two components in the mixture.

To illustrate the use of this, consider the diagrams in Fig. 10.7. The upper portion is a cross-section of a climbing dune in Wadi Merdum near Beni Ulid in Libya, the lower part of

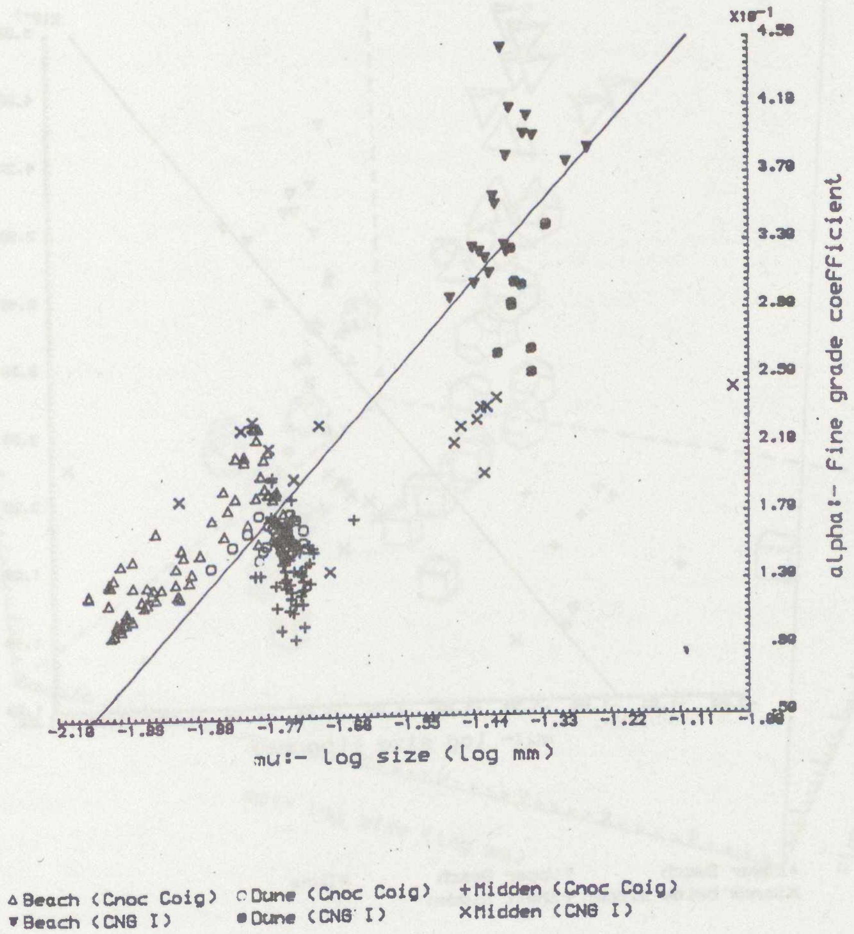

Fig. 10.3: Beach, dune and midden sands from Oronsay

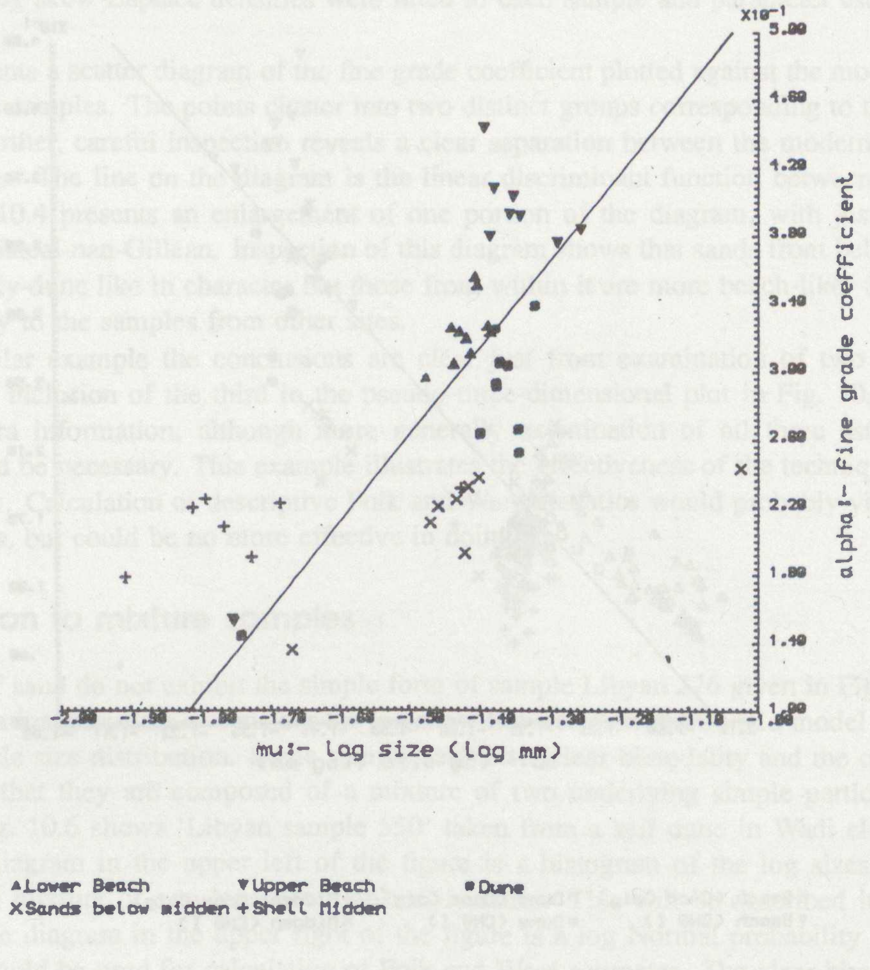

Fig. 10.4: Beach, dune and midden sand samples from CNG 1

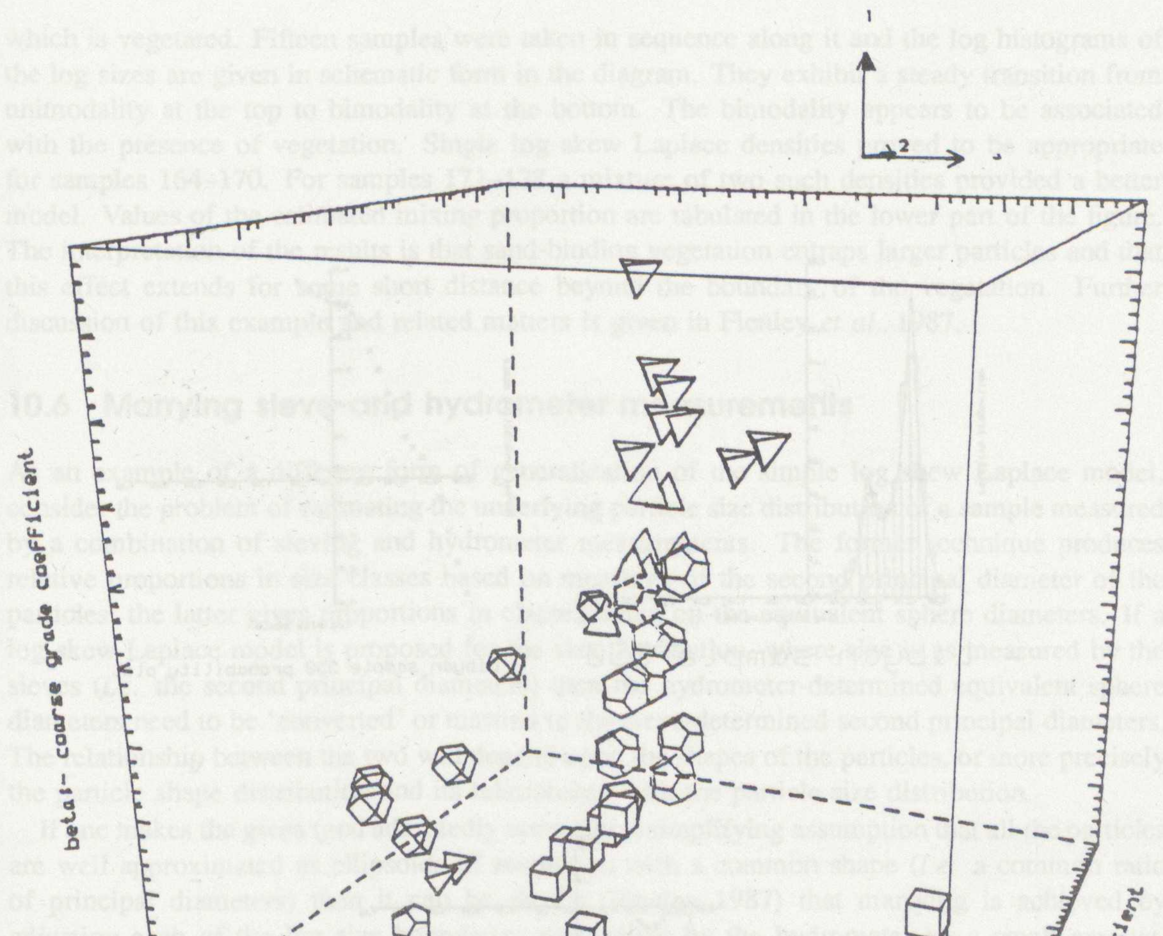

