Estimation of accumulated dose and its uncertainties: potential pitfalls in curve fitting

R.G. Lyons†, B.J. Brennan‡ and P.L. Hosking*
†Department of Geography and Oceanography, Australian Defence Force Academy, University College, University of New South Wales, ACT 2600, Australia.
Departments of ‡Physics and *Geography, University of Auckland, New Zealand.

Introduction

The problems of curve fitting in general, and determining uncertainties in AD in particular, are recognized as being of significance in ESR and TL dating, as evidenced by the recurrence of the topic over recent years (Rendell, 1985; Berger and Huntley, 1986; Franklin, 1986; Berger *et al.*, 1987; Lyons, 1988; Grün and MacDonald, 1989; Berger, 1990). This discussion aims to clarify for non-statisticians some of the issues which have arisen, to focus attention on potential problems, and, moreover, to encourage available techniques to be applied critically rather than naïvely. It also draws attention to some of the recent developments in statistical theory.

Relevant statistical concepts

For simplicity, the following examples use linear fitting to demonstrate some important concepts, but the concepts do, of course, apply equally to any other curve.

(i) Leverage.

Points will have a greater or less effect on the fitted curve depending on their position on the growth curve. As noted by Franklin (1986), end points are more influential in determining the position of the line, particularly for extrapolation, than more central points (fig. 1; table 1). Depending on whether the extreme points are "good" or "bad", leverage is either a desirable or undesirable property: a 'good' point is one which has both a low uncertainty (high precision) and is a true representation of the physical reality (high accuracy), while a 'bad' point has a low precision and/or accuracy (see also fig. 2 and the discussion below on outliers).

Confidence in the accuracy of the natural non-irradiated signal intensity should be well-founded. The high leverage exerted by its position can therefore be justified provided it also has a high precision. The leverage of low dose data points is particularly important if the Y-values have equal variance. It is then highly desirable to reduce the variance of the natural signal by repeating the measurement and using the average value in the curve fitting procedure (see later, on replication of data points). If, on the other hand, the data are believed to have a constant *percent* error, then the low dose points are "good" and their leverage is a desirable property; the data set will be enhanced by increasing the number of low dose points.

It is unfortunately true that high dose points are generally more suspect as being further removed from the environmental situation; repeated measurements will improve their precision but not their intrinsic accuracy. Even where the high dose points have a small uncertainty, a serious problem arises if the model fitted does not fully represent the underlying physical behaviour because they will be "bad" points with respect to the fitted model, and will exercise undesirable leverage (see fig. 4).

Leverage by high dose points can be reduced by two means:

(a) inverse weighting to redress the implicit heavy weighting they receive by virtue of their position. (Note that this inverse weighting is in addition to any required by the unequal variances, or precision, of the data.) Carrying out a log-transformation to the data in order to fit a linear instead of an exponential model, as in Apers et al., (1981), effectively applies a weighting of $1/Y_i$ and is desirable, rather than to be avoided as suggested in Grün and MacDonald (1989). Note Grün and Rhodes (1991) in estimating the optimal distribution of data points for the dose response curve fitted their models to log-transformed data, thereby reducing the effect of leverage by the high dose points. Their conclusions are therefore only valid for similarly weighted data. If curve fitting procedures are used which do not inversely weight the data, leverage by high dose points will be excessive and lower maximum doses should be used.

(b) provision of an adequate range of intermediate points. This will reduce the effect of leverage of high dose points where the model is adequate but the precision of the data poor. Intermediate points also indicate whether the model is appropriate (see below).

This discussion should not be taken to imply that high dose points are not of value. On the contrary, they are essential to evaluate the appropriateness of the model and to provide robustness to the extrapolation.

(ii) Influence.

The poor quality of a high leverage point is often masked because it tends to draw the fitted curve towards itself (see figs 1 & 2). This phenomenon is known as *influence*, and has become a major focus for statisticians

