Journal of Structural Geology, Vol. 4, No. 1, pp. 93 to 103, 1982
Printed in Great Britain

0191-8141/81/010093-11 $03.00/0
© 1982 Pergamon Press Ltd.

Convolution of fabric data to determine probability distribution

E. C. KALKANI and R. R. B. VON FRESE

Department of Geotechnical Engineering and Department of Geosciences,
Purdue University, West Lafayette, IN 47907, U.S.A.

(*Received* 9 *January* 1981; *accepted in revised form* 23 *October* 1981)

Abstract—To investigate the higher order, multimodal characteristics of fabric data populations, five filter operators are constructed on the surface of the reference hemisphere for convolution with fabric data. These operators include spherical unit disk, triangle, sinc, sinc^2 and Gaussian functions. To illustrate filter performance, two fabric data sets, taken from the lower Schoonover sequence of the Northern Independence Mountains, Nevada, are convolved with the filters. The results show that the spherical unit disk, triangle and sinc operators are especially useful for identifying characteristics of the population density distribution and related probability contours for unimodal data sets. The Gaussian and sinc^2 operators, however, provide increased resolution of the higher frequency components for multimodal fabric data sets, along with relative statistical probability contours from which inferences on the probability distribution of the population can be made. In general, the convolution process is a very effective tool for analyzing the multimodal behaviour of fabric data, because of the ease and the flexibility with which filter operators may be designed.

INTRODUCTION

FABRIC DATA registered over the surface of a reference hemisphere are usually presented as contours of a statistically approximate population density distribution on an equal-area projection. An extensive literature is available on the methods of presenting fabric data and their density distributions in which the principles are explained together with their limitations and applications. Included here are the works of Kamb (1959), Hertweck & Krückeberg (1962, 1963), Robinson (1963), Turner & Weiss (1963), Nobel & Eberly (1964), Spencer & Clabaugh (1967), Adler *et al.* (1968), Warner (1969), LaFountain (1970), Sander (1970) and Burger (1972). Contouring techniques have been extensively described by Harbaugh & Merriam (1968), while statistical considerations have been presented by Scheidegger (1965), Watson (1970) and Kohlbeck & Scheidegger (1977). Fabric diagram construction by computer has been optimized in terms of economy, reproducibility and accuracy by Kalkani & von Frese (1976, 1979, 1980) using a technique which unifies the construction completely as a function of reference hemisphere geometry.

The method of Kalkani & von Frese (1979, 1980) considers the fabric elements in the light of their impingement points on the gridded surface of the reference hemisphere. The population densities are determined by centering a spherical counting area on all fabric impingement points and counting the number of times a grid position is covered by the counting tool. Using a spherical counting area which is equal to 1% of the area of the reference hemisphere, the population density value (%) at a spherical grid point is given by $100(n/N)$, where N is the number of fabric elements considered and n is the number of times the grid point was covered by the counting tool. Contours of equal population density are then determined on the surface of the reference sphere and projected in an equal-angle or Lambert equal-area mode as the fabric diagram.

The calculation of population densities by the method of Kalkani & von Frese (1979, 1980) is equivalent to convolving the impingement points of the fabric data with a convolution operator or function defined by a spherical disk of unit thickness and area. The spherical unit disk operator is but one of many convolution operators, however, which can be designed to investigate a variety of fabric data features.

In the present paper, the filtering properties of the spherical unit disk operator are compared with spherical triangle, sinc, sinc^2 and Gaussian operators. For each case, the convolution is performed directly on the surface of the reference hemisphere using fabric data taken from the lower Schoonover sequence of the Northern Independence Mountains, Nevada. Contours of equal density of the convolved function, which represent the frequency distribution of the original data, are produced with variations that are related to the extent and nature of the convolving function. The location of maximum values in clusters of high population density for the filtered data is determined with an accuracy which depends on the fineness of the grid employed. The concept can be expanded to determine probability or statistical boundaries including percentages of the population.

ANALYTICAL CONSIDERATIONS

The density of fabric elements per unit solid angle on the surface of the reference hemisphere is a physical quantity which is sampled by recording the attitude of each fabric element. The distribution of directions in space assumed on a lattice in the θ-dip, ϕ-strike domain on the surface of the reference hemisphere forms the two-dimensional sample function $f_s(\theta, \phi)$ which is a close representation of the true distribution function $f(\theta, \phi)$. The true distribution can be reconstructed with considerable accuracy by simple interpolation between grid

points. In cases of sparsely sampled functions, however, the reconstruction is not as accurate, and more sophisticated mathematical techniques have to be used to determine $f(\theta, \phi)$.

A common approach in one-dimensional space to recover the original function from the sampled one is convolution. To extend the technique to the two-dimensional space of the surface of the reference hemisphere, the N data points forming the data function $F_d(\theta, \phi)$ are regarded as unit impulses, (i.e. delta functions $\delta(\theta - \theta_d, \phi - \phi_d)$, convolved with an appropriate function $g(\eta, \xi)$, called a filter, to obtain sample values $f_s(\theta, \phi)$ at the lattice points. Accordingly, the data function is given by

$$F_d(\theta, \phi) = \sum_{d=1}^{N} \delta(\theta - \theta_d, \phi - \phi_d), \quad (1)$$

and the filtered output by

$$f_s(\theta, \phi) = F_d * g = \int_0^{2\pi} \int_0^{R_0} F_d(\theta - \eta, \phi - \xi)\, g(\eta, \xi)\, d\eta\, d\xi, \quad (2)$$

where the θ and ϕ spherical coordinates correspond to dip and azimuth on the surface of the reference hemisphere of radius $R = 1$, and η, ξ are local spherical coordinates used to integrate over the range $0 \leq \xi \leq 2\pi$ and $0 \leq \eta \leq R_0$, where R_0 is a spherical radius calculated below. To compute the integral, the filter $g(\eta, \xi)$ is centered at consecutive data points θ_d, ϕ_d on the surface of the reference hemisphere. The products $F_d(\theta_d, \phi_d) \cdot g(\eta_g, \xi_g)$ are summed up for each grid point θ_g, ϕ_g of a spherical lattice on the reference hemisphere within the base of $g(\eta, \xi)$. This procedure is repeated for each fabric element until the entire data space has been covered. Hence, the convolution integral is the sum of products of corresponding values of $F_d(\theta, \phi)$ and $g(\eta, \xi)$ where the filter has been centered successively on each of the fabric data points.