Fig. 10.5: Caisteal nan Gillean sand samples

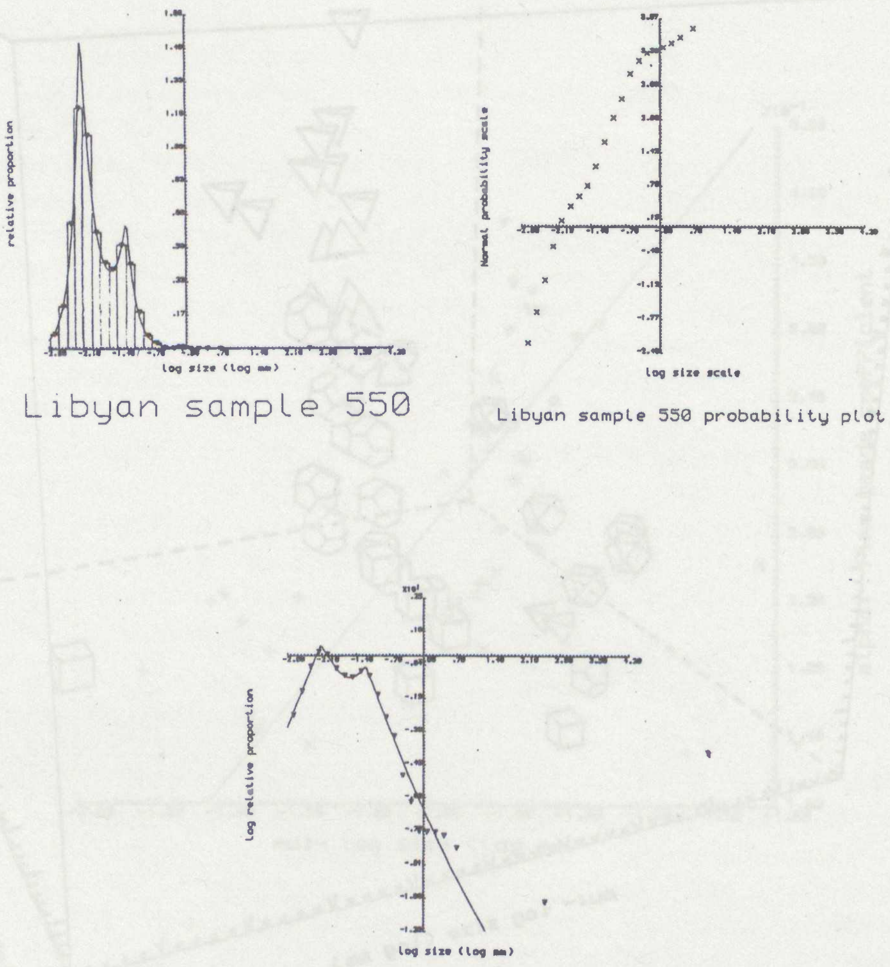

Fig. 10.6: Libyan sample 550

which is vegetated. Fifteen samples were taken in sequence along it and the log histograms of the log sizes are given in schematic form in the diagram. They exhibit a steady transition from unimodality at the top to bimodality at the bottom. The bimodality appears to be associated with the presence of vegetation. Single log skew Laplace densities proved to be appropriate for samples 164–170. For samples 171–178 a mixture of two such densities provided a better model. Values of the estimated mixing proportion are tabulated in the lower part of the figure. The interpretation of the results is that sand-binding vegetation entraps larger particles and that this effect extends for some short distance beyond the boundary of the vegetation. Further discussion of this example and related matters is given in Flenley *et al.* 1987.

