Indian Journal of Geology
Vol. 60, No. 1, p. 1-10, 1988

THE SORTING HYPOTHESIS AND NEW MATHEMATICAL MODELS FOR CHANGES IN SIZE DISTRIBUTION OF SAND GRAINS

*Address by Chief Guest**

J. K. GHOSH
Indian Statistical Institute, 203, B. T. Road, Calcutta-35

INTRODUCTION

There is a long tradition of interaction between the statisticians of the Indian Statistical Institute and geologists in and outside the Institute. I have myself been active in several research projects of this kind. In one such project (Ghosh *et al.*, 1981) we applied classical and stratigraphic ranking of the Gondwana units in an area in Andhra Pradesh. In another (Ghosh *et al.*, 1987) we set up a stochastic model for occurrence of earthquakes of different magnitudes as well as a method for prediction, applying these to data on earthquakes in the north east part of India and Japan. However the problem that I have found most fascinating is the study of changes in the distribution of size of sandgrains, as measured by changes in weight frequency, brought about by flowing water, in nature or experimental flumes. Our main findings will be presented briefly in this paper. Some of the work to be reviewed has not been published before.

The paper is organised as follows. We first present graphically some of the relevant data pertaining to samples collected from the bed of the river Usri in Bihar and experiments on suspension and deposition in flumes, carried out in Uppsala and Calcutta (Ghosh *et al.*, 1981, 1986). The following questions will then be posed and partially answered. Why does the grain size distribution approach normality as one moves away from the source of the river Usri (Fig. 1) ? How does one theoretically calculate the grain size distribution of suspension at a given height in a flow given only the size distribution in the bed and the flow parameters and how does one explain the unimodality, symmetry, or approximate normality in the suspension distribution, even though these were not present in the bed (Figs. 2, 3, 4) ? Finally, how does one model deposition, given the distribution of sand released in suspension and the parameters of the flow as it decelerates ? In course of answering these questions we develop a sorting hypothesis for normality which is based on what we call sorting theorems for normality; this is quite different from Kolmogorov's explanation based on the Central Limit Theorem. We also provide

*Delivered at the Sixty-third Annual General Meeting held on 13.10.1987

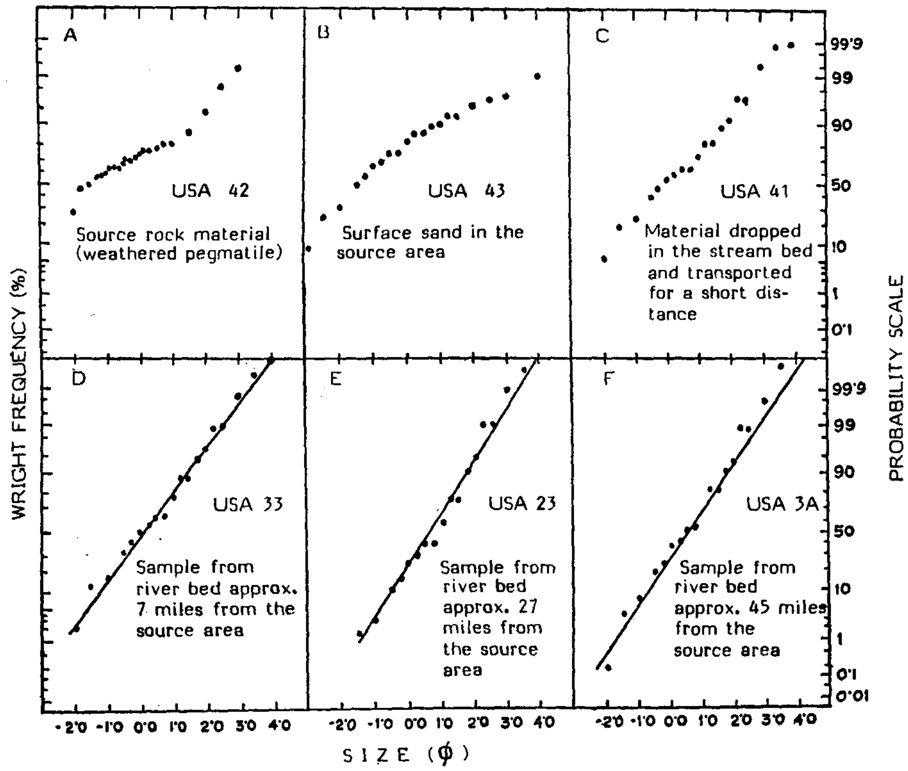

Fig. 1. Lognormality of the Usri data.

