PROBABILITY DENSITY PLOT FOR FISSION-TRACK GRAIN-AGE SAMPLES

MARK T. BRANDON

Department of Geology and Geophysics, Kline Geology Laboratory, P.O. Box 208109, New Haven, CT 06520-8109, U.S.A.

Abstract—This paper outlines the statistical basis, construction and interpretation of probability–density (PD) plots as applied to the analysis of mixed distributions of fission-track (FT) grain ages. Such distributions can be viewed as being composed of a mixture of elemental component distributions, with each component characterized by a unique average FT age. The original PD plot of Hurford *et al.* is shown to have a clear statistical basis, as defined by the Gaussian-kernel method of density estimation. This background is used to develop and justify an improved version of the PD plot. Three modifications are recommended: (1) as originally suggested by Galbraith, the PD plot should be constructed using the transform variable z, which is approximately proportional to the logarithm of FT age. By using z_i to represent the FT grain ages τ_i, each component distribution within a mixed FTGA sample becomes approximately Gaussian-distributed. This approximation is shown to work well for FT dating of zircon grains, which generally have relatively high uranium content and high track densities, and less so for FT dating of apatite grains; (2) estimation of the PD plot is optimized by setting the width of the Gaussian-kernel equal to $\alpha(\mathrm{SE}(\hat{z}_i))$, where α is a scaling factor with an optimal value of ~ 0.6 and $\mathrm{SE}(\hat{z}_i)$ is the standard error of the FT estimate for z for the ith grain. This arrangement ensures the best compromise between resolution and smoothness for the final PD plot; (3) probability density is estimated as a function of z but it is best presented as a function of FT age τ. This objective is accomplished by transforming the z coordinates of the PD plot to τ and plotting τ on a logarithmically-scaled axis, which ensures that the original scaling of the PD plot is preserved. With these modifications, the component distributions in a PD plot will appear as symmetric Gaussian-shaped peaks and the area beneath each peak will be proportional to the relative size of the component. Several examples are given that illustrate the general concepts behind the PD plot and the advantages of the recommended modifications. Copyright ● 1996 Elsevier Science Ltd

1. INTRODUCTION

The external-detector method of fission-track (FT) dating provides estimates of FT age for each mineral grain dated in a rock sample (pp. 75–76 in Wagner and Van Den Haute, 1992). The result is a fission-track grain-age (FTGA) distribution which can be used to check for "contaminant" grains in a volcanic tuff or to interpret depositional age and provenance for an unreset sandstone (e.g. Hurford *et al.*, 1984; Cerveny *et al.*, 1988; Brandon, 1992; Brandon and Vance, 1992; Garver and Brandon, 1994a,b). In this context, unreset means that the dated detrital minerals, usually zircon or apatite, retain FT ages related to cooling or magmatic events that occurred in the source region from which the grains were eroded.

In the statistical literature, distributions that contain more than one component distribution are called *mixed distributions*. As used here, a *component distribution* in a mixed FTGA distribution represents an idealized group of mineral grains that have identical annealing and etching properties and that started to acquire their FT grain ages at the same

instant of time. In this sense, the components of a mixed FTGA distribution are the fundamental elements of that distribution. In practice, our resolution of the component distributions, especially those that make up a small fraction of the total distribution, is obscured by sampling errors, related mainly to the stochastic nature of radioactive decay and the finite size of the sample.

One approach to the analysis of a mixed FTGA sample is to decompose the sample into a set of unique component distributions. Several statistical procedures have been proposed for this task. Seward and Rhoades (1986) present a method based on cluster analysis. The binomial peak-fit method of Galbraith (1988) uses a maximum-likelihood procedure to find a best-fit solution assuming binomially-distributed components (also see Galbraith and Green, 1990 and Galbraith and Laslett, 1993). The Gaussian peak-fit method of Brandon (1992) uses a non-linear least-squares procedure to find a best-fit solution assuming Gaussian-distributed components.