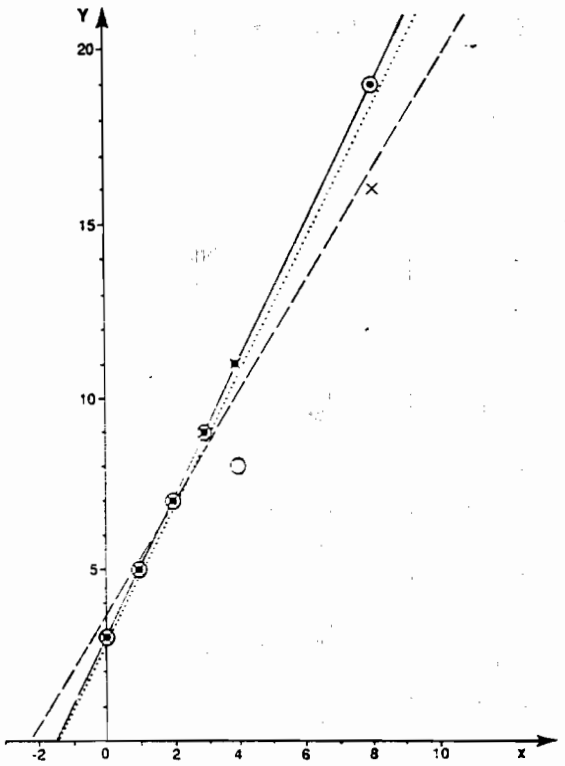

Figure 1.
Leverage in regression curve fitting. The solid line is the model line Y=2x+3, fitted to the infilled circles. Where an extreme point is "in error" (dashed line, fitted to the crosses), the regression line shows pronounced leverage with both AD and slope being significantly affected. Because an extreme point draws the line towards itself (influence), its poor quality is not necessarily obvious and it may not have a high residual. When a central point is in error, the effect on AD and slope is reduced, involving a sub-parallel shift instead of leverage (dotted line fitted to open circles). The effects of the same absolute or percentage change on the lowest point is given in table 1.

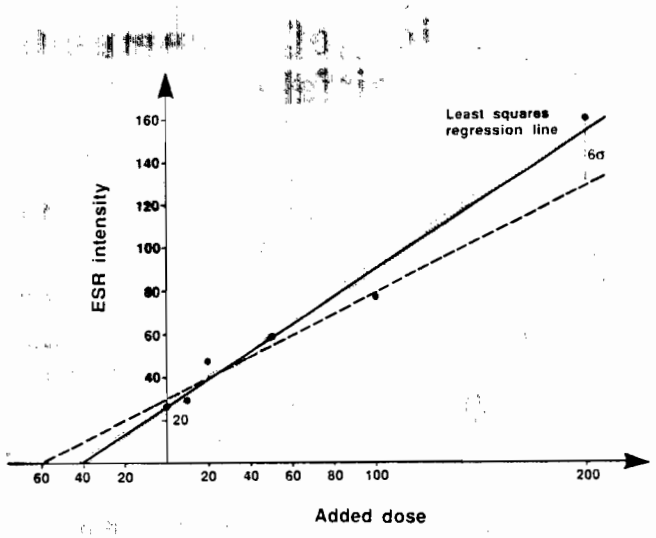

Figure 2.
Outliers produced by jackknifing reflect the quality of the associated jackknifed (omitted) point: when a poor quality high leverage point is omitted a high quality outlier is produced and vice versa. The data points have been produced for the model line (dashed line, y=30+0.5x, AD=60) with randomly generated errors, s=5, assigned to the Y values. To demonstrate the effect of leverage on a scale which can be graphically shown, the high leverage point has been deliberately assigned a value which places it 6σ from the model line (in fact, if such a point arose experimentally and its deviation from the line were known, it would no longer be regarded as a member of the data set). However, note that the regression line to all data points (solid line) passes close to the poor quality point by reason of its influence; the quality of a point cannot be determined by visual examination nor by residuals (table 2) but must be independently known. Omitting the poor quality high leverage point in the jackknifing procedure produces an outlier which lies too close to the model line to be distinguished on this scale. The dotted lines are the jackknife fits when the other points are omitted and, although closely grouped, give poor estimates of the model parameters.

in the last decade (Cook and Weisberg, 1982). An important consequence of influence is the difficulty of determining whether high leverage points are 'good' and hence whether the associated estimates are biased.

A plot of the fitted curve together with the data points serves only as a first visual test of the goodness of fit. Because of influence, an apparently good fit is, in itself, no guarantee of the quality of any particular data point nor that the best curve has been fitted. Mathematically, this means that a high or low residual cannot be used to identify a poor or high quality data point, respectively (table 2). Any attempt to 'improve' a data set by eliminating high residual point(s) is at best misguided and potentially disastrous. Thus, tests such as Chauvenet's test for identifying outlying data points (Misaqi, 1975) are not sufficient grounds for rejecting high leverage data points, as these tests essentially depend on some form of evaluation of residuals. The quality of a high leverage data point should be assessed independently of its relationship with the fitted curve, for example, by its variance (which can then be included as a weighting factor) or other experimental evidence.