a new probabilistic interpretation of the Rouse equation as corresponding to the steady state of a (Markovian) diffusion process with a reflecting lower boundary and a repelling upper boundary. This interpretation, in turn, leads to replacement of the reflecting boundary condition by a new partly discrete boundary condition and hence to alternatives to the Rouse equation which exhibit a "damping" effect. There is some experimental evidence to indicate that unless one introduces a "damping" in the Rouse equation, it will provide poor fit to data when the flow velocity is not high.

THE DATA

As indicated earlier, the data will be presented graphically. The size is shown in the ϕ-scale. The symbol ϕ will be also used to indicate the magnitude of size in the ϕ-scale. Thus the weight frequency for a particular size will be shown as $c(\phi)$. Sometimes the weight frequency is actually a relative weight frequency, adjusted to make the total over all ϕ equals one.

In Fig. 1, which relates to samples from the bed of Usri, weight frequency is plotted on a normal probability scale so that a

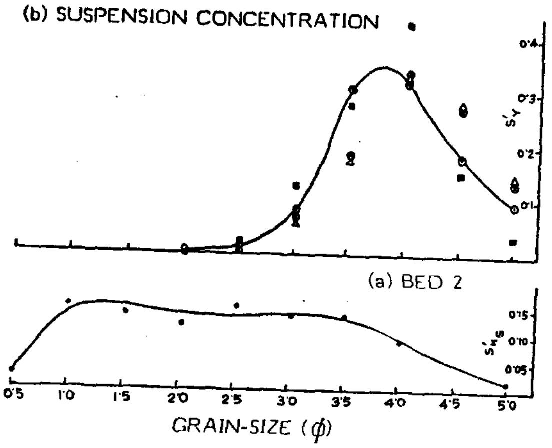

Fig. 2

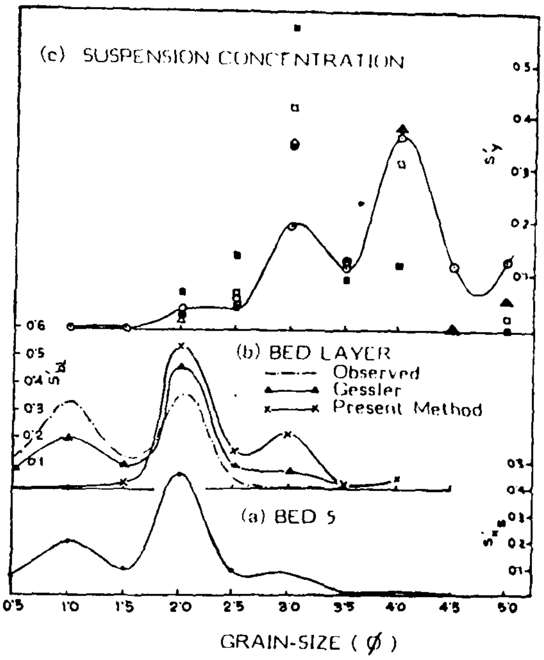

Fig. 4

Fig. 2-4. Approximate normality in suspension concentration as contrasted to the size distribution in the bed materials.

