Mathematical Geology, Vol. 4, No. 3, 1972

Mathematical Techniques for Paleocurrent Analysis: Treatment of Directional Data[1]

J. S. Rao[2] and Supriya Sengupta[3]

Statistical procedures for (1) sampling, (2) testing the existence of a preferred direction, and (3) testing homogeneity of two-dimensional directional data, which have been developed by the authors for paleocurrent studies, are presented. It is well known that conventional methods of statistical analysis are not applicable to directional data (e.g., crossbedding and ripple-mark directions, grain lineations, etc.) which are "circularly distributed" on a compass dial. A sampling technique for directional data has been developed using the circular measures of dispersion and approximate ANOVA of G. S. Watson. On the basis of a pilot survey, it is possible to compute the minimum sample size required for estimating, with a desired precision, the mean paleocurrent direction of a formation. The optimum allocation of sample size between and within outcrops also can be accomplished at a minimum cost. The procedure described for testing uniformity (or lack of preferred direction) is based on the arc lengths made by successive sample points and is simple to use if the sample size is moderate. A table of critical points and a numerical example are given after a description of the test procedure. Finally, the procedures for testing the homogeneity of directional data from several geological formations are described by (1) tests for equality of the resultant directions (polar vectors) and (2) tests for equality of dispersions. With these tests it is possible to determine whether the paleocurrent directions from different geological formations belong to significantly different populations.
KEY WORDS: directional data analysis, new statistical tests, sampling, statistics, orientation data, paleocurrent analysis, sedimentology.

INTRODUCTION

The two well-known methods for the determination of paleocurrent are (1) measurement of the directional or vectorial properties of the sediment and (2) mapping of the scalar properties which exhibit a systematic variation in the direction of sediment transport. This paper is concerned with the techniques of handling the directional data only. Following is a list of the directional properties of sediments which provide useful clues to paleocurrents (Pettijohn, 1962):

---

[1] Manuscript received 25 November 1971; research sponsored by the Air Force Office of Scientific Research, AFSC, USAF, under Grant No. AFOSR 71-2009. The United States Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.
[2] Department of Mathematics, Indiana University (USA).
[3] Geological Studies Unit, Indian Statistical Institute (India).

Planar structures
Crossbedding of different types and different inclinations
Linear structures
Striation and groove casts
Flute casts
Grain lineation
Fossil lineation
Parting lineation
Ripple marks
Rib-and-furrow
Complex structures
Convolute bedding

All the directional properties listed can be treated as vectors because they have direction as well as magnitude. Direction of linear structures is given by the attitude or orientation of the property concerned. With planar elements as crossbedding, direction is given in three dimensions by the azimuth and inclination of the foreset. Vector magnitude, for planar as well as linear features, can be determined arbitrarily by assigning unit weight to each observation (Steinmetz, 1962).

The importance of these properties as clues to paleocurrents is known. However, in using one or more of these vectorial properties for paleocurrent determination on a regional scale, one is faced with some procedural problems relating to the collection, summarization, and interpretation of data. Efficient handling of these problems requires statistical methods of analysis. Some problems have been discussed by Pettijohn (1962, p. 1448–1490).

Some graphic as well as mathematical methods for sampling and summarization of data were studied as early as 1938 by Reiche. A comprehensive review of the early works is given by Pettijohn (1962) and Potter and Pettijohn (1963). These include, among others, the pioneering efforts of Olson and Potter (1954) and Raup and Miesch (1957).

It must be emphasized here that vectorial data similar to those listed, which are spread circularly on a compass dial, pose special statistical problems. Although the directions can be measured as angles with respect to some arbitrary origin, the arithmetic mean of these values fails to provide a representative measure of the mean direction, and the usual standard deviation of measurements cannot be applied as a measure of dispersion for such data. Under some special conditions where the spread of the observations on the circumference is restricted, the circle may be cut open at the other end to get a line, and the circular distribution may satisfactorily be approximated by a linear normal distribution [e.g., Agterberg and Briggs (1963) claim this can be done if the angular data does not exceed 57° on either side of the mean vector]. We do not know how good artificial linearization is, and the field data in most practical situations does exceed the limits.

