

A Bayesian Approach to the Evaluation of Equivalent Doses in Sediment Mixtures for Luminescence Dating ¹

D. S. Sivia^{*}, C. Burbidge[†], R. G. Roberts^{**} and R. M. Bailey[‡]

^{*}*Rutherford Appleton Laboratory, Chilton, Oxon, OX11 0QX, UK*

[†]*Scottish Universities Environmental Research Centre, Glasgow, G75 0QF, UK*

^{**}*School of Earth and Environmental Sciences, University of Wollongong, NSW 2522, Australia*

[‡]*School of Geography and the Environment, University of Oxford, OX1 3TB, UK*

Abstract. The optically stimulated luminescence from minerals is proving to be a very useful dating technique in archaeology and physical geography. In this work we study the analysis of the relevant data from a Bayesian viewpoint, comparing some simple age and noise models.

INTRODUCTION

Carbon dating is a well-established technique in archaeology [1], but its use is conditional upon the availability of suitable organic material and is limited to the last 60,000 years. Luminescence from minerals, when thermally or optically stimulated, provides an alternative dating procedure. It relies on the release of energy from trapped electrons accumulated, over time, through excitations induced by the ionising radiation in the environment [2]. The strength of the luminescence signal indicates when a piece of pottery was last fired, or a sediment exposed to sunlight prior to deposition [3], for example, assuming that the ‘bleaching’ process fully reset the mineral clock to zero; an environmental radiation study is also required to convert the measurements into actual dates.

We consider just one aspect of the problem which can be phrased generically as follows: Given N measurements $\{x_k\}$, of the ‘equivalent dose’ of laboratory radiation D_e , with associated error-bars $\{\epsilon_k\}$, and background information and assumptions I , what is the underlying (age) distribution $F(x)$ of the sample? By $F(x)$ we mean that the probability density that the ‘true’ value for *aliquot* k is \hat{x}_k , $\Pr(\hat{x}_k|F(x), I)$, is $F(\hat{x}_k)$. It can be related to the measurement x_k by using marginalisation and the product rule of probability [4]:

$$\Pr(x_k|\epsilon_k, F(x), I) = \int \Pr(x_k, \hat{x}_k|\epsilon_k, F(x), I) d\hat{x}_k = \int \Pr(x_k|\hat{x}_k, \epsilon_k, I) F(\hat{x}_k) d\hat{x}_k, \quad (1)$$

¹ Presented at the 24th international workshop on *Maximum Entropy and Bayesian Methods* in science and engineering, **Garching bei München, Germany, 25-30 July, 2004**, and to be published in the American Institute of Physics Conference Proceedings.

where unnecessary conditioning symbols have been dropped on the far right, and the definition of $F(x)$ substituted.

In this paper, we analyse published test data [5] using a mixture model:

$$F(x) = \sum_{j=1}^M A_j h_j(x), \quad (2)$$

where A_j is the relative contribution from dose component j , $h_j(x)$, so that $\sum A_j = 1$, and M is a ‘few’. In particular, we compare alternative simple choices for the ‘intrinsic shape function’, $h_j(x)$, and different assumptions about the noise characteristics, $\{\epsilon_k\}$.

THE INTRINSIC SHAPE FUNCTION

The integral of Eq. (1) is easy if $h_j(x) = \delta(x - \mu_j)$. With this δ -function choice and the usual assumption of independent Gaussian noise, of variance ϵ_k^2 , the likelihood function for the $\{x_k\}$ becomes a product of the N terms for the individual x_k :

$$\Pr(x_k | \epsilon_k, \{A_j, \mu_j\}, M, I) = \sum_{j=1}^M \frac{A_j}{\epsilon_k \sqrt{2\pi}} \exp\left[-\frac{(x_k - \mu_j)^2}{2\epsilon_k^2}\right], \quad (3)$$