Assessing model aptitude

The choice of model is of importance in any application involving extrapolation. Models which are satisfactory for interpolation (such as high order polynomials) may be completely unsatisfactory for extrapolation. It is not intended here to discuss the appropriateness of any particular model, but simply to identify some relevant statistical procedures and pitfalls. Note that Readhead (1984, 1988) and Prescott (1983) propose a regenerative method of estimating AD in TL dating in which the curve fitting utilizes interpolation rather than

extrapolation; it is therefore much less vulnerable to incorrect model choice.

The phenomenon of leverage can be used informally to test the appropriateness of a model. If the AD estimated by the model shows a systematic trend as additional higher value points are progressively added to the calculation, then the model is suspect, and the effect of leverage may be extremely serious. (This technique is used by Grün (1990) to demonstrate the inappropriateness of the simple saturating exponential model to describe the growth curve of the g=2.0007 centre in corals.)

Just as a low residual at a single point is not necessarily an indicator of a 'good' point, neither is a low residual sum of squares (SSE) an indicator of a 'good' model, as is sometimes assumed. Introducing more parameters into a model will *always* reduce the residual sum of squares, because SSE is an estimate of the variation in the data left unexplained by the model - the more complex the model, the closer the fit and the "better" the correspondence. A better indicator of the fit of a model is the mean sum of squares (MSE) , as this allows both for the number of parameters estimated and the number of data points:

$$MSE = SSE/(n-p)$$

where,

SSE = residual sum of squares (error sum of squares)
n = no of data points
p = no of parameters fitted (incl. intercept term)

An equally useful indicator is the root mean square error, or the standard error of estimate (RMS), which is simply the square root of MSE but has the advantage that it is in the same units as the dependent variable.

The most common method to assess model adequacy, which takes into account all of the above, is to consider the ratio of MSR/MSE, where:

$$MSR = SSR/(p-1)$$

where,

SSR = regression sum of squares
MSE = mean sum of squares, as above.

This ratio has an F distribution, F(p-1, n-p). Thus the F-statistic (or the associated probability of rejection) is a useful measure for comparison of models. The Student's t-test is the corresponding test for the linear case.

Estimation of uncertainty in AD

The estimate of uncertainty in AD is a particularly thorny problem. Because the data for experimentally determined additive growth curves may not satisfy some of the basic statistical assumptions, caution is needed in applying standard statistical techniques.

Four methods have been suggested in the ESR and TL literature: regressing x on Y (Rendell, 1985); quasi-likelihood techniques (Berger *et al.*, 1987); constrained linear regression (Lyons, 1988); and jackknifed regression (Grün and MacDonald, 1989). Which method is considered the most appropriate will depend on the nature of the data.

The first of these methods, regressing x on Y, contravenes the basic assumption that the error is associated only with the dependent variable and that the independent variable is error-free. As a result the 'regression may be so greatly changed as to be meaningless' (Williams, cited in Rendell, 1985). While quasi-likelihood methods are statistically established techniques and their use is to be recommended where possible, it is also true that they are generally 'computationally demanding' (Seber, 1977) and may place unrealistic demands on both the quantity and quality of the data, requiring 'at least 15 points (Berger *et al.*, 1987), each with an uncertainty of <5%): meeting these requirements must be carefully assessed. Constrained linear regression (Lyons, 1988), which forces the regression line to pass through the natural signal intensity, does not have these restrictions on data quality and quantity but requires the natural signal to be extremely well-defined. It is only applicable to a linear model and thus its usefulness is limited. Re-sampling techniques such as jackknifing (see below), may be applied to any model to obtain uncertainty estimates: their validity with respect to regression is, however, the topic of debate because two basic statistical assumptions may not be satisfied by regression data:

(i) independent random data points. This is untrue for additive growth curve data, where the independent variable (added irradiation) is not randomly selected.
(ii) equal variance, i.e. the total *absolute* experimental uncertainty is the same for each point regardless of its magnitude. While this assumption is supported by some (Franklin, 1986; Scott and Sanderson, 1988), others would regard a constant *percentage* error for each data point as more reasonable (Berger, 1984; 1987).

