AN OPTIMUM HIERARCHICAL SAMPLING PROCEDURE
FOR CROSS-BEDDING DATA[1]

J. S. RAO[2] AND SUPRIYA SENGUPTA

Indian Statistical Institute, Calcutta-35, India

ABSTRACT

The commonly used techniques for hierarchical or multistage sampling of cross-bedding foreset azimuths (for paleocurrent study) are based on the conventional analysis of variance. It is now well known that the classical method of analysis of variance (ANOVA), which partitions the sum of squares of the observations, can not be indiscriminately applied for the analysis of circularly distributed directional data. An efficient and economical sampling technique for cross-bedding data has been developed using the circular measures of dispersion and the approximate ANOVA of Watson. The technique is illustrated here with the help of the pilot-survey data of the fluviatile Kamthi formation. The following sampling problems have been solved: (1) the minimum sample size required for estimating, with a desired precision, the mean direction of a formation, (2) the optimum allocation of the samples between and within the outcrops that would allow efficient sampling at minimum cost.

INTRODUCTION

Cross-bedding foreset dip directions are known to provide one of the most dependable clues to the paleocurrent. For efficient and economical sampling of foreset dip directions in a profusely cross-bedded formation, it is essential to know the minimum number of observations which would give a mean direction with specified precision for the whole formation.

An early attempt to provide a suitable, single stage sampling plan for cross-bedding was made by Reiche (1938) who suggested the technique of determination of the "flatness point" in a curve of the "cumulative vector directions" plotted against the number of measurements. At a later date, Potter and Olson (Potter and Olson 1954; Olson and Potter 1954) following a hierarchical sampling plan, measured the variability associated with different levels of sampling (between areas, between outcrops, between beds, and within beds). Their observations in the Pennsylvanian Mansfield formation showed largest variation between outcrops

within townships and least variation within cross-bedded units. Potter and Pettijohn (1963) have also recommended the analysis of variance (ANOVA) for development of an efficient sampling plan.

Raup and Miesch (1957) have suggested a simple field method for determination of the most efficient and economical sampling scheme. Their method is based on Stein's two-stage sampling procedure where the minimum number of measurements required to obtain a "significant average" (an average which lies within previously specified confidence limits) is estimated from the standard deviation of the first fifty measurements.

It has been stressed by a number of earlier authors (e.g., Watson 1966; Sengupta and Rao 1966; Rao 1969) that the conventional statistical techniques are of little use in the analysis of circularly distributed directional data like cross-bedding foreset azimuths. Although the directions can be measured as angles with respect to some arbitrary origin, the arithmetic mean fails to provide a representative measure of the mean for such data, and the usual standard deviation of measurements) using the approximate ANOVA of Watson (1956, method of ANOVA, which partitions the sums of squares of the observations, can not be indiscriminately applied in the analysis

of circularly distributed directional data. These facts put into doubt the sampling plans of the earlier authors (e.g., Potter and Olson 1954; Potter and Pettijohn 1963; Raup and Miesch 1957) which were based on the conventional ANOVA.

It is now well known that the direction of sample resultant provides a reasonable estimate of the population mean direction for data having unimodal circular distribution, and the length of the resultant can be used to obtain an appropriate measure of dispersion. The purpose of the present paper is to develop an efficient sampling technique which takes into consideration the circular measures of dispersion (instead of the usual standard deviation of measurements) using the approximate ANOVA of Watson (1956, 1966). Admittedly, the methods adopted here are approximate, but these approximations can be shown to hold good with a fair degree of accuracy.

Attempts will be made to answer three major questions in the present paper: (1) how many observations should be taken within a geological formation so that the mean direction is reasonably close to the true value of the mean, (2) how should the sampling effort be distributed over two stages (between outcrops and within outcrops) so that the overall estimate has the desired precision, and (3) if cost (or time) considerations could be profitably introduced, what is the procedure which would minimize the cost (or time) involved. The questions posed here are the usual questions faced in sampling and have been handled in several standard texts on the subject. The novelty of the present approach lies essentially in the use of the circular measures of dispersion and the corresponding ANOVA instead of the classical ANOVA based on variances.