10.6 Marrying sieve and hydrometer measurements

As an example of a different form of generalization of the simple log skew Laplace model, consider the problem of estimating the underlying particle size distribution of a sample measured by a combination of sieving and hydrometer measurements. The former technique produces relative proportions in size classes based on measures of the second principal diameter of the particles; the latter gives proportions in classes based on the equivalent sphere diameters. If a log skew Laplace model is proposed for the size distribution, where size is as measured by the sieves (*i.e.* the second principal diameters) then the hydrometer-determined equivalent sphere diameters need to be 'converted' or married to the sieve-determined second principal diameters. The relationship between the two will depend upon the shapes of the particles, or more precisely the particle shape distribution and its relationship with the particle size distribution.

If one makes the gross (and admittedly unrealistic) simplifying assumption that all the particles are well approximated as ellipsoids of revolution with a common shape (*i.e.* a common ratio of principal diameters) then it can be shown (Flenley 1987) that marrying is achieved by adjusting each of the log size boundaries determined by the hydrometer by a small amount. This adjustment is the same for each of the log size boundaries. Its precise size is determined by the shape of the particles (*i.e.* by the ratio of their principal diameters). The shape of the particles is of course unknown, but the shift required can be incorporated as an extra parameter and estimated by maximum likelihood from the data at the same time as the other three parameters of the log skew Laplace density. In effect, the basic model is modified to describe the additional feature of the combination of measuring techniques.

An example of this is given in Figs. 10.8 and 10.9. Sample A18 from Robin Hood's Cave in Creswell Crags contains an appreciable proportion of material finer than the smallest sieve used, sixty-three microns. The fine material was sized using the standard 'hydrometer method'. Fig. 10.8 gives the log histogram of the data, without any adjustment to correct for the differences introduced by the use of two distinct techniques. Also shown is the best-fitting log skew Laplace density. The broken vertical line marks the smallest sieve size, 63 microns. The largest size boundary determined by the hydrometer was just to the right of this, 67 microns. The size range between the two is clearly too wide as is evidenced by the markedly low relative proportion apparently recorded for this size range.

Estimation by maximum likelihood of the optimal amount by which to adjust the hydrometer determined boundaries shows that they should each be reduced by a factor of 0.9588, that is by 0.0421 log millimetres on the logarithmic scale. Fig. 10.9 shows the resulting log histogram and newly calculated log skew Laplace density. The marked anomaly at the join between hydrometer-determined and sieve-determined boundaries has been smoothed out. Admittedly the fine fraction is less well-modelled than the coarser, this is perhaps a result of the formation