Re-sampling techniques

Re-sampling techniques require constructing subsets of data from the initial complete data set and noting how the parameter of interest varies with the different subsets. For example, in 'simple delete-one jackknifing', as used in the program FITT supplied by Grün and MacDonald, the subsets of the data are obtained by omitting each data point in turn from the total set. The uncertainty in AD is calculated from the resulting distribution of AD_i. (While this description is conceptually apt, it should be noted that the mathematical theory and execution is more complex.)

Reasonably comprehensible descriptions of re-sampling techniques for the general user have been written by

Miller (1974) and Efron (1982). More sophisticated treatments can be found in journals such as the *Annals of Statistics* and the *Journal of the American Statistical Association*. In particular, Vol 14 No. 4, *Annals of Statistics*, contains a substantive invited paper by Wu (1986), including a comprehensive reference list, and a series of related discussion papers.

Re-sampling techniques such as jackknifing have wide application in experimental analysis. However, most of the theoretical work on re-sampling is based on the two assumptions outlined above and thus is not necessarily valid for regression (Wu, 1986; Carroll and Rupert, 1986; Hinkley, 1977). The application of resampling techniques to regression data is currently the subject of much debate in statistical research (see *Annals of Statistics*, Vol.14, No. 4) and 'more work is needed and is being done in this area' (Shao, 1986). Points of concern are potential bias in the parameter estimates, possibly unduly inflated (conservative) or skew estimates of the associated uncertainties (Hall, 1986), underestimation of uncertainties (Duncan, 1978, cited in Seber and Wild, 1989), and the disproportionate effect of outlying data points on the results. Re-sampling techniques can also be computationally demanding, and methods are being developed to bring the computing requirements within manageable limits for large data sets (not a problem with TL and ESR applications!) It should be remembered that statisticians almost always work with very large data sets, which tends to minimize the problems mentioned above. For the small data sets used in AD determinations, the computational demands are modest but these problems may be of much greater practical significance (Hinkley, 1977).

Developments in re-sampling techniques for estimating uncertainties in regression applications are significant. Wu (1986) suggests the use of a 'weighted sharpened delete-n jackknife' to cope with some of the problems raised by the the nature of regression data and non-satisfaction of the above assumptions. In this method, instead of only one point being omitted to form each

Table 1.
Leverage due to position: effects on AD are much greater for extreme points. Model line is Y = 2x + 3. The second set of figures for the central and bottom points have the same percentage deviation from the true Y value as the top point.

	"Bad" pt	Change in y	y-intercept	Slope	AD	% change in AD
model	none	0	3	2	1.5	
top	(8, 16)	-3	3.63	1.63	2.23	+49
middle	(4, 8)	-3	2.73	1.93	1.42	-5
	(4,9.26)	16%	2.84	1.96	1.45	-3
bottom	(0, 0)	-3	1.83	2.23	0.82	-45
	(0,2.53)	16%	2.82	2.04	1.38	-8

Table 2.
Residuals are poor indicators of data quality because of influence of extreme points. The model and data are as in fig. 2. AD_i are the jackknife estimates omitting the ith point. AD* is calculated from the average slope and the average Y-intercept from the jackknifed slope and intercept estimates. a) extreme high leverage point , poor quality but with a low residual, AD_{200} is an outlier and the closest estimate of AD, AD* is a very poor estimate of AD b) extreme point is very high quality (lies on the model line), AD_{200} is an outlier and the worst estimate of AD. Note that, in the absence of information on the quality of a high leverage point, the best estimate of AD will be obtained by fitting the data to all the points, suitably weighted to allow for leverage, as discussed in the section on the Application of the delete-one jackknife. Note also that the jackknife estimates of the uncertainty in AD given here, are substantially higher than those which would be calculated by the standard formula for independent data set because the resampled data sets used in jackknifing are not independent.

x	0	10	20	50	100	200
e_i	-3.5	-5.7	7.5	4.0	-2.8	30.0
residual	0.66	-2.98	8.79	0.99	-12.97	5.50
AD_i	39.7	42.5	34.6	39.8	41.1	58.9
	AD_{model} = 60.0	AD_{all} = 40.2		AD_{av} = 42.7 ± 17.0		AD* = 35.3
e_i	-3.5	-5.7	7.5	4.0	-2.8	0
residual	3.21	-6.08	2.54	3.52	5.85	-1.99
AD_i	55.7	66.2	57.3	61.4	59.4	51.7
	AD_{model} = 60.0	AD_{all} = 59.6		AD_{av} = 58.7 ± 10.1		AD* = 58.5

subset as in the delete-one jackknife, several (n) points selected at random are omitted. This increases the independence of the subsets and the estimated experimental uncertainty is then largely independent of the underlying theoretical distributions. It should also be readily extendable to non-linear situations (Rao and Prasad, 1986). The weighted sharpened delete-n jackknife is conceptually equivalent to the common practice in physics and chemistry, where the uncertainty of the estimate of a parameter is calculated from repeated determinations (i.e. from a number of independent data sets) in accordance with the central limit theorem.

