Ancient TL

A periodical devoted to Luminescence and ESR dating

Web site: http://www.aber.ac.uk/ancient-tl
On plotting OSL equivalent doses

R. Galbraith

Department of Statistical Science, University College London, Gower Street, London, WC1E 6BT, UK (email rex@stats.ucl.ac.uk)

(Received 27 November 2009; in final form 18 December 2009)

Abstract
This article is motivated by some recent discussion of the use of so-called “probability density” plots of OSL equivalent doses. Such graphs are not advocated in the statistics literature. I try to explain what they are doing, why they are easy to mis-interpret and why they are not to be recommended. I include discussion of the meaning of dose frequency distributions, statistical research on the problem of estimating a frequency distribution when observations from it have added errors, and the possible role of dose histograms, in addition to radial plots, as data displays.

Introduction
There has been some recent discussion of the use of so-called “probability density” (PD) plots for displaying single grain, or single aliquot, OSL equivalent doses, and it was suggested to me that I might contribute to this. PD plots are used quite widely, as can be seen by perusing articles to be published in Quaternary Geochronology arising from the 12th International Luminescence and Electron Spin Resonance dating conference. Some years ago they were used by the fission track community to display single grain fission track ages. I criticised them then on several grounds: they do not estimate the true age distribution, modes in a PD plot do not necessarily correspond to discrete component ages, they obscure good information by combining it with bad, and their reliability was untested (Galbraith, 1998). They have been largely abandoned by the fission track community — I suspect mainly because they have not been found useful in practice.

In principle those criticisms also apply to OSL equivalent doses, though the popularity of PD plots here suggests that some people do consider them to be useful. However, they do not appear to have been advocated in the statistics literature. In this article I will try to explain what I think PD plots are doing, why they are difficult to interpret, and what alternatives there might be. Some of these ideas are also in a book chapter (Roberts and Galbraith, in press) which is to appear, though it was originally written in 2006.

What are the data?
We have a set of bivariate observations — an equivalent dose and its standard error for each of \( n \) quartz grains or aliquots, where \( n \) might be as low as 20 or 30 or as high as several hundred. A general feature of such data is that both the observed doses and their standard errors vary. Usually they vary together, with a higher standard error associated with a higher dose, the main exception to this being when the observed doses are close to zero.

A natural candidate for a graph is therefore some sort of bivariate plot; and a particularly useful one is a radial plot, which most readers will be familiar with. Descriptions of this method can be found in Galbraith et al. (1999), Galbraith (2005), Roberts and Galbraith (in press) and in other references cited there, so I will not deal with them further here. It is worth emphasising, though, that radial plots have optimal statistical properties (Galbraith, 1988) — they display the data as informatively as is possible and without distortion. They have also been found to be powerful in practice and can reveal features not otherwise apparent. Regardless of what other plots are also made, I would recommend researchers to look carefully at a radial plot of their equivalent doses.

A radial plot, though, does not provide an explicit picture of the frequency distribution of equivalent doses, which is perhaps why researchers may want to see some sort of frequency curve. However, we need to think carefully about what the frequency distribution of doses is and whether it has a useful scientific meaning. Sometimes it does not, and it is a strength of the radial plot that it does not force this interpretation on the reader.

Dose frequency distributions
The information in a relative frequency distribution of observed equivalent doses is complicated. It contains mixtures of received doses, natural variation and estimation errors — some of which are multiplicative and some additive — and does not simply relate to the relative numbers of grains in some real population that have received each possible dose. It is much more complicated than, for example,
a frequency distribution of heights of men or weights of babies.

Consider a hypothetical situation where we have a sample of single grain equivalent doses from a field sample of quartz that have been measured essentially without error (i.e., with negligible standard errors). The doses received in nature may differ between grains for various reasons, such as differing burial history or partial bleaching. Furthermore, even if each grain had experienced the same radiation dose in nature, the measured doses (even though measured exactly) would vary because of natural variation in luminescence properties between grains. Different scenarios will typically produce different dose distributions — for example unimodal distributions with low dispersion (perhaps representing only natural variation in luminescence), mixtures of two or three such component distributions, or highly heterogeneous, asymmetric or multimodal distributions.

What would knowing the shape of the frequency distribution of the doses tell us?

If we are lucky, it might indicate the type of sample or scenario we have. But before going further, there is another question: does this frequency distribution represent a natural phenomenon or is it largely a result of the process of grain selection and measurement? In the latter case it may be of more limited interest, and possible inferences from the data may also be more limited.

For example, suppose that each grain in our sample had essentially experienced one of two alternative burial histories, so each had received one of two radiation doses (e.g., by mixing of grains from two juxtaposed sedimentary strata that differ significantly in age). We could fit a two component mixture to estimate those doses. But would the estimated mixing proportions reflect anything other than artifacts of the experimental procedure, particularly grain selection? After all, only a small fraction of grains in a sample actually produce a measurable luminescence signal and these could be a highly non-random subset. Nevertheless, the component doses themselves should still be meaningful. The same applies to mixtures of more than two components — what do the mixing proportions represent? And by extension, what do the relative frequencies of different doses represent? In particular, does the most frequent dose in a sample have any special scientific relevance or meaning? These are questions for practitioners. The answer to the last one may sometimes be yes and sometimes no.

For aliquots comprising several grains, the concept of a dose frequency distribution is more complicated. It makes some sense if all grains in the same aliquot have experienced the same burial history. Then any differences in “true” single grain doses within an aliquot (had they been observed) would presumably just be due to differences in luminescence properties, and the aliquot equivalent dose would be representative of the burial history. But if grains in the same aliquot had different burial doses, or had experienced different amounts of partial bleaching, then the aliquot dose distribution would be much harder to interpret.

Incidentally, I have seen researchers fit finite mixtures (say with two or three component doses) and then choose the dose with the largest mixing proportion to be the relevant one. This seems like bad logic, especially for samples composed of partially bleached sediments. The relevant dose might be that corresponding to the youngest grains, and these could easily be a minority of the sample. This type of reasoning arises when looking at humps and bumps in frequency distributions too.

Histograms and kernel density estimates

Continuing with the case where our equivalent doses are measured without error, suppose that we do want a picture of the shape of the dose distribution. This could be provided (if there were enough grains) by a well-drawn histogram, which is essentially a graph of relative numbers of grains falling into different dose intervals (bins). Histograms are of course very familiar and widely used. A possible alternative is a kernel density estimate (KDE). This is a continuous curve that is an estimate of the probability density function (assumed to be continuous) of the distribution that the observations are supposedly a random sample from.

Denote the sample of true doses by \(x_1, x_2, \ldots, x_n\) and imagine that they were drawn randomly from a distribution with probability density function \(f(x)\). Now think of a histogram of these with equal bin widths. For a large enough sample, and small enough bin widths, this will give an idea of the shape of \(f(x)\). The area of each rectangle, and in this case also its height, is proportional to the number of observations falling in that bin, and (suitably scaled) is an estimate of the relative numbers in that interval in the population.

Now imagine drawing a histogram by starting with a bin at the extreme left (with no data in it) and sliding that bin continuously along the \(x\) scale. At each value of \(x\) draw a point at height equal to the number of data values in the bin centered at \(x\). The points will
trace out a curve that goes through the top middle points of the histogram rectangles plus more in between. That curve is a kernel density estimate of \( f(x) \) — in this case, using a rectangular “window”. If you increase the bin width the curve will be smoother but may lose shape features, and if you decrease the bin width the curve will resolve more shape features but be more erratic. Choice of bin width is a compromise between these two.

Rather than using a rectangular window, many kernel density estimates use a Gaussian window, which does not have discontinuities at the ends. The curve you then get is equivalent to drawing a Gaussian probability density function centered at each data value (each with the same standard deviation \( b \) which is chosen by you) and then summing them point-wise. This is simply a data smoothing method — as is counting up numbers in a histogram bin — there is no probability interpretation of this Gaussian window.

The quantity \( b \) is called the bandwidth of the window, and is analogous to the bin width of the histogram: the larger \( b \) is, the smoother the curve but the less resolution in shape there is. Actually there are many types of window around — nowadays they are called kernel functions — including triangular and cosine, but the principle is the same.

Statisticians have found that the shape of the window does not make much difference to the shape of the density estimate. What really matters is the bandwidth, which is a compromise between how much smoothing and how much resolution in shape you want. Choice of bandwidth usually depends on the sample size, with smaller bandwidths used for larger samples. This is like choosing the bin sizes for a histogram. Note that any smoothing distorts the data and loses information. A kernel density estimate is always a biased estimate of \( f(x) \) and in statistical terms the choice of bandwidth is a compromise between reducing bias and reducing variance. There is some theory about how to choose a bandwidth in order, for example, to minimise mean squared error (which is variance plus squared bias). In general large samples are needed to get reasonably informative kernel density estimates.

As estimates of density functions, KDEs enjoy some theoretical advantages over histograms (Wand and Jones, 1995, p5). The main disadvantage of histograms in this regard is that their shape can depend on where the first bin starts as well as on the bin width. Being continuous, KDEs give an impression of high precision, even for small sample sizes — but often a spurious impression. They have been developed by statisticians for over 50 years and are a useful exploratory tool, but they are not often used to present scientific data. One reason, I think, is that a histogram is better for this purpose. A histogram explicitly displays proportions of observations in various intervals as areas, whereas a KDE displays relative frequencies as a continuous curve. A KDE does not so easily lend itself to visual comparisons or simple calculation; it emphasises humps and bumps in the frequency curve, many of which have no significance; and it hides information, particularly relating to sample size and variability. As a general-purpose graph, a histogram is nearer to the raw data, easier to use and more convincing.

**PD plots**

Now let us return to the situation where the standard errors are non-negligible and variable. Denote the observed doses and their standard errors for \( n \) grains by \((y_i, s_i)\) for \( i = 1,2,...,n\).

A PD plot is constructed by replacing each \( y_i \) with a Gaussian probability density function centered at \( y_i \) and having standard deviation \( s_i \), and then adding these point-wise to obtain a continuous curve. Its construction is similar to that of a KDE, but with a different kernel function (with a different bandwidth) for each observation. The plot has some intuitive sense: you can think of it as plotting for each candidate dose, the “popularity” of that dose, as voted for by the \( n \) grains in the sample, where each grain spreads its vote (unequally) over several neighbouring doses, with more uncertain grains voting for a wider range of doses. Note that popularity comes both from frequency (\( y_i \)s close together) and precision (small \( s_i \)). Does a particular dose have any special scientific meaning simply because it is measured with high precision? Surely not.

The name “probability density plot” suggests that it is a plot of a probability density. An immediate question is: what probability density? The answer is: that of an equal mixture of \( n \) Gaussian distributions, where the \( i \)th component has mean \( y_i \) and standard deviation \( s_i \). In other words, a PD plot is plotting the probability density function of a random variable \( z \) constructed as follows: choose one of the \( n \) \( y_is \) at random and add to it a Gaussian random error with standard deviation \( s_i \).

A second question is whether the random variable \( z \) (and its associated distribution) is of any interest. To understand this, it is useful to think of a simple statistical model. Suppose that for each given \( s_i \), the observed dose \( y_i \) is generated by the equation
where $x_i$ is randomly drawn from a distribution with probability density function $f(x)$ and $e_i$ is randomly drawn from a Gaussian distribution with mean 0 and standard deviation $s_i$. Intuitively, $x_i$ represents the "true" dose (i.e., measured without error) for grain $i$ and $e_i$ is the error in estimating $x_i$ (i.e., the difference between $y_i$ and $x_i$). Neither $x_i$ nor $e_i$ is observed. The function $f(x)$ is unknown and our aim is to estimate it, or at least some of its features.

This type of model is familiar. If we postulated a parametric form for $f(x)$, such as Gaussian, we would have a version of the central age model. But here we are trying to let the data tell us something about $f(x)$ without assuming a specific form. In the previous section we were essentially thinking about how to do this if we could directly observe the $x_i$s.

Under this model, we can now think of obtaining a value of $z$ by first choosing one of the $n$ $x_i$s at random, adding a random $e_i$ to it to get $y_i$, and then adding another Gaussian random error to $y_i$ to get $z$. So the distribution of $z$ (i.e., the PD plot) does depend on the $n$ $x_i$s, which have been sampled from $f(x)$. But it also depends on the $n$ $s_i$s — doubly so because two independent random errors, each with standard deviation $s_i$, have been added to $x_i$. Its usefulness in practice will depend on whether it provides recognisable and useful information about $f(x)$.