The usual procedure with a sample of N_0 data points is to consider an operator with its range of variables within a spherical area a equal to a fraction $1/K$ of the area of the reference hemisphere A. The spherical area a along with the spherical radius R_0 can be computed from the following equations (see also Kalkani & von Frese 1976, 1979):

$$A = 2\pi R^2, \quad (3)$$

$$a = (1/K)A, \quad \text{(where } K > 1, \text{ usually } K = N_0), \quad (4)$$

$$a = 2\pi R^2 / N_0, \quad (5)$$

$$a = \pi(2R \sin(R_0/2))^2 \quad \text{and} \quad (6)$$

$$R_0 = 2 \sin^{-1}(\sqrt{1/2N_0}). \quad (7)$$

For the case of N observations such that $N \neq N_0$ and $K \neq N_0$, the convolution operator has to be normalized for comparison with the standard diagram that is constructed using $N_0 = 100$ and the spherical unit disk operator. The normalization includes multiplication of the convolving function by the ratio N_0/N to account for $N \neq N_0$, and multiplication by the ratio $(R_0/R_k)^2$, where R_k is the spherical radius for the area $a = (1/K)A$. The convolving function is also adjusted so that its volume, which corresponds to a solid of revolution about its centerline (normal to the surface of the reference hemisphere), is equal to the volume V_0 generated by the spherical unit disk operator. Introducing the above normalization, the resulting function $f_s(\theta, \phi)$ at the grid point θ_g, ϕ_g for successive positions of the convolving function $g(\eta, \xi)$ will be:

$$f_s(\theta_g, \phi_g) = \sum_{i=1}^{N} (1/N)(N_0/N)(R_0/R_k)^2 g(\eta, \xi)_i. \quad (8)$$

The sum of the products $f_s(\theta_g, \phi_g)$, times the area around each grid point within the θ, ϕ lattice, corresponds to the total volume V_t, which is equal to V_0 because of the normalization over N observations.

The convolving function $g(\theta, \phi)$ must be symmetrical about its centreline at $\eta = 0$, $\xi = 0$, because the convolution process is assumed not only in two directions, but in all possible directions on the spherical surface. To simplify the mathematical descriptions, the function $g(\eta, \xi)$ is considered in cross-section in a two-dimensional (x, y) coordinate system and expressed as $y(x)$, where x varies from 0 to 1 as indicated in Fig. 1, and $y(x)$ forms a solid of revolution around the y-axis. The volume of this solid is V_0 as previously described. According to the Pappus–Guldinus theorem, the solid of revolution formed by a planar area revolved about an axis in its plane which does not intersect the area, is equal to the product of the area times the length of the path traced by the centroid of the area. Accordingly, the volume V_0 is computed as follows:

$$V_0 = 2\pi \int_0^1 R \sin(x/R)\, y(x) dx, \quad (9)$$

where R is the radius of the reference hemisphere and $y(x)$ is the function for the convolution operator. These elements along with the differential area $y(x)dx$ and radius of revolution $R \sin(x/R)$ are illustrated in Fig. 1.

The five operators chosen for the filtering function $y(x)$

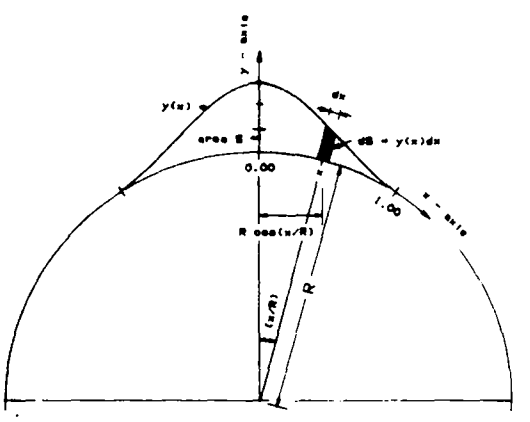

Fig. 1. Schematic diagram indicating the convolving function $y(x)$ and the characteristic dimensions necessary to calculate the volume of the solid of revolution for the shaded area around the y-axis.

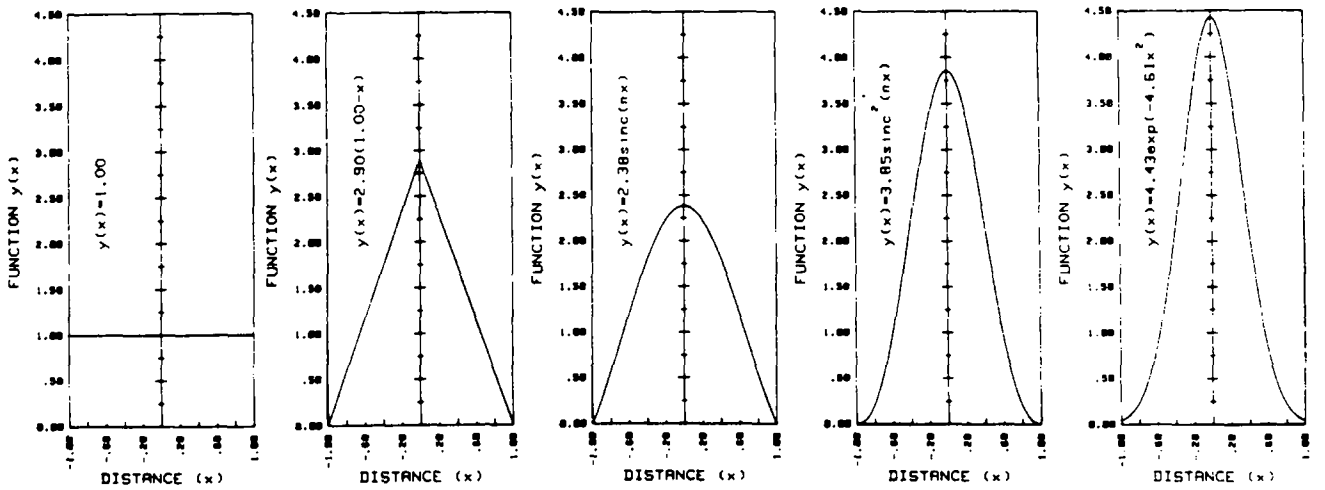

Fig. 2. Cross-sections for the five spherical operators $y(x)$ (unit disk, triangle, sinc, sinc^2, and Gaussian) used in the convolution process. The volumes of the resultant solids of revolution around the centerline are equal to the volume of the solid of revolution generated by the spherical unit disk operator.