The inadequacy of conventional statistical measures (arithmetic mean, standard deviation, etc.) in the analysis of circularly distributed vectorial data (also called directional data or orientation data) having an arbitrary point of origin, was outlined by Jizba in 1953 and Chayes in 1954. The problem of development of adequate statistical techniques for directional data received the enthusiastic attention of several workers since then (Pincus, 1956; Curray, 1956; Durand and Greenwood, 1958; Watson, 1956, 1966; Watson and Irving, 1957). A general review of the important publications on this subject is given in Miller and Kahn (1962). Unfortunately, however, in spite of these pioneering efforts inappropriate statistical measures have been or are being utilized. Although recognizing the inadequacy of arithmetic means and variances, some authors have continued to use conventional analysis of variance, whereas others have used conventional statistical tests, such as Student's t, as a test of homogeneity of directional data. Seemingly, statisticians have failed to communicate their findings in a manner readily understandable by geologists.

Attempts have been made by the authors during the last few years to critically examine the available statistical techniques for sampling as well as for testing the homogeneity of circularly distributed directional data. In some situations, where the conventional techniques have proved inadequate, efforts have been made to develop new procedures for the treatment of directional data. The purpose of this paper is to give an account of these statistical techniques in a form readily usable by geologists. Although illustrated with the help of the crossbedding data, these techniques are universally applicable to any form of directional (vectorial) data.

SAMPLING OF DATA

In a formation with a large number of outcrops, where each outcrop contains a profuse amount of directional features of a particular type (crossbedding, ripple marks, grain lineation, or any other), one is faced with the problem of the number of measurements necessary to estimate the mean direction. Clearly the answer will depend on several factors, for instance, the precision with which one wants to estimate the mean direction as well as on the amount of dispersion within the formation. In other words, the question is, what is the minimum number of observations which would give the mean direction with a specified precision for the formation, that is, a mean for which the confidence limits are set in advance by the geologist? It is also important to have an idea about the allocation of samples, i.e., the optimum number of

Table 1. Value of Concentration Parameter κ_0 Required for γ_N to Attain Given Precision

Semiangle of confidence, deg	Confidence level, $(1-\alpha)$	Concentration value required, κ_0
	0.90	354.6292
5	0.95	504.0610
	0.99	870.6682
	0.90	88.7598
10	0.95	126.1125
	0.99	217.9236
	0.90	22.1772
20	0.95	31.5221
	0.99	54.4496

observations which should be collected from an exposure and how many such exposures should be sampled in a locality.

The techniques used so far for hierarchical or multistage sampling of crossbedding foreset azimuths are based on the conventional analysis of variance. However, we know that the classical method of analysis of variance cannot be indiscriminately applied for the analysis of circularly distributed directional data.

This problem has been discussed by Rao and Sengupta (1970), who have developed an optimum hierarchical sampling technique for crossbedding data, using the circular measures of dispersion and the approximate ANOVA for circularly distributed data (Watson, 1956, 1966). The sampling problems solved for the crossbedding data are (1) the minimum sample size required for estimating, with a desired precision, the mean direction of a formation and (2) the optimum allocation of samples between and within the outcrops that would allow efficient sampling at minimum cost. Solutions have been provided for estimation with the semiangles of confidence of 5, 10, and 20°, at confidence levels of 0.90, 0.95, and 0.99 for each situation (Table 1).

The following is a summary of sampling procedures outlined by Rao and Sengupta (1970). The method given here will be valid for any directional variables used in paleocurrent work.

(a) Before the actual sampling is undertaken, it is necessary to conduct a pilot survey for a small number of representative samples of the directional element for the formation concerned. An equal number of observations from each outcrop facilitates computation. The computational procedure adopted assumes that within the i^{th} outcrop the observations ϕ_{ij} have a circular normal distribution (CND) with a mean direction $(\gamma+\zeta_i)$ and a concentration parameter ω. The ζ_i's have a CND with mean direction zero and concentration β so that the overall formation mean is γ. Suppose we visited n outcrops and took m observations from each of them, making a total sample of size $N = mn$ for the formation. Let ϕ_{ij} denote the j^{th} observation from the i^{th} outcrop ($j = 1, \ldots, m$; $i = 1, \ldots, n$).