GEOLOGY OF THE AREA AND THE PILOT SURVEY

The technique of analysis developed here is illustrated with the help of the data obtained from the Permo-Triassic, fluvial Kamthi formation. Three members of the Kamthi (lower, middle, and upper) were sampled near Bheemaram, in southeast India for this purpose. Depositional environments of these rocks have been discussed in detail elsewhere (Sengupta 1970). In short, the upper and the middle Kamthi sandstones represent the younger and older point-bar and channel-bar deposits of the Kamthi river. A part of the middle Kamthi member also includes the levee deposits adjacent to the main river channel. The lower Kamthi sandstones represent cutoff meander channels within the older floodplain deposits of the river. An earlier statistical analysis of a large sample of the Kamthi data has indicated that the population means of the cross-beddings as well as their dispersions are significantly different in the three Kamthi members (Sengupta and Rao 1966).

For the purpose of the present work, a "pilot survey" was conducted and a set of 280 fresh measurements were taken in about 42 square miles of the Kamthi formation, following a rigid sampling scheme. A minimum number of widely spaced outcrops were measured from the three Kamthi members. Although the choice of an outcrop was partly guided by its accessibility, this was not allowed to be a significant source of bias in the analysis. The number of outcrops measured in each case was broadly proportional to the exposed area of the Kamthi member concerned. Thus, six outcrops were measured from the lower Kamthi member (area: approximately 8 square miles), eight from the middle Kamthi member (area: approximately 12 square miles), and fourteen from the upper Kamthi member (area: approximately 22 square miles). A fixed number of ten measurements of foreset-dip directions were made in each outcrop by random sampling.

The model and method of analysis described below give a general idea of the sampling plan to be adopted under two distinctly different conditions: (1) when the cross-beddings are profuse and the foreset-dip directions are fairly consistent, as in the middle and the upper Kamthi, and (2) when the cross-beddings are less frequent and the dispersions are relatively higher, as in the lower Kamthi.

The method of analysis discussed here (and illustrated with the Kamthi data) is of a general nature and can be utilized by anyone having similar problems involving circular data.

THE MODEL AND THE METHOD OF ANALYSIS

We will consider a hierarchical or multi-stage sampling scheme. For our purposes it is sufficient to consider a situation with only two levels, though the results of this and the next section can easily be extended to the case where there are more levels of classification. Suppose, as in the present case, the first-stage units are called "outcrops" and the second-stage units, "observations." Further, suppose that we have taken n first-stage units or outcrops and m second-stage units or observations from each of these first-stage units. It is, however, not essential to choose the same number of second-stage units from each first-stage unit though this will considerably simplify the computations as well as the expressions involved. Let ϕ_{ij} denote the jth observation from the ith outcrop ($j = 1, \ldots, m; i = 1, \ldots, n$) of the formation. We assume

$$\phi_{ij} = \gamma + \xi_i + \epsilon_{ij}, \qquad (1)$$

where γ = the mean direction for the formation, ξ_i = deviation due to the ith outcrop from the overall (or formation) mean direction, and ϵ_{ij} = error term. The deviations at each level are supposed to average out to the mean of the next higher level. For example, if the number of observations are increased indefinitely in an outcrop, then their resultant direction would approach the particular outcrop mean direction. More precisely, we assume that within the ith outcrop, the observations ϕ_{ij} have a "circular normal distribution" (CND), with mean direction $(\gamma + \xi_i)$ and a concentration parameter ω. It may be recalled that a random angle ϕ, with reference to any arbitrary vector in two dimensions, is said to have a CND with parameters γ and κ if it has the density

$$\frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(\phi - \gamma)}, \qquad 0 \leq \phi < 2\pi. \quad (2)$$

Here γ is the population-mean angle and κ is a measure of concentration, larger values of κ standing for more concentration around the mean direction γ. While ϕ_{ij}'s for any fixed i have a CND with parameters $\gamma + \xi_i$ and ω, the ξ_i's themselves have a CND with the mean direction as the zero direction and a concentration parameter, β. In other words, the different outcrop means have a CND whose mean is the formation mean γ and whose concentration parameter is β. It may be remarked that most of our analysis would remain valid even if the assumptions of circular normality do not hold strictly for the data in question.

In a particular formation in whose mean direction we are interested, suppose we visited n outcrops and took m observations from each of them, thus making a total sample of size $N = mn$ for the whole formation. We then compute the lengths of the resultant $R_i(i = 1, \ldots, n)$ where

$$R_i^2 = \left(\sum_{j=1}^{m} \cos \phi_{ij} \right)^2 + \left(\sum_{j=1}^{m} \sin \phi_{ij} \right)^2 \quad (3)$$

for the n individual outcrops as well as the length of the overall resultant, R, where

$$R^2 = \left(\sum_{i=1}^{n} \sum_{j=1}^{m} \cos \phi_{ij} \right)^2 + \left(\sum_{i=1}^{n} \sum_{j=1}^{m} \sin \phi_{ij} \right)^2 \quad (4)$$