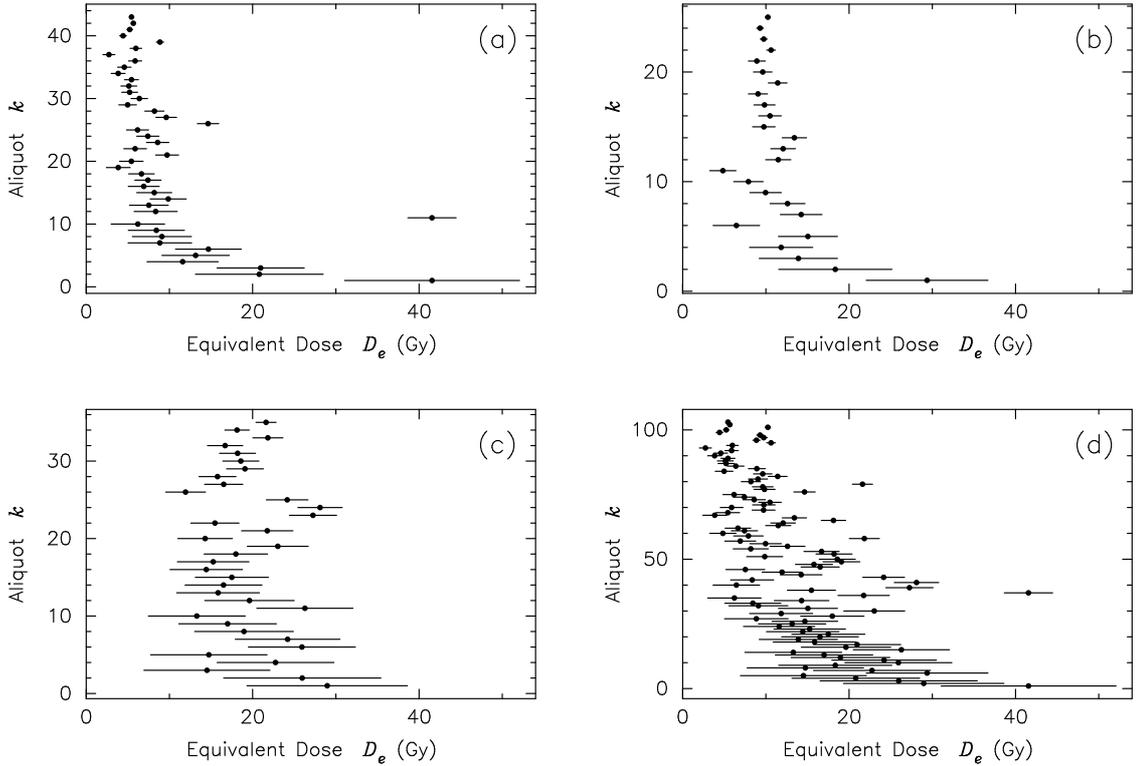


FIGURE 1. (a)-(c) Data obtained by irradiating quartz grains with a laboratory beta-dose of 5, 10 and 20 Gy respectively; (d) the amalgamated measurements for the 103 (= 43 + 25 + 35) grains.

TABLE 1. The evidence for the number of δ -function components, M , in Fig. 1.

M	(a)	(b)	(c)
1	-243.3	-62.4	-112.0
2	-174.2	-63.7	-109.0
3	-129.0	-65.0	-109.5
4	-125.0	-66.0	-110.1
5	-123.8	-67.4	-110.5

The only pdf still outstanding is the prior for the model parameters, $\Pr(\{A_j, \mu_j\}, M|I)$. The simplest assignment is a uniform one in a suitable range: $0 \leq A_j \leq 1$, with the sum subject to normalisation, $0 \leq \mu_j \leq D_{max} \sim 50$ Gy and $1 \leq M \leq 5$ (say). The best estimate of the D_e values and proportions of the age components is given by the allowed maximum of the likelihood function if their number is taken as being known, at least in the ‘quadratic approximation’, whereas the probabilistic evidence for M itself is given by the average value of the likelihood [4].