There is a conspicuous lack of published theory about this. I've never seen a proper statistical study of PD plots, or even a reference to such a study — indeed I have never seen them advocated in a statistics journal. But there is some published research in statistics journals on how to estimate $f(x)$. One result of this is that the data $(y_i, s_i)$ in general contain very little information about the shape of $f(x)$. This is a warning against giving much credence to locations and relative heights of peaks in any estimate of $f(x)$. I summarise this research below.

My own experience from looking at PD plots, both with real and with simulated data, is that they are not uninformative but nor are they very informative, and their shape can be greatly affected by the $s_i$s. If we observed the same doses, but with different precisions, the curve can look very different. Often $s_i$ tends to increase with $y_i$. This alone will tend to produce a highly positively skewed curve with the highest peak or peaks near the left hand (lower dose) end. That is, one can often guess its general shape even without seeing any data. In general a high peak will be partly a result of several $y$s being close together but partly also a result of $s$s being relatively small. Conversely, if there are a substantial number of low-precision (large $s$s) grains in the sample, as there often are, these will tend to smooth out the whole curve and dilute the information from the high precision grains. Examples of these effects in the context of fission track ages can be seen in Galbraith (1998).

The force of these effects will of course be less if all or nearly all of the $s$s are small compared with differences between $y$s. In that case the distribution of $y$s will not differ greatly from that of the $x$s and a PD plot may be similar to a kernel density estimate (based on the $x$s) having the same average bandwidth.

In the above model $s_i$ is unrelated to $x_i$. But usually in practice the standard error tends to increase with dose. Sometimes the relative standard error is approximately independent of the dose. Then equation (1) would apply better with $(y_i, s_i)$ equal to the estimate and standard error of the natural log of the dose. But a PD plot of log doses would look very different in shape, and may have different numbers, locations and relative heights of peaks, compared to using a linear dose scale. Which scale should be used and why?

Some pitfalls
I don't think I have ever seen a paper where the author presents a PD plot and then comments that its shape may be reflecting the differing estimation errors rather than how the equivalent doses vary. Nearly always it is interpreted, implicitly or explicitly, in terms of which doses are predominant or indicated. This is understandable, because the graph invites one to do this, but it is misleading. Here are examples of possible pitfalls.

- You draw a PD curve and find that it has a high peak near the left hand end at a dose that plausibly corresponds to the burial dose of that sample (perhaps inferred from other information). So you present the PD plot as if it were pointing to that as the burial dose, or as support for that value. There may well be some relation between where the highest peak occurs and the burial dose — it may even agree closely sometimes — but it is not a reliable one. Often the PD curve is likely to have its highest peak near the left end simply because of the scale on which it is drawn and the nature of the error distributions. Sometimes also the burial dose may be reflected in only a minority of grains, and may not appear as a peak at all.

- In the previous scenario you might argue that in this case the PD graph gives the “right” answer, so it is
useful here. How do you know it is the right answer? Presumably from some other information. Then what use is the PD plot? It’s not good enough that the location of a peak in a PD plot might sometimes agree with the burial dose. As Lewis Carroll famously wrote, a stopped clock is right twice every day. A better approach would be to say “The PD graph suggests such and such. How can I investigate that hypothesis more seriously?”

• You look at the grains sitting under a peak of the PD curve and use these to estimate the burial dose, or some dose of interest. Or likewise, you use grains under different peaks to estimate different mixture components and their standard errors. This is like a so-called “classification” method of estimating mixture components. Such methods are known to be biased — sometimes very biased — and to provide unreliable standard errors. Fortunately there are more reliable methods available, such as maximum likelihood estimation.

• Among lots of information and data in a paper are several PD plots, and a commentary that refers to these to support the discussion of some phenomenon or theory of interest. The proposed theory may well be right, but logic tells us that if a PD plot does not reliably estimate the true dose distribution (which it does not) then those graphs do not support the theory.

An important aspect of this is that even if the writer is able to avoid such pitfalls, it may still be hard for readers to do so.

How can we get a picture of \( f(x) \)?

Suppose we have the scenario given by equation (1) and we want to estimate the function \( f(x) \). A PD plot will not do this, so how can we do it? There are two general statistical approaches: parametric and non-parametric. The central age and minimum age models are examples of parametric methods. These assume a specific form for \( f(x) \), but with unknown parameters that represent quantities of interest which are then estimated from the data. The idea of using a non-parametric method is to see if the data can tell us what shape \( f(x) \) has without imposing a particular form.

There has been some research on this. An important general result is that the data \((y_i, x_i)\) contain a very limited amount of information about the shape of \( f(x) \), as opposed to its location and dispersion (e.g., Goutis, 1996; Madger and Zeger, 1996; Wand and Jones, 1995, p160). The same data can easily arise from quite different \( f(x) \)s.

Several methods have been suggested. One is the “non-parametric maximum likelihood estimate” (NPMLE). This turns out to yield a discrete probability distribution concentrated on a relatively small number of values — that is, it estimates \( f(x) \) as a set of \( k \) different values and their probabilities, where \( k \) is quite small compared with \( n \) (Laird 1978). When \( f(x) \) is assumed to be continuous it is arguable that it would be nice if the estimate of \( f(x) \) were also continuous. To this end Madger and Zeger (1996) proposed a smoothed version of the NPMLE (called SNPMLE). This assumes that \( f(x) \) is a mixture of \( k \) Gaussian distributions (where \( k \) is unknown) each having a standard deviation greater than or equal to some known value \( b \). The thinking behind this is that you can produce a wide variety of different shapes by mixing enough Gaussian distributions in differing proportions. The condition on the standard deviations is necessary in order to guarantee convergence to a solution. You could call it a semi-parametric method. The SNPMLE converges to a mixture (with differing mixing proportions) where, again, \( k \) is relatively small and all components of this mixture have the same standard deviation, equal to \( b \). The value of \( b \) is chosen empirically to achieve a desired amount of smoothing, like a bandwidth of a kernel density estimate — the larger \( b \) is, the smoother the graph. Other methods have been proposed by Goutis (1996) and Newton (2002). These methods are all computationally intensive to implement. More recent work includes Delaigle and Meister (2008), Staudenmeyer et al. (2008) and Wang et al. (submitted) so theoretical progress is being made in this area.

The general message seems to be: it is hard to infer the shape of an underlying distribution when observations from it have added errors, even when these errors have known standard deviations. A more fruitful approach might be to ask: what specific features of \( f(x) \) do we really want to know? Then try to ascertain these by appropriate statistical modelling.

Improving a histogram

To help interpret a histogram of single grain equivalent doses, Roberts and Galbraith (in press) suggest adding a scatter plot of \( x_i \) against \( y_i \). This is illustrated in Figure 1 for a sample of 82 single quartz grains. Olley et al. (2004) reported that these grains were transported by wind on to the bed of Lake St. Mary (in semi-arid south-eastern Australia) within the last 40 years. Many of the observed equivalent doses are close to zero and some are negative.
The histogram has a positively skewed shape for doses below 1 Gy and three more extreme values around 2 and 3 Gy. The scatter plot shows that several grains have standard errors greater than 0.5 Gy, which is quite large compared with differences between the dose estimates, and two of the extreme values have standard errors greater than 1 Gy, so could, in principle be consistent with the values for some of the lower dose grains. It must be emphasised that this graph is simply a plot of the raw data; the histogram in particular is a summary of the *y*<sub>i</sub>s and should not be interpreted as a graph of the *x*<sub>i</sub>s. The scatter plot helps with this by drawing attention to the *s*<sub>i</sub>s associated with each *y*<sub>i</sub>. In fact the histogram is better viewed as an adjunct to the (*y*<sub>i</sub>, *s*<sub>i</sub>) scatter plot, showing the marginal distribution of *y*<sub>i</sub>, rather than the other way round.

This example is presented here simply to illustrate the method. It is unusual in having several negative and near-zero equivalent doses; but their presence serves to remind us that the *y*<sub>i</sub>s are not the *x*<sub>i</sub>s (the true doses) but just *estimates* of them. For example, the smallest *y*<sub>i</sub> is -0.35 Gy. Because *x*<sub>i</sub> cannot be negative, we can deduce that this *y*<sub>i</sub> underestimates its *x*<sub>i</sub> by at least 0.35 Gy. The *s*<sub>i</sub> for this grain is 0.45 Gy, indicating that its *x*<sub>i</sub> could still be as large or larger than -0.35 + 2 × 0.45 = 0.55 Gy. In general, a histogram of *y*<sub>i</sub>s need not look like a histogram of the corresponding *x*<sub>i</sub>s.

Graphs like Figure 1 were used, in conjunction with radial plots, by Arnold et al. (2009) to compare a number of samples of differing origin. We found the scatter plot useful for revealing the relationship between *s*<sub>i</sub> and *y*<sub>i</sub>. Sometimes there was a strong positive correlation, indicative of multiplicative errors, and other times there was little or no correlation (especially for small *y*<sub>i</sub>, such as in Figure 1 here) suggesting that the main sources of error were additive. My co-authors also found the histograms useful for indicating some general characteristics of the sample.

Figure 2 shows a radial plot of the same data. This uses a linear dose scale rather than the usual log scale, the latter not being possible with negative estimates. It is not easy to draw this scale in such a way as to accommodate the three extreme values and at the same time to show the rest of the data in detail. Here I have drawn it so as to see the main data clearly; and radii through the three values greater 2 Gy go off the *D*<sub>4</sub> scale. The two values near 2 Gy are seen here to be almost completely uninformative — you hardly notice them — and the vast majority of points (all but about 10) are consistent with having zero dose.

In this example the information in the radial plot is so clear that very little further analysis is necessary. Possible further analysis might be to try to ascertain whether the burial dose is actually zero or some positive value close to zero and perhaps calculate an upper confidence limit for it. This would require proper statistical modelling. In fact Olley et al. (2004) estimated a burial dose of 0.1 Gy with a confidence interval that included 0 Gy.
Figure 3. An alternative radial plot of the data in Figure 1 using the modified log transformation $d = \log(D_e + 1)$. The two $D_e$ values near 2 Gy are again plotted as filled circles.

For data containing zero or negative estimates, an alternative to using a linear dose scale is to use a modified log transformation given by $d = \log(D_e + a)$ for some suitable $a$. That is, add $a$ Gy to each observed dose and then take logs. The standard error of $d$ is then approximately $se(D_e)/(D_e + a)$. Figure 3 illustrates this method for $a = 1$. The dose scale is now non-linear, calculated from the formula $D_e = e^d - a$, and there is no difficulty in including the extreme values on it.

The message from Figure 3 is very similar to that from Figure 2. It looks a bit different because the estimates are plotted with respect to relative, rather than absolute, standard errors; in particular, the three extreme values are more prominent. This method is useful when the data contain some near zero doses and some larger ones.

Choice of bin width

A reviewer raised the question of what bin width to use for the histogram, particularly in relation to consistency of presentation and also whether the standard errors should be used to determine it.

In Figure 1 I have used bins of width 0.2 Gy, located so that 0 Gy comes in the middle of a bin (and consequently, so do 1, 2 and 3 Gy). A reasonable alternative would be to have 0 Gy at the edge of a bin. General guidelines tell us to use smaller bins for larger sample sizes and to try to achieve a reasonable amount of smoothing without losing too much detail, but there is no hard and fast rule. It is helpful to use friendly values; 0.2 Gy is better than 0.23 Gy, say. If I had used 0.1 Gy there would be twice the number of bins with smaller numbers in each, while 0.4 Gy would produce half the number of bins and more data pooled. Here 0.2 Gy seems about right. My personal preference is to err on the side of more bins rather than fewer, so as to reveal more of the raw data.

A histogram does not have to have equal width bins of course. For very highly skewed data it is sometimes suggested to have wider bins in the tail (drawn so that the area of a rectangle, not its height, is proportional to the number of observations in the bin). But equal bin widths are easier to understand and are nearly always used for routine presentation. In Figure 1 it is much better to show the three large values in separate bins rather than combining them into one long bin. Incidentally, if you look at these actual doses in the scatter plot you can see that they do not fall in the middle of each bin; the histogram just tells you that the points are somewhere in the bin, not necessarily in the middle.