based on all the N observations. Now it is well known (Watson 1966; Rao 1969) that if R is the length of the resultant based on p observations from a CND with concentration parameter κ, then $2\kappa(p - R)$ is approximately distributed as a χ^2 with $(p - 1)$ degrees of freedom if κ is large. Using this fact, one can have the following approximate ANOVA for the angular data (Watson 1966). Here

$$\bar{m} = \frac{1}{n - 1} \left(N - \frac{\sum_{1}^{n} m_i^2}{N} \right) \quad (5)$$

in case a variable number of m_i observations are taken from the ith outcrop. This $\bar{m}$ is a

weighted average of the m_i's, and if $m_i = m$ for all outcrops then $\bar{m} = m$.

From table 1 we can test the significance of between-outcrop variations by using the F-test

$$F = \frac{\left(\sum_1^n R_i - R\right)}{\left(N - \sum_1^n R_i\right)} \cdot \frac{(N - n)}{(n - 1)}, \quad (6)$$

which has a F distribution with $(n - 1)$ and $(N - n)$ degrees of freedom under the usual ANOVA set up or under the so-called Model I of ANOVA. Under Model II, the same technique and computations may be used to estimate the between-outcrop and

again have a CND with a concentration parameter

$$\kappa = \frac{1}{(n\hat{\beta})^{-1} + (N\hat{\omega})^{-1}}. \quad (7)$$

These results for the circular (or two-dimensional) data are completely analogous to those of Watson (1966) and Watson and Irving (1957) for the three-dimensional directions.

ESTIMATING THE FORMATION-MEAN DIRECTION WITH A SPECIFIED PRECISION

Suppose we want to estimate the formation mean γ with a semiangle of confidence

TABLE 1

ANOVA TABLE FOR ANGULAR DATA

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

within-outcrop concentrations—β and ω, respectively—under the assumptions we have made in the beginning of this section. If the test in (6) shows no significant variation between the sites (or outcrops), then the between-outcrop variation may be ignored or, in other words, the concentration parameter β may be taken to be infinity. If, on the other hand, the F-test gives a significant result, then the estimates of β and ω, say $\hat{\beta}$ and $\hat{\omega}$, may be obtained by equating the mean squares (MS) in column 4 of table 1 with the corresponding expectations in column 5.

Now if all the $N = mn$ observations are used to get the overall resultant and its direction $\hat{\gamma}_N$ is used to estimate the mean direction of the formation, then this will

ψ_0 at a confidence level $1 - a$, $(0 < a < 1)$. That is, if $\hat{\gamma}_N$ denotes the direction of the overall resultant based on N observations, we want the estimate $\hat{\gamma}_N$ to lie within the interval $\gamma - \psi_0$ to $\gamma + \psi_0$ in $(1 - a)$ percent of the cases. This would clearly depend on the distribution of $\hat{\gamma}_N$, which from the earlier section is circular normal with concentration given by (7). Thus the concentration parameter κ associated with the distribution of $\hat{\gamma}_N$ would determine the confidence interval for γ. But since in our case the problem is one of finding the sample size required in order to have a confidence interval of specified width and confidence level, we try to determine a value of κ, say κ_0, required for $\hat{\gamma}_N$ to lie in the specified interval with the given probability. Thus we

must first answer the question: How large should κ be such that a semiangle of confidence ψ_0 around γ carries a probability of $(1 - a)$ percent? If κ were less than 7, one can refer to Watson's (1966) chart and read out the required κ. But this range of κ values is too small to be of use in most practical situations of our interest where the κ values often exceed 100. For such large values of κ, the following normal approximation to the CND can profitably be used. If ϕ is a random angle having a CND with parameters γ and κ, then it is known that the distribu-

standard normal distribution (e.g., $\lambda_a = 1.96$ for $a = 5$ percent). Comparing (8) and (9), $\sqrt{\kappa_0}\, \psi_0 = \lambda_a$ or

$$\kappa_0 = \frac{\lambda_a^2}{\psi_0^2}. \qquad (10)$$

We give a short table of values of κ_0 (table 2) required to attain some commonly specified precisions as given by ψ_0 and a.