Another approach is the probability density (PD) plot of Hurford *et al.* (1984). The PD plot is no substitute for the peak-fitting methods outlined

above, but it does provide a useful statistically-based tool for presentation and interpretation of FTGA data. Hurford *et al.* (1984) suggest that the probability density distribution of a FTGA sample is approximated by

$$\hat{f}(\tau) = \frac{1}{n} \sum_{i=1}^{n} G[\tau; \tau_i, SE(\tau_i)^2] \qquad (1)$$

where τ_i and $SE(\tau_i)$ are the FT age and standard error for the ith grain with $i = 1$ to n. The symbol $\hat{}$ used here and throughout the text denotes an estimator of a distribution parameter. The function $G[\cdot]$ is the Gaussian probability density function (PDF)

$$G[x; \mu, \sigma^2] = \frac{e^{-0.5((x - \mu)/\sigma)^2}}{\sigma\sqrt{2\pi}} \qquad (2)$$

where x is the function variable and μ and σ^2 are its parameters (note that the semicolon is used to separate variables from parameters). Hurford *et al.* (1984) provide no justification for equation (1). But it turns out that their procedure is similar to a well-established method of density estimation called the Gaussian-kernel method. Silverman (1986, pp. 13–19, 35–74) presents a detailed analysis of the theoretical basis and practical application of this method, and other density-estimation methods as well.

My objective here is to examine the statistical basis, construction and interpretation of the PD plot, as it relates to graphical analysis of mixed FTGA samples. The paper is divided into four parts. Section 2 provides a general review of density estimation using the Gaussian-kernel method. Section 3 reviews an "approximate" Gaussian method and an "exact" binomial method for representing FTGA sample distributions. The advantage of the Gaussian method is that it produces a transformed sample distribution where each component is approximately Gaussian distributed. This feature is useful for the successful implementation of the Gaussian-kernel method. The binomial method is used for comparison to determine when the Gaussian method is sufficiently accurate for general use. Section 4 summarizes the procedure for constructing a PD plot. Section 5 presents some typical PD plots to show how they are interpreted and how they compare with the numerical results of the Gaussian peak-fit method.*

2. KERNEL METHODS FOR DENSITY ESTIMATION

A *PD plot* is an estimate of the probability density of the FTGA population from which the dated grains were drawn. A *FTGA sample* refers to a group of stratigraphically-related FT grain ages. A more specific definition depends on the problem to be addressed. For instance, a FTGA sample will usually come from one or more rocks collected at a single locality, but, in some cases, a sample might represent a group of grain ages collected at many localities from a specific stratigraphic horizon. To construct a PD plot, we need a method of estimating probability density using only the observed FT data and avoiding assumptions about the number of component populations present in the sample. The kernel-function method (Silverman, 1986, pp. 13–19, 34–74) provides a general solution for this problem.

2.1. *Kernel functions*

To introduce the kernel method, let us define an arbitrary continuous random variable X with a population PDF given by $f(x)$. Our observations of X are based on samples x_i where $i = 1$ to n. An estimate of the probability density at x is given by (Silverman, 1986, p. 15)

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} K[(x_i - x); h], \qquad (3)$$

where $K[\cdot]$ is the kernel function (as yet unspecified) and h is a parameter that indicates the "width" of the kernel function. A relatively simple kernel function is the box function,

$$K[(x_i - x); h] = 1/h \quad \text{if } |x_i - x| \leq 0.5h \qquad (4)$$
$$K[(x_i - x); h] = 0 \quad \text{if } |x_i - x| > 0.5h,$$

where h is equal to the width of the box. This version of the kernel method is called the *naive estimator* (Silverman, 1986, pp. 11–13) and is closely related to the conventional histogram, as illustrated in Fig. 1(a). A conventional histogram is calculated by evaluating equation (4) at even increments of h along the x axis, whereas the naive estimator produces a *continuous histogram* because the histogram bar is allowed to slide continuously along the x axis.