What about consistency of presentation? In fission track analysis it is standard practice to measure about 100 track lengths and present them in a histogram with 1 micron bins on a scale that goes from 0 to 20 microns. This is possible because unannealed track lengths have a very tight distribution with mean about 16 microns and standard deviation about 1 micron. You never see a track longer than 20 microns. When tracks are heated they shorten and become more variable in length; the distribution shifts towards zero and becomes more dispersed and skewed. It tells us something about the thermal history that the grain has experienced. This consistency of presentation is a huge advantage and greatly outweighs other criteria for choosing bin widths. Many such histograms are shown in articles and at conferences and it is possible to compare them, not only within the same presentation, but also between different articles and talks, even in different journals and conferences.

Is such a thing possible for equivalent dose distributions? I don't think so. Samples may have doses ranging between, say, 10 and 80 Gy. Using 0.2 Gy bins there would do no smoothing at all. Using bins of width 4 or 5 Gy might be reasonable there but would be useless for the data in Figure 1. But there is some scope for consistency of style, including axis labels and terminology. This is a matter for general discussion. Sometimes it may be useful to plot equivalent doses on a log scale, which raises further questions about style. OSL equivalent doses are far more complicated than fission track lengths! They are more like fission track ages, but more complicated
than them too. Fission track ages are routinely presented in radial plots but not in histograms.

Should the \( s_i \) be used to determine the bin width? No. The histogram is a graphical display of the observed doses (the \( y_i \)). The standard error \( s_i \) tells us something about how close \( y_i \) is likely to be to its \( x_i \), but this has nothing to do with choosing the bin width for a histogram of \( y_i \). If we had a larger sample size we would want smaller bins (regardless of the \( s_i \)) to get a better summary of the data.

This point serves to emphasise that a histogram of \( y_i \) is not the same as a histogram of the \( x_i \) and should not be seen as such. If the \( s_i \) are all small compared with differences between \( x_i \) then the two will be similar. If the \( s_i \) are non-negligible, then all of the previous discussion and theory is telling us that we just don't have much information about the frequency distribution of \( x_i \). We have some information about its location and dispersion; which is what the central age model is extracting, and we can try to extract other information using parametric models such as the minimum age models. There are non-parametric methods for estimating this frequency distribution but they do not yield either PD plots or histograms of \( y_i \).

**Summary**

When OSL equivalent doses are observed with non-negligible and differing standard errors they are not easy to compare. A radial plot will display them informatively and without distorting their message. I recommend looking at a radial plot in addition to any other graphs that might be made.

Research has shown that such data contain little information about the form of the underlying frequency distribution of true doses; quite different underlying distributions can easily give rise to the same observed data. Several methods have been suggested for trying to estimate such an underlying distribution, though little is known about how useful they are in practice. A question to consider is what use this frequency distribution would be if it were known. If only some of its features or parameters were of interest then a more fruitful approach might be to try to estimate these directly.

A histogram of observed doses will reflect features of the single grain error distributions and the relationship between observed doses and their standard errors, as well as variation in true doses. In order to interpret it without pitfall it is necessary to add further information, such as an adjacent scatter plot of standard errors against doses. Together these can provide a useful description of the data, but will typically not provide a true picture of the underlying dose distribution.

PD plots also depend on the error distributions and their relationship with dose — more so than histograms because effectively two independent errors are added to each true dose. There appears to be no rationale or justification for them in the statistics literature. They too do not provide an estimate of the underlying dose distribution. All you can really do with them is look and see where peaks occur. These may or may not reflect features of the true dose distribution, which in turn may or may not reflect events in nature.

Perhaps their biggest difficulty, though, is that it is hard to avoid the types of pitfalls mentioned above. The reader is faced with a continuous curve that looks meaningful; but it does not mean what it appears to mean and there is no reliable way to extract what we want from it. Someone once said that Wagner’s music is better than it sounds. Indeed it may be. But PD plots are not as good as they look. I don’t recommend them.

**Acknowledgments**

I am very grateful for useful comments and suggestions from Geoff Duller, Jane Galbraith, Bert Roberts and Ann Wintle.

**References**


**Reviewer**

A.G. Wintle
An alternate form of probability-distribution plot for $D_e$ values

G. W. Berger

Desert Research Institute, 2215 Raggio Parkway, Reno, NV 89512, USA
(e-mail: glenn.berger@dri.edu)

(Received 17 November 2009; in final form 15 March 2010)

Abstract
The use of probability density distribution (PDD or PD herein) plots for summarizing visually the distributions of paleodose ($D_e$) values (even when accompanied by a displayed ranking of $D_e$ values) has been criticized over the past decade. Here the suitability of this plot is revisited and an alternate form is proposed for creating a more realistic statistical representation of such data. The alternate plot (TPD, Transformed-PD) is generated by use of a logarithmic transform of $D_e$ values and use of relative, rather than absolute, errors in $D_e$ values. The radial plot also employs such parametric transforms. Examples are given and discussed of distributions of $D_e$ values from both multiple-grain and single-grain SAR (Single Aliquot Regenerative dose) experiments for which a youngest-age interpretation is required. The PD and TPD representations of these data are compared with each other and with the corresponding radial plots. These examples illustrate both the systematic advantages of the TPD plot compared to the conventional PD plot (as heretofore used with $D_e$ data sets) and some limitations of the TPD plot. These limitations can be viewed as either minor or major, depending upon the data set. Generally, the use of the TPD plot (together with ranked $D_e$ values) is an improvement over the comparable use of the PD plot.

Historical context and objective
The PD plot has been used for decades in fission-track (FT) dating (e.g., Hurford et al., 1984; Brandon, 1996) and in $^{40}\text{Ar}/^{39}\text{Ar}$ (or ‘Ar-Ar’) dating (e.g., Deino and Potts, 1992; Morgan and Renne, 2008). In these two dating methods PD plots have been used in combination with a superimposed graph of the ranked age estimates and their error bars to represent visually age distributions derived from single-grain analyses. In luminescence dating, the PD plots have been used to display temporal distributions of age estimates from suites of samples within a given geographic region (e.g., Stokes et al., 2004), although Bayesian representations of such temporal distributions may be more valid (e.g., Rhodes et al., 2003). An even broader use of PD plots has been to represent visually detrital zircon (single-grain) age distributions (‘age spectra’), to infer episodicity of continental processes (e.g., Condie and Aster, 2009; Condie et al., 2009), usually using data-handling approaches promoted by Sircombe (2004) and Sircombe and Hazelton (2004).

In the last decade PD plots have been employed to replace histograms in the representation of distributions of multiple-grain and single-grain $D_e$ values (e.g., Jacobs et al., 2003; Duller and Augustinus, 2006; Feathers et al., 2006; Berger et al., 2009; Pietsch, 2009; Porat et al., 2009) derived from the SAR (Murray and Wintle, 2003) procedure and its later modifications (e.g., inclusion of IR-wash steps: herein, all modified versions are termed ‘SAR’). It therefore seems reasonable to advocate the replacement of the conventional (constant-bin-width) histogram to illustrate dating-result distributions, by some alternate plot so that information on the relative precision of separate ages (FT and Ar-Ar) or $D_e$ values (SAR luminescence) can be presented visually, and so that the presence or absence of relative structure (clustering of data) can be illustrated more effectively.

For some time Galbraith (e.g., 1998, 2005) has advocated replacement of the PD plot in FT dating with the radial plot, based on some sound statistical arguments concerning the generation and propagation of analytical errors in FT age calculations and on other aspects of FT dating. His criticisms of PD plots have been imported into luminescence dating to argue against the use of PD plots for $D_e$ distributions (e.g., Better-Jensen et al., 2003; Lian and Roberts, 2006; Duller, 2008).

The alternate PD plot proposed here minimizes or circumvents the major weakness of the conventional PD plot. The main weakness is the over-emphasis of the statistical significance (relative probability) of
low \(D_e\) values in a distribution of a range of \(D_e\) values. As shown below, this over-emphasis is caused by two assumptions: that \(D_e\) values arise from a statistically Normal (Gaussian) distribution, and that the errors in \(D_e\) values are not proportional to the \(D_e\) values. The proposed alternate PD plot (when accompanied by ranked \(D_e\) values and errors) can communicate more intuitively the relative statistical significance of various apparent components (subpopulations) of the measured \(D_e\) values than can the conventional PD plot. That is, the alternate PD plot can quickly and more ‘accurately’ (in the statistical relative-probability sense) indicate relative structure (or lack of it) in the distribution compared to the conventional PD plot. This alternate PD plot can in turn motivate the selection of various quantitative methods for age calculation. The alternate form of the PD plot does not replace the radial plot as an accurate visual representation of the statistics of each \(D_e\) value (when these are greater than zero), but provides for a visually intuitive display of paleodose values (when they are greater than zero). For paleodose values near or less than zero, the conventional PD plot is still useful (e.g. Berger, 2009; Pietsch, 2009), whereas (with \(D_e\) values less than zero) the radial plot is not.

Some basic concepts

At the heart of the choice for use or not of the PD plot is consideration of two variables: the appropriateness of the density function (kernel density estimator or kernel) and of the smoothing parameter (bandwidth or data-window width) (e.g. Brandon, 1996; Silverman, 1986; Wand and Jones, 1995). There are many kernel functions (e.g. uniform or box), triangular, quartic, triweight, biweight, Gaussian, cosine, lognormal, Gamma) in use with various types of data (e.g. Silverman, 1986; Wand and Jones, 1995). While many of these kernels are statistically ‘suboptimal’ (Wand and Jones, 1995), some such as the Gaussian kernel are not suboptimal by much. Thus, “the choice between kernels can be made on other grounds, such as computational” ease (Wand and Jones, 1995, p. 31). Essentially then, use of a Gaussian kernel is a form of data smoothing. As discussed by several authors (e.g. Silverman, 1986; Wand and Jones, 1995), “smoothing methods provide a powerful methodology for gaining insights into data” (Jones et al., 1996) without highly sophisticated mathematics.

Jones et al. (1996) review the effects of bandwidth choice on smoothing. They show that the Gaussian kernel can over-smooth the density estimate somewhat or ‘seriously’, depending on the data example. At the other extreme, the shape of the conventional histogram is highly sensitive to the choice of bin width (analogous to bandwidth for kernel density estimators) and of the placement of the bin edges. There are many disadvantages of histograms (e.g. low ‘efficiency’) compared to kernel estimators (e.g. Wand and Jones, 1995, p.7). An extensive discussion on the use of histograms in Earth Science is provided by Vermeesch (2005).

The appropriateness of bandwidth choice has been discussed in some of the literature concerned with the use of PD plots to represent age spectra for single-grain-zircon (for example) data. A Gaussian function has been deemed apt for representing the kernel of such data. The bandwidth of the Gaussian kernel is based on the standard deviation of each datum. Thus the bandwidth can change at each data point, which provides a distinct improvement over histograms, that employ a constant bin width. However, in (for example) zircon-age-spectra applications (e.g. Condie et al., 2009) bandwidth choice usually comes down to consideration of the age resolution of individual analyses (e.g., 1 Ma or 20 Ma). In the examples and discussion of \(D_e\) distributions below, the historic choice of a Gaussian kernel (explicitly defined below) is maintained as this seems to be a reasonable representation of each \(D_e\) value derived from multiple-grain and single-grain SAR experiments.

To summarize, this communication outlines how the kernel and bandwidth of PD plots for \(D_e\) distributions can be transformed to a more realistic visual and statistical representation of the relative structure in such distributions. Implicitly, there are many experimental variables that affect the statistics of any population of \(D_e\) values. Of course deconvolution computational methods (e.g. Minimum Age Model or MAM, Galbraith et al., 1999) may still be required to calculate geologically accurate component \(D_e\) values.

Radial and PD plot

Galbraith (1988) introduced the radial plot as “useful for exploratory, diagnostic, or descriptive purposes or to supplement more formal estimation and hypothesis-testing methods”. This is essentially the reason others have used the PD plot. Clearly, the radial plot offers advantages over the PD plot, particularly when the PD plot is unaccompanied by a display of ranked data with error bars. In Berger et al. (2009), the PD plot offers these uses (exploratory, diagnostic or descriptive). Those plots motivated the hypothesis-testing exercise of calculating weighted means and standard errors of subsets (for different samples) of \(D_e\) values, which in turn led to some accurate youngest age estimates.
Figure 1: A radial plot for a mixture of two distinct populations of artificial data (open and filled circles), having distinct means (represented by dashed lines). This is a copy of Figure 3 in Galbraith (1988).