(b) Sine and cosine values are computed for each direction ($\bar{\phi}_{ij}$) measured. The length of the resultant for each outcrop is obtained as follows:

$$R^2_i = \left(\sum_{j=1}^{m} \cos \phi_{ij}\right)^2 + \left(\sum_{i=1}^{m} \sin \phi_{ij}\right)^2$$
$$= C^2_i + S^2_i$$

where R_i is the outcrop resultant for the i^{th} outcrop, and m is the number of observations within each outcrop.

(c) The overall resultant R for all outcrops is given by

$$R^2 = \left(\sum_{i=1}^{n} C_i\right)^2 + \left(\sum_{i=1}^{n} S_i\right)^2$$

where n is the number of outcrops surveyed, and C_i and S_i are as defined in step (b).

(d) The ANOVA for the directional data is computed, where $\hat{\omega}$ and $\hat{\beta}$, the estimates for within outcrop and between outcrop concentration parameters, are obtained by equating columns (4) and (5) of Table 2.

(e) The optimum number of observations m^* to be taken at an outcrop is obtained from the relation

$$m^* = \sqrt{C_1 \cdot \hat{\beta} / C_2 \cdot \hat{\omega}}$$

where C_1 and C_2 are the costs for reaching an outcrop and taking an observation within an outcrop, respectively. The geologist should have a rough idea of the relative cost (C_1/C_2), say as 10:1 or 20:1.

Table 2. ANOVA Table for Circular Data

Source of variation (1)	df (2)	*SS* (3)	*MS* (4)	*E(MS)* (5)
Between outcrops	$n-1$	$\sum_1^n R_i - R$	$(\Sigma R_i - R)/(n-1)$	$\frac{1}{2}\left(\frac{1}{\omega}+\frac{m}{\beta}\right)$
Within outcrops	$\overset{n}{\underset{1}{N}}-n$	$N-\Sigma R_i$	$(N\text{-}\Sigma R_i)/(N-n)$	$\frac{1}{2\omega}$
Total	$N-1$	$N-R$	...	...

(f) The optimum number of outcrops to be sampled n^* is given by the equation:

$$n^* = \kappa_0[(1/\hat{\beta})+(1/m\cdot\hat{\omega})]$$

where κ_0 is the concentration for a desired confidence level $(1-\alpha)$ and the semiangle of confidence (ψ_0) and is obtained from Table 1. The geologist should decide how much precision he needs, i.e., ψ_0 and at what confidence level, and look in Table 1 for κ_0.

An illustration of the application of this sampling technique based on actual field data of crossbedding azimuths from the Kamthi Formation near Bhimaram (Bheemaram), India, has been given by Rao and Sengupta (1970).

SUMMARIZATION OF DATA

This section briefly touches on methods of obtaining summary measures for the data and their graphic presentation.

Computation of Resultant Direction

For the directional data circularly distributed on a compass dial on either side of true north (360°), it is obvious that the usual method of arithmetic averaging leads to erroneous conclusions (e.g., arithmetic mean of 20° and 340° is 180°). It is accepted that a meaningful measure of average in the examples of these directions is given by the direction of the vector resultant of the sample, treating each observation as a unit vector with components cos α_i, sin α_i. That is, corresponding to the sample $\alpha_i, \ldots, \alpha_n$ we compute

$$V = \sum_1^n \cos \alpha_i \qquad W = \sum_1^n \sin \alpha_i$$

and take

$$\gamma = \tan^{-1} (W/V)$$

as the sample mean direction. Where grouping of data cannot be avoided, we compute the mean direction in a similar fashion, i.e., compute

$$V = \sum_{i=1}^n n_i \cos x_i \qquad W = \sum_{i=1}^n n_i \sin x_i$$

$$\hat{\gamma} = \tan^{-1} (W/V)$$

where x_i is the midpoint azimuth of the i^{th} class interval, n_i is the number of observations in the i^{th} class, and $\hat{\gamma}$ is the azimuth of the resultant vector. The quadrant in which this $\hat{\gamma}$ lies is determined by the signs of V and W. We also may mention that grouping of data should be avoided wherever possible. One can then draw better conclusions as the ungrouped data are more precise.