Although the relative merits of the various methods of jackknifing and other re-sampling techniques are still a matter for discussion, progress is substantial and very promising (Shao and Wu, 1989).

Application of the delete-one jackknife

Meanwhile, even though the application of the simple delete-one jackknife to regression problems is not universally accepted by the theoreticians, it may well be argued that 'something is better than nothing'. Certainly, the availability and convenience of the FITT program for simple delete-one jackknifing, supplied by Grün and MacDonald, is a strong argument for its use. However, bearing in mind the concepts discussed above, and the emphatic comment by Wu (1986) that 'routine and blind use of any re-sampling technique is problematic', several points arise:

1) Treatment of outliers.

The problem of leverage may lead to outliers, or values which are substantially different to most of the values, in the set of AD_i generated and hence to highly dubious estimates of both AD and the uncertainty in AD. Depending on the quality of the point omitted to obtain the outlying estimate, the outlier will represent either the best (poor quality point omitted) or the worst (high quality point omitted) estimate of AD (fig. 2 & table 2). It is essential to evaluate the quality of the outlier-producing point independently (NB it cannot be assessed simply from examination of the residual) in order to interpret the jackknife data correctly. A point must not be discarded simply because it produces an outlier AD estimate unless its exclusion is justified on other grounds. The outlying estimate may be the closest estimate of AD if it arises from excluding a poor quality high leverage point.

The above comments are not an objection to identifying and, possibly, eliminating outliers, nor to the sometimes large estimates of uncertainty that may arise if the data set itself contains outliers, as was mistakenly interpreted by Grün (1990). Rather, we are concerned that outliers in the resulting estimates of AD and curve parameters should be closely scrutinized for their significance.

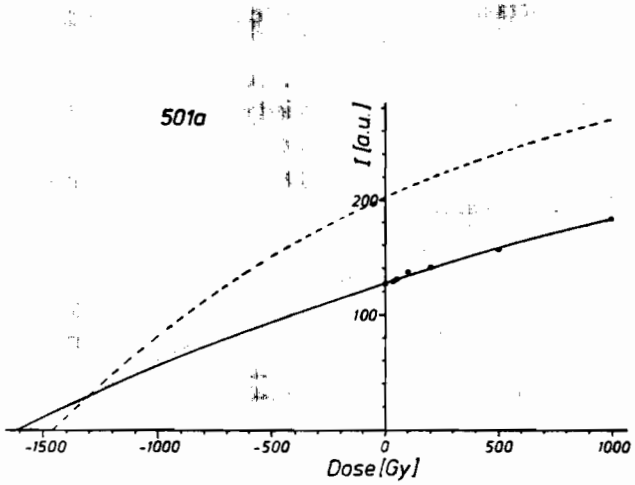

Figure 3.
Estimation of AD from jackknife estimates and the effect of duplicating points (from Grün and MacDonald, 1989; the labelling of the 3 lines has been corrected, with permission). The solid curve is the best fit to all points. The parameters obtained by averaging the estimates from jackknifing for each curve parameter may give rise to a curve with little correspondence to the data points (dashed curve). The dotted curve is the result of averaging the estimates from jackknifing after adding one additional (assumed) point at 1 kGy. Contrary to Grün and MacDonald's statement, this "improvement" is not due to the "(better) choice of radiation steps", but to the forced inclusion of the 1 kGy value, by duplication, in all the jackknifed estimates i.e. by the elimination of the corresponding outlier.