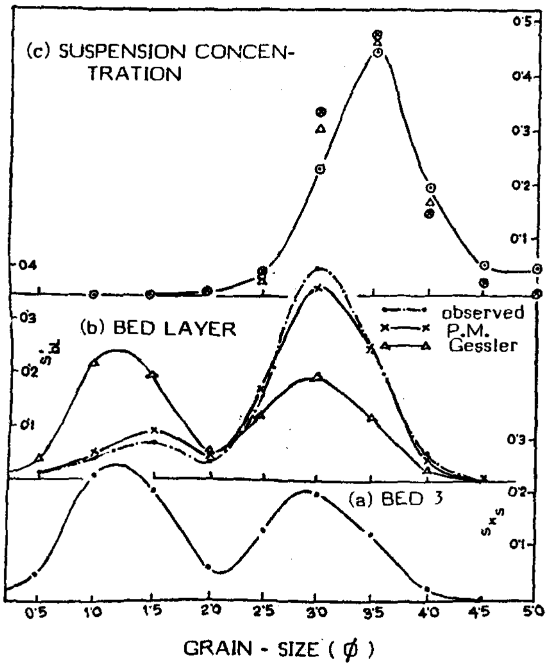

Fig. 3

normal distribution will be indicated if the data falls on a line. The approach to linearity and hence to normality is clear from Fig. 1.

Figs. 2, 3 and 4 relate to data and predicted values for experiments in suspension, reported in Ghosh *et al.*, 1981 and Sengupta, 1975, 1979.

The rather unusual bed in Fig. 4 was chosen to check that features in the bed if they are sufficiently pronounced will be reflected in the suspension distribution. This was an indirect confirmation of the suspen-