Thus if $\hat{\gamma}_N$ is to attain the given precision, the number of outcrops n and the number of observations in each, namely, m, should

TABLE 2

VALUE OF κ_0 REQUIRED FOR $\hat{\gamma}_N$ TO
ATTAIN A GIVEN PRECISION

ψ_0 Semiangle of Confidence..........	$(1-a)$ Confidence Level	κ_0 Concentration Value Required
5°........	0.90	354.6292
	0.95	504.0610
	0.99	870.6882
10°........	0.90	88.7598
	0.95	126.1125
	0.99	217.9236
20°........	0.90	22.1772
	0.95	31.5221
	0.99	54.4496

tion of $\phi^* = \sqrt{\kappa}\,(\phi - \gamma)$ approaches a standard normal distribution for large κ. Therefore, for given $a(0 < a < 1)$, we want a κ for which

$$(1 - a) = P_\kappa\,(\gamma - \psi_0 \leq \phi \leq \gamma + \psi_0)$$

$$= P(-\sqrt{\kappa}\,\psi_0 \leq \sqrt{\kappa}\,(\phi - \gamma)$$
$$\leq \sqrt{\kappa}\,\psi_0) \qquad (8)$$

$$= P(|\phi^*| \leq \sqrt{\kappa}\,\psi_0)\,.$$

But since ϕ^* is distributed as $N(0, 1)$, we have

$$P(|\phi^*| \leq \lambda_a) = (1 - a)\,, \qquad (9)$$

where λ_a is the a percent point for the

be such as to give a κ value as shown in table 2. In other words, from (7) m and n should be such as to satisfy the equation

$$\kappa_0 = \frac{1}{(n\hat{\beta})^{-1} + (mn\hat{\omega})^{-1}}, \qquad (11)$$

in order that the overall resultant direction have a specified accuracy, where κ_0 is determined by ψ_0 and a and is obtained from table 2. The equation (11) still does not completely specify what m and n should be. Several sets of m and n values are possible, of which we may choose one depending on the convenience and availability of outcrops. For example, if a maximum number

of only n_0 outcrops are available, then one can obtain the value of m from (11) as

$$m = \frac{1}{\hat{\omega}\left(n_0\kappa_0 - \frac{1}{\hat{\beta}}\right)}.$$

If, on the other hand, a large number of outcrops are available and one wants to take only one observation per outcrop (i.e., $m = 1$), then we need to visit

$$n = \frac{1}{\kappa_0}\left(\frac{1}{\hat{\beta}} + \frac{1}{\hat{\omega}}\right)$$

outcrops. Between these two extremes, one has several choices for m and n, all of which lead to the same precision for the overall resultant direction.

We will now introduce cost considerations into the picture and then try to arrive at an optimum choice of m and n, that is, a choice which minimizes the total cost of the operation. We generally have a rough idea as to the cost (or time) involved in visiting an outcrop and in taking a single observation. Let C_1 be the cost involved in reaching the outcrop (location, identification, and transportation costs) and C_2 be the cost of making an observation within the outcrop. Then the total cost involved in the operation may be expressed as

$$C = C_0 + nC_1 + nmC_2 , \qquad (12)$$

where C_0 is the total overhead cost. Now we might ask for a choice of m and n, which, besides assuring us of a specified precision, minimizes the cost involved in the entire operation. It may be remarked that we do not need to have a precise idea about the costs C_1 and C_2. What is needed for our purposes is only a rough idea about the relative costs, or the ratio C_1/C_2.

It is clear from (7) that the number of outcrops, n, way be expressed as

$$n = \kappa_0\left(\frac{1}{\hat{\beta}} + \frac{1}{m\hat{\omega}}\right). \qquad (13)$$

Substituting this in the cost function (12), we have

$$C = C_0 + n(C_1 + mC_2)$$

$$= C_0 + \kappa_0\left(\frac{1}{\hat{\beta}} + \frac{1}{m\hat{\omega}}\right)\cdot(C_1 + mC_2)$$

$$= \text{terms not containing } m \qquad (14)$$

$$+ \kappa_0\left(\frac{C_1}{m\hat{\omega}} + \frac{mC_2}{\hat{\beta}}\right),$$

which when minimized with respect to m gives

$$m^2 = \frac{C_1\hat{\beta}}{C_2\hat{\omega}}, \quad \text{or} \quad m = \sqrt{\frac{C_1\hat{\beta}}{C_2\hat{\omega}}}. \qquad (15)$$

Thus the optimum number of observations that should be taken from each outcrop is given by (15). Substituting this in (13), we get the corresponding value of n, the optimum number of outcrops. If m and n are chosen this way, the minimum cost associated with this operation (or survey) would be

$$C_{\min} = C_0 + \kappa_0\left(C_1 + \frac{\sqrt{C_1C_2\hat{\beta}}}{\hat{\omega}}\right)$$

$$\cdot\left(\frac{1}{\hat{\beta}} + \sqrt{\frac{C_2}{C_1\hat{\beta}\hat{\omega}}}\right) \qquad (16)$$

$$= C_0 + \kappa_0\left(\sqrt{\frac{C_1}{\hat{\beta}}} + \sqrt{\frac{C_2}{\hat{\omega}}}\right)^2.$$

The optimum value of m given in (15) can also be found graphically by plotting the cost (14) as a function of m. Such graphs can be made for several choices of C_1/C_2 and $\hat{\beta}/\hat{\omega}$.