Table 1. Expressions of the convolving function $y(x)$, values of coefficients C_n, revolving areas S, revolving radii r and revolution volumes V_0, for x ranging from 0 to 1

n	$y(x)$	C_n	S	r	V_0
1	C_1	1.00000	1.00000	0.45969	2.88837
2	$C_2\,(1.00 - x)$	2.89964	1.44982	0.31707	2.88837
3	$C_3\,\mathrm{sinc}(\pi x)$	2.38387	1.40524	0.32713	2.88837
4	$C_4\,\mathrm{sinc}^2(\pi x)$	3.85139	1.73856	0.26441	2.88837
5	$C_5\,e^{-4.605x^2}$	4.42582	1.82337	0.25211	2.88837

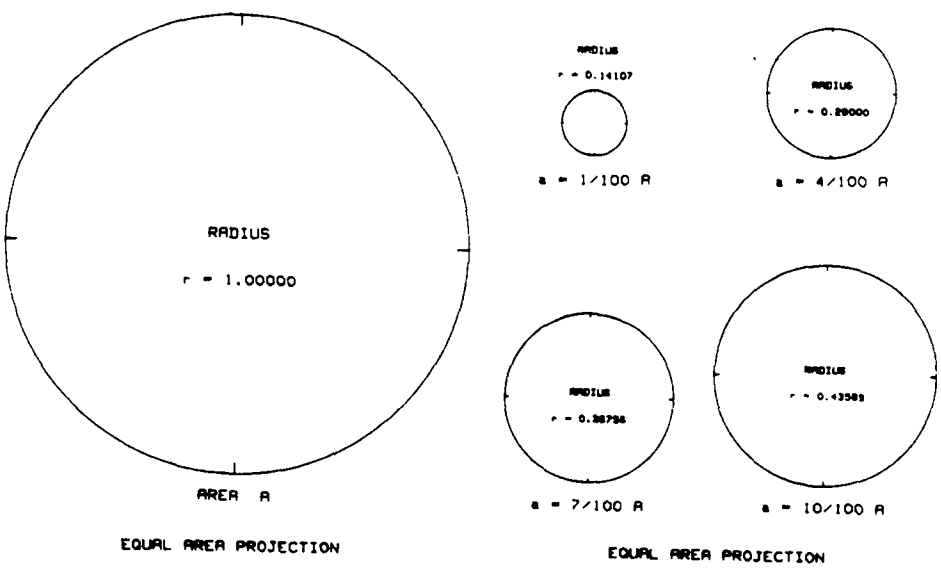

Fig. 3. Equal area projection of the area of the reference hemisphere (left) and areas of influence used for each of the filtering functions (right), as fractions of the area of the reference hemisphere.

in this study are illustrated in cross-section in Fig. 2, along with their algebraic expressions. The expressions of these functions also are given in Table 1, along with the coefficients required to produce constant volumes of revolution equal to V_0 for x in the range 0 to 1, the revolved areas S, and the radii of revolution r. The values of $y(x)$ are introduced into the convolution process by setting $x = R_r/R_k$, where R_r is the spherical distance of the position of datum point θ_d, ϕ_d from the grid point θ_g, ϕ_g and R_k is the spherical radius of the maximum direction of the convolving function for $a = (1/K)A$. Four values of $1/K$ are assumed in this study, namely $1/K = 1/100$, 4/100, 7/100 and 10/100. To facilitate comparisons, the equal-area projection of these areas and their respective radii are shown in Fig. 3, along with a projection of the area of the reference hemisphere. The expressions for R_r and R_k are:

$$R_r = \cos^{-1}[\cos\theta_d \cos\theta_g + \sin\theta_d \sin\theta_g \cos(\phi_d - \phi_g)], \quad (10)$$

$$R_k = 2\sin^{-1}\sqrt{(1/2K)}. \quad (11)$$

The values of $y(x)$ are divided by the number of observations N and normalized for N_0 and K as described previously, so that the total volume V_t of the function $f_s(\theta, \phi)$, produced on the reference hemisphere, will be equal to V_0.

APPLICATION OF CONVOLUTION

Two fabric data sets, taken from the lower Schoonover sequence of the Northern Independence Mountains, Nevada, are investigated to demonstrate the filtering properties of the different convolution operators. The fabric data sets are illustrated in Fig. 4, where the left diagram gives the poles of the normals to the axial planes of the folds and the right diagram shows the positions of the fold axes. For the poles to the axial planes of the folds, the number of observations is $N = 120$, whereas for the positions of the fold axes $N = 150$.

The convolution process was performed using the filtering functions previously described and indicated in Table 1. The results of the convolution process are shown as density contours for all five filtering functions with $a = (1/100)A$ and $a = (4/100)A$ in Fig. 5 and $a = (7/100)A$ and $a = (10/100)A$ in Fig. 6. In each of the figures, the left two panels of diagrams correspond to the poles for the axial planes of the folds, and the right two panels to the positions of the fold axes. The areas of influence range from $(1/100)A$ to $(10/100)A$. The lower limit corresponds to $a = (1/100)$ for the usual fabric diagram of $N = 100$, where a single observation will produce a density equal to one per (1/100) of the area of the reference hemisphere, whereas $a = (10/100)A$ is an upper limit above which the variation of radius of influence increases slowly. The equal intervals between the areas of influence indicate the effect of the convolution process in defining the position of the maximum distribution value. The contour values in Figs. 5 and 6 indicate the loci of equal density, which correspond to the number of observations per unit area of the reference hemisphere. For the cases considered here, the density contours in all diagrams are normalized to $N = 100$ observations and to the unit area $a = (1/100)$ of the area of the reference hemisphere. Also, the total volumes of the convolution operators have been normalized to the volume of the solid of revolution of the spherical unit disk operator.