Figure 1(a)-(c) shows three sets of data obtained by irradiating well-bleached quartz grains with doses of 5, 10 and 20 Gy, respectively, from a laboratory source of beta rays; (d) is their amalgamation [5]. Since these test measurements pertain to single dose components, by design, their spread is indicative of the intrinsic shape function. That is to say, if we analyse each of Fig. 1(a)-(c) with an acceptable choice for $h_j(x)$ in Eq. (2), but an unknown number of components, then we ought to find most evidence for $M=1$. Carrying out the calculation for the δ -function case of Eq. (3), we find that the noise is not sufficient to account for the scatter in Fig. 1(a) but is adequate for the measurements in (b) and (c). The formal results for $\ln[\Pr(\{x_k\}|M, \{\epsilon_k\}, I)]$ are given in Table 1.

Since a δ -function model, and the noise in the measurements, cannot account for the spread in Fig. 1(a), a reasonable way of proceeding is to try an intrinsic shape function that has an inherent width. The simplest of these is a Gaussian, because the integral of Eq. (1) is then still straightforward; Eq. (3) just becomes

$$\Pr(x_k|\epsilon_k, \{A_j, \mu_j, \sigma_j\}, M, I) = \sum_{j=1}^M \frac{A_j}{\sqrt{[2\pi(\epsilon_k^2 + \sigma_j^2)]}} \exp\left[-\frac{(x_k - \mu_j)^2}{2(\epsilon_k^2 + \sigma_j^2)}\right], \quad (4)$$

where σ_j^2 is the variance of the j th component. As a first attempt, we could set all the widths equal to a constant, σ , or even a constant fraction of the equivalent doses, $\sigma\mu_j$. Indeed, the latter option was chosen in the original work [5] by using a fixed width in $\ln(D_e)$. While working on a logarithmic axis has merit, as the equivalent dose appears to be a scale parameter, there is a competing argument: since a difference in age of one hundred years can be just as important on an absolute scale of a thousand years as it is on ten thousand (say), the D_e display characteristics of a location parameter. There is also the mathematical difficulty of taking logarithms of negative D_e when dealing with very noisy measurements, especially since some zero-age grains are not unexpected in natural samples. We opted to allow variable widths, in the range $0.5 \leq \sigma_j \leq 5$ Gy, with a uniform prior, and used a Markov Chain Monte Carlo (MCMC) algorithm to handle the strong non-linearities in the calculation [6]. The probabilistic evidence for the number

TABLE 2. The evidence for the number of Gaussian components, of variable widths, in Fig. 1.

M	(a)	(b)	(c)	(d)
1	-147.2	-63.5	-107.2	-358.9
2	-125.6	-65.0	-108.1	-348.5
3	-124.5	-67.1	-109.2	-346.2
4	-124.8	-68.7	-110.2	-345.1
5	-125.7	-70.1	-111.5	-344.6

of components in the data of Fig. 1, $\ln [\Pr (\{x_k\} | M, \{\epsilon_k\}, I)]$, is given in Table 2. An estimate of the equivalent dose distribution $F(D_e)$, or $F(x)$, for the data of Fig. 1(d) is shown in Fig. 2(a). It was obtained by summing up the Gaussian contributions from all the MCMC samples displayed in Fig. 2(b). The width associated with each point in this two-dimensional ‘cloud plot’ is not shown, of course, and the number of components was integrated out by our MCMC algorithm using the prior $\Pr (M | I) \propto 1/M!$ for $M \geq 1$; this Poisson assignment, with mean unity, simply reflected our expectation that M was no more than a few without imposing an explicit upper bound.

A ROBUST NOISE MODEL

Although a Gaussian model returns a far higher evidence for $M = 1$ for the data of Fig. 1(a) than does a δ -function, it’s still not adequate as far as the noise is concerned; at least two are preferred. Admittedly, the second component merely makes a low-level contribution to $F(x)$ at higher equivalent doses and its need might be eliminated by using an asymmetric intrinsic shape function; for example, with a log-normal distribution (as in [5]). This seems odd, however, as the data of Fig. 1(b) are slightly better explained by Eq. (3) than Eq. (4) and because a single δ -function is not too unreasonable for Fig.