Computation of Dispersion

Because the usual measure of dispersion, e.g., the standard deviation is not applicable to directional data, an alternative measure of dispersion is required. Variability or scatter within a sector is represented by the magnitude of length of the resultant vector R, where $R = \sqrt{W^2+V^2}$. A useful measure of concentration of azimuths is the consistency ratio (R/n of Reiche, 1938, p. 913), expressed in terms of percent, i.e., $L = (R/n)\times 100$. L has been termed "vector magnitude" by Curray (1956) and "vector strength" by Pincus (1956). Equivalently (for distributions which are unimodal), the quantity $(n-R)$ provides an excellent measure of dispersion of the sample directions; this is large if the observations are widely scattered and small if they are consistent.

Graphical Presentation of Data

The observed directions within an area can be graphically represented in the form of a rose diagram (a circular histogram). The resultant direction (vector resultant) is usually represented by an arrow at the center of the diagram, and the length of the arrow is made proportional to the vector strength. The two-dimensional moving-average method of representation of vector resultant directions of crossbedding has been used by Potter (1955) and Pelletier (1958). Moving averages emphasize the major trends of sediment transport by smoothing the local variations. This method has been recommended by Potter and Pettijohn (1963, p. 274), who also have suggested different types of maps for presentation of directional data.

SIMPLE TEST FOR UNIFORMITY

Testing uniformity or lack of a preferred direction in the observed data is an important first step in analyzing the directional data. If there is no significantly preferred direction, there is little use in computing the mean direction, or in any further tests on such a mean direction. Several tests for uniformity are available and a discussion of these tests along with a comparison of their large-sample efficiencies may be found in Rao (1969, in press). A simple test for uniformity which is useful for moderate-sample sizes, has been introduced by the senior author and is described here.

Suppose $\alpha_1, \ldots, \alpha_n$ are n directions in two dimensions measured say in angles from 0 to 360°. If these observations are symmetrically scattered around the circumference of the circle, i.e., equispaced on the circumference, this can be considered as evidence in favor of uniformity. On the other hand, if these observations tend to cluster in one or more directions, this may be considered as evidence in favor of a preferred direction (or directions). Using this simple idea, we will construct a test of uniformity based on the sample arc lengths i.e., the distances between the successive observations on the circumference, say $T_1, \ldots, T_n$. The n observations clearly divide the circumference into n arcs; let $T_1, \ldots, T_n$ be the lengths of these n arcs. These sample arc lengths $T_1, \ldots, T_n$ are easy to compute. Let the observations $\alpha_1, \ldots, \alpha_n$ be numerically arranged, and let us call the smallest of the lot $\alpha_1{}^*$, the second smallest $\alpha_2{}^*$, etc., and the largest $\alpha_n{}^*$. The arc lengths are given by

$$T_1 = \alpha_2{}^* - \alpha_1{}^*,\ T_2 = \alpha_3{}^* - \alpha_2{}^*, \ldots, T_n = \alpha_1{}^* - \alpha_n{}^* + 360$$

It is simple to understand this (especially the definition of T_n), if the observations are represented as points on the circumference of a unit circle.