Galbraith (1988) presented an artificial data set and compared a radial plot and a PD plot to elucidate visually the relative structure in that data set. The data set represented a mixture of two populations, one with mean +0.5 and the other with mean -0.5. In Figure 1 one of these two populations is represented by open circles, the other, by filled circles, and the respective means, by the two dashed lines. Without these graphically displayed distinctions (filled circles and dashed lines), this radial plot merely suggests the existence two populations, but does not clearly resolve them. Plotted as a "weighted histogram" (a PD plot) in Figure 2, these same artificial data also suggest a bimodal distribution. Inexplicably, Galbraith then states that "the weighted histogram is superficially attractive ... but does not point to the true mixture as informatively as Figure 2 does" (his radial plot without filled circles and dashed lines). He makes other qualitative remarks disparaging this PD plot, but it seems that Figure 2 is nearly (or for this data set, 'equally') as informative visually as his radial plot.

Alternate PD-plot formulation

The conventional PD plot of $D_e$ values uses the Gaussian kernel:

$$P_i(D_e) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp \left[ -\frac{(D_e - D_e^i)^2}{2\sigma_i^2} \right]$$  \hspace{1cm} (1)

for an aliquot or grain $i$, having paleodose $D_e^i$ and absolute error $\sigma_i$. Equation 1 gives the probability of observing a particular $D_e$ value within a range centered on the $D_e^i$ value. When summed over all accepted-data $i$ values, equation 1 gives smoothed PD-plot curves. Note that in this form, the sum of $P_i(D_e)$ does not give a 'normalized' probability density, where the sum is divided by the number of items. Singhvi et al. (2001) give an example of a normalized PD plot.

There have been two main criticisms of this conventional PD plot. One is that because for luminescence data the error in $D_e$ is "often proportional to the equivalent dose" (e.g. Duller, 2005, Analyst© Software, Appendix 3), then the PD plot derived from equation 1, using only absolute error estimates, inaccurately represents the relative probabilities of $D_e$ values. It has been recognized for some time (e.g. Berger et al., 1987, Appendix A; Galbraith, 2003) that errors in luminescence data can be constant-relative, not "absolute". The second main criticism is that $D_e$ values from SAR data sets are more likely to represent lognormal distributions (e.g., Galbraith et al., 1999; Galbraith, 2003) than Gaussian distributions. The conventional PD plot does not take this distinction into account whereas the radial plot does.

A more technically vague criticism (Galbraith, 1998,
2005, in reference to FT data, but adopted by Bøtter-Jensen et al., 2003, in reference to \( D_e \) values) of the use of the conventional PD plots is that “the presence of several estimates with low precision can obscure information, even when there are other high precision estimates in the sample” (Galbraith, 2005, p.195). If this remark is meant to apply only to PD (and other smoothed) plots that lack a companion graph of data points and their error estimates, then it makes sense. Otherwise, this remark applies also to several (most?, any?) pooled estimates, such as the “sound statistical method” (Galbraith, 1988, p.125) of calculating weighted (by inverse variance) means. The use of weighted (by inverse variance) means can be a statistically appropriate tool for some \( D_e \) data sets (e.g., Berger et al., 2009; Pietsch, 2009). However, as illustrated with examples below, most \( D_e \) data sets probably require a modified form of weighted mean calculation, or use of a central age model or minimum age model calculation.

There are many examples in the literature (e.g., Arnold and Roberts, 2009, and citations therein) for which neither the deconvolution methods nor radial plots are able to ‘resolve’ meaningful age components, nor do they help resolve ‘information’ from discretely displayed \( D_e \) values. Plotting of ranked \( D_e \) values (with errors) together with PD plots certainly ‘resolves’ individual data points. Thus there are many examples for which smoothed distributions of \( D_e \) values occur and for which the radial plot does not provide insight into the geological significance of the \( D_e \) values. This is probably because there is no geological significance to many such values. These ‘smoothed’ \( D_e \) values likely reflect only some unrecoverable grain-transfer history, manifesting an interrupted series of daylight exposures somewhere between the start and end of the journey to the final resting place. In this context, it seems then that such technically vague criticisms as mentioned above should be avoided because they can be self-contradictory. Rather, the aforementioned two main criticisms should be addressed. The proposed alternate form of the PD plot is intended at least to minimize the effects of the above two main shortcomings in the PD plots.

The construction of the radial plot (e.g., Galbraith et al., 1999) is based on considerations of the Poisson statistics of luminescence (and FT) measurements, and employs a logarithmic transform. In luminescence, the radial plot assumes that \( D_e \) distributions often resemble lognormal distributions. In particular, for luminescence, one can then employ the transforms:

\[
Z = \ln(D_e), \quad Z_i = \ln(D_{ei}), \quad \text{and} \quad \sigma_i(Z_i) = \sigma_i/D_e
\]

For luminescence, the radial plot assumes that \( D_e \) distributions often resemble lognormal distributions. In particular, for luminescence, one can then employ the transforms:

\[
Z = \ln(D_e), \quad Z_i = \ln(D_{ei}), \quad \text{and} \quad \sigma_i(Z_i) = \sigma_i/D_e
\]

In other words, the \( D_e \) values in equation 1 are replaced by their natural logarithms and the absolute errors are replaced by their corresponding relative errors. As indicated by Galbraith (2003), the appropriate \( \sigma \) would be the standard error of the logarithm of the paleodose, but as he also indicates, this is effectively the relative standard error of the paleodose. When these transformations are applied to equation 1, the alternate (transformed) probability kernel is obtained:

\[
T – P_i(Z) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp \left[ -\frac{(Z_i - Z)^2}{2 \sigma_i^2} \right]
\]

The resultant summation over all accepted-data aliquots or grains provides the alternate PD plot (a Transformed-PD or TPD plot).

Use of equation 2 instead of equation 1 addresses the criticism of Duller (2005, Analyst© Software, Appendix 3), as shown in Figure 3. Duller gave an example of 4 artificial \( D_e \) values all having the same relative error of 10%. He showed that the resultant conventional PD plot (dashed curve in Fig. 3) tends to over-emphasize the relative significance (probability) of the lowest two \( D_e \) values. However, application of equation 2 to these same artificial data generates the solid curve in Figure 3, accurately representing their relative probabilities.
Figure 4: Relative probability plots, with a ranked series of $D_e$ values and their errors, for a multiple-grain SAR experiment on an irrigation-ditch sediment sample from northern New Mexico (after Berger et al., 2009). Quartz grains of 150-212 μm diameters were employed. Only 22 of 48 aliquots met the standard data-acceptance criteria. Each aliquot contains 10-20 grains.

It is perhaps worth noting that Brandon (1996) employed a logarithmic transformation and relative errors in his use of the Gaussian kernel, but Galbraith (1998) found fault mainly with Brandon’s choice of bandwidth estimation and with the nature of Brandon’s error assumptions for FT dating.

It is also worthwhile pointing out that PD plots constructed from (‘unlogged’) equation 1 can usefully represent $D_e$ distributions containing negative $D_e$ values (e.g. Pletsch, 2009; Fig. 10 in Berger, 2009) such as can arise from analysis of modern-age or very young samples, whereas TPD (and radial) plots cannot represent such $D_e$ data. How can $D_e$ values less than zero arise? An ideal zero-age sample would be expected to produce a distribution of $D_e$ values (e.g., from single grains) described approximately by a Gaussian (‘bell curve’) centered on $D_e = 0$. Negative values can arise from statistical fluctuations in the luminescence signal within the regions of the shine curves selected for ‘signal’ and ‘background’, such that $L_0/T_0$ can become negative for some grains or aliquots. Not only the TPD plots, but also the usual (logged) deconvolution methods (e.g. MAM and Central-Age-Model or CAM) will fail for distributions containing negative $D_e$ values, as reviewed by Arnold et al. (2009). They conclude that one has to resort to other calculation methods, and suggest use of an ‘unlogged’ version of MAM. Berger et al. (2009) found that the straightforward use of weighted mean (by inverse variance) $D_e$ values (and standard error of the weighted mean) is sufficiently accurate for their very young samples (e.g. generating a single-grain quartz age estimate of 92.3±9.6 a compared to an historical age of <127 a).

While graphic-representational and computational ambiguity is not uncommon with the use of multiple-grain SAR $D_e$ values (e.g. Arnold and Roberts, 2009), it is also not uncommon with the use of single-grain $D_e$ values, as the examples below illustrate. Rather than reflecting limitations of graphical methods of display or of statistical deconvolution methods of computation, difficulties in interpretation of such data sets more likely reflect unavoidable geological or empirical complexities. For example, positively-skewed ‘smeared’ single-grain $D_e$ distributions (assuming that each grain hole emits at most a single-grain signal) can merely reflect conditions of non-episodic mixing of grains having arbitrarily different daylight exposure histories. Hopefully, at least the youngest age information can be resolved, graphically or by deconvolution, or by both.

Weighted mean calculations for $D_e$ data
As mentioned, the usual weighted mean calculation (e.g. Topping, 1962) employs weighting by inverse variance of absolute errors. This is appropriate for data having independent errors and for which the

Figure 5: Radial plot of the data in Figure 4. The center of the ±2 $\sigma$ lowest bar passes through the weighted mean $D_e$ value derived from the 5 filled-circle data points in Figure 4. For comparison purposes, a thin line is drawn to the approximate center of the peak at ~17 Gy in Figure 4. The ±2 $\sigma$ top bar is drawn to the approximate center of the broad peak at ~45 Gy in Figure 4.
Figure 6: Relative-probability plots for the single-grain quartz SAR results from the sample in Figure 5. Only 63 of 1000 grain-hole stimulations provided acceptable $D_e$ values using the normal data-acceptance criteria. The small ‘wiggle’ in the dashed curve at ~6 Gy is an artifact of the spacing of curve-fit data points used for this spline-curve fitting.

The examples of $D_e$ subsets of Berger et al. (2009) and Pietsch (2009) approximate such data. However, many subset (and whole) $D_e$ distributions (examples below) have errors that are roughly proportional to the $D_e$ values. Moreover, as mentioned above, $D_e$ distributions often resemble lognormal distributions. Therefore, the formulae for calculation of weighted means of $D_e$ values in general should be modified to use $\ln(D_e)$ values and weighting by inverse relative errors. Such a modification or transformation, appropriate for the general statistics of most $D_e$ values, can provide a convenient tool for quick estimates (via spreadsheets) of a mean of a data subset, yet that can be more accurate statistically than the conventional weighted mean calculation. In some of the examples below, this relative error weighted mean (REWM) calculation is employed.

Examples of PD, TPD and radial plots

In the following examples, the PD plot is represented by a dashed line and the TPD plot, by a solid line. Also, the PD and TPD curves are plotted along with ranked $D_e$ values and their 1σ absolute errors. The maximum ‘heights’ of the PD and TPD plots have been scaled (not normalized) to approximate equivalence for easy visualization of their similarities and differences.

Figure 7: Radial plot of the data in Figure 6. The ±2σ lowest bar is centered on the weighted mean calculated for the lowest peak in Figure 6. The ±2σ top bar is centered at the ~65 Gy peak in Figure 6.

The first example compares the graphical representation by the three plots (PD, TPD, radial) of $D_e$ distributions obtained from both multiple-grain and single-grain SAR experiments on quartz sand from the oldest sample (STFW05-2, historical age ≈450 a) in the study of Berger et al. (2009). $D_e$ values from the use of 0.4 mm aliquots (~10-20 grains each) are shown in Figure 4. Notice that both the PD and TPD plots visually denote a small (5 data points) cluster of $D_e$ values, the weighted (by inverse variance) mean of which was used by Berger et al. (2009) to estimate a minimum age of 487 ± 74 a (from mean $D_e = 1.65 ± 0.25$ Gy), comparable to the known probable historical age of the sample. However, only the TPD plot draws visual attention to two other (probably geologically meaningless) subgroups of $D_e$ values, one at ~16 Gy and one around 45-50 Gy. This example shows how the TPD plot more realistically represents the relative structure in the population of $D_e$ values than does the conventional PD plot.

In the more accurate statistical representation of these data in a radial plot (Fig. 5), the 3 apparent groups so clearly visualized in the TPD plot are also quite apparent. Though not stated in Berger et al. (2009), the MAM-4 estimate of the youngest-age $D_e$ values in Figure 4 is 2.03 ± 0.28 Gy and the MAM-3 estimate is 1.88 ± 0.45 Gy, neither of which is significantly different from the aforementioned weighted mean of 1.65 ± 0.25 Gy. Incidentally, the REWM estimate for the 5-point cluster of $D_e$ values in Figure 4 is 1.78 ± 0.25 Gy, somewhat closer to the
Figure 8: Relative-probability plots for the single-grain quartz SAR results from a fluvial-sand sample of Antarctica (Berger, unpublished data). Only 56 of 3000 grain-hole stimulations yielded acceptable $D_e$ values (using normal data acceptance criteria). Estimated youngest-age $D_e$ values are presented in the text.