The box function is one of many possible kernel functions (Silverman, 1986, pp. 13–19, 34–43). We focus on the Gaussian PDF which is a useful kernel function when a sample is composed of one or more Gaussian-distributed components. The Gaussian-kernel function is defined by

$$K[(x_i - x); h] = G[(x_i - x); 0, h^2], \qquad (5)$$

where the width of the kernel h is equal to the "standard deviation" of the kernel function. When written out in full, the density estimated by the

---

Gaussian-kernel method is given by

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} G[(x_i - x);0,h^2] . \qquad (6)$$

The standard error for $\hat{f}(x)$ can be estimated in the usual manner (Whittle, 1958; Silverman, 1986, p. 36).

$$\text{SE}(\hat{f}(x))^2 = \frac{1}{n(n-1)} \sum_{i=1}^{n} G[(x_i - x;0,h^2) - \hat{f}(x)]^2, \qquad (7)$$

because each x_i is an independent observation.

Figure 1 shows an application of the Gaussian-kernel method as specified by equations (6) and (7). In this case, $h = \alpha\sigma$ where α is a scaling factor and σ is the standard deviation of the sample distribution (this procedure is justified below).

2.2. *Properties of the Gaussian-kernel function*

To understand the advantages of the Gaussian PDF as a kernel function, we need to examine the expected value of the estimator $\hat{f}(x)$ defined by equations (3) and (6). Silverman (1986, pp. 36–37)
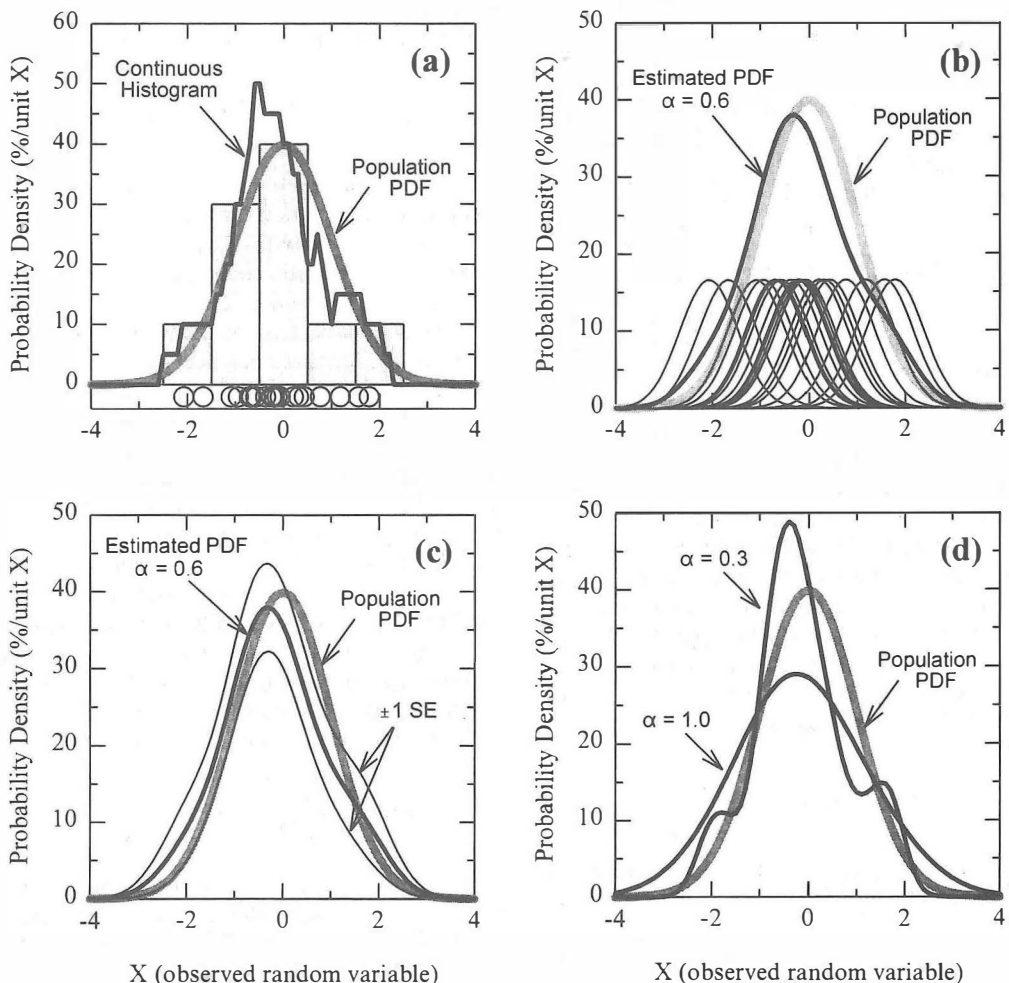