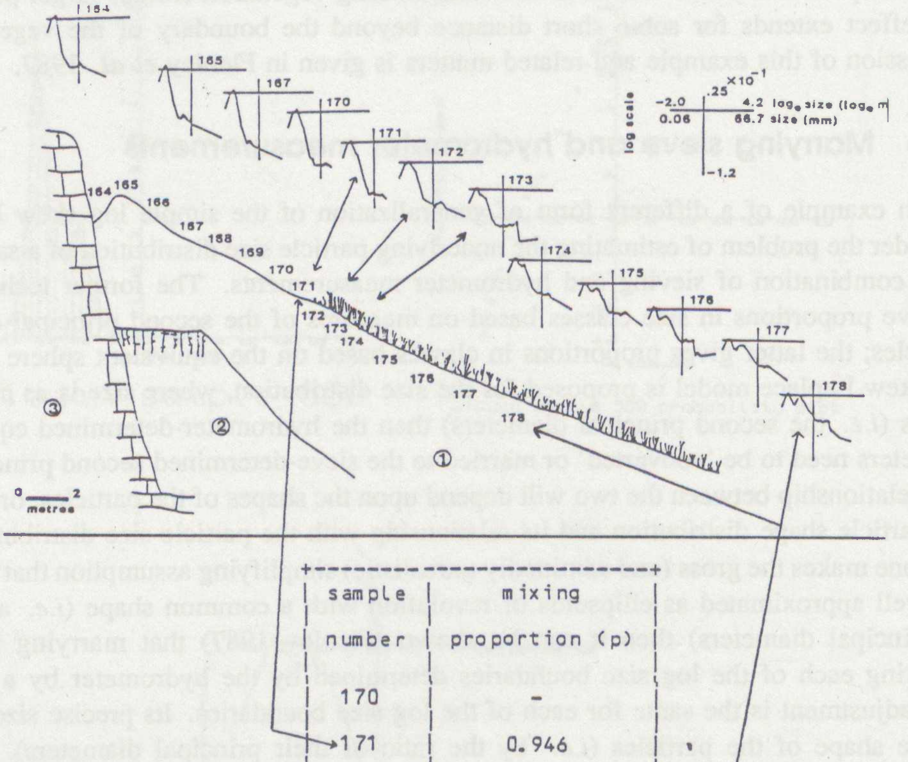

Fig. 10.7: Partially vegetated climbing dune

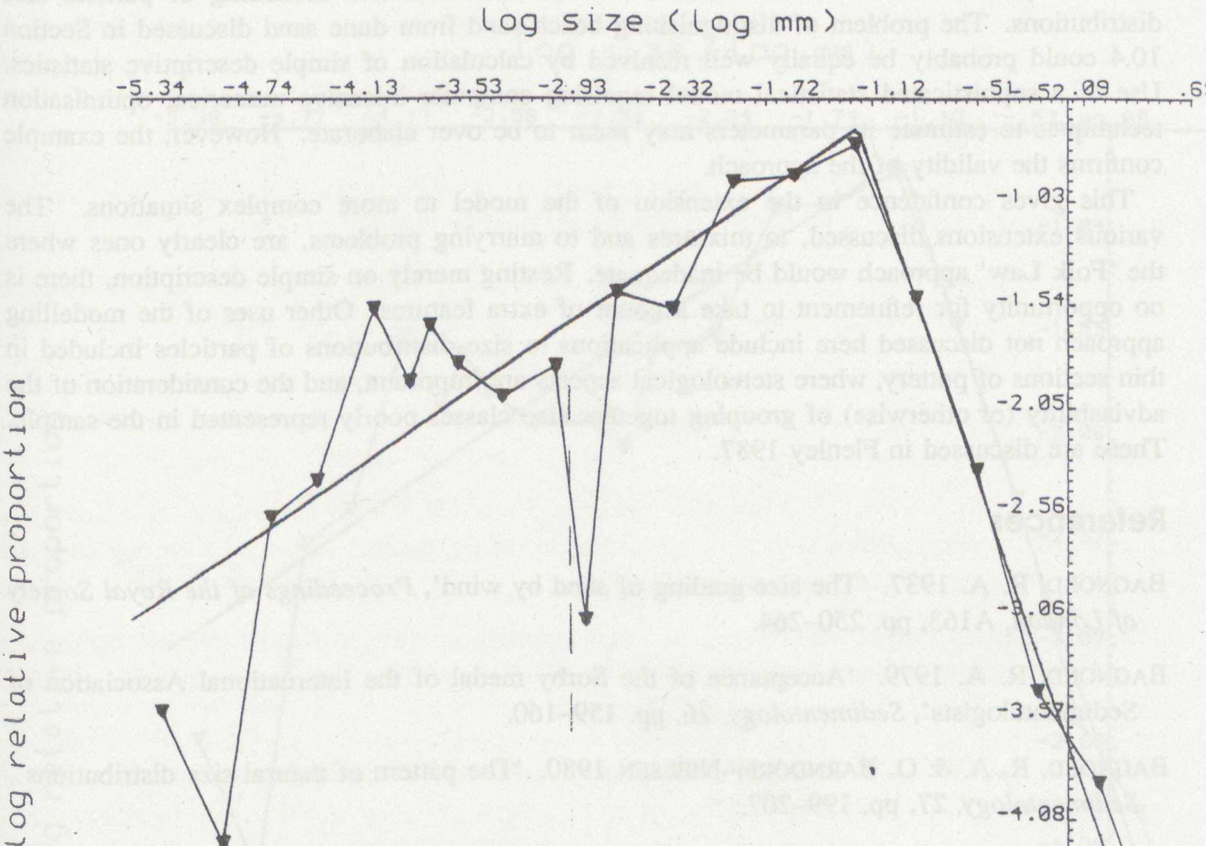

Fig. 10.8: Creswell Sample A18: sieves and unadjusted hydrometer

of this portion of the sample being governed by rather different processes from those at larger size ranges.