Under the uniformity hypothesis, the expected length of an arc is $(360/n)^\circ$ because there are n observations to the 360° of the circumference. The test consists in comparing each of the observed arc lengths $T_1, \ldots, T_n$ with $(360/_n)$. The proposed test statistic which is one such measure of discrepancy between $(T_1, \ldots, T_n)$ and $(360/n)$ is half the sum of absolute derivations

$$\begin{aligned} U_n &= \tfrac{1}{2}\sum_{i=1}^{n} |T_i - (360/n)| \\ &= \tfrac{1}{2}[|T_1 - (360/n)| + \ldots + |T_n - (360/n)|] \end{aligned}$$

Clearly from what we said earlier, smaller values of U_n indicate agreement with the hypothesis of uniformity or lack of preferred direction. On the other hand, if U_n is too large as indicated by the table of critical points, there is reason to reject the hypothesis of uniformity and conclude that there is indeed a preferred direction. Under the hypothesis of uniformity the density function of U_n, say $f_n(u)$, is given in Rao (1969):

$$\begin{aligned} f_n(u) &= (n-1)!\sum_{j=1}^{n-1} \binom{n}{j}(u/2\pi)^{n-j-1}\ \{\phi_j(nu)/[(n-j-1)!\ n^{j-1}]\} \\ &\qquad\qquad \text{for } 0 \leqslant u \leqslant 2\pi[1-(1/n)] \\ &= 0 \qquad \text{otherwise} \end{aligned}$$

where $\phi_j(x)$ is the density function of the sum of j independent uniform random variables on $[0, 2\pi]$ and has the expression

$$\phi_j(x) = [1/2\pi\cdot(j-1)!]\sum_{k=0}^{\infty}(-1)^k\binom{j}{k}\ \langle (x/2\pi) - k\rangle^{j-1}$$

with the notation $\langle x\rangle = x$ if $x > 0$ and $\langle x\rangle = 0$ if $x \leqslant 0$. The following table (Table 3) gives the critical points of U_n for sample sizes $n = 2(1)20$ and for three levels of significance $\alpha = 0.01$, 0.05, and 0.10. If for a given sample size n and level α, the calculated value of U_n exceeds the tabulated critical point

Table 3. Critical points $U_0(\alpha,n)$ (in Degrees) for Statistics U_n

n \ α	0.01	0.05	0.10
2	178.20	171.00	162.00
3	219.24	193.68	174.24
4	221.04	186.48	171.72
5	212.04	183.60	168.84
6	206.04	180.72	166.32
7	202.68	177.84	164.88
8	198.36	175.68	163.44
9	195.12	173.52	162.36
10	192.24	172.08	161.28
11	189.72	170.28	160.20
12	187.56	169.20	159.48
13	185.76	167.76	158.40
14	183.96	166.68	157.68
15	182.16	165.60	156.96
16	180.72	164.88	156.60
17	179.64	164.16	155.88
18	178.20	163.08	155.16
19	177.12	162.36	154.80
20	176.04	161.64	154.44

$U_0(\alpha,n)$, we reject the hypothesis of uniformity. The critical points have been given in terms of degrees for ready applicability.

Example: we give here an example of the following crossbedding azimuths that were observed in a particular outcrop as 20, 35, 350, 120, 85, 345, 80, 320, 280, and 85°.

It is required to know whether these azimuths indicate a preferred direction of paleocurrent. The arc lengths $\{T_i\}$ made by these observations on the circle are easily seen to be 15, 45, 5, 0, 35, 160, 40, 25, 5, and 30° and the fixed arcs are of length 360/10 = 36° in this example. Therefore

$$U_{10} = (\tfrac{1}{2})\sum_{i=1}^{10}|T_i - 36|$$

$$= 137°$$

This value of 137°, for $n = 10$, is not significant even at the 10-percent level of significance as the critical point in this example is only 161.28°. Therefore, we conclude that the observations could have come from a uniform distribution.

TESTS FOR HOMOGENEITY OF DIRECTIONAL DATA

The problem of comparison of directional data belonging to two or more populations has led to many interesting discussions, because Student's t and similar conventional tests are not valid for data having circular distribution (Court, 1952; Potter and Pettijohn, 1963; Krumbein and Graybill, 1965; Watson, 1966). A standard test for comparing the mean directions of several circular normal populations with the same concentration parameter, can be constructed from Table 1 (see, for instance, Watson, 1966). The statistic

$$F = [(\sum_1^n R_i - R)/(N - \sum_1^n R_i)] \qquad [(N-n)/(n-1)]$$

which follows an F distribution with $(n-1)$ and $(N-n)$ degrees of freedom, tests the equality of mean directions of the n populations.