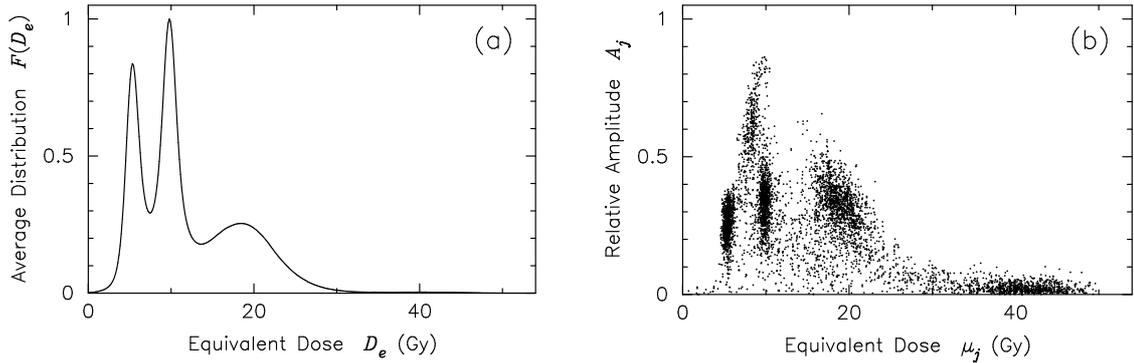


FIGURE 2. (a) The equivalent dose distribution for the data of Fig. 1(d), scaled vertically to a maximum value of unity, obtained by averaging over the parameters of the Gaussian components model. (b) The corresponding MCMC samples of $\{A_j, \mu_j\}$, with the $\{\sigma_j\}$ suppressed and M marginalised out (using a Poisson prior with $\langle M \rangle = 1$).

1(c). It appears that something quirky occasionally happened in the measurements of Fig. 1(a), but we don't really understand what. Rather than opting for a more elaborate $h_j(x)$, therefore, we could revert to the simpler δ -function choice and try a noise model of greater robustness.

As in [7], let's make the conservative assumption that the quoted error-bars represent the most optimistic estimate of the uncertainties; that is to say, the $\{\varepsilon_k\}$ will be treated as lower bounds of the 'true', but unknown, Gaussian noise $\{\hat{\varepsilon}_k\}$. The latter can be eliminated from the analysis, in the usual fashion, through the use of marginalisation and the product rule:

$$\Pr(x_k|\hat{x}_k, \varepsilon_k, I) = \int \Pr(x_k, \hat{\varepsilon}_k|\hat{x}_k, \varepsilon_k, I) d\hat{\varepsilon}_k = \int \Pr(x_k|\hat{\varepsilon}_k, \hat{x}_k, I) \Pr(\hat{\varepsilon}_k|\varepsilon_k, I) d\hat{\varepsilon}_k, \quad (5)$$

where unnecessary conditioning symbols have been dropped on the far right and, as before, \hat{x}_k is the true value for aliquot k . In the previous section, where the given error-bars were believed, we had implicitly assigned $\Pr(\hat{\varepsilon}_k|\varepsilon_k, I) = \delta(\hat{\varepsilon}_k - \varepsilon_k)$. To express complete ignorance about $\hat{\varepsilon}_k$, a scale parameter argument would suggest that a Jeffreys' prior should be used (as long as $\hat{\varepsilon}_k \geq \varepsilon_k$); this was done in [7]. Being improper without an upper bound on $\hat{\varepsilon}_k$, however, this is not suitable for model selection. So, we opted to assign

$$\Pr(\hat{\varepsilon}_k|\varepsilon_k, I) = \frac{\varepsilon_k}{\hat{\varepsilon}_k^2}, \quad (6)$$

for $\hat{\varepsilon}_k \geq \varepsilon_k$, and zero otherwise, as the closest form to the Jeffreys' prior that is easy to integrate and normalise. In conjunction with a Gaussian pdf for $\Pr(x_k|\hat{\varepsilon}_k, \hat{x}_k, I)$, Eq. (5) leads to the likelihood function $\Pr(x_k|\hat{x}_k, \varepsilon_k, I)$ shown by the solid line in Fig. 3; for comparison, the corresponding pdf with the Jeffreys' prior is plotted with a dotted line and the Gaussian, where the error-bars are believed, is drawn with a dashed line. The slowly decaying tails offer protection against the skewing effect of 'outliers' since the penalty for a large misfit is less severe. With δ -functions for $h_j(x)$, Eqs. (1), (2), (5) and