The expressions for the optimum values of m and n, say m^* and n^*, would, in general, contain a fractional part, in which case it may be rounded off to the next higher integer so that the final estimate would be at least as accurate as required. In fact, if this is done and, say, m_0 and n_0 are the optimum values so arrived at, then $\hat{\gamma}_N$ will have a precision parameter

$$\kappa = \frac{1}{(n_0\hat{\beta})^{-1} + (m_0n_0\hat{\omega})^{-1}}, \qquad (17)$$

and hence the semiangle of confidence for $\hat{\gamma}_N$ will be

$$\psi* = \frac{\lambda_\alpha}{\sqrt{\kappa}}. \qquad (18)$$

We illustrate these ideas and methods in the following section by applying them to a specific situation.

ANALYSIS OF KAMTHI POPULATIONS

Earlier studies (Sengupta and Rao 1966) have shown that the cross-bedding foreset-dip directions of the three Kamthi members belong to three significantly different populations. Thus we have a simple hierarchical

observations from each. The ANOVA table for these $N_U = 140$ observations belonging to the UK is given below. The data and the details of computation are not presented for reasons of brevity. The overall resultant based on all the 140 observations is $R_U = (82.5159, -22.5615)$, with a length $R_U = \sqrt{(82.5159)^2 + (-22.5615)^2} = 85.5447$. The sum of the lengths of individual resultants is

$$\sum_{i=1}^{14} R_{iU} = 91.8074.$$

From this, we form the ANOVA table for the UK data (see table 3). The value of the F-statistic, 1.2593, is insignificant at the 5

TABLE 3

ANOVA TABLE FOR UK DATA

Source	df	SS	MS	E(MS)	F
Between outcrops.........	13	6.2627	0.4817	$\frac{1}{2}\left(\frac{1}{\omega_U}+\frac{10}{\beta_U}\right)$	$F=1.2593$
Within outcrops..........	126	48.1926	0.3825	$\frac{1}{2\omega_U}$	....
Total.................	139	54.4553	....	....	....

classification within each Kamthi member with two levels of sampling in each, namely, outcrops and observations.

As mentioned in the beginning, for the present analysis a total of twenty-eight outcrops were sampled, out of which fourteen are from the upper Kamthi (UK), eight from the middle Kamthi (MK), and the rest from the lower Kamthi (LK). From each outcrop we have taken a total of ten observations at random though, as we remarked earlier, it is not essential to have the same number of observations within each outcrop. However this will facilitate computations.

We shall first deal with the UK, and the analysis of the other two populations proceeds on exactly similar lines. We shall use the subscript U to denote the quantities corresponding to the UK population. We then have $n_U = 14$ outcrops with $m_U = 10$

percent level with 13 and 126 degrees of freedom, the tabulated critical value being 1.83.[3] Thus the variation between outcrops is not significant in the UK population and hence β, the measure of concentration between UK outcrop means, may be taken to be infinity. In other words, all the UK outcrops may be taken to have the same mean direction which is also the UK formation-mean direction. If $\beta = \infty$, then estimate of ω, say $\hat{\omega}_U^*$, may be obtained by pooling the between and within the sum of the squares and the degrees of freedom. This gives $\hat{\omega}_U^* = 2.5526$.

Also equation (7) for κ reduces to

$$\kappa = N\hat{\omega}_U^*. \qquad (19)$$

Therefore, the number of observations N

[3] This situation needs special treatment not discussed earlier.

which one should take in order to attain a semiangle of say 10° with 95 percent confidence (which corresponds to a value of $\kappa_O =$ 126.1125 from table 2) is

$$N = \frac{\kappa_O}{\hat{\omega}_U^*} = \frac{126.1125}{2.5526} = 49.40 , \quad (20)$$

or roughly fifty observations. Thus, in the UK, one can estimate the formation-mean direction to within $\pm 10°$ with 95 percent confidence on the basis of fifty observations.