MAM estimates than is the conventional weighted mean estimate of 1.65 Gy. The main difference between these weighted mean calculations and the MAM calculations is that the MAM models embrace (within 2σ) the sixth data point, shown outside the shaded bar in Figure 5.

The PD and TPD representations of the single-grain quartz $D_e$ values for this sample are shown in Figure 6. It is clear in Figure 6 that of the two plots, only the TPD plot indicates the true statistical significance (relative probability) of several high-value $D_e$ points. This TPD plot also draws more visual attention to the subgroup of $D_e$ values just on the high-side edge of the prominent youngest-age $D_e$ probability peak, thus hinting at the presence of more than one ‘Gaussian’ there. A comparable visual resolution of the $D_e$ values is attained with the corresponding radial plot (Fig. 7). Thus this single-grain example shows clearly the inadequate visual representation of the relative significances (statistical probabilities) of $D_e$ values provided by the conventional PD plot, and the statistically improved representation with use of the TPD plot.

The second example is of the representation of single-grain quartz $D_e$ data for quartz from fluvial sand in the McMurdo Dry Valleys of Antarctica (Berger, unpublished data). In Figure 8, as in Figures 4 and 6, the TPD plot changes significantly the visual representation of the relative probabilities of the $D_e$ values, and de-emphasizes the lowest $D_e$ values compared to the PD plot. This change appears to lead to a significant ambiguity in estimation of a youngest-age $D_e$ from this data set. On the one hand, the lowest-$D_e$ peak in the PD plot suggests that the filled-circle data (under bar A) could represent a statistically separate subgroup, for which a conventional (inverse variance weighting) weighted mean of 14.0 ± 1.5 Gy (n=10, internal standard error [SE], Topping, 1962) can be calculated. On the other hand, with use of the TPD plot these same data points do not appear to denote a separate subgroup. In this case only a subjectively selected range of $D_e$ values under bar B is 28.2 ± 2.0 Gy (external SE). However, close inspection (see $D_e$ values and error bars in Fig. 8) suggests that the REWM calculation is more appropriate. In this case, the REWM under bar A is 14.9 ± 1.6 Gy (internal SE) and under bar B, 28.2 ± 2.0 Gy (external SE). Thus use of the conventional weighted mean can be misleading (bar B) and use of the REWM can be uninformative. In these and all examples, only the largest (most conservative) of the two calculable SE values (Topping, 1962) have been selected. For the data in Figure 8 the MAM-3 estimate of 19.9 ± 3.1 Gy falls between the REWM estimates, and is more appropriate than either.

Thus this example shows that in general for single-grain data sets, the TPD plot is likely to give a more...
A statistically probable visual representation of the distribution of $D_e$ values than would the PD plot, but that statistical methods (e.g., MAM) other than visually-guided weighted mean calculations are likely required to assign confidence to any quantitative estimate of a youngest-age value from such distributions. This is further emphasized in the corresponding radial plot (Fig. 9). Here there is no clear visual resolution of sub-population $D_e$ values, except that which can be estimated by other means (use of either the TPD plot or the MAM computations).

The final single-grain quartz example is from a fluvial deposit in Arizona. In Figure 10, as above, there is a significant difference in visual representation when the TPD plot is employed. Not only is there a significant de-emphasis placed on the contribution of the 3 lowest-$D_e$ values, but also there is an apparently greater ambiguity introduced into the potential choice of $D_e$ values for youngest-age $D_e$ estimation. This example was chosen because it is known that the field-sampling and sample-storage/shipment (by a ‘third party’) created a laboratory sample for which it was difficult or impossible to exclude all grains that were exposed to daylight during sampling. Thus it is likely that the 3 lowest-$D_e$ data points are ‘contaminant’ quartz grains, and should be excluded from youngest-age estimation. Certainly they are not feldspar grains.

Unlike in Figures 4 and 6, the TPD plot in Figure 10 (more so than in Fig. 8) provides little or no visual guidance for the selection of subgroups of $D_e$ values for calculation of estimates of the last-daylight-exposure age. While the PD plot might suggest a statistically probable grouping under bar A (delimited roughly by the 50-60%-of-maximum limits of the peak), the TPD plot suggests that the range of this subgroup should be extended to that of either bars B or C. The corresponding conventional weighted means are: 16.1 ± 0.6 Gy (bar A, n=108); 17.6 ± 0.6 Gy (bar B, n=157); and 17.9 ± 0.6 (bar C, n=165). However, as mentioned, the use of REWM is more appropriate for such data (lognormal distribution, similar relative errors). Thus the respective REWM’s are: 18.6 ± 0.6 Gy; 22.9 ± 0.8 Gy; and 24.3 ± 0.9 Gy. Excluding the 3 lowest data points, the MAM-3 estimate for this distribution is 17.0 ± 1.7 Gy. This example therefore shows that for such a distribution the MAM provides probably the most statistically accurate estimate for the youngest-age grouping of $D_e$ values, but that the TPD main peak can denote visually roughly where the youngest-age estimate may be. In this case, a REWM estimate for a subgroup (bar A) of $D_e$ values close to the ‘half-Gaussian’ (peak low-side, e.g., Pietsch, 2009) part of the TPD peak (the peak is asymmetrical) is comparable to the MAM estimate.

Thus in this example, considerations of both the details of the sample-handling history and of the deconvolution approach are necessary for appropriate interpretation of embedded youngest-age information. These inferences are affirmed by the corresponding radial plot (Fig. 11). Here also an appropriate interpretation requires use of either the MAM computation or guidance from a TPD plot.
preferably the former, but coupled with before-hand knowledge of imperfections in field sampling.

In this context of the complexities of interpretation of $D_e$ distributions and of the topic of field-sample contamination, consideration of the effects on such distributions of the details of field sampling may be worth examining in the future. There are many examples of the presence of inexplicably ‘too-young’ $D_e$ values in radial plots (e.g., Jacobs et al., 2008a, 2008b). While unaccounted effects of microdosimetry may contribute to this presence, perhaps the use of ‘brute-force’ tube-sampling methods may also contribute via unobserved translocation of daylight-exposed grains into the interior length of the tubes. More attention needs to be directed to this aspect of single-grain dating.

Conclusions

The proposed transformed (using logarithms and relative error estimates) form of the conventional (unlogged) PD plot addresses the perceived main shortcomings of the PD plot when displaying $D_e$ distributions. The Transformed-PD (TPD) plot, employing the same parametric transforms as used in the radial plot, can place a statistically more realistic emphasis on the higher-$D_e$ data points. That is, like the radial plot, it also reveals meaningful relative structure in $D_e$ distributions, providing a display which can motivate the selection of quantitative methods for age calculation. Moreover, the TPD plot (accompanied by ranked $D_e$ values and error estimates) is visually easy to understand, especially when dealing with ‘partial bleaching’ (mixed age) populations for which a youngest-age estimate is required. Motivated by such plots, selection of $D_e$ values for age estimation can vary from use of the straightforward calculation of weighted means with standard errors (e.g., Topping, 1962) to the use of more sophisticated methods such as the minimum-age, central-age or mixed-age models (Galbraith et al., 1999; Arnold et al., 2009). However, for many groupings of $D_e$ values, transformed weighted mean calculations (using relative errors and ln[$D_e$] values) are more appropriate than are conventional weighted mean calculations.

Because the TPD plot employs a logarithmic transform of $D_e$ values, this plot (and the radial plot) cannot be used, unlike the PD plot, to represent $D_e$ distributions having negative values, such as are routinely encountered with modern-age or very young samples. For those samples, the PD plot is still useful for visual representation of the relative structure of the $D_e$ distribution and for selection of an appropriate subset of $D_e$ values for use in weighted mean calculation (e.g., Berger, 2009) or use in another approach (Pietsch, 2009). These subset calculations in turn can provide sufficiently accurate estimates of a youngest age. For either type of plot, it remains important to display concurrently a ranked series of $D_e$ values with their estimated errors.

In this presentation, it is implicit that a ‘one-size-fits-all’ style of graphical display of $D_e$ distribution data is unsuited to the variety of data likely to be generated from geological settings.

Acknowledgements and postscript

I thank Dr. Ashok Singhvi for careful comments that helped me to clarify this presentation. A spreadsheet form of TPD-curve-data calculation is available from the author. It will generate data suitable for plotting smooth curves via other software.

References


Condie, K.C., Belousova, E., Griffin, W.L., Sircombe, K.N. (2009) Granitoid events in space and time: Constraints from igneous and...


**Reviewer**

A.K. Singhvi
Decomposition of UV induced ESR spectra in modern and fossil dental enamel fragments

R. Joannes-Boyau and R. Grün

Research School of Earth Sciences, The Australian National University, Canberra, ACT 0200, Australia (e-mail renaud.joannes-boyau@anu.edu.au)

(Received 11 December 2009; in final form 22 April 2010)

Abstract

Using an automated simulated annealing (SA) procedure, spectrum decomposition on angular measurements of tooth enamel fragments shows that UV irradiation of modern human and fossil bovid samples results in distinctively different ESR spectra. In the fossil sample, UV irradiation generates qualitatively identical spectra to natural. The amounts of non-oriented CO$_2^-$ radicals in the modern and fossil samples are about 35% and 9%, respectively. The two oriented CO$_2^-$ radicals, R$_1$ and R$_2$, attributed to orthorhombic and axial types, show a ratio of 64:36 in the modern and 34:66 in the fossil sample. R$_1$ is also observed in the natural fossil sample, while the axial type was either absent or too small to be identified in a $\gamma$-irradiated fossil sample. We could not observe a measurable UV induced signal after 7 months of sunlight and laboratory light exposure, respectively. The clear difference between $\gamma$ and UV induced signal raises the possibility of using UV lights for dating protocols.

Introduction

Radiation induced signals in tooth enamel can be used for the evaluation of past radiation doses in retrospective dosimetry (e.g. Ikeya et al. 1984; Romanyukha et al. 1994), and dating (e.g. Grün, 1989, Grün et al. 2008). These two application areas differ with respect to the dose ranges evaluated and the age of the materials. While retrospective dosimetry deals with a dose range from a few mGy to 5 or 10 Gy on modern teeth, dating deals with old teeth (usually between several thousand and up to several million years old) and a dose range from a few Gy to several thousands of Gy. The radiation induced signals in fossil teeth are qualitatively different from those of modern teeth as any unstable signals in fossil teeth have partly or completely faded over geological times.

Previous studies suggested that the main ESR signal generated by gamma radiation can be attributed to two categories of CO$_2^-$ radicals, one anisotropic (AICOR) and the other with no preferential orientation, also called non-oriented CO$_2^-$ radicals (NOCOR) (e.g. Callens et al. 1995; Brik et al. 2000; Ishenko et al. 2002; Grün et al. 2008; Joannes-Boyau and Grün, 2009). In fossil teeth it was found that 9% of NOCORs were present in the natural sample, while 40% were present in the laboratory $\gamma$-irradiation component (Joannes-Boyau et al. in press, submitted). Those values differ from modern teeth, which have up to 80% NOCORs after $\gamma$-irradiation (Vorona et al. 2007; Rudko et al. 2007). However, the calculation method used for these studies differs from ours which could lead to systematic errors. Two studies by Joannes-Boyau et al. (in press, submitted) demonstrated that two types of AICORs (axial and orthorhombic) contribute to the ESR signal of fossil tooth enamel. In the $\gamma$-induced spectra, the axial form (R$_2$) was undetected (either absent or negligible) and only the orthorhombic form (R$_1$) contributed to the anisotropic ESR spectral components (Joannes-Boyau et al. in press).

Studies on retrospective dosimetry suggested that UV may contribute significantly to the overall ESR intensity (Liidja et al. 1996; Nilsson, 2001; El-Faramawy 2005). Brik et al. (1998) and Vorona et al. (2007) showed that UV induced spectra in modern teeth contained significantly less NOCORs than those by $\gamma$-irradiation. However, the relative depletion of NOCORs could have been the result of a combination of UV radiation and heating during the experiment. UV exposure is usually associated with significant heating but it is not known whether samples were cooled during these experiments. The occurrence of methyl radicals in the ESR spectra of Nilsson (2001) indicates a possible thermal influence, but these radicals could have been induced by UV irradiation itself.