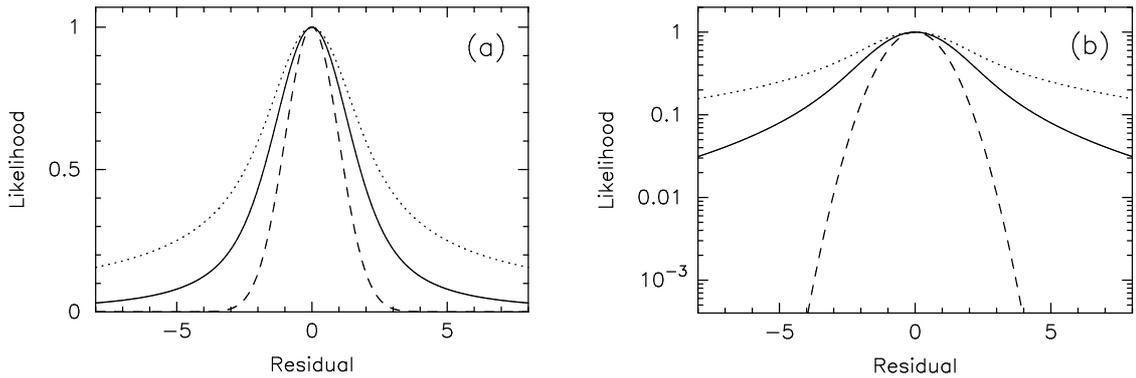


FIGURE 3. (a) The likelihood, $\Pr(x_k|\hat{x}_k, \varepsilon_k, I)$, plotted as a function of the residual, $(x_k - \hat{x}_k)/\varepsilon_k$, when ε_k is believed (dashed) compared with when it's used as a lower bound (solid); the Jeffreys' prior solution for the latter case is also shown (dotted), and all three have been scaled vertically to unity. (b) The same likelihood functions plotted on a logarithmic axis.

TABLE 3. The ‘lower-bound noise’ evidence for the number of δ -function components, M , in Fig. 1.

M	(a)	(b)	(c)	(d)
1	-122.4	-65.5	-117.5	-366.3
2	-123.8	-67.3	-118.9	-354.8
3	-125.6	-69.3	-120.4	-351.3
4	-126.8	-70.9	-121.2	-351.7
5	-127.9	-72.3	-122.1	-352.0

(6) yield the following likelihood for an individual datum:

$$\Pr(x_k | \hat{\epsilon}_k \geq \epsilon_k, \{A_j, \mu_j\}, M, I) = \sum_{j=1}^M \frac{A_j \epsilon_k}{\sqrt{2\pi} (x_k - \mu_j)^2} \left\{ 1 - \exp\left[-\frac{(x_k - \mu_j)^2}{2\epsilon_k^2}\right] \right\}. \quad (7)$$

The resultant probabilistic evidence for the number of components in the data of Fig. 1, $\ln[\Pr(\{x_k\} | M, \{\epsilon_k\}, I)]$, is given in Table 3. The best estimates of the $\{A_j, \mu_j\}$ for the $M=3$ model and their standard marginal error-bars, in the quadratic approximation, for the data of Fig. 1(d) are shown in Fig. 4(a); the ‘known’ D_e values and proportions of the three dose components are marked by asterisks. The related MCMC samples are displayed in Fig. 4(b).

DISCUSSION

Let’s consider first the individual sets of data, where we knew that $M=1$ by design but pretended to be ignorant. Of the three combinations of intrinsic shape function and noise model tried, the greatest evidence for the correct answer was from Eq. (7) for the measurements in Fig. 1(a). This is not surprising from a visual inspection, as the data show far more scatter than the quoted error-bars can accommodate and in a skewed

