

W.J. Baggaley  
Physics Department, University of Canterbury, Christchurch,  
New Zealand.

ABSTRACT

A problem commonly encountered in meteor studies is the need for a realistic interpretation of statistical data which contains uncertainties in two variables. A simple solution is provided and for illustration applied to diffusion-height data.

1. LINEAR REGRESSION

The statistical analysis of data consisting of two variables, each subject to error, in order to determine the existence of any linear relationship between the variables often occurs in observational meteor studies. For a scatter diagram in Cartesian coordinates two regression lines can be recognized: the regression of  $y$  on  $x$  minimizes the residuals in  $y$  whereas the regression of  $x$  on  $y$  minimizes the residuals in  $x$ . The former line of slope  $m_1$  is the best estimate of the data trend if uncertainties in  $x$  for each data point are negligible while the latter regression line of slope  $m_2$  represents the best estimate if the  $x$  uncertainties are dominant. In many situations significant uncertainties are present in both coordinate values and since no convenient procedure appears to be readily available for estimating the best linear fit in this case, the simple but unrealistic approach is often adopted of taking the gradient to be equal to the arithmetic mean of  $m_1$  and  $m_2$ .

In a recent work Ross (1979) shows how a solution may be obtained in a straightforward way: if the coordinate uncertainties  $\Delta x$  and  $\Delta y$  are the same for all data points and defining  $\beta$  as  $(\Delta x/\Delta y)$  then the gradient,  $m$ , of the best fit line is given by a solution of the quadratic equation

$$m^2\beta^2 - m(\beta^2m_2 - 1/m_1) - 1 = 0$$

For the special case  $\beta=0$   $m=m_1$  while for  $\beta=\infty$   $m=m_2$  as expected.

## 2. METEOR DATA

For illustration of the method, consider the problems concerned with the analysis of scatter diagrams obtained in studies of the height ( $h$ ) variations of ambipolar diffusion coefficients ( $D$ ) obtained from radio-meteors. Such studies yield estimates of the atmospheric scale-height  $H$ : for an isothermal atmosphere and with  $D$  inversely proportional to atmospheric density then

$$h = H \log_e D + \text{const.}$$

so that the slope of the  $h$  versus  $\log_e D$  plot yields an estimate of  $H$ . Considerable dispersion of the data is a feature of such measurements and Weiss (1955), Murray (1959) and Greenhow and Hall (1961) considered the problem of extracting valid statistical parameters from their scatter diagrams. Greenhow and Hall reported a height uncertainty of 4 km and an uncertainty in  $D$  of a factor of 1.3 derived from echo profile characteristics. Thus  $\Delta x = 4.0$ ,  $\Delta y = 0.26$ . The regression line slopes for errors in height only and in  $D$  only are given by the authors (their table 1) as 6.3 km ( $m_1$ ) and 13.2 km ( $m_2$ ) respectively. Hence  $m = 7.46$  km for the best estimate of the scale-height  $H$  compared with the value  $7.8 \pm 0.3$  obtained by the approximate procedure adopted by Greenhow and Hall.

It is suggested that the simple method described here will be a fruitful aid in analyzing statistical data.

## REFERENCES

- Ross, A.W. (1979) *Amer. J. Phys.* (in press)  
 Greenhow, J.S. and Hall J.E. (1961) *Planet. Space Sci.* 5, 109.  
 Murray, E.L., (1959) *Planet. Space Sci.* 1, 125.  
 Weiss, A.A. (1955) *Aust. J. Phys.* 8, 280.

Summary of discussion. This problem has been discussed many times in the literature. See Wald, A. (1940, *Annals Math. Statistics II*, pp. 284-300), for a modern discussion and historical summary. The method proposed here changes the x-y scales by normalizing to the errors so that error ellipses are transformed into error circles and the line is that which minimizes the perpendicular distance from each point to the line. An equivalent formula is given by Worthing, A.G. and Geffner, J. ("Treatment of Experimental Data", John Wiley and Sons, New York, 1943).

J. Jones (1970, *Planetary Space Sci.* 18, p. 1836) has analysed the data of Greenhow and Hall by a similar but less general method.

The method can be extended to more than two variables but is not capable of dealing with the important case where one of the variables exhibits a cutoff due to a sensitivity threshold of the measuring equipment.