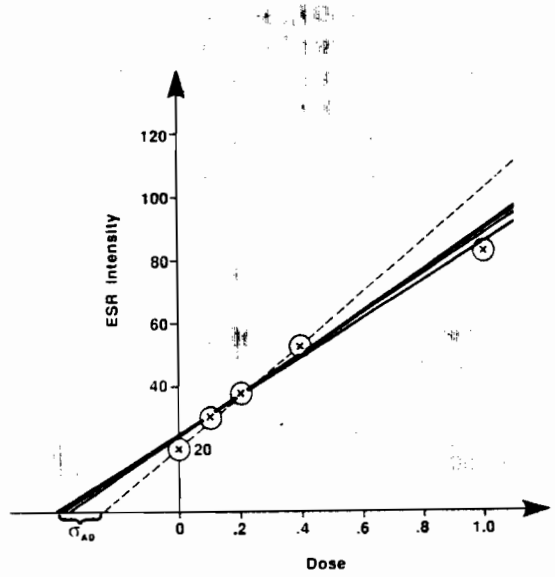

Figure 4.
Model aptitude cannot be tested by the effect of duplication on jackknife uncertainty estimates. The data is saturating exponential but the fitted model is linear; the extreme point gives rise to an outlier. Duplicating this point (or including another high value point close to it) effectively prevents its elimination and will reduce s_{AD} without improving the fit; s_{AD} from jackknifing cannot be used to determine model fit.

The vulnerability of the jack-knife technique to outliers in the parameter estimates and the effective weighting it gives to extreme points when extrapolation is being used, mean that it must be applied with caution, and due consideration given to appropriate weighting factors. Its use should also be restricted to high quality data if the resulting estimates of uncertainty are to be realistic.

2) Estimation of AD.

The best estimate of the curve parameters should be calculated by fitting all the data. It should *not* be derived by numerically averaging the estimates of the curve parameters obtained in the jackknifing procedure (e.g average Y-intercept/average slope for a linear model), as these variables are not independent. Nor should it be obtained from the average of the AD_i from jackknifing (table 2) because these AD_i are not of equal value due to leverage effects. For example, a poor quality high leverage point may lead to severe bias in the estimate of AD, because it will give an AD significantly different from the true estimate. This will have a disproportionate effect on the average value of AD, both because of the size of the discrepancy and because it is inherently a poor quality estimate and therefore should not receive equal weighting in the averaging process (Carroll and Rupert, 1986).

If either of these averaging methods is used, the AD estimate may not only be biased but the resulting curve an obviously poor fit, as shown in fig. 3 from Grün and MacDonald (1989). More sophisticated averaging techniques have been proposed to deal with these difficulties: Hinkley (1977) and Siminoff and Tsai (1986) propose using 'pseudovalues' to downweight high leverage data when calculating the jackknifed estimates. These effectively weight the data to allow for leverage effects when each jackknifed estimate is calculated, thus reducing the bias. The problem is more complex when the data points have unequal variances (Freedman, 1986), as eliminating a high variance point yields an inherently 'better' estimate than one where a low variance point has been omitted, but should be amenable to the same technique of pseudovalues using the Fisher information matrix (Wu, 1986).

3) Replication and distribution of data points.

The magnitude of the uncertainty estimated by jackknifing depends heavily on the distribution of the data points. Replication of points will reduce the estimate markedly, and spuriously, as it effectively prevents the elimination of the replicated data point in the jackknifing procedure (fig. 4). Instead of including any replications individually when jackknifing is used to estimate uncertainty in AD, they should be averaged and the resulting single point given an increased weighting because of its reduced uncertainty. Note that if other techniques, such as quasi-likelihood methods, are used, averaging of replicated points is not required. Indeed, except when resampling techniques are used, maximum value is obtained from the data if all points are included in the curve fitting process individually

Any apparent reduction in AD uncertainty as estimated due to duplicating data will be largely spurious and the AD estimate itself will be biased. The 'demonstration' in Grün and MacDonald (1989) of dramatically improved uncertainty limits and 'better' curve fitting as a result of duplication is incorrect and seriously misleading (fig. 3). On the contrary, the example demonstrates, not that there has been a 'wrong choice of radiation steps', but that a good distribution of points over the whole range of the independent variable is vital to enable the validity of the model to be assessed. Even without duplication at precisely the same dose, a low AD uncertainty will be estimated by jackknifing if there is more than one high leverage point with a high precision, no matter how inappropriate the model is. The simple delete-one jackknife itself cannot be used to assess the validity of a model, as even a very poor model will give a small uncertainty in AD provided the experimental uncertainty on each point is low. (A test for model aptitude has been given in a previous section.)

Cautionary measures

The following procedures should help to avoid some of the serious pitfalls in curve fitting and error estimation outlined above.