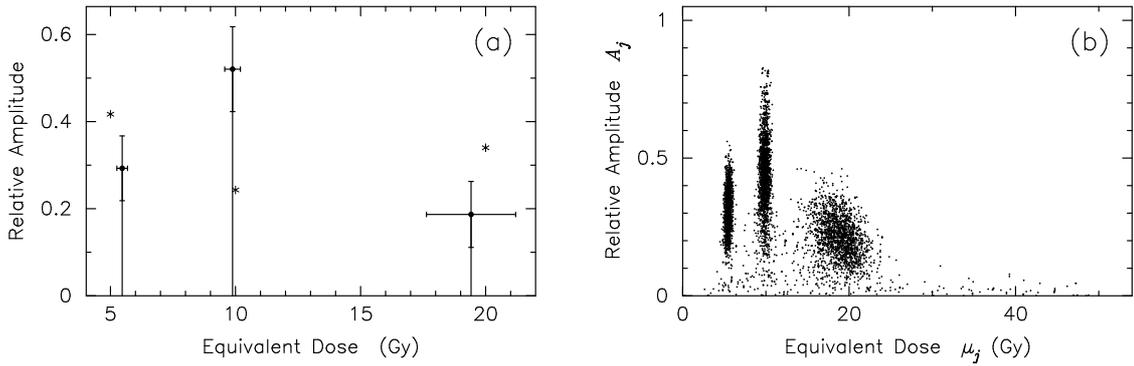


FIGURE 4. (a) The best estimates and their marginal error-bars, in the quadratic approximation, of the parameters for the $M=3$ δ -function model for the data of Fig. 1(d) using the robust likelihood of Eq. (7); the asterisks mark the correct (known) D_e values and proportions of the three dose components. (b) The MCMC samples of $\{A_j, \mu_j\}$, with M marginalised out using a Poisson prior with $\langle M \rangle = 1$.

manner. The most ‘well-behaved’ case is that of Fig. 1(b), and so probability theory is happiest with the simplest analysis of Eq. (3): believe the $\{\epsilon_k\}$ and assume δ -functions. The measurements in Fig. 1(c) show significant spread, but the quoted noise-levels are fairly large and there is nothing weird. There is some preference for the Gaussians of Eq. (4) over δ -functions, therefore, but little reason to doubt the $\{\epsilon_k\}$.

In a real problem, there might be a mixture of components and it’s our task to infer their number and nature. The amalgamation of the previous data in Fig. 1(d) presents a challenge then, in that each of the contributions is best analysed under different assumptions. Although ever more complicated formulations of the problem are always possible, pragmatically we’re likely to use something akin to Eq. (4) or Eq. (7). As seen from Figs. 2 and 4, both tell a similar story. Technically the measurements favour the likelihood of Eq. (4) to that of Eq. (7), but the results from the latter seem cleaner (based on the visual appearance of the cloud plots, and a knowledge of the true answer). At least on the basis of this example, there seems to be merit in the simplicity and conservatism of the robust noise analysis.

A comparison of the inferred parameters of the mixture model with their ‘known’ values indicates that a respectable job has been done for the equivalent doses, but their proportions are not correct. The best-fit $\{A_j\}$ are not too misleading, however, when their uncertainties are taken into account. Our central conclusions are consistent with those of [5], but we’ve shown that there’s an alternative set of simple assumptions, based on allowing for the possibility of ‘outliers’ in the data, that can be helpful for the analysis of luminescence dating measurements.

REFERENCES

1. Arnold, J. R., and Libby, W. F., *Science*, **110**, 678–680 (1949).
2. Aitken, M. J., *An Introduction to Optical Dating*, Oxford University Press, Oxford, 1998.
3. Stokes, S., *Geomorphology*, **29**, 153–171 (1999).
4. Sivia, D. S., *Data Analysis: a Bayesian Tutorial*, Oxford University Press, Oxford, 1996.
5. Roberts, R. G., Galbraith, R. F., Yoshida, H., Laslett, G. M., and Olley, J. M., *Radiation Measurements*, **32**, 459–465 (2000).
6. Skilling, J., *BayeSys and MassInf*, Maximum Entropy Data Consultants Ltd, Cambridge, 2003.
7. Sivia, D. S., “Dealing with Duff Data,” in *MaxEnt ’96*, edited by M. Sears, V. Nedeljkovic, N. E. Pendock, and S. Sibisi, University of Witwatersrand, Johannesburg, 1997, pp. 131–137.