10.7 Conclusions

The examples discussed above illustrate the value of statistical modelling of particle size distributions. The problem of distinguishing beach sand from dune sand discussed in Section 10.4 could probably be equally well resolved by calculation of simple descriptive statistics. Use of a sophisticated statistical model requiring computer intensive numerical optimisation techniques to estimate its parameters may seem to be over elaborate. However, the example confirms the validity of the approach.

This gives confidence in the extension of the model to more complex situations. The various extensions discussed, to mixtures and to marrying problems, are clearly ones where the 'Folk Law' approach would be inadequate. Resting merely on simple description, there is no opportunity for refinement to take account of extra features. Other uses of the modelling approach not discussed here include applications to size distributions of particles included in thin sections of pottery, where stereological aspects are important, and the consideration of the advisability (or otherwise) of grouping together size classes poorly represented in the sample. These are discussed in Flenley 1987.

References

BAGNOLD, R. A. 1937. 'The size-grading of sand by wind', *Proceedings of the Royal Society of London*, A163, pp. 250–264.

BAGNOLD, R. A. 1979. 'Acceptance of the Sorby medal of the International Association of Sedimentologists', *Sedimentology*, 26, pp. 159–160.

BAGNOLD, R. A. & O. BARNDORFF-NIELSEN 1980. 'The pattern of natural size distributions', *Sedimentology*, 27, pp. 199–207.

BARNDORFF-NIELSEN, O. 1977. 'Exponentially decreasing distributions for the logarithm of particle size', *Proceedings of the Royal Society of London*, A353, pp. 401–419.

CHRISTIANSEN, C., P. BLAESILD, & K. DALSGAARD 1984. 'Re-interpreting 'segmented' grain-size curves', *Geological Magazine*, 121, pp. 47–51.

EHRLICH, R. 1983. 'Size analysis wears no clothes or have moments come and gone?', *Journal of Sedimentary Petrology*, 53, p. 1.

FIELLER, N. R. J. & D. D. GILBERTSON 1985. 'Skew log Laplace distributions and the discrimination of shoreline environments at Dawlish Warren, Devon', *Proceedings of the Ussher Society*, 6, pp. 271–272.

FIELLER, N. R. J., D. D. GILBERTSON, & W. OLBRICHT 1984. 'A new method for the environmental analysis of particle size data from shoreline environments', *Nature*, 311, pp. 648–651.

FIELLER, N. R. J., D. D. GILBERTSON, & D. A. Y. TIMMINS 1987. 'Sedimentological analysis of the shell-midden sites', *in* P. R. Mellars, (ed.), *Excavations on Oronsay: Prehistoric Human Ecology on a Small Island*, Edinburgh University Press.