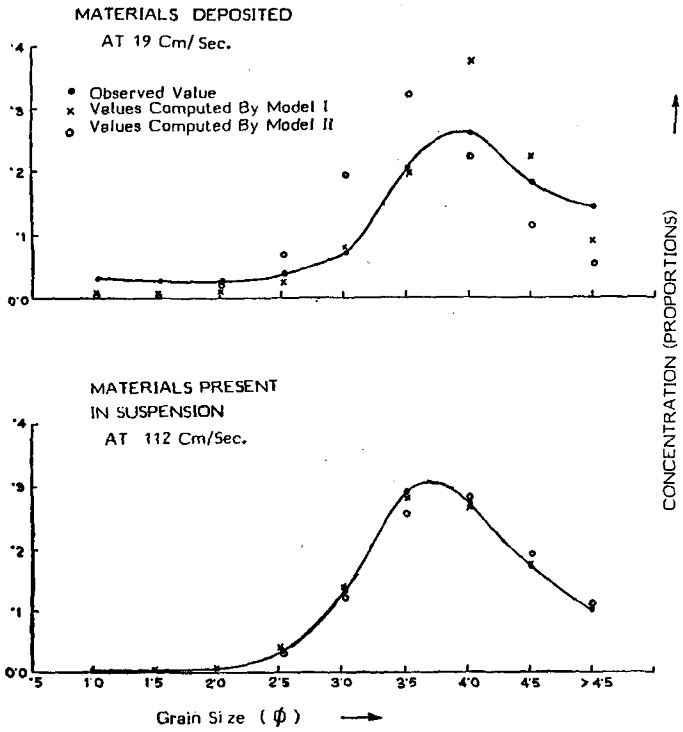

Fig. 5

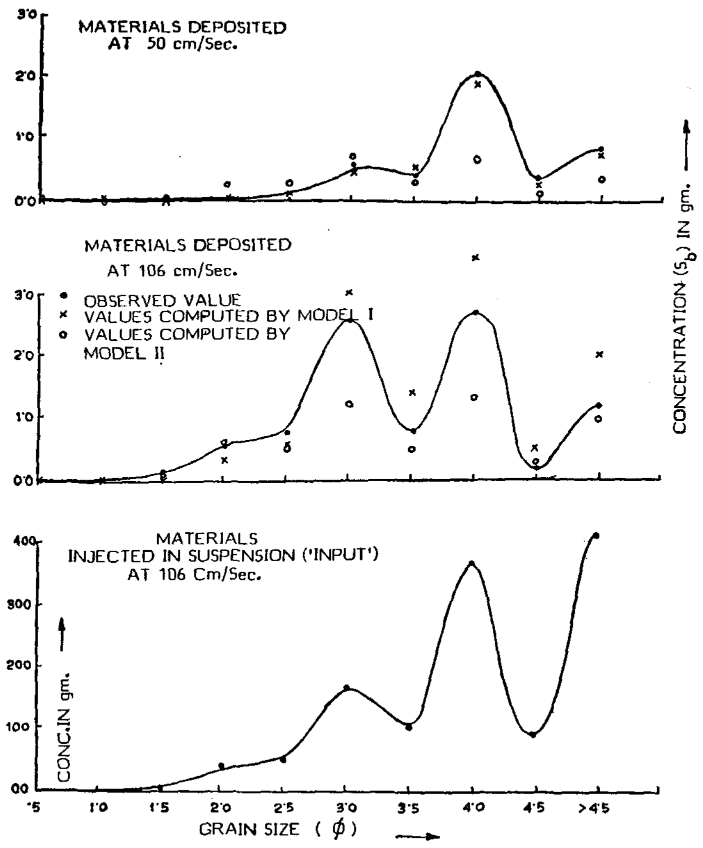

Fig. 6

Fig 5-6. Experimental data on deposition.

sion models in which both the flow and the bed determine the suspension distribution.

Figs. 5 and 6 exhibit data on deposition and corresponding values calculated from the model (Ghosh *et al.*, 1986).

THE SORTING HYPOTHESIS FOR NORMALITY

Kolmogorov's original explanation in 1941 for normality postulated that changes in size occur owing to the breaking up of grains, and successive break-ups are independent so that one may appeal to the Central Limit Theorem. In a somewhat simplified form the argument runs as follows. Consider a grain of Size X_0 units, in some unit of length. After one unit of time, owing to say crushing, the grain breaks into smaller particles. A typical first generation particle will have size $X_0.X_1$, where $0<X_1<1$ represents the relative decrease in size. After n units of time, a typical nth generation particle will have size $X_0.X_1, \ldots, X_n$, $0<X_i<1$. So in the ϕ-scale this becomes $\log X_0+\log X_1+\ldots+\log X_n$, which would be approximately normally distributed by the Central Limit Theorem for sums of independent random variables. The weak link in this chain of argument, at least as applied to the Usri data is the hypothesis that change in size distribution is due to breaking up of grains. In the case of the Usri data sorting caused by transportation is the main source of change in distribution of size. But other models that have been proposed using effects of breakage, erosion, and transportation make so many special assumptions that they seem to assume what is to be proved or explained and lack the generality of Kolmogorov's argument. We show below how a sorting hypothesis alone can lead to a very general argument for normality.

The argument rests not on the Central Limit Theorem but on another fundamental theorem in statistics which says that under certain very general assumptions a conditional or so called posterior distribution is approximately normal. To fix ideas, consider a simple example first. Let O be a random variable with probability density function (p.d.f.) π (O) and suppose given ϕ, $X_1, \ldots, X_n$ are (conditionally) independent and normal with mean ϕ and variance 1. One may think of $X_1, \ldots, X_n$ as a sample from a normal population with ϕ and variance 1 and (in the manner of a Bayesian) the population itself as a sample from a collection of hypothetical populations over which O has p.d.f. π (ϕ). Then if π is a smooth function and n is large the conditional density $\pi(\phi/X_1, \ldots, X_n)$ of ϕ given $X_1, \ldots, X_n$ is approximately normal with mean equal to $\bar{X}=(X_1+\ldots+X_n)/n$ and variance tending to zero.

$$\text{Here, } \pi(\phi/X_1,\ldots,X_n)=\frac{\pi(\phi)\prod_{i=1}^{n} f\phi(X_i)}{\pi(\phi)\prod_{i=1}^{n} f\phi(X_i)\,d\phi} \cdots(1)$$

$$\text{where } f\phi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\phi)^2}$$

In the special case where $\pi(\phi)$ itself is normal with mean μ and variance ι^2, easy direct calculation shows the above ratio is a normal density with mean $\lambda\mu+(1-\lambda)\bar{X}\rightarrow\bar{X}$ and variance $\lambda\iota^2$, where $\lambda=(n\iota^2+1)^{-1}\rightarrow\propto O$ as $n\rightarrow\propto$. In a non-mathematical way the reason why such a result holds may be explained as follows. Given $X_1, ..., X_n$, the conditional distribution of ϕ is "sorted out" to lie in a small neighbourhood of $\bar{X}$ and there the logarithm of the conditional density can be well approximated by a quadratic. But this is the same as saying that the conditional density has a normal approximation with mean $\bar{X}$.