However, the evidence for accepting the hypothesis $\beta = \infty$ is not strong, the F-value being not too small. Therefore, one may take into account the between-outcrop variation also, in which case we follow the procedure as outlined in the earlier section. We find the estimates, $\hat{\beta}$ and $\hat{\omega}$, by equating columns 4 and 5 of table 3, and we get

$$\hat{\omega}_U = 1.3072 \text{ and } \hat{\beta}_U = 50.4032 \quad (21)$$

for the UK formation. If we now wish to estimate the formation mean with $\pm 10°$ accuracy with 95 percent confidence (which corresponds to a κ value of 126.1125), the values m and n should satisfy the equation (7), that is,

$$126.1125 = \frac{n_U}{(\hat{\beta}_U)^{-1} + (m_U \hat{\omega}_U)^{-1}} ,$$

or $\qquad (22)$

$$n_U = 126.1125 \, [(50.4032)^{-1} + (m_U \, 1.3072)^{-1}] .$$

Throughout our present study we assume a cost ratio of $C_1/C_2 = 10/1$, which seems to be reasonable in our case. Now appealing to (15), the optimum number of observations, say m_U^*, to be taken within each UK outcrop is given by

$$m_U^* = \sqrt{\frac{C_1}{C_2} \frac{\hat{\beta}_U}{\hat{\omega}_U}} = \sqrt{\frac{10 \times 50.4032}{1 \times 1.3072}} \quad (23)$$

$$= 19.6 ,$$

or roughly twenty observations from each outcrop. Substituting this value of $m_U^* = 20$ in (22),

$$n_U^* = 126.1125 \left(\frac{1}{50.4032} + \frac{1}{26.1440} \right) \quad (24)$$

$$= 7.3258 ,$$

or roughly 7.[4] Thus in UK if one takes seven outcrops and twenty observations from each, this would, besides giving us the required accuracy, also minimize the overall cost of estimating the formation mean. In fact, if we took seven outcrops and twenty observations from each, the accuracy parameter κ for the overall resultant direction is, from (17), $\kappa_U = 7/\{(50.4032)^{-1} + [20(1.3072)]^{-1}\} = 120.5033$, and hence the estimated mean direction has the semiangle of confidence

$$\psi_U = \frac{\lambda_{0.05}}{\sqrt{\kappa_U}} = \frac{1.96}{\sqrt{120.5033}} \quad (25)$$

$$= 0.1785 \text{ radians} ,$$

or 10° 13' with 95 percent confidence.

Similar analysis can be done for the other two Kamthi populations. For the eight outcrops of the MK with ten observations from each, we have ANOVA table 4. For 7 and 72 df, the F-value of 4.5660 is quite significant even at 1 percent level, the tabulated value being 2.14, and therefore the several site (outcrop) means differ significantly in the MK. Our model is thus more appropriate in this situation. Equating columns 4 and 5 of table 4, we get the estimates $\hat{\beta}$ and $\hat{\omega}$ for the MK, namely,

$$\hat{\omega}_M = 2.0000, \text{ and } \hat{\beta}_M = 5.6085 . \quad (26)$$

Now if we are to estimate the overall mean direction with a semiangle of $\pm 10°$ with 95 percent confidence as in the earlier case, the n_M and m_M should satisfy the equation

$$126.1125 = \frac{n_M}{(\hat{\beta}_M)^{-1} + (m_M \hat{\omega}_M)^{-1}}$$

or $\qquad (27)$

$$n_M = 126.1125 \, [(5.6085)^{-1} + (m_M \, 2.0000)^{-1}] .$$

[4] Earlier we suggested rounding off such fractions to the next higher integer. But this need not be fol-

But from equation (16) and an assumed cost ratio of 10/1, the optimum number of observations to be taken in a MK outcrop is

$$m_M^* = \sqrt{\frac{10 \times 5.6085}{1 \times 2.0000}} = 5.295 , \quad (28)$$

or roughly 6. Now from (27) the number of outcrops to be visited in the MK is

$$n_M^* = 126.1125 \left(\frac{1}{5.6085} + \frac{1}{12.0000}\right) \quad (29)$$

$$= 32.99 ,$$

or roughly 33. Thus for minimizing the cost of the operation, we need to take thirty-

Finally for the six outcrops of the LK, where again we have taken ten observations in each, we have the ANOVA shown in table 5. Here again the F-value is significant at 5 percent level with 5 and 54 df, the tabulated value being 2.40. Thus the site (outcrop) means in LK do vary significantly. Therefore, equating columns 4 and 5 of table 5, we have

$$\hat{\omega}_L = 1.3154, \text{ and } \hat{\beta}_L = 6.1058 . \quad (31)$$

If we wish to estimate the LK formation mean by means of the direction of the over-all resultant, again with an accuracy of