(i) A sufficient number of data points should be determined to allow for unusual points to be identified and the response curve to be well-defined. It would be unlikely that reliable results would be obtained with fewer than 8-10 observations (3-4 per parameter fitted), although this is very much dependent on the complexity of the model, the quality of the data and the method of analysis employed.

(ii) Leverage should be taken into account by concentrating data in the lower dose range, with some points at higher doses to give an indication of how the dose response curve behaves at higher levels. This concentration is particularly effective if the data has a strong linear component at lower doses but departs from linearity at higher doses, as for example in the linear/exponential model proposed by Levy (1989). Where the model is well defined, precise high dose data is extremely valuable, but where the model is less reliably known, even very precise high dose data should be downweighted to counteract the effects of leverage (Fox et al., 1980).

(iii) Allowance should be made for unequal variance in the Y-values, if it is considered significant. For example, the data can be weighted by the reciprocal of the variance for each data point, if it is known. Transforming the data by taking the log of the dependent variable and fitting a linear regression to the transformed data is mathematically equivalent to weighting an exponential data set for a constant percent error (de Levie, 1986).

(iv) The natural signal intensity should be evaluated as reliably as possible and weighted heavily in the curve fitting, as this point is fundamental to the concept of ESR and TL dating.

(v) The data should be checked for general form. Only if the data are linear (see Franklin, 1986) and natural signal intensity sufficiently well defined, can constrained weighted linear regression be used. The importance of the choice of model is reinforced by the example given in Grün and MacDonald (1989), who caution that apparently linear data may give very different results when fitted with an exponentially saturating curve, although the difference is probably not so severe if a combined linear/exponential curve is fitted.

(vi) If the data are of sufficiently high quality and quantity (>15 points, <5% uncertainty), quasi-likelihood methods as described by Berger *et al.* (1987), may be used to fit either linear or non-linear models.

(vii) If the data are non-linear and its quality/quantity deficient for likelihood techniques, an estimate of the uncertainty in AD can be obtained by the jackknife technique. Its interpretation should include a careful consideration of the potential problems of influence and leverage, and appropriate weighting to minimize their effects, particularly if the data have unequal variances. If jackknifing is used to estimate the uncertainty in AD, averaged values should be used for replicated points, and the variance in the mean value be included in the weighting. The jackknifed AD_i estimates should be checked for outliers, their significance assessed and any implications for calculating the best estimate considered. AD itself should be derived from the best fit to all data points and not by averaging the jackknife estimates of parameters.

(viii) In all curve fitting, particularly jackknifing, the data should be plotted along with their fitted curve to check for oddities, bearing in mind the possible effects of influence. Whether a significantly better fit has been achieved by using a model with more parameters can be checked by using the tests outlined above in the section on assessing model aptitude. It should be noted that even if a more complex model fits the curve more closely, extrapolation may still not be valid (e.g. if polynomials are fitted): for extrapolation, in contrast to interpolation, it is essential that the curve form expresses the underlying physical properties.

It is important to be pragmatic about the power of the techniques chosen: there are times when the quality of data may not justify the use of sophisticated techniques whose basic assumptions may not be satisfied, and less demanding techniques, even visual fitting, may be appropriate. The major source of systematic error is still likely to be the use of an incorrect model, and this cannot be detected by routine uncertainty calculations. Thus, at this stage, major progress in improving the reliability and accuracy of ESR and TL ages estimates is most likely to arise from theoretical advances in our knowledge of the appropriate curve forms to use.

Acknowledgements

We are grateful to Dr Rainer Grün for making the FITT progam available and initiating the use of re-sampling techniques to address a difficult problem in ESR and TL dating. We would like to thank our colleagues in the Department of Statistics, University of Auckland, particularly Prof. George Seber, Prof. Alistair Scott and Dr Alan Lee, as well as Dr Robert Gentleman, from McMaster University, for helpful discussions on the statistical principles involved. Thanks are also due to Prof. John Prescott for his meticulous refereeing; his comments have substantially clarified the arguments in this paper.