We can now state the following fundamental theorem in statistics. Let ϕ be a random variable with smooth density $\pi(\phi)$ and given ϕ, $X_1, ..., X_n$ are independent and identically distributed with quite a arbitrary but smooth density $f_\phi(\times)$. Then the conditional density $\pi(\phi/X_1, ..., X_n)$ of ϕ given $X_1, ..., X_n$ is approximately normal with mean $\hat{\phi}$ and variance tending to zero. Here $\pi(\phi/X_1, ..., X_n)$ is defined as above and $\hat{\phi}$ is a function of $X_1, ..., X_n$ satisfying the equation

$$O=\frac{d}{d\phi}\left(\prod_{i=1}^{n} f_\phi(X_1)\right)_{\phi=\hat{\phi}}$$

i.e., $\hat{\phi}$ is the so called maximum likelihood estimate of ϕ. The reason why this theorem holds is again due to the kind of "sorting out" of the distribution of ϕ given $X_1, ..., X_n$, explained in the previous paragraph. A recent reference where a precise statement as well as refinements are provided is in Ghosh *et al.* (1982).

We now provide an explanation along these lines for the normality observed in Fig. 1. Suppose we are making observations at a distance d from the source of Usri. Assume that all the sand being observed here have been transported from the source in n units of time or less. For a grain of size ϕ the displacements are random variables $X_1, ..., X_n$. For fixed ϕ, $S=X_1+...+X_n$ is approximately normal, under reasonable assumptions, by the Central Limit Theorem. The O-distribution at the source may be taken to be $\pi(\phi)$, Then the ϕ-distribution at distance d is just the conditional distribution of ϕ given S=d. This conditional distribution is not the same as the conditional distribution given $X_1, ..., X_n$, but an application of the arguments in Ghosh *et al.* (1982) will show that this conditional distribution also exhibits a sorting effect and, as a consequence of that, its logarithm can be approximated by a quadratic. This amounts to a normal approximation. From a non-mathematical point of view, sorting takes place because given the distance d one expects only certain grain-sizes to occur in abundance. The conditional distribution must, then, be concentrated there.

A few technical remarks are in order. If one assumes the X_i's to have identical distribution, then the variance of the conditional distribution will tend to decrease with d and so the sorting effect will increase. Data does not support this. It is partly for this reason that identical distribution was not assumed above. A better modelling will require the distribution to change with the distance travelled. One would then need to introduce a Markov process. Possibly the sorting effect can be established in this framework.

The main advantage of our explanation over Kolmogorov's has been indicated earlier. Crushing is not the reason for change in size distribution. The main disadvantage of our method as compared with Kolmogorov's is that the logarithmic scale for size does not appear so naturally. In our argument, any scale, say ψ, will do provide the absolute value of derivative of ψ with respect to ϕ is neither too low nor too large.

Finally it must be pointed out that a major drawback of this explanation as well as those of Kolmogorov and others is that when the explanation has been given and the mystery cleared up we do not seem to have learnt anything of fundamental importance for the sorting process itself. The modelling discussed in the next two sections is more useful from that point of view.

MODELLING SUSPENSION DATA AND THE SORTING EFFECT

In the following model the effect of the bed-layer is ignored to some extent, the effect of this is not serious. For details see Ghosh *et al.* (1981). Denote the distribution or rather the density of bed by $\pi(\phi)$. Let $f_\phi(y)dy$ be the probability that a grain of size ϕ will eventually end up at $(y-\frac{1}{2}dy, y+\frac{1}{2}dy)$. (To be more precise $f_\phi(y)$ may be taken to be the limit (as $t \to \infty$) of $f_\phi(t, y)$, the latter on multiplication by dy giving the probability of ending up in $(y-\frac{1}{2}, y+\frac{1}{2}dy)$ after time t). These notations have been made consistent with those in the previous section so that the similarities become clear. The height 'y' now plays to some extent the role played earlier by the distance 'd'.

Once again the formula for size distribution at height y is the same as that of the Bayes formula (1) ;

$$\pi(\phi/y)=\pi(\phi)\, f_\phi(y)/(\text{Total of numerator over } \phi).$$

For four choices of $f_\phi(y)$—two of which were new—the above is shown in Figs. 2, 3, 4. If $f_\phi(y)$ again shows strong "localisation" or sorting effects, one would get approximate normality around the mode of $\pi(\phi/y)$. If $f_\phi(y)$ is determined by the Rouse equation, the details are worked out for two beds (partly theoretically, partly numerically) in Ghosh and Mazumder (1981).