The aim of the present study was to assess the influence of UV light on fossil teeth enamel, focusing particularly on dose estimation.
Materials and methods
The experiments were carried out on a tooth enamel fragment of a modern human (to avoid ethical problems, one of our own teeth was analysed) and on a fossil bovid from the archaeological site of Holon estimated at around 200,000 years old (for more details see Porat et al. 1999). A long lamella was separated from the fossil tooth (H2) using a dental diamond saw and a series of consecutive fragments were extracted and used for a range of heating and irradiation experiments (e.g. Grün et al. 2008; Joannes-Boyau and Grün 2009; Joannes-Boyau et al. in press, submitted). The modern human sample (MH) was used to validate the modus operandi for the Holon study and to compare results with previous studies conducted by Liidja et al. (1996), Nilsson et al. (2001), El-Faramawy (2005), Vorona et al. (2007) and Rudko et al. (2007).

The methodology established by Joannes-Boyau et al. (in press, submitted) was used for the present study. X, Y and Z denote configurations, x, y and z, the main axes of the measured fragment (Fig. 1). T1, B1 and B2 are positions in the measured or simulated ESR spectra, and R1, R2, R3 and B2 the fitted Gaussian components (for more details see Joannes-Boyau et al., submitted). The concentrations for the radicals were derived from the double integration of the fitted lines to account for changes in the line width. For the features in the measured ESR spectra, T1-B1 and B2, it was not possible to carry out double integrations; their angular variations were derived from their intensities (for more details see Joannes-Boyau et al., in press, submitted).

Both fragments H2 and MH were each mounted in three separate Teflon holders containing a Parafilm mould and were incrementally measured by rotating them around their three major axes. We used the following configurations: X: rotation around the axis perpendicular to the dentine-enamel junction, Y: around the axis of tooth growth and Z: perpendicular to X and Y (Fig. 1). The sample holders were inserted in a Bruker ER 218PG1 programmable goniometer and measured with a Bruker Elexsy E500 ESR spectrometer in 10° increments over 360° with the following measurement conditions: 2 mW microwave power, 0.1 mT modulation amplitude 12 mT sweep width with a sweep time of 21 s. The spectra were accumulated over 50 consecutive measurements. The sample was measured before and after irradiation.

The fragments were exposed to UV light for 168 hours using a Hoenle UVASPOT 400T lamp emitting UVA/B at 400 W, producing incident energy of 5.05±0.1 mW cm⁻² at the sample location. The temperature was recorded in 2 min intervals with a thermolog controller near the surface and at the base of the fragments, located near the cooling plate. The measured temperatures were 21±2°C at the base of the sample and 33±2°C at the top. The γ-irradiation was carried out with a ¹³⁷Cs-source for 100 min, which corresponds to an approximate dose of 187 Gy (the experiment is described in Joannes-Boyau et al., in press).

Extraction of the isotropic components of the UV induced spectra
The first step in spectra decomposition of H2 consists of subtracting the natural signal from the UV irradiated spectra. The alignment of the two spectra is the key to avoid artefact signals, for that matter the isotropic methyl lines are used as markers for alignment. Previous studies by Joannes-Boyau et al. (in press, submitted) have shown that non-CO₂ components are found in the natural signal and in the irradiated signal (Fig. 2a). Those signals described as isotropic lines have to be removed before decomposing the ESR spectrum at each angle. In the fossil sample, the isotropic lines in the UV-induced spectra were found to have a similar signature as in the natural (combined into one signal, named Isotropic Combined signal (IC)) (Fig. 3), but increased with UV exposure. No methyl line was created by UV exposure. A wide line, W₂, at g=2.0051 (see Joannes-Boyau et al. submitted),

![Figure 1: Direction of axes and configurations used for the measurement of the tooth enamel fragment (from Joannes-Boyau and Grün 2009).](image-url)
Figure 2: Decomposition of the measured spectra (for more details see Joannes-Boyau et al. in press). A: (Top) Isotropic lines in the natural spectra. A simulated signal is fitted into the natural measured spectra to extract the isotropic component (IC is the residual offset for clarity). (Bottom) Position of the isotropic components that when merged form the Iso-combined spectra (IC). B: (Top) Comparison of the measured and simulated spectra. The residual is offset for clarity and corresponds to the subtraction of the simulated spectra from the measured. (Bottom) Decomposition of the measured spectra using the SA decomposition with four Gaussian components representing the anisotropic lines $R_1$, $R_2$, $R_3$ and $B_2$.

Figure 3: Comparison of the isotropic signal found in the modern and fossil tooth natural spectra IC (from Joannes-Boyau et al., submitted) (a), fossil tooth UV irradiation spectra (b), modern human tooth UV irradiation spectra (c) and modern human tooth natural spectra (d) respectively. The symbol $S$ indicates the unknown isotropic line that appears with UV exposure described in the text with a g-value around $g=2.0093$.

Figure 4: Comparison of the influence of different exposure on the merged ESR signal of modern human tooth enamel fragment. (a) Native signal N.S.; (b) 7 months laboratory lights (c) 7 months sunlight exposure (d) 168 hours UV lamp exposure.
attributed to a combination of \( \text{SO}_2^- \) and \( \text{CO}^- \) radicals, increased. \( \text{SO}_2^- \) radicals which occur at \( g=2.0056 \) are most likely responsible for this increase (Bouchez et al. 1988; Schramm and Rossi 1999, see also the fitting of Grün 2002). The \( \text{SO}_2^- \) radical is known to be temperature sensitive. It is likely that the \( \text{SO}_2^- \) radicals were created at the surface of the sample where the temperatures were slightly higher due to the UV exposure. A new isotropic line ($) at \( g=2.0093 \), with a line width of around 0.1 mT (Fig. 3) was created by UV radiation. This new, unknown component does not interfere with the main signal and fades rapidly (its intensity is negligible after three months storage at ambient temperature). The NOCORs were removed using the same amount found in the natural signal corresponding to 9% of the total intensity. After subtracting all aforementioned components (natural, isotropic lines and NOCORs), only UV-induced anisotropic components remain.

**Spectrum Decomposition**
The anisotropic components of the irradiation spectra were decomposed with an automated simulated annealing (SA) procedure which is particularly well suited to separating overlapping signals (Fig. 2b) (for more details see Joannes-Boyau et al. in press, submitted). SA is a Monte Carlo method used for combinatorial optimisation problems (for details see Metropolis et al. 1953; Kirkpatrick et al. 1983; Černý 1985; Mossegard and Sambridge 2002; Bodin and Sambridge, 2009). The spectra were decomposed with four Gaussian lines which had the same prescribed \( g \)-value range limits used in the previous study to avoid unrealistic solutions outside the regions for the \( \text{CO}_2^- \) radicals in hydroxyapatite (see above). No restrictions were set on the intensity although the line widths were kept between 0.10 mT and 2 mT to avoid aberrations. Our SA procedure is able to randomly generate a large number of synthetic spectra defined by a linear combination of four Gaussian lines. Each simulated spectra is compared to the measured spectra in terms of a least square misfit.

**Results and discussion**
The modern human tooth (MH) contains a wide, isotropic signal centred at 2.0044 ± 0.0005 prior to any irradiation, also called the native signal (NS) (Liidja et al. 1996; see Figs. 3 and 4). Due to the intensity and width of the signal compared to the background, the measured \( g \)-value is not very precise. The spectra confirm that the enamel fragment was not exposed to any measurable ionising radiation. The ESR spectra after 168 hours UV exposure is qualitatively different from the natural, as the shapes of the two spectra clearly differ (Fig. 4).

![Figure 5: Angular variation of ESR signals induced by UV exposure in a modern human fossil tooth enamel fragment. Summary of all decomposition results of the radical concentration (obtained by double integration) variation over 360° (a) X-configuration, (b) Y-configuration and (c) Z-configuration of \( R_1 \), \( R_2 \), \( R_3 \) and \( B_2 \) component. Note that on the Z-configuration \( R_3 \) shows a 90° symmetry.](image-url)
multiplied difference while (2001). This makes us confident that temperature was found in our measurements, contrary to Nilsson et al., 2007). The exception is that no methyl radicals appear in the UV induced spectra. The T1B region is slightly shifted to 270:230, 100:140, 300:260, 0.024 and 0.017, respectively (Fig 3). These residuals are approximately 40% in X, Y and Z configurations, respectively, indicating different orientations of the radicals in the crystal structure. Similar values were found by Joannes-Boyau and Grün (2009) in the sample from Holon with offsets of -26°, 7° and -5° ± 10° in X, Y and Z configurations, respectively. The angular differences between R1 and R2 of the two teeth point to differences in the tooth formation of the two species.

The results obtained on MH are similar to those described in the literature (e.g. Callens et al. 1995, 1998; Lidija et al. 1996; Nilsson et al., 2001; El-Faramawy 2005; Vorona et al. 2007; Rudko et al., 2007). The exception is that no methyl radicals were found in our measurements, contrary to Nilsson et al. (2001). This makes us confident that temperature was well controlled by the cooling plate used in the present study and that the anisotropic signal extracted from the fossil tooth is UV rather than temperature induced. After 3 months, the UV induced spectra show major changes. The NOCORs have faded by 15% of their intensity but still represent 30% of the initial spectra. The T1-B1 region is slightly shifted to higher g-values and shows a depletion of 3 to 5% of its original signal. The ratio between R1:R2 has changed to 59:41 respectively. The ratio disparity after 3 month could be attributed to the disappearance of some of the R1 species and at the same time the appearance of new R2 radicals. However, because of the small amount of variation (between 3 to 5%), we cannot conclude that any transfer process took place within the 3 months.

The fossil bovid tooth

In Z-configuration the angular variations of the natural and UV spectra show little difference while those of the γ-spectra are significantly muted (Fig 6). This can be attributed to the much higher contribution of the NOCORs in the γ-component (around 40%) compared to the natural (around 9%, see Joannes-Boyau et al. in press). This also implies that the relative distributions of the NOCORs in the natural and UV components are approximately the same; otherwise the angular variation would change dramatically such as the γ-induced variation, since NOCORs remains constant at all angles (see Joannes-Boyau et al., in press).

Fig. 7 compares a natural spectrum to those of the γ- and UV components. While it is not possible to scale the natural spectrum to fit the γ-component (Fig 7a), due to a distinctively different distribution of NOCORs (Joannes-Boyau et al. in press), the natural spectrum can be easily scaled into the UV component (Fig. 7b). This implies a closely similar distribution of the different types of CO2 radicals in the natural and UV components. Scaling the natural spectrum by a factor of 1.75 into the natural + UV, l over the UV components. Scaling the natural spectrum by a factor of 1.75 into the natural + UV, l over the UV components.

Table 1: Results of the decomposition of the anisotropic components of the modern tooth after UV-irradiation

<table>
<thead>
<tr>
<th>Angle (°)</th>
<th>Maximum conc. AIC ORs</th>
<th>Relative conc. NOCORs</th>
</tr>
</thead>
<tbody>
<tr>
<td>150</td>
<td>0.11</td>
<td>0.07</td>
</tr>
<tr>
<td>190</td>
<td>0.12</td>
<td>0.09</td>
</tr>
<tr>
<td>230</td>
<td>0.13</td>
<td>0.11</td>
</tr>
<tr>
<td>270</td>
<td>0.13</td>
<td>0.14</td>
</tr>
</tbody>
</table>

The native signal grows slightly with irradiation, at the same time that CO2 radicals appear in the spectra. Subtracting the native signal (multiplied by a factor 1.2) plus an isotropic component (IC MH in Fig. 3) which is similar to the one used in the natural sample from Holon (IC Natural, see Joannes-Boyau et al., submitted), yields typical CO2 radicals. The amount of NOCORs is around 35%.

![Table 1: Results of the decomposition of the anisotropic components of the modern tooth after UV-irradiation](image-url)
Figure 6: Angular variation of the natural, UV induced and γ-induced spectra in Z-configuration on fossil tooth enamel fragments.