Fig. 1. Examples of density estimation for a synthetic sample generate by 20 random draws from a Gaussian PDF with $\mu = 0$ and $\sigma = 1$. Part (a) shows the sample distribution (circles at the bottom of the plot) and two estimated histograms. The first is a conventional histogram and the second is a continuous histogram which is estimated using a box-shaped kernel function [naive estimator, see equations (3) and (4)]. Note that the continuous histogram intersects the center of each bar in the conventional histogram because the two plots are identical at those points. Part (b) shows the estimated PDF as determined by the Gaussian kernel method [equation (6)] using an optimal value of $h = 0.6$ ($h_{opt} = \alpha\sigma = 0.6$ given an expected standard deviation $\sigma = 1$ and an optimal scaling factor of $\alpha = 0.6$; see Section 2.3 for details). One way to describe the Gaussian-kernel method is to envision that the kernel function converts each observation into a Gaussian, as illustrated in (b). Density is estimated by averaging the densities indicated by the "observed Gaussians" at each point along the x axis. (Note that the "observed Gaussians" shown here are displayed at 25% of their true amplitude.) Part (c) shows the ± 1 standard error (SE) envelope for the estimated density as determined by equation (7). Part (d) demonstrates the result of increasing and decreasing α from its optimal value.

shows that

$$E(\hat{f}(x)) = \int_{y=-\infty}^{y=+\infty} K[(x-y);h]f(y)\,dy$$
$$\equiv K[x;h]*f(x)\,, \quad (8)$$

where $E(\cdot)$ indicates the expected value. This equation shows that $E(\hat{f}(x))$ is equal to the convolution of the kernel function $K[x;h]$ with the population density function $f(x)$. The second expression in equation (8) is the convolution integral and the third expression shows an equivalent form of this integral where "*" is the convolution operator. Those readers not familiar with the mathematical operation of convolution will find a good introduction in Brigham (1988, pp. 50–73).

As a rule, the convolution of two arbitrary Gaussian functions results in a third Gaussian function (e.g. Silverman, 1986, p. 37). We start with a simple example: a population PDF represented by a single Gaussian $f(x) = G[x;\mu,\sigma^2]$. Convolution of the population PDF with the Gaussian-kernel function (5) gives

$$E(\hat{f}(x)) = G[x;0,h^2]*G[x;\mu,\sigma^2] = G[x;\mu,\sigma^2+h^2]\,. \quad (9)$$

This result demonstrates that if $f(x)$ is Gaussian-distributed around a mean μ then $E(\hat{f}(x))$ will also be Gaussian-distributed around the same mean μ. In fact, the only change introduced by the convolution is that $E(\hat{f}(x))$ has a larger variance than $f(x)$: $\sigma^2 + h^2$ vs σ^2, respectively. This conclusion is illustrated in Fig. 1(b) by the slightly larger width of the estimated PDF relative to the population PDF.