A problem of this type, originating from actual field data, was presented by Sengupta and Rao (1966) and Sengupta (1970). The existing tests for directional data were found unsuitable for comparison of the crossbedding foreset dip directions belonging to three different members of the Kamthi Formation near Bhimaram because the observations showed a wide divergence from circular normality and also the three formations had significantly different concentrations. We wished to know if the crossbedding dip directions observed in the three different members (lower, middle, and upper) of the fluviatile Kamthi Formation belong to three significantly different populations. In other words, did the direction of sediment transport significantly change with time during Kamthi sedimentation? Visual comparison of the data was not enough, because the shift in the resultant directions of the three members was small, and spreads of the total data of crossbedding azimuths for the three Kamthi members were overlapping.

Two large sample homogeneity tests or H tests were proposed by the senior author for testing the equality of polar directions and the equality of dispersions of the directional data (Rao, *in* Sengupta and Rao, 1966; Rao, 1969). These tests *do not* assume any specific circular normal distribution for the observations and are generally valid provided the samples are not too small. Besides, the large sample test for equality of mean directions can be applied although the populations have different concentrations. Through further studies, the large-sample efficiency of Rao's H test for testing equality of mean directions was shown to be equal to Watson's F test; if both the tests can be applied for a given data. Efficiencies of some of the existing tests for uniformity also have been analyzed and compared (Rao, 1969). The practical procedures for applying the homogeneity tests in the example of paleocurrent (e.g., crossbedding) data are given. Reference may be made to Rao (1969) for detailed discussions of the theory of these tests and for other statistical methods developed by him.

TEST FOR EQUALITY OF MEAN DIRECTIONS (POLAR VECTORS)

Suppose our problem is to test for equality of polar vectors of k populations of angular variables (e.g., crossbedding azimuths). Suppose a sample of size n_i is taken from the i^{th} population. Let ϕ_{ij} stand for the j^{th} observation in the i^{th} sample ($j = 1, \ldots, n_i$; $i = 1, \ldots, k$). Let us denote the cosine and sine components of ϕ_{ij} by x_{ij} and y_{ij}, i.e., $x_{ij} = \cos \phi_{ij}$ and $y_{ij} = \sin \phi_{ij}$. The following steps describe the test. (a) Sine and cosine values (x_{ij} and y_{ij}) are computed for each direction ϕ_{ij} measured.

(b) The *means* of cosine values and sine values are computed for the i^{th} sample. Let x_i and y_i denote the means of cosine values and sine values, respectively, for the sample of size n_i from the i^{th} population, i.e.,

$$x_i = \sum_{j=1}^{n_i} x_{ij}/n_i \qquad y_i = \sum_{j=1}^{n_i} y_{ij}/n_i$$

(c) The sample variances of the cosine and sine values [$S(cc^i)$ and $S(ss^i)$, respectively], and the sample covariance between the cosine and sine values [$S(cs^i)$] from the i^{th} sample are computed as follows:

$$S(cc^i) = \sum_{j=1}^{n_i} (x_{ij}-x_i)^2/(n_i-1)$$

$$S(ss^i) = \sum_{i=1}^{n_i} (y_{ij}-y_i)^2/(n_i-1)$$

$$S(cs) = \sum_{j=1}^{n_i} (x_{ij}-x_i)\,(y_{ij}-y_i)/(n_i-1)$$

(d) Let the population mean direction in the i^{th} population be represented by γ_i. Then a consistent estimator of $\tan \gamma_i$ is given by

$$T_i = y_i/x_i$$

(e) The estimated variance of T_i, say S_i^2, is given by

$$S_i^2 = \frac{1}{n_1}\left\{\frac{S_{SS}^{(i)}}{x_i^2} + \frac{y_i^2\, S_{CC}^{(i)}}{x_i^4} - \frac{2y_i\, S_{CS}^{(i)}}{x_i^3}\right\}$$