A computer program was developed to perform fabric data convolution which is available from the authors on request. For Figs. 5 and 6, the convolution values were calculated at each grid point of an (18 × 72) array over the surface of the reference hemisphere (18 dip-elements and 72 azimuth-elements, which corresponds to a grid interval

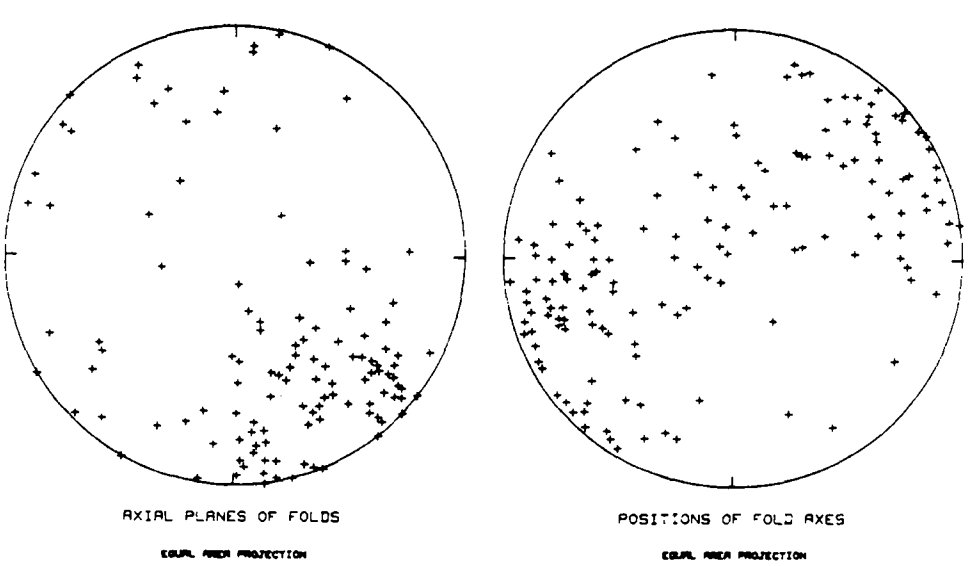

Fig. 4. Equal area projections of two fabric data sets which indicate, for the lower Schoonover sequence of Northern Independence Mountains, Nevada (courtesy of Dr. Russ Dyer, Stanford), the poles to the axial planes of the folds (left) and the positions of the fold axes (right).

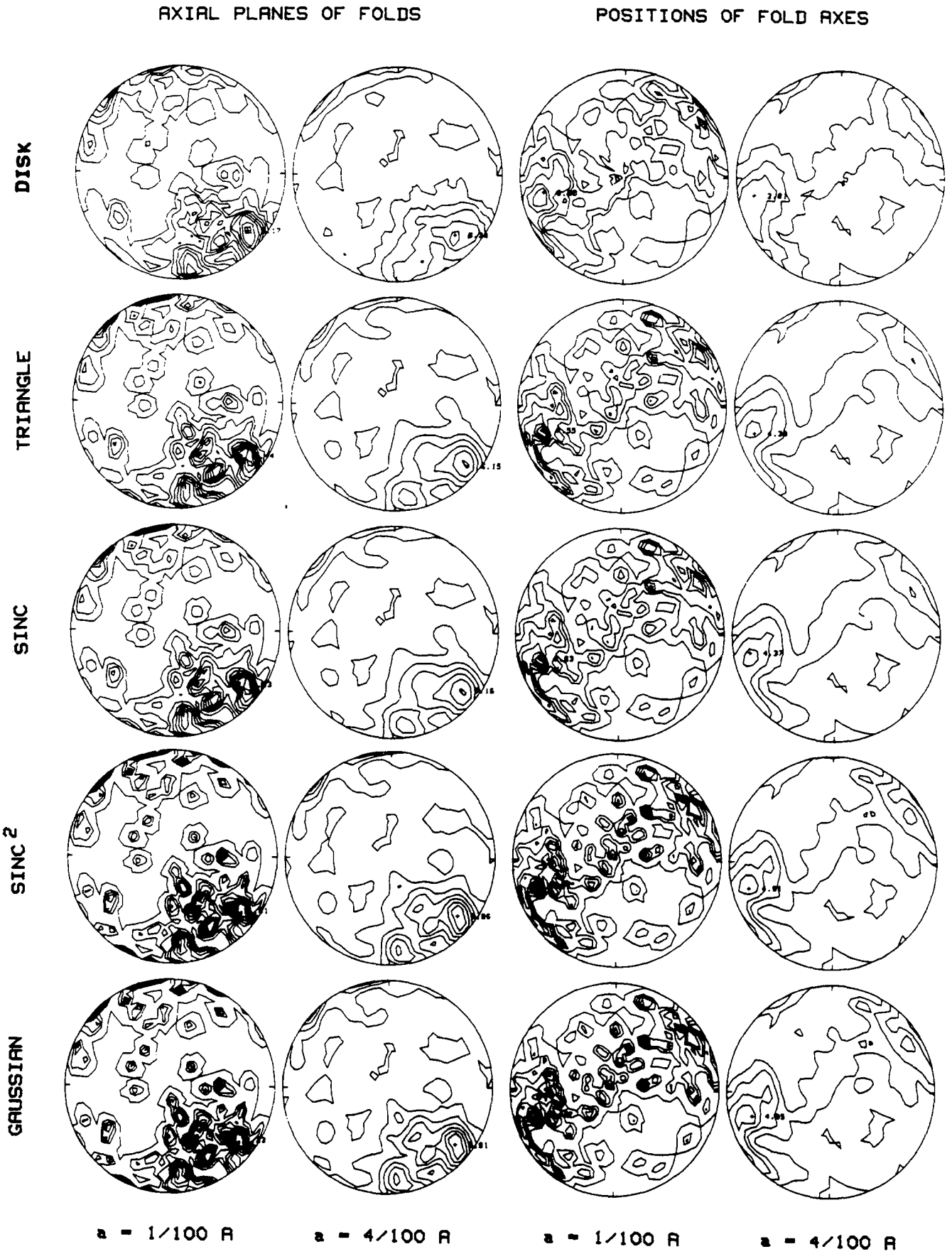

Fig. 5. Density contours at 1% contour interval of the two data sets of Fig. 4, for all filtering functions indicated in Fig. 2, with areas of influence a equal to 1/100 and 4/100 of the area of the reference hemisphere A.