Now consider a mixed distribution composed of $j = 1$ to m Gaussian component distributions with the jth component represented by the random variable X_j with a distribution $f_j(x) = \pi_j G(x;\mu_j,\sigma_j^2)$. The parameters u_j, σ_j and π_j are the mean, standard deviation and fractional size of X_j where $\Sigma_{j=1}^{m}\pi_j = 1$. The population PDF is now given by

$$f(x) = \sum_{j=1}^{m}\pi_j G[x;\mu_j,\sigma_j^2]\,. \quad (10)$$

For this case, the expected density estimated by the Gaussian-kernel method is

$$E(\hat{f}(x)) = G[x;0,h^2]*\sum_{j=1}^{m}\pi_j G[x;\mu_j,\sigma_j^2] \quad (11)$$

$$= \sum_{j=1}^{m}\left(G[x;0,h^2]*\pi_j G[x;\mu_j,\sigma_j^2]\right) = \sum_{j=1}^{m}\pi_j G[x;\mu_j,\sigma_j^2+h^2]\,.$$

This result represents a natural extension of equation (9). We now find that if X_j is a mixture of one or more Gaussian component distributions, then $E(\hat{f}(x))$ will contain the same components. Each component will remain Gaussian distributed around the same mean but its variance will be increased by h^2.

Grain ages determined by FT dating, and other dating methods as well, generally follow a more complex type of mixed distribution. To illustrate, consider a set of Gaussian-distributed random variables, designed as X_{jk}, with mean values $\mu_{jk} = \mu_j$ and standard deviations σ_{jk}. To understand the interrelation of the random variables X_{jk}, let us focus on a subset of variables designated by $j = 1$. All random variables within X_{1k} will have the same mean μ_1 but different standard deviations σ_{1k}. For the FTGA problem, X_{1k} is equivalent to a single-component distribution because all grains within the distribution have a common mean, but variations in uranium concentration and numbers of counted tracks cause the standard deviation to vary from grain to grain.

For this case, the population PDF is

$$f(x) = \sum_{j=1}^{m}\sum_{k=1}^{n_j}\pi_{jk}G[x;\mu_j,\sigma_{jk}^2]\,, \quad (12)$$

where n_j is the number of Gaussians in the jth component distribution and π_{jk} is the fractional size of each Gaussian with $\Sigma_{j=1}^{m}\Sigma_{k=1}^{n_j}\pi_{jk} = 1$. The fractional size of the jth component is now given by $\pi_j = \Sigma_{k=1}^{n_j}\pi_{jk}$.

Most isotopic-dating methods provide estimates for σ_{jk} in the form of analytical uncertainties. Using our general notation, an estimated grain age is $\hat{x}_j$ with the uncertainty given by $SE(\hat{x}_j)$, the standard error of $\hat{x}_j$. For a mixed distribution, $SE(\hat{x}_j)$ is an estimate of the standard deviation of the jk Gaussian from which $\hat{x}_j$ was drawn. We see below that this relationship is useful despite the fact that we have no specific knowledge about which Gaussian was sampled.

Now we consider the expected density estimated by the Gaussian-kernel method for the population PDF given by equation (12). Because X_{jk} is a mixed Gaussian population, equation (11) remains applicable, although it must be modified to conform to the notation used in equation (12).

$$E(\hat{f}(x)) = \sum_{j=1}^{m}\sum_{k=1}^{n_j}\pi_{jk}G[x;\mu_j,\sigma_{jk}^2+h^2]\,. \quad (13)$$

This result shows that our general conclusion still holds: $E(\hat{f}(x))$ contains the same set of Gaussians as present in the population PDF, but the variance of each of the Gaussians is increased by h^2.

At this point, it is important to distinguish between the different standard deviations we are using: h, σ_{jk}, σ_j, s_j and W_j. For the Gaussian-kernel method, h corresponds to the standard deviation of the kernel. Note that h is only used to control the width of the kernel and should not be viewed as a statistical parameter. Next are σ_{jk}, which represent the standard deviations of all Gaussians in the population distribution, and σ_j, which represent the standard deviations of the component distributions. Remember that each component distribution contains a mixture of Gaussians with the same mean but

different standard deviations. Thus, by definition,

$$\sigma_j^2 = \frac{1}{\pi_j} \sum_{k=1}^{n_j} \pi_{jk} \sigma_{jk}^2 \,. \tag{14}$$

It is interesting to note that if we could randomly sample the jth component population, then the standard errors for those observations $SE(\hat{x}_j)$ could be used to estimate σ_j,

$$s_j^2 = \frac{1}{n} \sum_{i=1}^{n} SE(\hat{x}_j)^2 \,, \tag{15}$$

where $E(s_j) = \sigma_j$.