(f) Let us consider the hypothesis

$$H_0\colon \tan \gamma_1 = \tan \gamma_2 = \ldots = \tan \gamma_k$$

The following H statistic (see Rao, 1965) can be used to test the hypothesis H_0 or equivalently the homogeneity of the T values. Compute

$$H = \left\{\sum_1^k \frac{T_i^2}{S_i^2} - \left(\sum_1^k \frac{T_i}{S_i^2}\right)^2\right\} \Big/ \left(\sum_1^k \frac{1}{S_i^2}\right)$$

(g) Under some general conditions this statistic H has a chi-square distribution with $(k-1)$ degrees of freedom, when the hypothesis H_0 is true. A significant value of H would lead to the rejection of the hypothesis H_0, and we may conclude that the polar vectors are different. A simple remark is in order. Because $\tan \gamma = \tan(\pi+\gamma)$, the hypothesis H_0, as stated, does not distinguish between pole and the antipole. But this is not a drawback of the technique because wide differences in polar vectors can easily be determined by a simple examination of the data.

TEST FOR EQUALITY OF DISPERSIONS

The following procedure is adopted for testing the equality of dispersions of k populations of angular variates. As before let x_i and y_i denote the means of cosine and sine values of the i^{th} sample, respectively, and let $S_{CC}(i)$, $S_{SS}(i)$, and $S_{CS}(i)$ denote the sample variances and covariance of the cosine and sine values.

(a) A measure of concentration (that is, the reciprocal of dispersion) for the i^{th} population is given by

$$U_i = x_i^2 + y_i^2$$

(b) The asymptotic estimated variance of U_i is obtained as follows:

$$S_i^{*2} = 4/n_i \{x_i^2 S_{CC}^{(i)} + y_i^2 \, S_{SS}^{(i)} + 2x_i . y_i \, S_{CS}^{(i)}\}$$

(c) The homogeneity test may be used again to test the homogeneity of $U_1, U_2, \ldots, U_k$ or, in other words, the hypothesis that the concentrations in the k populations are equal. Compute

$$H = \left\{\sum_1^k \frac{U_i^2}{S_i^{*2}} - \left(\sum_1^k \frac{U_i^2}{S_i^{*2}}\right)^2\right\} \Big/ \left(\sum_1^k \frac{1}{S_i^{*2}}\right)$$

H is distributed as χ^2 with $(k-1)$ degrees of freedom under the hypothesis of equality of dispersions.

(d) A significant value of H would lead to the rejection of the hypothesis and would lead to the conclusion that the concentrations (or, equivalently, dispersions) in the various populations are different.

Example: in the illustration given by Sengupta and Rao (1966), the statistic H, if computed separately for T and U values of the crossbedding azimuths from three different members of the Kamthi Formation, gave significant results in both examples showing that the population direction of crossbeddings as well as their dispersions are significantly different in the three Kamthi units. This led to the conclusion that despite repeated oscillations and many local changes in flow direction, the shift in the direction of sedimentation with time in the Kamthi river was a significant one.

ACKNOWLEDGMENTS

We are grateful to Prof. P. C. Mahalanobis and Prof. C. R. Rao for their kind interest and encouragement in this work and to Dr. J. K. Ghosh for going through an earlier version of the manuscript. This work has developed from a program of study of the Gondwana deposits of the Pranhita–Godavari Valley undertaken by the Geological Studies Unit of the Indian Statistical Institute at the initiative and encouragement of Dr. P. L. Robinson.

REFERENCES

Agterberg, F. P., and Briggs, G., 1963, Statistical analysis of ripple marks in Atokan and Desmoinesian rocks in the Arkoma Basin of east-central Oklahoma: Jour. Sed. Pet. v. 33, no. 2, p. 393–410.

Chayes, F., 1954, Effect of change of origin on mean and variance of two-dimensional fabrics: Amer. Jour. Sci. v. 252, no. 9, p. 567–570.