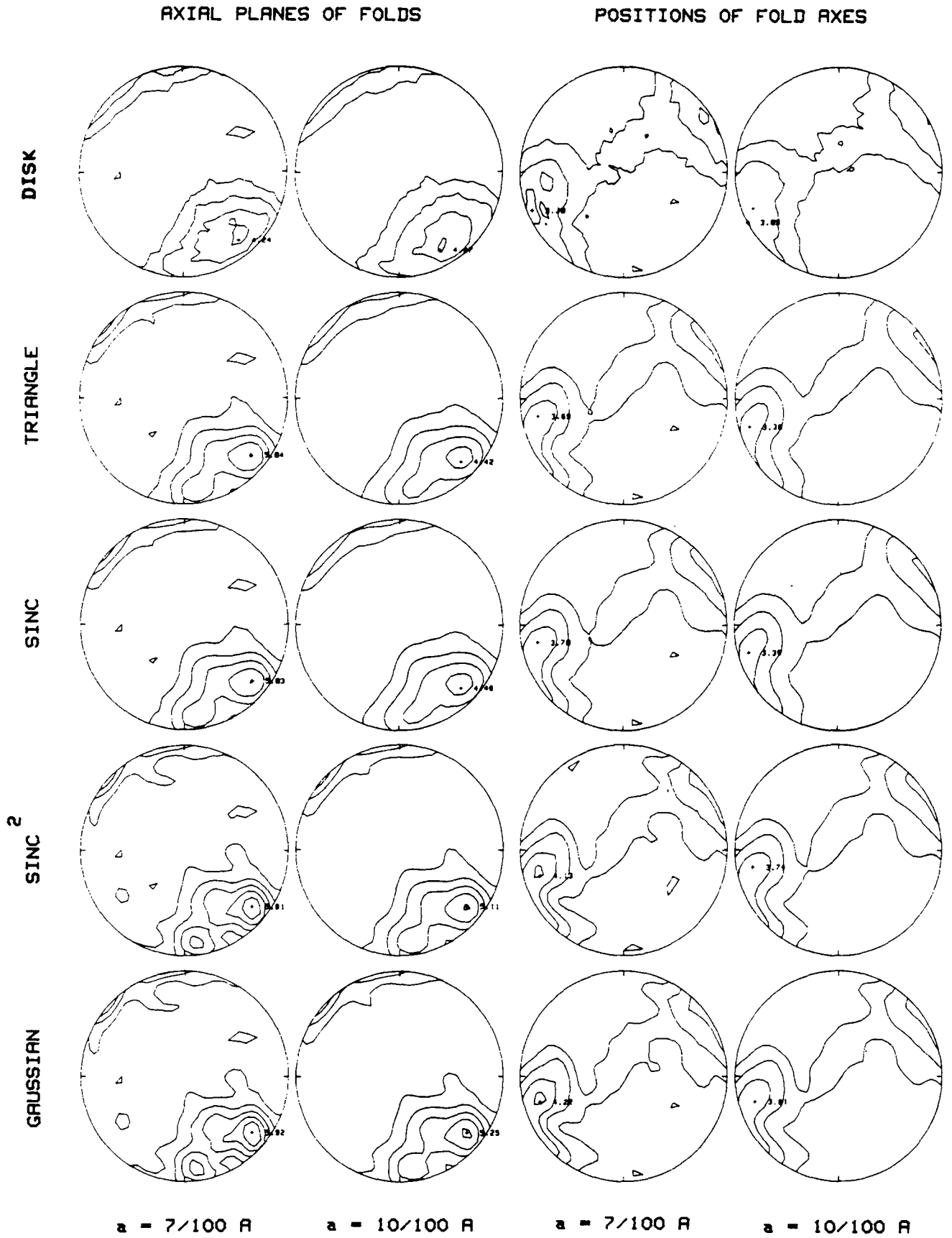

Fig. 6. Density contours at 1% contour interval of the two data sets of Fig. 4, for all filtering functions indicated in Fig. 2, with areas of influence a equal to 7/100 and 10/100 of the area of the reference hemisphere A.

of 5°). The density contours were produced on the surface of the reference hemisphere and projected onto the meridial plane in an equal-area mode. The maximum grid value encountered was printed and its position also indicated on each diagram. The position and magnitude of the maximum value depends on the filtering function used and its area of influence, as well as the fineness of the grid.

Comparison of the diagrams in Figs. 5 and 6 show a number of interesting results which include the following:

(1) Application of the spherical unit disk operator tends to average the positions of the mean, median and mode of the distribution, and indicates the position of the maximum at the centre of the distribution.

(2) Application of the spherical triangle and sinc operators yield similar fabric diagrams, with weak concentrations of contours indicating two or more clusters in the distribution.

(3) Application of the sinc^2 and Gaussian operators give similar fabric diagrams, with distinct concentrations of contours indicating two or more clusters in the distribution.

(4) Increasing the area of influence for all filtering functions lowers the high frequency cut-off, and removes some of the smaller, and perhaps distracting details of the density distribution of the fabric data.

POPULATION PARAMETERS

Due to the normalization process, the volume of the density diagram in each density drawing is constant, and equal to the volume of revolution of the spherical unit disk function for $a = (1/100)A$. The peak or peaks of the individual clusters of the distribution vary in position and magnitude as the filtering function varies, and as the area of influence of the filtering function increases. For this part of the study, only areas of influence equal to $a = (7/100)A$ and $a = (10/100)A$ are considered, because the resulting density distributions are closer to a Gaussian distribution of the form $y(x) = \alpha \exp(-\beta x^2)$. Inferences about the probability distribution of the population will be drawn from the distribution of the sample, which in our case is sufficiently large.

The distribution of fabric data, as is the case for most natural distributions, follows a binomial distribution, which at the limit where the number of observations N becomes large, approaches a normal distribution. The resultant distributions for the two fabric data sets are compared to a normal or Gaussian distribution with maximum at $x = 0.00$ and 1/100 of the maximum at $x = \pm 1.0$. The volume of the solid of revolution for the Gaussian curve is calculated and the relative elevation contours are chosen to include 10, 20, ..., 90% of the population for each cluster. The values of the resulting distributions in Fig. 6 can be contoured at levels that are calculated as the product of a normalization factor (see Appendix) times the values of the normal curve for which the included percentages of the observations are predetermined at 10% contour intervals. Accordingly, the resulting contour levels will correspond to loci of points which include 10, 20, ..., 90% of the population for each cluster. These are here called statistical probability contours and are created on the surface of the reference hemisphere and, subsequently, projected onto the meridial plane.

Statistical probability contours, which correspond to loci of points from 90% of the population (peripheral contour), to 10% of the population (contour closest to the position of the maximum) are shown in Figs. 7 and 8. These results are contoured at 10% intervals for all filtering functions and areas of influence equal to 7/100 and 10/100 of the area of the reference hemisphere. The diagrams of Fig. 7 correspond to statistical contours for the cluster with the largest peak as recorded on the density diagrams of Fig. 6, while the statistical contours of Fig. 8 are normalized to the cluster with the second largest peak on the density diagrams of Fig. 6.