The standard deviations of the component distributions in $E(\hat{f}(x))$ are represented by W_j in order to distinguish them from the population standard deviations σ_j. Stated in a more casual fashion, W_j is a measure of the width of the jth peak as revealed in the PD plot.

Equations (13) – (15) show that

$$W_j^2 = \sigma_j^2 + h^2 \tag{16}$$

and

$$\hat{W}_j^2 = s_j^2 + h^2 \,. \tag{17}$$

2.3. Optimal value for h

A critical issue for the application of the Gaussian-kernel method is the selection of an optimal value for h, designated as h_{opt} (Silverman, 1986, pp. 40–48). This problem involves a trade-off between the increased resolution provided by a small h and the increased precision provided by a large h. As used here, resolution refers to our ability to distinguish closely spaced component distributions, and precision refers to the uncertainty of the estimated densities that make up the PD plot [e.g. equation (7)]. Let us consider the case of a sample of size n drawn from a single Gaussian distribution with a known standard deviation σ. For this case, Silverman (1986, p. 45) shows that the approximate mean integrated square error for the density estimate is minimized when

$$h_{opt} = \alpha\sigma = \left(\frac{4}{3n}\right)^{0.2} \sigma \approx 1.06n^{-0.2}\sigma \,. \tag{18}$$

Note that the variable α is introduced here as a scaling factor. Equation (18) indicates that h_{opt} is linearly proportional to σ and relatively weakly dependent on n (Fig. 2). Figure 1(c) shows $\hat{f}(x)$ as estimated using $\alpha = 0.6$, which is the optimal value indicated by equation (18) for $n = 20$. Figure 1(d) shows the effect of increasing and decreasing α from its optimal value.

Equation (18) can be used for mixed distributions as long as we have *a priori* knowledge of the standard deviations of the Gaussians that make up the mixed distribution. As noted above, $SE(\hat{x}_j)$ is an estimate of the standard deviation of the distribution from which
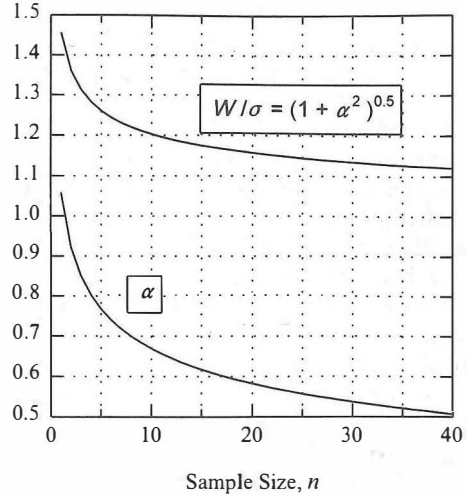

Fig. 2. The lower curve shows equation (18) which gives the optimal value of α as a function of sample size n. The upper curve shows $W/\sigma = \sqrt{1 + \alpha^2}$ which describes the increase in variance introduced by the Gaussian-kernel method [see equations (16), (17) and (21) for details].

$\hat{x}_j$ was drawn. Thus, we can change equation (6) to define a modified version of the Gaussian-kernel method,

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} G[\hat{x}_j - x; 0, \alpha^2 SE(\hat{x}_j)^2] \,. \tag{19}$$

Note that when $\alpha = 1$, equation (19) becomes identical to the original PD plot equation (1) proposed by Hurford et al. (1984). The estimator for the standard error is now

$$SE(\hat{f}(x))^2 = \frac{1}{n(n-1)}$$

$$\times \sum_{i=1}^{n} \left[G(x_j - x; 0, \alpha^2 SE(x_j)^2) - \hat{f}(x) \right]^2 \,. \tag{20}$$

equation (17) becomes

$$\hat{W}_j = \sqrt{1 + \alpha^2} s_j \,. \tag{21}$$

Inspection of Fig. 2 shows that as n increases from 1 to 40, $\sqrt{1 + \alpha^2}$ decreases from 1.46 to 1.17. In other words, as the size of a component distribution increases, we are permitted to use a smaller α which results in a smaller W_j and better resolution of the components in the PD plot.