Table 2: Angular variation

<table>
<thead>
<tr>
<th></th>
<th>T1-B1</th>
<th>B2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural</td>
<td>Z</td>
<td>0.6</td>
</tr>
<tr>
<td></td>
<td>Y</td>
<td>0.62</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>Z</td>
<td>0.44</td>
</tr>
<tr>
<td>Natural+γ</td>
<td>Y</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>Z</td>
<td>0.61</td>
</tr>
<tr>
<td>Natural+UV</td>
<td>Y</td>
<td>0.61</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>0.30</td>
</tr>
<tr>
<td></td>
<td>Z</td>
<td>0.66</td>
</tr>
<tr>
<td>Natural</td>
<td>Y</td>
<td>0.68</td>
</tr>
<tr>
<td>(anisotropic)</td>
<td>X</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>Z</td>
<td>0.76</td>
</tr>
<tr>
<td>γ</td>
<td>Y</td>
<td>0.78</td>
</tr>
<tr>
<td>(anisotropic)</td>
<td>X</td>
<td>0.45</td>
</tr>
<tr>
<td>UV</td>
<td>Z</td>
<td>0.67</td>
</tr>
<tr>
<td>(anisotropic)</td>
<td>Y</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>X</td>
<td>0.37</td>
</tr>
</tbody>
</table>

Fig. 9 shows the stacks of all spectra of the anisotropic components of the natural sample as well as of the UV and γ-induced (see also Tables 2 to 4). The γ-irradiation stacks, especially the Y-configuration, show a very different pattern to the corresponding UV stacks (Fig. 8, middle column, rows B and C). The T1-B1 complex is significantly narrower in the γ-irradiation spectra than in the other two. The angular variations of the T1-B1 and B2 positions (marked in Fig. 6) of the γ-irradiation spectra are significantly more pronounced in all configurations than of the natural and UV spectra (Table 5). At the same time, it is unclear why UV exposure induces a smaller amount of NOCORs than γ-irradiation. Perhaps this is due to the energy difference between UV emissions (in the range of 3 to 10 eV) and γ-rays (>600 keV). So far no explanation can be clearly proposed, however, local saturation of radicals could induce significant differences in the signals.

Fig. 10 and Tables 2 to 4 summarise the results of the decomposition. Note that the R2 component was not identifiable in the γ-irradiation spectra (Joannes-Boyau et al. in press) while it is present in both natural and UV spectra. The g-values of the four components are in a similar range and do not show any significant shifts for the natural and UV irradiated signals. Their line widths appear slightly narrower in UV than in the natural, but this parameter always shows the largest deviations in repeated SA runs (Joannes-Boyau et al. submitted). The average and relative radical concentrations of natural spectra...
Figure 8: Stacks of fossil bovid enamel natural spectra multiplied by a factor of 1.75 (row A), natural + UV irradiation (row B) and residual from the subtraction of the natural (x1.75) from the natural + UV irradiation (row C) for the three configurations X, Y and Z, column 1, 2 and 3 respectively.
Figure 9: Stacks of fossil bovid enamel: natural spectra (row A), natural anisotropic (row B), natural + UV irradiation (row C), UV exposure anisotropic spectra (row D) and γ-irradiation anisotropic spectra (row E) for the three configurations X, Y and Z, column 1, 2 and 3 respectively.
Figure 10: Summary of the angular behaviour of the anisotropic components of the natural spectra as well as the UV irradiation spectra decomposed with four Gaussian components and with the γ-irradiation spectra decomposed with three Gaussian components (see Joannes-Boyau et al., in press).
are somewhat different to the UV induced (compare Table 2 with 3). The angular variations for all components of UV spectra were slightly higher than the natural, but were smaller than the γ-irradiation spectra (compare Tables 2, 3 and 4). This clearly implies that UV radiation causes a significantly different radical distribution in the enamel than γ-irradiation. The fading observed on the UV irradiated sample is visible in the enamel for modern human tooth after three months is not as evident for the fossil bovid tooth. The difference in signal intensity between the last UV exposure and the fading test conducted 3 months later is in the range of 3 to 4% of the total intensity. Prima facie, it appears that the diminution is induced by the fading of some NOCORs. However, the spectra depletion falls into the error range and therefore does not allow any affirmation.

Average R₁/R₂ ratios in natural and UV components are different in the three configurations (compare Tables 2 and 3). Nonetheless, when normalising on the total radical concentration, R₁/R₂ in the natural (35:65) is virtually the same as the UV ratio (34:66). Based on the ESR intensity of the fragment, we have estimated that the UV lamp is the γ-equivalent of 3.1 ± 0.3 mGy/min. Since the natural and UV spectra show closely similar radical distributions (in contrast to γ-irradiation), UV irradiation could be more suitable than γ-irradiation for the establishment of dose response curves in ESR dating. While fading, unknown intensity attenuation and energy calibration will complicate this approach at the present time, systematic correlations between UV exposure and the equivalent dose from γ-irradiation could lead to the use of photons instead of γ-rays for dating purposes. Further, studies should be undertaken on the effect of UV on the equivalent dose (Dₑ) assessment, however, because UV irradiation is a very slow process, local saturation could happen during exposure. For that reason, a specific protocol should be designed to measure the sample while irradiating, which would complicate the experiment greatly.

The high radical concentrations induced by UV raise the possibility that sunlight or laboratory light induced radicals may interfere with dose estimations.
However, intermittent exposure to daylight for more than 50 years while laughing about ones own bad jokes did not induce any measurable CO$_2$$_2$ radicals in MH, neither did exposure to indirect sunlight or laboratory light over 7 months in the fossil sample (Fig. 4). Samples are normally shielded behind window glass, known to block UV, and it is therefore unlikely that archaeological samples are exposed to direct UV sunlight over extended periods of time.

Summary
UV and γ-irradiation induce very different compositions of CO$_2$$_2$ radicals in tooth enamel. In a modern sample, UV generated 35% of NOCORs and a mix of 64:36 of orthorhombic to axial radicals. In the fossil sample, UV generated 9% NOCORs and a mix of 34:66 of orthorhombic to axial radicals. While there are some differences between the natural and UV components in the various configurations, the overall radial distribution of the UV and the natural is the same. This is in contrast to γ-irradiation component of the fossil sample, which had about 40% AICORs and no axial radicals. While the UV components in the modern samples showed strong fading over three months (18 to 20% of the spectra) with possible transfer process between R$_1$ and R$_2$, the fading in the fossil sample was small (3 to 4%).

Blocked sunlight and laboratory light exposure over 7 months had no measurable influence on the samples.

Conclusions
Like γ-irradiation, UV irradiation induces significant differences between modern and fossil samples. While this change must have an impact on dose estimations, this aspect has never been systematically investigated in ESR dating. In the fossil sample, UV generates a similar mixture of CO$_2$$_2$ radicals as found in the natural while the γ-irradiation response is completely different. This could make UV irradiation the choice for the establishment of dose response curves.

Acknowledgments
We are very grateful to T. Bodin, Research School of Earth Sciences, Australian National University (ANU), Canberra, for helping with the design of the simulating annealing program used for this work. We thank N. Manson, Research School of Physical Sciences, ANU, for helpful comments. We are grateful to F. Callens and H. Vrielinck, Gent, for their considered advice in the earlier stages of this study. Aspects of this study were supported by the ARC funded DP0664144 Microanalysis of human fossils: new insights into age, diet and migration. We thank Kathryn Fitzsimmons, RSES, for critical comments.

RG is grateful to the Institut des Sciences humaines et sociales du CNRS, Bordeaux, and the Laboratoire d’Anthropologie des populations du Passé, Université de Bordeaux I, for their kind hospitality in the writing-up stage of this manuscript. RJB and RG would like to thank Jean-Jacques Bahain at the Institut de Paléontologie Humaine for reviewing the paper in such a short period of time.

References


Reviewer
Jean-Jacques Bahain
Thesis Abstracts

Author: Yorinao Shitaoka
Thesis Title: Contribution of luminescence dating to the Paleolithic studies in East Asia
Grade: Ph.D.
Date: March 2005
Supervisor: Masayoshi Mizuno, Tsuneto Nagatomo
Address: Department of Literature, Nara University, Misasagi 1500, Nara, Japan

The objective of this study is to apply optically stimulated luminescence (OSL) dating to archaeological sediments in order to establish a chronological framework of paleolithic archeology in East Asia. Moreover, thermluminescence (TL) dating of Japanese marker–tephras (e.g. Aira-Tanzawa pumice, AT) was carried out. 64 samples were collected from 27 locations from Nihewan basin, China, Imjin–Hantan basin, Korea, Russian Far East (the Maritime region), Sakhalin and Japan. The paleodose of each sample was estimated using quartz coarse grains, polymineral fine grains and quartz fine grains.

For make luminescence measurements, an original TL/OSL system (NRL–99–OSTL) was developed. It has two assemblies for temperature control, one can raise the temperature to 500 °C and the other is cooled at −150 °C by flowing liquid nitrogen. The TL/OSL system automatically measures the TL and OSL up to 32 aliquots. For OSL measurements, a stimulation source that consists of 32 blue light emitting diodes (LEDs) at 470 nm (Nichia Chemical Industries Co., Ltd., Japan) is installed. This LED unit is exchangeable with another unit with 32 infrared LEDs at 890 nm (Hamamatsu photonics Co., Ltd., Japan). Luminescence signal is detected by a photomultiplier tube (Hamamatsu R1140P) which is cooled to −20 °C through two condensing lenses and optical filters.

OSL ages of archeological sediments were obtained between 220 ka and 6 ka. TL ages of marker tephras ranging between 530 ka and 26 ka were obtained. These luminescence ages are in agreement with the various paleoenvironmental data. These ages will contribute to understand the paleo–human activity and paleolithic industry in East Asia.

Author : Ya-Wen Chen
Thesis Title: Luminescence dating on Quaternary sediments: Cases of active structures in western Taiwan
Grade: Ph.D.
Date: July 2008
Supervisor: Yue-Gau Chen, Andrew S. Murray
Address: Department of Geosciences, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei 106, Taiwan, Republic of China

Hundred thousand years is only a short period in the history of earth, however, it is the most momentous period for human being. If scientists would like to predict the future, to learn the trend of the past is the key and reliable chronological data are needed. This study presents the feasibility and application on luminescence methods. After multiple tests this method has been confirmed to provide confident results to interpret regional tectonics. Sequence stratigraphy and continuous chronological framework, especially from the luminescence dating, are integrated to investigate the accommodation space for specific drill site in southwestern coastal plain of Taiwan. It concludes the tectonic subsidence has been on-going since very late Pleistocene with a depo-centre located further southwestwards. Also a tectonic uplift due to detachment westward migration is proposed to occur after 100 ka. Furthermore, with the anticipation to diminish losses during earthquake disaster, multiple dating methods, in particular focusing on luminescence dating method, are employed in the studies to explore the details of active faults. The Chiuchiungkeng Fault and Chelungpu Fault have been both tested in this study to deduce the long-term slip rate and fault behavior. Based on the proper experiment criteria for each different area, luminescence dating method is acceptable to be adopted for the Quaternary researches in western Taiwan.
The purpose of this thesis is to find a new method for dating raised-beach deposits in order to create more accurate sea-level curves for the Antarctic Peninsula. Reconstructions of the Antarctic Peninsula Ice Sheet (APIS) at the Last Glacial Maximum remain poorly constrained due to a lack of geologic data, which has resulted in a variety of ice-sheet models for the APIS during the Late Pleistocene and Holocene. Sea-level data can be used to estimate the past thickness and volume of ice sheets by comparison with geophysical models of the response of the Earth to ice and water loading. The few sea-level curves that do exist along the Antarctic Peninsula are taken from areas of known tectonic activity or are limited by the use of radiocarbon dating. Radiocarbon dating is limited by the availability of organic material, the poorly constrained radiocarbon reservoir for Antarctica, and the possibility of reworking. We test the use of optically stimulated luminescence (OSL) in dating cobbles from raised beaches within the South Shetland Islands of the Antarctic Peninsula. OSL is commonly used for sediment, but the application to rocks is in its infancy. Methods were developed to isolate quartz grains from the shielded undersides of cobbles from raised beaches last exposed to sunlight in the intertidal zone.

Of the 12 samples obtained from the shores of Maxwell Bay, 8 contained enough material for OSL analysis. The corrected OSL ages for the beaches range from 0a (modern beach) to 2048a +/- 123 (9.0m beach) and are in agreement with calibrated radiocarbon ages from the same deposits. In addition, the OSL ages are all internally consistent in that ages obtained from higher ridges are always older than lower ridges and ages obtained from the same ridges overlap one another. In conjunction with their elevation, the OSL ages were used to reconstruct sea-levels for the South Shetland Islands. Because of the large radiocarbon reservoir in Antarctica, OSL is shown to be a more precise method for dating raised beaches than radiocarbon techniques.