Comparison of Fig. 7 and Fig. 8 shows a number of results that principally include the following.

(1) The spherical unit disk operator gives a single maximum derived from averaging the positions of the median, mode and mean of the distribution. Hence, the statistical probability contours that correspond to this cluster of the density distribution are valid for the total set of the observations.

(2) The spherical triangle and sinc operators each indicate two maxima of the population, which are more pronounced for the sinc^2 and the Gaussian operators. For these cases, the statistical probability contours between clusters are overlapping, while at the peripheries they represent true probabilities around the cluster for which the statistical contours are drawn.

CONCLUSIONS

Application of convolution to fabric data on the surface of the reference hemisphere produces density diagrams with high frequencies attenuated, and density distributions that depend on the convolving or filtering function as well as the area of influence of the function on the surface of the reference hemisphere. Use of the spherical unit disk function produces a maximum at the centre of the distribution and statistical probability contours which are symmetrical about the central peak. The spherical triangle and sinc functions, on the other hand, yield similar diagrams, while the sinc^2 and the Gaussian functions give population density diagrams with increased resolution of the multimodal components of the fabric data. For these cases, second order clusters of the fabric data can be identified, as well as statistical contours for each individual cluster which indicate the probability of occurrence of fabric data, where no overlapping of the clusters is present.

For geotechnical applications, the convolution process applied to fabric data with appropriate filter operators

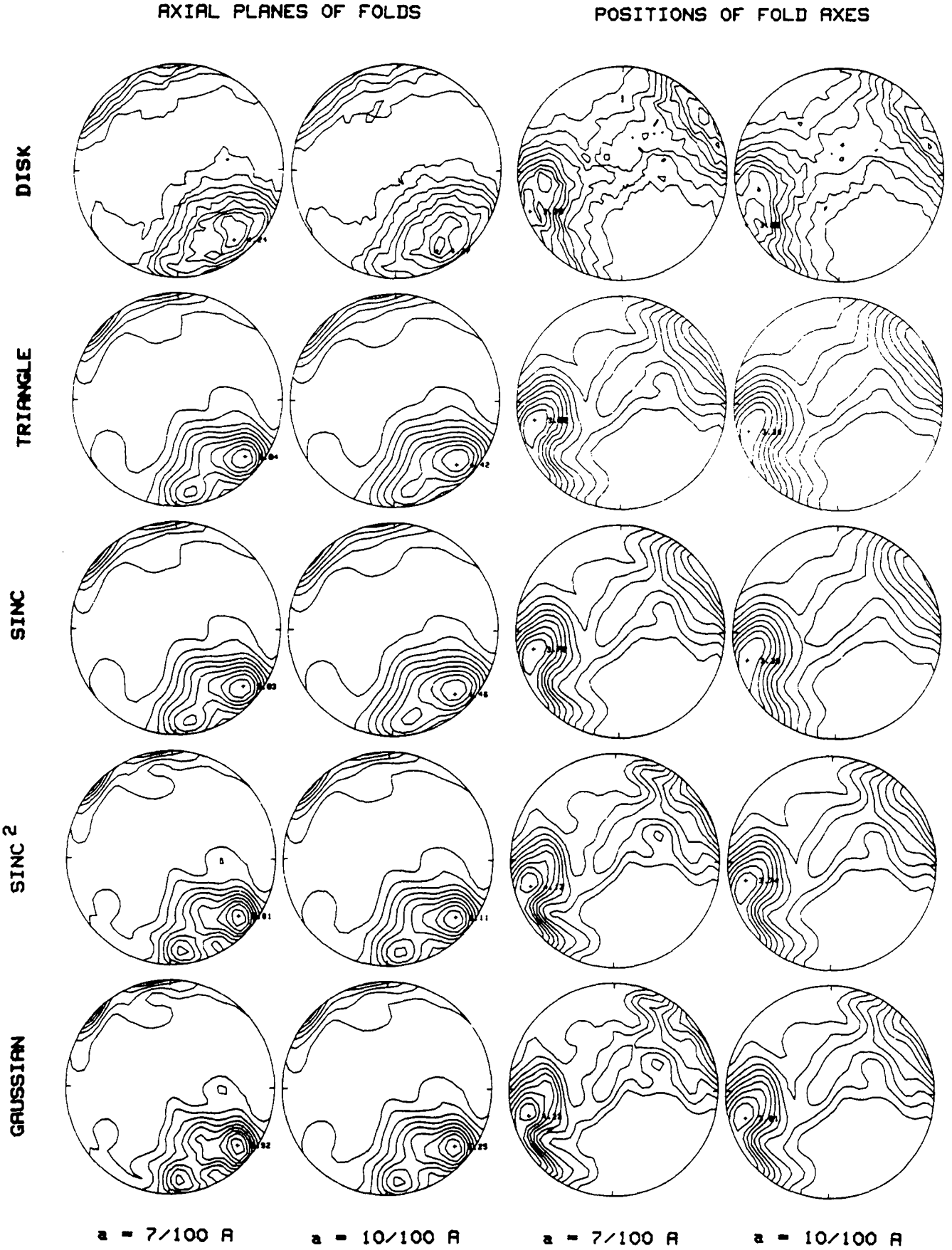

Fig. 7. Statistical contours at 10% probability interval of the two data sets of Fig. 4, for all filtering functions indicated in Fig. 2, with areas of influence a equal to 7/100 and 10/100 of the area of the reference hemisphere A, after normalizing to the highest cluster peak value (highest cluster maximum is indicated).