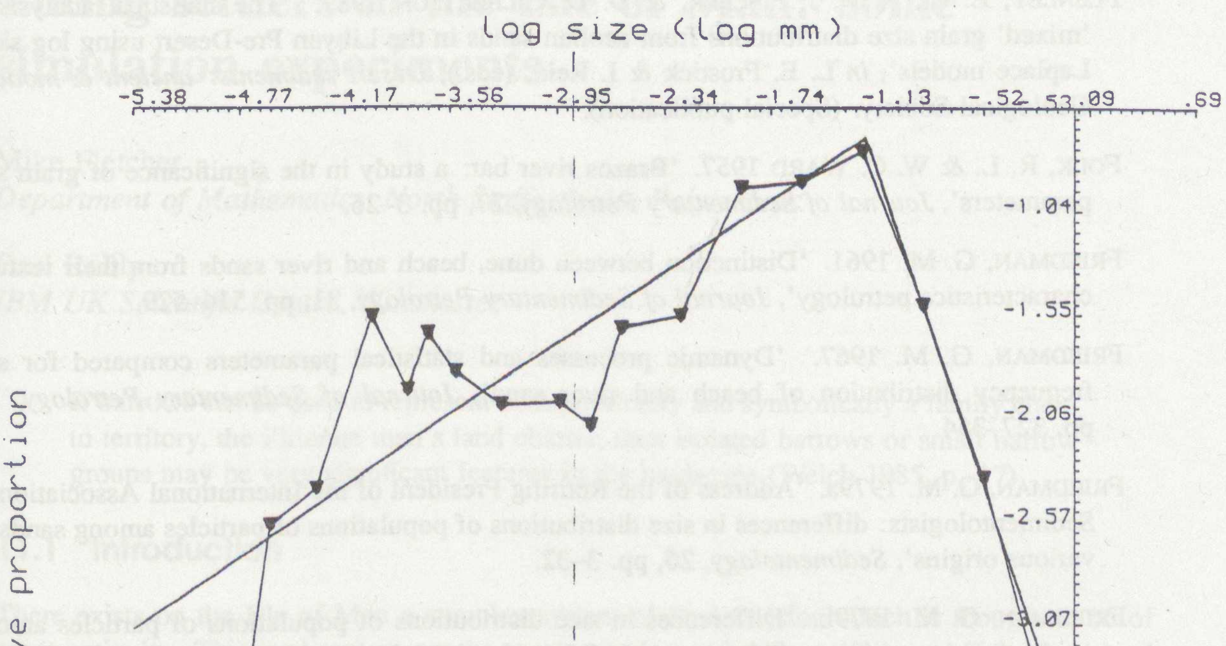

Fig. 10.9: Creswell Sample A18: sieves and adjusted hydrometer

FLENLEY, E. C. 1985. 'Use of mixture distributions in the modelling of sand particle sizes', *in Proceedings of the International Workshop on the Physics of Blown Sand*, pp. 633–648, Department of Theoretical Statistics, University of Aarhus, Denmark.

FLENLEY, E. C. 1987. *Statistical analysis of particle size data.* PhD thesis, University of Sheffield.

FLENLEY, E. C., N. R. J. FIELLER, & D. D. GILBERTSON 1987. 'The statistical analysis of 'mixed' grain size distributions from aeolian sands in the Libyan Pre-Desert using log skew Laplace models', *in* L. E. Frostick & I. Reid, (eds.), *Desert sediments: ancient & modern*, Geological Society. (Special publication).

FOLK, R. L. & W. C. WARD 1957. 'Brazos river bar: a study in the significance of grain size parameters', *Journal of Sedimentary Petrology*, 27, pp. 3–26.

FRIEDMAN, G. M. 1961. 'Distinction between dune, beach and river sands from their textural characteristics petrology', *Journal of Sedimentary Petrology*, 31, pp. 514–529.

FRIEDMAN, G. M. 1967. 'Dynamic processes and statistical parameters compared for size frequency distribution of beach and river sand', *Journal of Sedimentary Petrology*, 37, pp. 327–354.

FRIEDMAN, G. M. 1979a. 'Address of the Retiring President of the International Association of Sedimentologists: differences in size distributions of populations of particles among sands of various origins', *Sedimentology*, 26, pp. 3–32.

FRIEDMAN, G. M. 1979b. 'Differences in size distributions of populations of particles among sands of various origins: addendum to IAS Presidential Address', *Sedimentology*, 26, pp. 859–862.

INMAN, D. L. 1952. 'Measures for describing the size distribution of sediments', *Journal of Sedimentary Petrology*, 22, pp. 125–145.

KADDAH, M. T. 1974. 'The hydrometer method for detailed particle-size analysis: 1. graphical interpretation of hydrometer readings and test of method', *Soil Science*, 118, pp. 102–108.

OLBRICHT, W. 1982. *Modern Statistical Analysis of Ancient Sand*, Master's thesis, Department of Probability and Statistics, University of Sheffield.

WYRWOLL, K. & G. K. SMYTH 1985. 'On using the log hyperbolic distribution to describe the textural characteristics of eolian sediments', *Journal of Sedimentary Petrology*, 55, pp. 471–478.