The rest of this section will be devoted to a stochastic interpretation of the Rouse equation first introduced in Ghosh and Mazumder (1981) may be written as

$$\varepsilon(y)\frac{df_\phi(y)}{dy}+\nu(\phi)\, f_\phi(y) = O$$

where $\nu(\phi)$ is the settling velocity and $\varepsilon(y)$ is the turbulence diffusion coefficient and is proportional to U_{max}, the maximum velocity, maximum being taken over a vertical direction. Even though the flow as well as the displacement is three dimensional, we are considering only the vertical direction. It can be shown that in the present context of finding the stationary distribution, this suffices provided wall effects are negligible.

As with the π's, so with the f_ϕ's—it is common to use the symbol c in this context, where c stands for concentration. Note that f and c agree upto a constant of normalisation. So in the following $f_\phi(t, y)$ and $c_\phi(t, y)$ have been identified for convenience. Since ϕ will be kept fixed, dependance on ϕ is suppressed in the notation.

The Rouse equation is the steady state version of the turbulence diffusion equation

$$\frac{\delta c}{\delta t} = \frac{\delta}{\delta y}(\nu c) + \frac{\delta}{\delta y}(c')$$

with a reflecting boundary condition at the bed height which we regard as the lower boundary of the process. The diffusion equation can be rewritten as

$$\frac{\delta c}{\delta t} = \frac{1}{2} \frac{\delta^2}{\delta y^2} (ac) - \frac{\delta}{\delta y} (bc)$$

where $a \equiv a(y) = 2\varepsilon(y)$, $b \equiv b(y) = \varepsilon'(y) - \nu$ and c', ε' etc. indicate derivatives with respect to y.

This is the forward equation of a Markov process and has the following interpretation : If it is at y at time t, then locally the particle has a small displacement in time dt which is normal with mean b(y)dt and variance a(y)dt. It can be shown that the upper boundary is repelling, so the particle can never get there. The behaviour at the lower boundary has to be specified to specify the stochastic diffusion process completely. Suppose we have reflection at the lower boundary. By the general theory of Markov processes and compactness of state space etc. there is a steady state or stationary distribution for the particle (over y). This satisfies the Rouse equation.

DEPOSITIONAL DATA AND AN IMPROVED MODEL FOR SUSPENSION

In course of modelling depositional data, vide Figs. 5, 6, we discovered that for low to moderate values of U_{max} the Rouse equation is not satisfactory. We shall only indicate here how it may be improved. Experimental and mathematical details as well as other aspects of interest to geologists and probabilists will be found in Bhattacharya and Ghosh ; Ghosh *et al.* (1986).

After studying the data for some time it became clear that more grains are left in suspension than would be expected from the Rouse equation which leads to an exponential decay of mass as one moves up vertically. To Introduce such a damping effect, we began by keeping the same diffusion equation but changing the boundary condition at the lower boundary as follows. We require

$$\varepsilon c' + (1-\alpha)\, c = O, \quad y = K_s$$

where $(1-\alpha)$ is the probability that a particle which comes in contact with the bed will be lifted and K_s is the roughness of the bed.

If there were a steady state under these conditions, the corresponding c would satisfy

$$\frac{\delta c}{\delta t} = O, \text{ and so}$$

$$\frac{d}{dy}(\tfrac{1}{2}ac) - bc = K \ldots (2)$$

$$\text{i.e.,} \quad \varepsilon \frac{dc}{dy} + \nu c = K$$

for a positive K, which can be found by comparing with the boundary condition. A reflecting boundary would have given K=O. The previous equation may be written as

$$\frac{1}{c} \frac{dc}{dy} = -\frac{\nu}{\varepsilon} + \frac{K}{c} \ldots (2a)$$

The introduction of the new boundary condition damps the relative decrease in c (namely the left hand side (2a)) by an amount equal to K/c. This formula gives a good fit but unfortunately Markov process theory tells us, there is no steady state—ultimately all particles will settle at the bottom because of the boundary condition. In a way this is known or felt to be true even by people who do not know Markov process theory but handle diffusions.