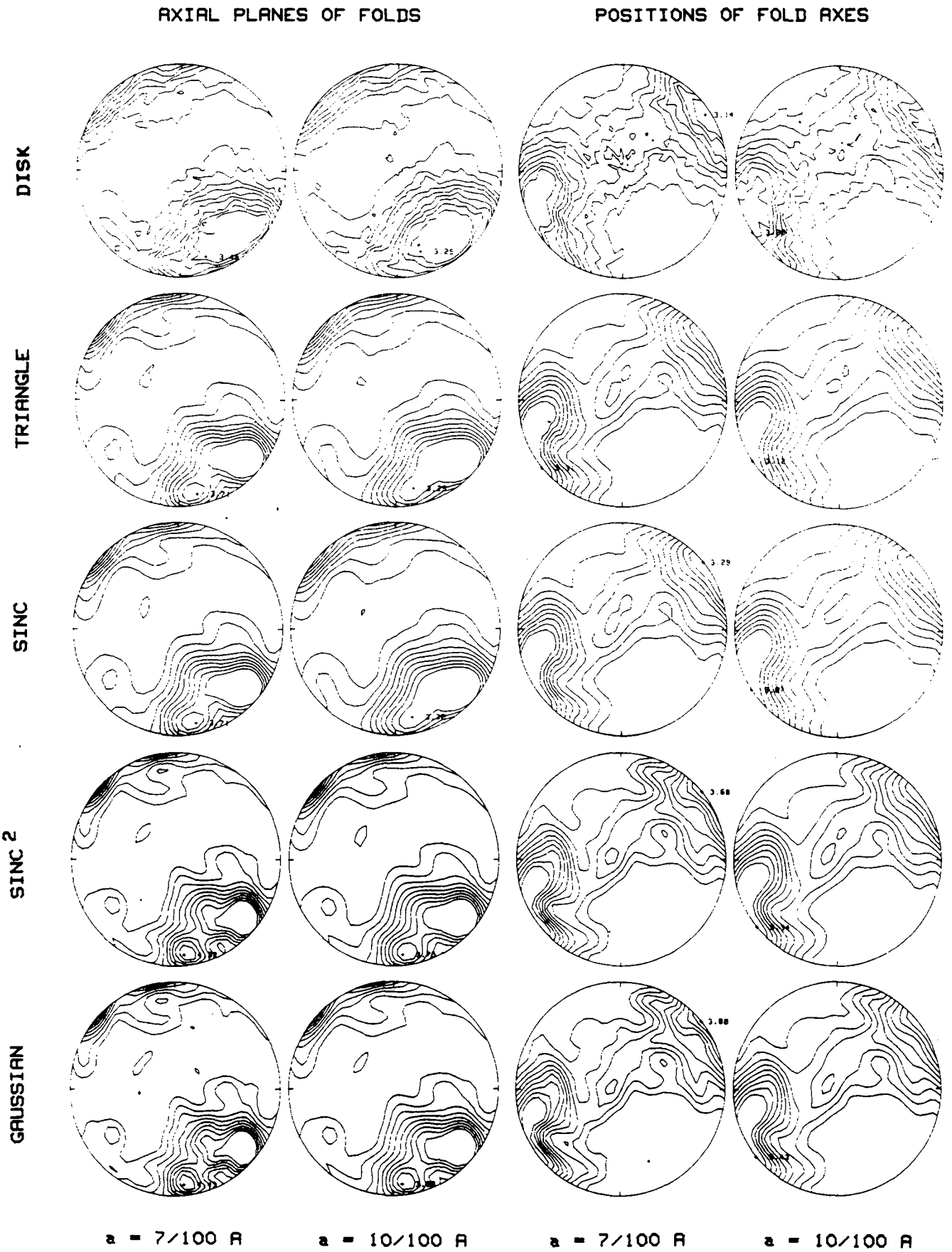

Fig. 8. Statistical contours at 10% probability interval of the two data sets of Fig. 4, for all filtering functions indicated in Fig. 2, with areas of influence a equal to 7/100 and 10/100 of the area of the reference hemisphere A, after normalizing to the second highest cluster peak value (second highest cluster maximum is indicated).

can provide an objective picture of the relative statistics for the data population. These procedures, for example, can be used by the design engineer to locate areas of the fabric diagram containing 66 or 90% of the data population, as well as to statistically differentiate between one or more clusters of the fabric data. The convolution process, in general, is a very effective approach for investigating the multimodal characteristics of fabric data populations, because appropriate filter operators frequently can be specified and designed with little difficulty. Accordingly, this approach has widespread application in kinematic and dynamic analyses of fabric data, where convolution with appropriately designed operators can be used to focus upon the higher order, multimodal behaviour of fabric data populations.

REFERENCES

Adler, R. E., Krückeberg, F., Pfisterer, W., Pilger, A. & Schmidt, M. W. 1968. *Elektronische Datenverarbeitung in der Tektonik.* Clausthaler Tektonische Hefte 8, Inst. Techn. Univ. Clausthal, Zellerfeld, W. Germany.

Burger, H. R. 1972. Computerized solution for calculating calcite compression and tension axes. *Bull. geol. Soc. Am.* **83**, 2439–2442.

Harbaugh, J. W. & Merriam, D. F. 1968. *Computer Applications in Stratigraphic Analysis.* John Wiley, New York.

Hertweck, G. & Krückeberg, F. 1962. Die statistische Auszahlung von Gefügediagrammen durch elektronische Rechenanlagen. *Anz. öst. Akad. Wiss.* **9**, 1–132.

Hertweck, G. & Krückeberg, F. 1963. Die Behandlung von Gefügediagrammen durch elektronische Rechenanlagen, *Neues Jb. Geol. Paläont. Mh.* 86–98.

Kalkani, E. C. & von Frese, R. R. B. 1976. A comparison of fabric diagrams in terms of reproducibility, accuracy and economy. *Bull. Ass. Engng Geol.* **13**, 297–313.

Kalkani, E. C. & von Frese, R. R. B. 1979. An efficient construction of equal-area fabric diagrams. *Comput. Geosci.* **5**, 301–311.

Kalkani, E. C. & von Frese, R. R. B. 1980. Computer construction of equal-angle fabric diagrams and program comparisons. *Comput. Geosci.* **6**, 279–288.

Kamb, W. B. 1959. Ice petrofabric observations from Blue Glacier, Washington, in relation to theory and experiment. *J. geophys. Res.* **64**, 1891–1909.

Kohlbeck, F. & Scheidegger, A. E. 1977. On the theory of the evaluation of joint orientation measurements. *Rock Mech.* **9**, 9–25.

LaFountain, L. J. 1970. Plotted and point contoured stereograms by computer, X–Y plotter, or microfilm devices. *Bull. geol. Soc. Am.* **81**, 1267–1272.

Nobel, D. C. & Eberly, S. W. 1964. A digital computer procedure for preparing Beta diagrams. *Am. J. Sci.* **262**, 1124–1129.

Robinson, P. 1963. Preparation of Beta diagrams in structural geology by digital computer. *Am. J. Sci.* **261**, 913–928.

Sander, B. 1970. *An Introduction to the Study of Fabric of Geologic Bodies.* Pergamon Press, New York.