Court, A., 1952, Some new statistical techniques in geophysics, *in* Advances in geophysics, 1: Academic Press Inc., New York, p. 45–85.

Curray, J. R., 1956, Analysis of two-dimensional orientation data: Jour. Geology, v. 64, no. 2, p. 117–131.

Durand, D., and Greenwood, J. A., 1958, Modifications of the Rayleigh test for uniformity in analysis of two-dimensional orientation data: Jour. Geology, v. 66, no. 3, p. 229–238.

Jizba, Z. V., 1953, Mean and standard deviation of certain geologic data—a discussion: Amer. Jour. Sci. v. 251, no. 12, p. 899-906.

Krumbein, W. C., and Graybill, F. A., 1965, Introduction to statistical models in geology: McGraw-Hill Book Co., New York, 475 p.

Miller, R. L., and Kahn, J. S., 1962, Statistical analysis in the geological sciences: John Wiley & Sons, Inc., New York, 483 p.

Olson, J. S., and Potter, P. E., 1954, Variance components of cross-bedding direction in some basal Pennsylvanian sandstones of the Eastern Interior Basin, pt. 1, statistical methods: Jour. Geology, v. 62, no. 1, p. 26–49.

Pelletier, B. R., 1958, Pocono paleocurrents in Pennsylvania and Maryland: Geol. Soc. America Bull. v. 69, no. 8, p. 1033–1064.

Pettijohn, R. J., 1962, Paleocurrents and paleogeography: Am. Assoc. Petroleum Geologists Bull., v. 46, no. 8, p. 1468–1493.

Pincus, H. J., 1956, Some vector and arithmetic operations of two-dimensional orientation variates with applications to geologic data: Jour. Geology, v. 64, no. 6, p. 533–557.

Potter, P. E., 1955, The petrology and origin of the LaFayette gravel, pt. I. Mineralogy and petrology: Jour. Geology, v. 63, no. 1, p. 1–38.

Potter, P. E., and Pettijohn, F. J., 1963, Paleocurrents and basin analysis: Springer-Verlag, Berlin, 296 p.

Rao, C. R., 1965, Linear statistical inference and its applications: John Wiley & Sons, Inc. New York, 522 p.

Rao, J. S., 1969, Some contributions to the analysis of circular data: Unpubl. doctoral dissertation, Indian Statistical Institute, Calcutta, 212 p.

Rao, J. S., in press, Bahadur efficiencies of some tests for uniformity on the circle: Ann. Math. Stat. v. 43.

Rao, J. S., and Sengupta, S., 1970, An optimum hierarchical sampling procedure for cross-bedding data: Jour. Geology, v. 78, no. 5, p. 533–544.

Raup, D. B., and Miesch, A. T., 1957, A new method for obtaining significant average directional measurements in cross-stratification studies: Jour. Sed. Pet. v. 27, no. 3, p. 313–321.

Reiche, P., 1938, An analysis of cross-lamination, the Coconino Sandstone: Jour. Geology, v. 46, no. 7, p. 905–932.

Sengupta, S., 1970, Gondwana sedimentation around Bheemaram (Bhimaram), Pranhita-Godavari Valley, India: Jour. Sed. Pet. v. 40, no. 1, p. 140–170.

Sengupta, S., and Rao, J. S., 1966, Statistical analysis of cross-bedding azimuths from the Kamthi formation around Bheemaram, Pranhita-Godavari Valley, Sankhya: Ind. Jour. Stat. v. 28B, p. 165–174.

Steinmetz, R., 1962, Analysis of vectorial data: Jour. Sed. Pet., v. 32, no. n. 4, p. 801–812.

Watson, G. S., 1956, Analysis of dispersion on a sphere: Royal Astron. Soc. Monthly Notices, Geophy. Supp. 7, p. 153–159.

Watson, G. S., 1966, The statistics of orientation data: Jour. Geology, v. 74, no. 5, pt. 2, p. 786–797.

Watson, G. S., and Irving, E., 1957, Statistical methods in rock magnetism: Royal Astron. Soc. Monthly Notices, Geophy. Supp. 7, p. 289–300.