TABLE 4

ANOVA TABLE FOR MK DATA

Source	df	SS	MS	E(MS)	F
Between outcrops..........	7	7.9907	1.1415	$\frac{1}{2}\left(\frac{1}{\omega_M} + \frac{10}{\beta_M}\right)$	$F = 4.5660$
Within outcrops...........	72	17.9985	0.2500	$\frac{1}{2\omega_M}$	....
Total.................	79	25.9892	....	....	....

three outcrops in the MK but only six observations from each. This larger number of outcrops is a consequence of the fact that in MK the between-outcrop variation is very significant. This scheme of visiting thirty-three outcrops and taking six observations from each would give an accuracy parameter of $\kappa_M = 33/[(5.6085)^{-1} + (6 \times 2.0000)^{-1}] = 126.1303$ for the resultant based on $33 \times 6 = 198$ MK observations. The MK mean can be estimated with a semiangle of

$$\psi_M = \frac{1.96}{\sqrt{126.1303}} \quad (30)$$

$$= 0.1745 \text{ radians} ,$$

or 10° with 95 percent confidence.

lowed as a hard and fast rule since that would often increase the sample size and we would have much more precision than we asked for.

$\pm 10°$ with 95 percent confidence, then the n_L and m_L should satisfy the equation

$$126.1125 = \frac{n_L}{(\hat{\beta}_L)^{-1} + (m_L \hat{\omega}_L)^{-1}}$$

or   (32)

$$n_L = 126.1125 \, [(6.1058)^{-1}$$

$$+ (m_L \, 1.3154)^{-1}] .$$

But the equation (16) involving costs tells us that

$$m_L^* = \sqrt{\frac{10 \times 6.1058}{1 \times 1.3154}} = 6.8125 , \quad (33)$$

or roughly seven observations should be taken from each LK outcrop. Substituting this value of $m_L = 7$ in (32),

$$n_L^* = 126.1125 \left[\frac{1}{6.1058} + \frac{1}{7(1.3154)}\right]$$

$$= 34.35 ,$$

or roughly 34. Thus we should sample thirty-four outcrops in the LK taking seven observations from each which leads to an accuracy parameter of

$$\kappa_L = \frac{34}{(6.1058)^{-1} + [7(1.3154)]^{-1}}$$

$$= 124.8246$$

for the distribution of the overall resultant direction. Then this sample mean direction based on $34 \times 7 = 238$ observations would be within $\psi_L = 1.96/\sqrt{124.8246} = 0.1754$ radians, or $10°3'$ of the true LK formation mean with 95 percent confidence.

ber of observations required for estimation, with a desired precision, of mean direction which will be reasonably close to the true value of the formation mean; and (2) the pattern of distribution of the samples between and within the outcrops that would allow sampling with optimum efficiency and at minimum cost.

The techniques used here in finding answers to these questions can be summarized as follows:

1. To start with, a small number of representative cross-bedding samples are obtained (a more or less uniformly spaced grid sample) for the population (formation) un-

TABLE 5

ANOVA TABLE FOR LK DATA

Source	df	SS	MS	E(MS)	F
Between outcrops..........	5	5.9950	1.1990	$\frac{1}{2}\left(\frac{1}{\omega_L} + \frac{10}{\beta_L}\right)$	$F = 3.1544$
Within outcrops..........	54	20.5235	0.3801	$\frac{1}{2\omega_L}$	....
Total.................	59	26.5185	....	....	....

SUMMARY AND CONCLUSIONS

The techniques so far used for sampling and analysis of cross-bedding data were based on the conventional ANOVA. It is now well known that for the circularly distributed directional data (like cross-bedding foreset azimuths) the technique of ANOVA is of little use. For such data, the direction of the sample resultant (instead of the arithmetic mean) provides a reasonable estimate of the population mean, and the length of the resultant (instead of the standard deviation) provides an approximate measure of dispersion.

An efficient and economical sampling technique for cross-bedding directions has been developed in the present work using the circular measures of dispersion. With the help of the pilot-survey data of the fluviatile Kamthi formation, it has been attempted to find out: (1) the minimum num-

der study. Observations from each outcrop within the population are recorded separately. Equal number of observations from each outcrop facilitates computations, although this is not an essential requirement.