Scheidegger, A. E. 1965. On the statistics of the orientation of bedding planes, grain axes, and similar sedimentological data. *Prof. Pap. U.S. Geol. Surv.* **525-C**, C164–C167.

Spencer, A. B. & Clabaugh, P. S. 1967. Computer program for fabric diagrams. *Am. J. Sci.* **265**, 166–172.

Turner, F. J. & Weiss, L. E. 1963. *Structural Analysis of Metamorphic Tectonites.* McGraw-Hill, New York.

Warner, J. 1969. FORTRAN IV program for the construction of PI diagrams with the UNIVAC 1108 computer. *Kansas Geol. Surv. Comput. Contrib.* **33**, 1–38.

Watson, G. S. 1970. Orientation statistics in the earth sciences. *Bull. geol. Instn Univ. Upsala* N.S. **2**, 73–89.

APPENDIX

1. *Theoretical curve*

The normal or Gaussian curve is of the form

$$y(x) = \alpha e^{-\beta x^2}, \tag{1}$$

with total volume of revolution V_0 for the range $x = 0$ to ∞, and partial volume V_1 for the range $x = 0$ to x_1 according to the equations:

$$V_0 = \int_0^{\infty} 2\pi x y(x)dx = \int_0^{\infty} 2\pi x \alpha e^{-\beta x^2} dx = \frac{-\pi\alpha}{\beta} e^{-\beta x^2}\Big|_0^{\infty} = \frac{\pi\alpha}{\beta}, \tag{2}$$

$$V_1 = \int_0^{x_1} 2\pi x y(x)dx = -\frac{\pi\alpha}{\beta} e^{-\beta x^2}\Big|_0^{x_1} = \frac{\pi\alpha}{\beta}(1 - e^{-\beta x_1^2}). \tag{3}$$

The differential volume of the solid of revolution can also be expressed in terms of the angle θ and the surface $y(x)\,dx$ integrated over $x = 0$ to ∞. For practical purposes, integration to infinity on the surface of the reference hemisphere is truncated at the distance where $y(x)$ becomes equal to 1/100 of the maximum value $y(x = 0)$. The differential volume for each increment $d\theta$ integrated over (2π) gives the total volume V_0 and partial volume V_1 as

$$V_0 = \int_0^{2\pi}\left(\int_0^{\infty} x\alpha e^{-\beta x^2} dx\right)d\theta, \tag{4}$$

$$V_1 = \int_0^{2\pi}\left(\int_0^{x_1} x\alpha e^{-\beta x^2} dx\right)d\theta, \tag{5}$$

which, after reducing the integrals, are equal to the values of equations (2) and (3).

2. *True distribution*

In the measured data we observe that the total volume of the distribution may consist of one or more clusters. We assume for each cluster normal distributions with cluster overlap within the distribution. If the normal curve of each cluster i is of the form

$$y(x) = \gamma e^{-\delta x^2}, \tag{6}$$

with γ and δ corresponding to the ith cluster, the following equations hold, for the total volume $(V_0)_i$ for the range of $x = 0$ to ∞ and partial volume $(V_1)_i$ for the range $x = 0$ to x_2, in terms of the equation (6)

$$(V_0)_i = \int_0^{2\pi}\left(\int_0^{\infty} x\gamma e^{-\delta x^2} dx\right)d\theta, \tag{7}$$

$$(V_1)_i = \int_0^{2\pi}\left(\int_0^{x_2} x\gamma e^{-\delta x^2} dx\right)d\theta. \tag{8}$$

The values of the total and partial volume in terms of the proportionality constant m_i are

$$(V_0)_i = m_i V_0, \tag{9}$$

$$(V_1)_i = m_i V_1. \tag{10}$$

The value of γ is constant and is the maximum value of the cluster, while the value δ is a function of θ and defines the shape of the distribution. This means that in plan the distribution is not a circle, but a wavy line that can be approximated by a circle. The expressions for $(V_0)_i$ and $(V_1)_i$ can be written after integration over x as

$$(V_0)_i = \int_0^{2\pi} \frac{\gamma}{2\delta} d\theta, \tag{11}$$

$$(V_1)_i = \int_0^{2\pi} \frac{\gamma}{2\delta}(1 - e^{-\delta x_2^2})d\theta = \int_0^{2\pi} \frac{\gamma}{2\delta} d\theta - \int_0^{2\pi} \frac{\gamma}{2\delta} e^{-\delta x_2^2} d\theta. \tag{12}$$

3. *Calculation of probability contours*

Introducing

$$y_1 = \alpha e^{-\beta x_1^2}, \tag{13}$$

$$y_2 = \gamma e^{-\delta x_2^2}, \tag{14}$$

as the contour heights of the specific volume V_1, where subscripts for x and y are (1) and (2) corresponding to the theoretical curve and the measured distribution respectively, and assuming the variation of δ with angle θ is small, or otherwise assuming δ constant and integrating, we can express equations (11) and (12) as

$$(V_0)_i = \frac{\pi\gamma}{\delta}, \tag{15}$$

$$(V_1)_i = \frac{\pi\gamma}{\delta} - \frac{\pi}{\delta} y_2, \tag{16}$$

and rewrite equation (3) as

$$V_1 = \frac{\pi\alpha}{\beta} - \frac{\pi}{\beta} y_1. \tag{17}$$

Using equations (9) and (10), the systems of equations {(15), (16)} and {(2), (17)} can be written as

$$\frac{\gamma}{\delta} = m_i \frac{\alpha}{\beta}, \tag{18}$$

$$\frac{\gamma}{\delta} - \frac{1}{\delta} y_2 = m_i \frac{\alpha}{\beta} - m_i \frac{1}{\beta} y_1, \tag{19}$$

so that by eliminating m_i we get

$$y_2 = \frac{\gamma}{\alpha} y_1. \tag{20}$$

In our case α is the maximum value of the convolving normal curve, γ is the maximum of the ith cluster of the resulting distribution and y_1 is the contour for which the included volume is equal to V_1. The values y_1 for the convolving normal function have been calculated for 10, 20, ..., 90% of the total volume, which correspond to the probability of occurrence of observations within the contours. Equation (20) indicates that y_2 is calculated in terms of y_1, where γ and α are known (i.e. y_1 and α are computed from the theoretical normal distribution, and γ is the maximum of the individual cluster).

The ratio γ/α is the normalization factor applied to produce the contours of Figs. 7 and 8 for the highest and second highest cluster peaks, respectively.