Ancient TL

www.ancienttl.org · ISSN: 2693-0935

Berger, G. and Huntley, D., 1986. *Linear regression of TL data.* Ancient TL 4(2): 26-29.
https://doi.org/10.26034/la.atl.1986.104

Linear regression of TL data

G.W. Berger and D.J. Huntley
Department of Physics
Simon Fraser University
Burnaby B.C.
Canada V5A 1S6

Introduction

Recently, Rendell (1985) has presented a comparison of five approaches to linear regression of TL signals versus applied dose. The fifth approach apparently represents the one used at S.F.U., but it is titled and summarized incorrectly. This note presents the technique we use, outlines how it differs from the other four, and states why we believe the latter are inappropriate.

Background

The history of attempts to develop rigorous techniques for linear regression has been outlined by York (1966). He proceeded to develop an exact solution to the generalized problem, following the initial approach of Deming (1943). York showed that because experimental uncertainties in the values of the variables X and Y will vary from point to point, the most general approach requires solving, iteratively, a recursive "cubic" equation in the parameter b (slope). This recursive equation (his equation 20) he called the "least squares cubic".

In application it is necessary to assign uncertainties to each X_i and Y_i, uncertainties which may vary from point to point. In practice this assignment is usually done by choosing weights w(Xi) and w(Yi) that are inversely proportional to the variances in the respective variables. This, and other special cases of the generalized equations of 1966 are discussed and illustrated graphically in a short paper by York (1967). In a later publication, York (1969) developed a generalized solution of this "least squares cubic" for the case of uncertainties in X and Y that are correlated.

In the general case one may, therefore, have considerable mathematical complexity. In TL work this complexity is fortunately absent, because there is no evident correlation between the uncertainties in the TL signal and those in the applied dose. Furthermore, random uncertainties in the applied dose may be negligible; this will be true if random errors in both the dose rate and irradiation times are insignificant. There remains the question of how to determine the variances in the TL intensities. One could measure these; however, since a rather large number of (>10) sample discs would need to be measured for each point, this approach is somewhat impractical. Instead, we argue that TL intensity variations are due to variations in the amount or distribution of matter on the sample disc, or are due to intrinsic brightness variations of sample grains. In these cases one would expect the variance of the TL intensity to be proportional to the square of the TL intensity.

Analysis of TL Data

Our approach is thus based on the following criteria:

(i) any random errors in the laboratory irradiation doses are insignificant,
(ii) uncertainties in the TL intensities are the same percentage of the intensity for all data points,
(iii) values for both the equivalent dose (D_{eq}) and its uncertainty are required.

These three criteria dictate the use of the following equations, derived from York (1966), where Y is applied dose and X is the TL signal:

$$Y = a + bX \tag{1}$$

where $a = D_{eq}$,

$$a = \bar{Y} - b\bar{X} \quad , \qquad b = \frac{\sum w(X_i)V_i^2}{\sum w(X_i)U_iV_i} \quad ,$$

$$\bar{X} = \frac{\sum w(X_i)X_i}{\sum w(X_i)} \quad , \qquad \bar{Y} = \frac{\sum w(X_i)Y_i}{\sum w(X_i)} \quad ,$$

$U_i = X_i - \bar{X}$ and $V_i = Y_i - \bar{Y}$.

The variances of a and b are given by

$$\sigma_a{}^2 = \sigma_b{}^2 \cdot \frac{\sum w(X_i)X_i{}^2}{\sum w(X_i)} \quad \text{and}$$

$$\sigma_b{}^2 = \frac{1}{n-2} \cdot \frac{\sum w(X_i)(bU_i - V_i)^2}{\sum w(X_i)U_i{}^2} \quad .$$

All sums are from i = 1 to n, where n is the number of data points. For a constant percent error p, the weighting factors are given by $w(X_i) = (pX_i)^{-2}$, but since p cancels out of all expressions, we use $w(X_i) = X_i^{-2}$.

It is not strictly correct to state, as Rendell does, that this approach treats dose as the dependent variable. The TL signal is still the dependent variable, but the requirements of weighting and error calculation necessitate a change from the usual notation. Rendell also states our equations for U and V incorrectly.

Discussion

None of the first four approached as described by Rendell meet the criteria stated at the beginning of the previous section and, therefore, we believe they are invalid. In particular, none of them provide an uncertainty in the equivalent dose, and the first three use inappropriate weighting factors.

Application of our method to the set of data given by Rendell yields D_{eq} = 10.95 ±1.12 and a slope of 2.07 ±0.15 (this is to be compared with the reciprocal of the slopes in her table). The large uncertainty in D_{eq} is expected because of the large extrapolation. In practice, we prefer to avoid the use of linear regression for such data and to apply a larger range of doses, even if it becomes necessary to use sublinear regressions (higher order polynomials or saturating exponentials). However, that is another issue and will not be discussed here.

Finally, it should be apparent from our comments and from York's work that it is not useful to invoke values of "correlation coefficients" to describe the quality of data sets. This data correlation coefficient (r in Rendell) supplies little or no useful information about the quality of the regression. What is useful, however, is a "goodness-of-fit" parameter, such as described by York (1969) (his $[S/(n-2)]^{1/2}$ parameter). This and analogous goodness-of-fit parameters (see also Brooks et al 1972) are used routinely in the assessment of isochrons in radioisotopic dating. Unfortunately, its utility depends on an independent knowledge of the uncertainties in each TL observation. Such a state of knowledge does not yet exist in TL work because the variablility of TL signals is dominated by unspecified grain-to-grain or disc-to-disc differences.

References

Brooks, C., Hart, S.R. and Wendt, I. (1972) Realistic use of two-error regression treatments as applied to rubidium-strontium data. Rev. Geophys. Space Phys., 10, 551-577.

Deming, W.E. (1943) Statistical adjustment of data. John Wiley and Sons, New York.

Rendell, H.M. (1985) Problems with linear regression as applied to TL data. Ancient TL, 3, 6-9.

York, D. (1966) Least-squares fitting of a straight line. Can. J. Phys., 44, 1079-1086.

---- (1967) The best isochron. Earth Planet. Sci. Letters, 2, 479-482.

---- (1969) Least squares fitting of a straight line with correlated errors. Earth Planet. Sci. Letters, 5, 320-324.

Reviewer's Comments

As the authors point out, there are some inaccuracies in Rendell's description of their method. They should have an opportunity to correct them, and to state the assumptions on which their method is based. However, since the scope of a newsletter such as 'Ancient TL' does not allow detailed justification of the method, it is clear that many readers will require to consult York's original paper in order to understand the rationale behind the method and to assess its usefulness. For example, one potentially confusing aspect is the formulation of the line, equation (1). Here Y looks like the dependent variable - hence the original mistake by Rendell. In fact if (1) is turned around to $X = -a + 1/bY$, then the estimate for b is just the inverse of the usual weighted least squares estimate for 1/b in the regression of X on Y. In view of this it is difficult to see, without access to York's paper, how the error estimates have been derived, and in what respect the approach differs from the standard 'X on Y' case (in their notation).

A further point to note is that N. Debenham's method, summarised as approach four in Rendell, does provide a measure of the uncertainty in D, given by the (asymmetric) intercepts of the one-sigma hyperbolic confidence bands for the regression of X on Y (in their notation). Readers could be given the (incorrect) impression that York's method is the only one meeting the criteria stated in the paper.

Morven Leese (Statistician, British Museum Research Laboratory)