The position of the southern African subcontinent in the mid-latitudes means this region was influenced by fluctuations in a number of atmospheric and oceanic climate circulation systems during the late Quaternary. Whilst the reconstruction of palaeoenvironmental and palaeoclimatic conditions in southern Africa has developed rapidly over half a century, our understanding remains limited by poor spatial coverage and sources of uncertainty within our existing data. The availability of terrestrial proxy archives is restricted by the arid nature of the environment. Sedimentary landforms, such as aeolian dunes and the silt, mudstone and tufa deposits associated with fluvial systems are vital sources of palaeoenvironmental information.

This thesis considers the dimensions of uncertainty in three key terrestrial proxy archives. Linear sand dunes, interdune water-lain deposits and tufa are used to reconstruct palaeoenvironmental conditions at three sites in Namibia. The uncertainty relates to sampling strategies, chronological control and palaeoenvironmental interpretation. Optically stimulated luminescence (OSL) dating has been applied to linear dunes in the west of the southern Kalahari linear dunefield and to interdune deposits in the northern Namib Sand Sea, whilst the utility of $^{234}$U-$^{230}$Th dating was tested for tufa deposits in the Naukluft Mountains.

This study demonstrates the influence of sampling strategy on the dunefield-scale record of linear dune accumulation; the choice of vertical sampling interval is important, a young bias is introduced in datasets with a shallow sampling bias, and small datasets demonstrate a reliance on individual sites. This study also provides a revision of radiocarbon based chronologies for water-lain units in the northern Namib Sand Sea. This adds further evidence to avoid radiocarbon-based humidity-proxy histograms using inorganic carbonates for reconstructing palaeoenvironments. This study establishes a rigorous methodology for increasing confidence in $^{234}$U-$^{230}$Th dating of fluvial tufa deposits. The record of...
environmental change preserved at these three sites provides insight into the dynamic response of these terrestrial proxy archives to Quaternary climatic fluctuations since MIS 5. Southern Kalahari linear dunes record a period of accumulation at the transition from MIS 2 to the Holocene, centred on ~10 ka. The Tsondab River progressively retreated eastward from MIS 5 to present. The Naukluf tufa include deposits of considerable antiquity, deposition of barrages prior to MIS 5 and some Holocene deposition inside the channel.

Marine sediments contain important archives of past ocean and climate changes, but at high latitudes, such as the polar regions, the absence of carbonate has prevented the construction of accurate chronological models. To get an age model, a method which does not rely on carbonate is needed. In this thesis we have investigated the potential of optically stimulated luminescence (OSL) dating to establish a chronology for our deep sea sediment cores from the Northern Pacific area. OSL dating makes use of the omnipresent quartz and feldspar grains in the sediment so there is no limitation in the presence of the dosimeter. Optical dating is a widely accepted dating method for terrestrial sediments, but only few studies have tested its reliability in the marine environment.

A luminescence age reflects the time that has elapsed since the sediment grains were last exposed to sunlight. The luminescence age equation contains two equally important factors: the equivalent dose and dose rate. The palaeodose is the total radiation dose that crystals (quartz, feldspar) have absorbed during burial, and the dose rate is the rate at which the sample was exposed to ionising radiation in the environment. Ionising radiation comes from the decay of natural radionuclides (the 238U, 232Th series and 40K) present in the sediment. Dividing the equivalent dose by the dose rate gives the luminescence age of the sample.

This study is focused on testing the reliability of the optical dating method when applied to marine sediments from the Northern Pacific ocean, more specifically in regions that are known for seasonal sea-ice. The aims of this thesis are (1) determine the accuracy of luminescence ages, e.g. by comparison with AMS 14C dating and marine oxygen isotope stratigraphy where possible, (2) produce a high resolution sequence of absolute ages which can describe changes in sedimentation rate through time, and (3) test whether it is possible to date back to marine isotope stage 5e (MIS 5e).

In this thesis we have made use of fine (4-11 μm) grains of quartz extracted from the marine sediment cores taken in the south-western Sea of Okhotsk, the Bering Sea and the Central Sea of Okhotsk. In the first study, a high resolution optical age dataset (64 samples) from the south-western Sea of Okhotsk showed continuous ages up to 24 ka and indicated clear sedimentation rate differences during glacial and interglacial periods. Also for the most northern located core in the Bering Sea (sub-arctic area), the OSL ages (12 samples) go back to 64ka and the ages are in good agreement with the 14C ages up to 25ka. The OSL ages from both of south-western Sea of Okhotsk and the Bering Sea are in good agreement with the 14C ages. In the third study in the central Sea of Okhotsk, a test of accuracy of the OSL ages back to MIS 5e was carried out. The OSL ages (40 samples) are in good agreement with the well-established oxygen isotope stratigraphy at this site back to the MIS 5e (130ka). In all these studies, the luminescence characteristics of the dosimeter are investigated in detail and extensive tests of the performance in the chosen measurement protocol are presented.

Despite the good agreement of the optical ages with the independent age control, it must be noted that the OSL ages are largely dependent on the water content. The evaluation of the appropriate water content model is discussed. It appears that the observed water content values, measured immediately after core extraction, seem to give the most accurate ages.

This thesis shows that OSL dating is a useful method for dating marine sediment in the North Pacific area. The data confirm that the OSL dating using fine-grain quartz that is distributed all over the ocean’s sediments has very great potential in the establishment of an absolute chronology for deep sea sediments; because luminescence dating does not depend on the presence of carbonate for AMS 14C dating, it is now likely that we can establish a
chronology is regions of the ocean that were previously undatable.

Author: Daniel Rufer  
Thesis Title: Characterization and age determination of Quaternary volcanism in the southern Ankaratra region (central Madagascar) through novel approaches in luminescence dating  
Grade: Ph.D.  
Date: October 2009  
Supervisors: Guido Schreurs (Bern), Alfons Berger (Copenhagen), Edwin Gnos (Muséum d'Histoire naturelle de Genève), Frank Preusser (Bern)  
Address: Institute of Geological Sciences, University of Bern, Switzerland

This work introduces two novel approaches for the application of luminescence dating techniques to Quaternary volcanic eruptions: crystalline xenoliths from lava flows are demonstrated to be basically suitable for luminescence dating, and a set of phreatic explosion deposits from the Late Quaternary Vakinankaratra volcanic field in central Madagascar is successfully dated with infrared stimulated luminescence (IRSL).

Using a numerical model approach and experimental verification, the potential for thermal resetting of luminescence signals of xenoliths in lava flows is demonstrated. As microdosimetry is an important aspect when using sample material extracted from crystalline whole rocks, autoradiography using image plates is introduced to the field of luminescence dating as a method for detection and assessment of spatially resolved radiation inhomogeneities.

Determinations of fading rates of feldspar samples have been observed to result in aberrant g-values if the pause between preheat and measurement in the delayed measurements was kept short. A systematic investigation reveals that the phenomenon is caused by the presence of three signal components with differing individual fading behaviour. As this is restricted to short pauses, it is possible to determine a minimal required delay between preheating and measurement after which the aberrant behaviour disappears.

This is applied in the measuring of 12 samples from phreatic explosion deposits from the Antsirabe – Betafo region in the Late Quaternary Vakinankaratra volcanic field. The samples were taken from stratigraphically correlatable sections and appear to represent at least three phreatic events, one of which created the Lac Andraikiba maar near Antsirabe. The obtained ages indicate that the eruptive activity in the region started in the Late Pleistocene between 113.9 and 99.6 ka. A second layer in the Betafo area is dated at approximately 73 ka and the Lac Andraikiba deposits give an age between 63.9 and 50.7 ka. The youngest phreatic layer is dated between 33.7 and 20.7 ka.

These ages are the first recorded direct ages of such volcanic deposits, as well as the first and only direct ages for the Late Quaternary volcanism in the Vakinankaratra volcanic field.

This illustrates the huge potential of this new method for volcanology and geochronology, as it enables direct numerical dating of a type of volcanic deposit which has not been successfully directly dated by any other method so far.

Author: Sally E. Lowick  
Thesis Title: Luminescence dating of long palaeoenvironmental records from the Alpine foreland  
Grade: Ph.D.  
Date: December 2009  
Supervisor: Frank Preusser  
Address: Institute of Geological Sciences, University of Bern, Switzerland

Optically stimulated luminescence (OSL) was applied with the aim of constraining the chronology of long palaeoenvironmental records from the Alpine foreland. Preliminary work on samples from the Valeriano Creek on the southeastern Alpine foreland identified anomalous fading in the feldspar fraction, but was able to successfully date both coarse and fine grain quartz from the same samples to beyond 200 ka. These ages helped to constrain the deposition of sediments in order to understand the environmental response to climate change, and allowed a comparison of fluvial dynamics north and south of the Alps. The work then moved to dating a long sedimentary core taken from Azzano Decimo in the Friulian foreland of northeastern Italy, from which the palynostratigraphy required chronological constraint. Due to desiccation of the core, only the consolidated silty sediments were sampled to ensure they had not been exposed to light, and a method was formulated for retrospectively calculating the water content of the samples. This method was then tested on fresh samples and results suggested that it was a reliable procedure. Dating of the polymineral fraction proved unsuccessful and so the work focused on the
fine grain quartz and this was measured principally using a modified single-aliquot regenerative-dose protocol. Despite all samples meeting the performance criteria usually set for assessing the reliability of the protocol, the samples suffered significant age underestimation beyond ~ 70 ka when compared to independent age constraint. The determination of dose rates using both high resolution gamma spectrometry and instrumental neutron activation analysis compared well, suggesting that these were reliable, and also confirmed the absence of any radioactive disequilibrium. As age underestimation in quartz is now frequently reported in the literature, the study then focused on investigating its origin by comparing the behaviour of the Azzano quartz with that from the Niederweningen mammoth site, north of the Alps, where OSL ages agree well with expected ages up to 200 ka. The application of both a single-aliquot regeneration and added-dose, and a sensitivity-corrected multiple aliquot regenerative protocol, were not able to overcome the age underestimation seen at Azzano. Both the stability of the OSL signal used for dating, and the source traps of this signal were investigated, and suggest that it should be reliable for dating. Although the quartz OSL dose response is generally expected to fit well to a single saturating exponential function, samples from both sites displayed an additional linear growth at high doses. Several avenues of investigation were unable to identify the origin of this component, although analysis of the growth displayed by burial doses confirmed that this phenomenon was also present in the natural signal and not only generated within the laboratory. In all comparative analyses of samples that display both a significant age underestimation and those that agree well with expected ages, it was not possible to detect any variation in their behaviour, or an indication of when equivalent dose determination could be proved reliable.
Bibliography
Compiled by Daniel Richter

From 1st November 2009 to 30th April 2010


Papers from the 12th International Conference on Luminescence and Electron Spin Resonance Dating (LED2008) held in Beijing, published in Volume 5/2-3 of Quaternary Geochronology


Berger, G. W., Ante, S., and Domack, E. (2010). Luminescence from glacimarine sediment-trap samples at the Antarctic Peninsula. *Quaternary Geochronology* 5, 244-249.


Announcements

Risø TL/OSL Reader software updates available on the web

You may now download the latest software for your TL/OSL reader from the Risø home page www.osl.risoe.dk. Once you reach the website, select the option on the left hand side of the screen for ‘The Risø TL/OSL reader’ and then the section entitled ‘Software’. As well as being able to download the latest release of the different software packages, there is also a section entitled “Revision History” where you can see the most important changes that have been made to the software since previous versions were released.

For instance the Sequence Editor has recently had a few small but very practical changes:

1. When you run a sequence, the grid with the commands does not disappear as it used to, but now it is shown in the lower part of the screen next to the usual communication log window. The cell being executed is marked, and you may freely inspect all the cells, but no changes can be made during the run.

2. In “Sequence Options” you may tick “Sequence file copy” (this is the default setting). When this is done, you automatically save a copy of the sequence file (*.SEC file) and this is automatically given the same name and path as the data file (*.BIN file). This file can be opened and used again, but not saved again without changing the name. The idea is that you always store a .SEC file as documentation for how the .BIN file was created. A detailed description of this system is available on the home page under the heading “Tech and Change Notes”

Torben Lapp
Radiation Research Division, Risø DTU, Denmark