What does the solution mean then ? As a sort of answer, suppose we consider the following partly discrete boundary condition, which can only be handled by people who know about Markov processes. When the particle comes to the lower boundary, it is partly reflected and partly takes a jump back into the process (in this way damping the effect of pure reflection). Let the jump have a density s(y). Then again Markov process theory tells us that a steady state solution exists but it can no longer be calculated directly from the forward equation. Non-standard calculations (Bhattacharyya and Ghosh) finally lead to an equation for the steady state

$$\frac{d}{dy}\left(\tfrac{1}{2}ac\right) - bc = \frac{a(K_s)c(K_s)}{2 \cdot A}\left[1 - \int_{K_s}^{y} s(u)\,du\right] \quad \ldots(3)$$

where the right hand side introduces a damping factor whose effect diminishes as you move up i.e. increases y.

The earlier solution (Ghosh *et al.*, 1981) providing good fit to data must be thought of as an approximation to this with a single K on the RHS for all y. The physical meaning of the jump condition is that when a particle is lifted up from the bed by a strong vortex then the vortex takes it up quite a bit from the bed. Using this boundary condition one can also study the process numerically as it evolves over time.

The only question that remains to be answered is why deposited grains reflect the distribution put in suspension. The answer is that deposition takes place when the velocity diminishes substantially and this prevents the sorting mechanism to operate. Apparently our experimental and theoretical findings about deposition is not in uniformity with the prevailing view among geologists that the coarse grains are deposited first. No such sorting was found by us.

REFERENCES

Bhattacharya R N and Ghosh J K (under preparation). Steady State Distributions of One Dimensional Distributions Under General Boundary Conditions.

Ghosh J K, Mazumder B S and Sengupta S, 1981. Methods of Computation of Suspended Load from Bed Materials and Flow Parameters. Sedimentology, v 28, p 781-791.

Ghosh J K and Mazumder B S, 1981. Size Distribution of Suspended Particles—Unimodality, Symmetry, and Lognormality, 'Statistical Distributions in Scientific Work' (ed C. Tailie and others), v 6, p 21-32, D. Reidel Publishing Company.

Ghosh J K, Saha M R and Sengupta S, 1981. Gondwana Stratigraphic Classification by Statistical Method, Down-To-Earth Statistics (ed D F Merriam), Syracuse University Geology Contributions, p 47-62.

Ghosh J K, Sinha B K and Joshi S N, 1982, Expansions for Posterior Probability and Integrated Bayes Risk. Proc 3rd Purdue Symp on Statistical Decision Theory and Related Topics 1 (ed S S Gupta and James D Berger), p 403-456.

Ghosh J K, Mazumder B S, Saha M R and Sengupta S, 1986. Deposition of Sand by Suspension Currents : Experimental and

Theoretical Studies. Jour Sed Petrol, v 56, p 56-66.

Ghosh J K, Krishnan T and Kutty T, 1987. A Stochastic Model for Earthquake and Prediction via d-value. Proc of the Symposium 'Earthquake Prediction—Present Status', Pune (under publication).

Sengupta S, 1975. Size-Sorting during Suspension Transportation Log-Normality and Other Characteristics. Sedimentology, v 22, p 257-273.

Sengupta S, 1979. Grain Size Distribution of Suspended Load in relation to Bed Materials and Flow Velocity Sedimentology, v 26, p 63-82.

The text figures are reproduced from the sources mentioned below, with permission of the authors.

Fig 1. after Sengupta, S., 1975, UNGI Report No. 40, Uppsala University, fig. 20.

Figs. 2-4. after Ghosh, J. K., Mazumder, B. S., and Sengupta, S., 1981. Sedimentology, v. 28, p. 784-785, figs. 2-4.

Figs. 5-6. after Ghosh, J. K., Mazumder, B. S., Saha, M. R. and Sengupta, S., 1984, IAHR 4th Intl. Symp. on Stochastic Hydraulics, Urbana, Univ. Illinois, Pre Symp. Proc. pp. 55-64, fig. 2-3.