2. Sine and cosine values are computed for each foreset-dip direction (ϕ) measured. The resultant direction for each outcrop is computed as in equation (3), that is,

$$R_i^2 = \left(\sum_{j=1}^{m} \cos \phi_{ij}\right)^2 + \left(\sum_{j=1}^{m} \sin \phi_{ij}\right)^2$$

$$= C_i^2 + S_i^2,$$

where R_i is the outcrop resultant for the ith outcrop and m = number of observations within each outcrop.

3. The overall resultant for all the outcrops is now computed as in (4), that is,

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

where n = number of outcrops surveyed and C_i and S_i are as in step 2 above, that is,

and

$$C_i = \sum_{j=1}^{m} \cos \phi_{ij} ,$$

$$S_i = \sum_{j=1}^{m} \sin \phi_{ij} .$$

4. The appropriate values are now filled in table 1.

5. The values for concentration within outcrop ($\hat{\omega}$) and concentration between outcrops ($\hat{\beta}$) are obtained by equating columns 4 and 5 of table 1.

6. The optimum number of observations

ability is smallest in case of the upper Kamthi member where the sediments were deposited as point bars and channel bars of a river, apparently having a fairly consistent direction of sediment transport. The situation requires sampling of only a few outcrops and collection of a comparatively greater number of observations within each outcrop—an easy plan to follow in the field.

The large, between-outcrop variation in lower Kamthi demands measurement of a large number of outcrops. The lenses of cross-bedded sandstones measured in the lower Kamthi are interpreted as remnants of cut-off meander channels caught up with-

TABLE 6

SCHEME OF SAMPLING OF KAMTHI CROSS-BEDDING DATA

Formation Member	No. Outcrops to be Sampled (n^*)	No. Observations Required within Each Outcrop (m^*)	Total No. Observations
Upper Kamthi..........	7	20	140
Middle Kamthi..........	33	6	198
Lower Kamthi..........	34	7	238

to be taken in an outcrop, m^*, is obtained from equation (15), that is,

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

where C_1 and C_2 are the costs for reaching an outcrop and taking an observation within an outcrop, respectively.

7. The optimum number of outcrops to be sampled, n^*, is obtained from equation (13), where κ_0, the concentration for a desired confidence level $(1 - a)$ and the semiangle of confidence, (ψ_0), is obtained from table 2.

This technique of analysis suggests the scheme of sampling for the Kamthi formation members as shown in table 6.

This shows that between-outcrop vari-

in the flood basin deposits—a pattern which can be expected to have little consistency.

The large between-outcrop variation in middle Kamthi is presumably related to the fact that a part of this member represents irregularly deposited levees adjacent to the main river channel.

REFERENCES CITED

BATSCHELET, E., 1965, Statistical methods for the analysis of problems in animal orientation and certain biological rhythms: Am. Inst. Biol. Sci. Mon. 57 p.

OLSON, J. S., and POTTER, P. E., 1954, Variance components of cross-bedding direction in some basal Pennsylvanian sandstones of the Eastern Interior basins: statistical methods: Jour. Geology, v. 62, p. 26–48.

POTTER, P. E., and OLSON, J. S., 1954, Variance components of cross-bedding direction in some basal Pennsylvanian sandstones of the Eastern Interior basins: geological application: Jour. Geology, v. 62, p. 50–73.

———, and PETTIJOHN, F. J., 1963, Paleocurrents and basin analysis: Berlin, Springer-Verlag, 296 p.

RAO, J. S., 1969, Some contributions to the analysis of circular data: Unpub. Ph.D. thesis, Indian Statistical Inst., Calcutta.

RAUP, O. B., and MIESCH, A. T., 1957, A new method for obtaining significant average directional measurements in cross-stratification studies: Jour. Sed. Petrology, v. 27, p. 313–321.

REICHE, P., 1938, An analysis of cross-lamination: the Coconino sandstone: Jour. Geology, v. 46, p. 905–932.

SENGUPTA, S., 1970, Gondwana sedimentation around Bheemaram (Bhimaram), Pranhita-Godavari Valley, India: Jour. Sed. Petrology, v. 40, p. 140–170.

———, and RAO, J. S., 1966, Statistical analysis of cross-bedding azimuths from the Kamthi formation around Bheemaram, Pranhita-Godavari Valley, Sankhyā: Indian Jour. Statistics, ser. B, v. 28, pts. 1, 2, p. 165–174.

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

——— 1966, The statistics of orientation data: Jour. Geology, v. 74, p. 786–797.

———, and IRVING, E., 1957, Statistical methods in rock magnetism: Royal Astron. Soc. Monthly Notices Geophys. Supp., v. 7, p. 289–300.