References

Apers, D., Debuyst, R., DeCanniere, P., Dejehet, F. and Lombard, E. 1981 A criticism of the dating by electron paramagnetic resonance (ESR) of the stalagmitic floors of the Caune de l'Aragoat Tautavel. In *Absolute Dating and Isotope Analysis in Prehistory- Methods and Limits* (eds DeLumley, H. and Labeyrie, J.) *Proc Pretirage*, 533-550

Berger, D.J. and Huntley, D.J. (1986) Linear regression of TL data. *Ancient TL* 4, 2, 26-29

Berger, D.J., Lockhart, R.A. and Kuo, J. (1987) Regression and error analysis applied to dose-response curves in thermoluminescence dating. *Nuclear Tracks and Radiation Measurements* **13**, 177-184

Berger, D.J. (1990) Regression and error analysis for a saturating-exponential-plus linear model. *Ancient TL* **8**, 3, 23-25

Cook, R.D. and Weisberg, S. (1982) *Residuals and influence in regression*. Chapman and Hall, NewYork

Carroll R.J. and Ruppert, D. (1986) Discussion on 'Jackknife, bootstrap and other resampling methods'. *The Annals of Statistics* **14**, 4, 1298-1300

De Levie, R. (1986) When, why, and how to use weighted least squares. *J. of Chemical Education* **63**, 1, 10-13

Efron, D. (1982) The jackknife, bootstrap and other resampling plans. CBMS-NSF Reg. Conf. Ser in *Appl. Math. No 38. SIAM*, Philadelphia

Fox, T., Hinkley, D., and Larntz, K. (1980) Jackknifing in nonlinear regression. *Technometrics*, **22** 29-33.

Franklin, A.D. (1986) Extrapolation errors in linear regression. *Ancient TL* 4, 2, 31-35.

Freedman, D.A. (1986) Discussion on 'Jackknife, bootstrap and other resampling methods', *The Annals of Statistics* **14**, 4, 1305-1307.

Grün, R. and MacDonald, P.D.M. (1989) TL/ESR dose response curves. *Applied Radiation and Isotopes* **40**, 10-12, 1077-1080.

Grün, R. and Rhodes, E.J. (1991) On the selection of dose points for saturating exponential ESR/TL dose response curves *Ancient TL* **9**, 3, 40-46

Hall, P. (1986) Discussion on 'Jackknife, bootstrap and other resampling methods', *The Annals of Statistics* **14**, 4, 1311.

Hinkley, D. V. (1977) Jackknifing in unbalanced situations. *Technometrics* **19**, 285-292.

Miller, R.G. (1974) The jackknife- a review. *Biometrika* **61**, 1-15.

Misaqi, F.L. (1975) Applications of statistics to radiation surveys in mines. *MESA Information Report 1020*, US Dept of Interior, Washington.

Levy. P.W. (1989) Principles determing the length of time materials can be dated by TL, ESR and other trapped charge build-up methods. In *Occasional Publication no 9*, The Res. Lab. for Archaeology and the History of Art, Oxford University, pp24-29.

Lyons, R.G. (1988) A simple statistical method for comparison of accumulated dose in ESR dating. *Nuclear Tracks and Radiation Measurements* **14**,1-2, 361-363

Prescott, J.R. (1983) Thermoluminescence dating of sand dunes at Roonka, South Australia. *PACT* **9**, 505-512.

Rao, J.N.K. and Prasad, N.G.N. (1986) Discussion on 'Jackknife, bootstrap and other resampling methods', *The Annals of Statistics* **14**, 4, 1320-1321.

Readhead, M.L. (1984) *Thermoluminescence dating of some Australian sedimentary deposits* Unpub PhD thesis Australian National University.

Readhead, M.L. (1988) Thermoluminescence dating study of quartz in aeolian sediments from South-Eastern Australia *Quaternary Science Reviews* **7**, 257-264.

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

Shao, J. (1986) Discussion on 'Jackknife, bootstrap and other resampling methods', *The Annals of Statistics* **14**, 4, 1322-1325.

Scott, E.M. and Sanderson, D.C.W. (1988) Statistics and the additive dose method in TL dating. *Nuclear Tracks and Radiation Measurements* **14**,1-2, 345-354.

Seber, G.A.F. (1977) *Linear regression analysis*. Wiley and Sons, New York.

Seber, G.A.F. and Wild, C.J. (1989) *Nonlinear regression*. John Wiley and Sons, New York.

Siminoff, J.S. and Chih-Ling Tsai. (1986) Discussion on 'Jackknife, bootstrap and other resampling methods', *The Annals of Statistics* **14**, 4, 1326-1328.

Wu, C.F.J. (1986) Jackknife, bootstrap and other resampling methods *The Annals of Statistics* **14**, 4,1261-1294.

PR. J.R. Prescott