When using equation (19) to construct a PD plot, we are forced to use a single value for α, and yet a mixed distribution typically contains an unknown number of components, with each component possessing a different and unknown size n_j. Ideally, we would want to match a different α for each component but there is no practical way to do this. I have found the following compromise provides a workable solution to this problem. In practice, the size of a typical component in a mixed distribution is

usually $n_j = 1 - 40$. Thus, I have adopted the convention of setting $\alpha = 0.6$ which is its average value over this range in n_j (Fig. 2). With regards to this compromise, three points are worth mentioning: (1) when the prescribed α value is poorly matched to the size of a specific component in a mixed distribution, the distortions introduced in the PD plot are limited to the representation of that component alone; (2) a component with $n_j < 20$ should use a larger α, but this requirement is offset by the fact that the sample standard deviation decreases as n_j decreases from $< \sim 20$ (i.e. small-sample bias). Simulations using the Student t distribution indicate that the need for a larger α when $n_j < 20$ is approximately compensated by the downward bias in the sample standard deviation; (3) the width of a peak or component distribution in a PD plot, as represented by W_j, is not strongly dependent on α when $\alpha < 0.7$. Thus, there is little advantage in fine-tuning α around its prescribed value of 0.6.

3. FT STATISTICS FOR A SINGLE-COMPONENT SAMPLE

Before using the Gaussian-kernel method, we need to consider if a real FTGA distribution is adequately approximated by a mixed Gaussian distribution, as defined by equation (12). More specifically, we are concerned with the expected form of the component distribution.

Let us start with a brief review of FT statistics. FT age τ is calculated using the deterministic formula (e.g. Hurford and Green, 1983; Galbraith, 1990).

$$\tau = \lambda^{-1} \ln(1 + \lambda \zeta g \rho_d \rho_s / \rho_i) , \qquad (22)$$

where λ is the total decay constant for ^{238}U $(1.55125 \times 10^{-10} \text{ yr}^{-1})$, ζ is the zeta-calibration constant, g is the geometry factor (0.5 for the external-detector method), ρ_d is the track density for the fluence monitor and ρ_s / ρ_i is the ratio of measured densities for spontaneous and induced tracks, respectively. For the external-detector method, ρ_s and ρ_i are estimated by counting the number of spontaneous and induced tracks, N_s and N_i, present within the same area A of a polished grain surface. N_s and N_i are inferred to have independent Poisson distributions (Galbraith and Laslett, 1985) with distribution parameters defined by $\rho_s = E(N_s/A)$ and $\rho_i = E(N_i/A)$.

For our purposes, all parameters on the right-hand side of equation (22) are known except for ρ_s and ρ_i. The parameters ζ and ρ_d are treated as constants because they have fixed values for all grains in a FTGA sample, assuming that grains from the same sample are irradiated at the same time and under identical conditions. Note that this assumption is not essential because the errors in ζ and ρ_d are usually much smaller than those associated with ρ_s and ρ_i and thus can be safely ignored.

Two methods are available for representing and estimating the distribution of a single-component FTGA sample. The *Gaussian method* is based on the fact that the logarithm of the ratio of two independent Poisson-distributed random variables is approximately Gaussian distributed, with the approximation improving as N_s and N_i increase. The general theory behind the logistic transform, as it is called, is reviewed by Cox (1970, pp. 30–35) and McCullagh and Nelder (1989, pp. 106–107). The implementation of the transform to FT dating (Bardsley, 1983; Galbraith, 1990) is accomplished by recasting the FT age equation

$$\tau = \lambda^{-1} \ln(1 + e^z) , \qquad (23)$$

as a function of a new parameter

$$z = \ln(\lambda \zeta g \rho_d) + \ln\left(\frac{\rho_s}{\rho_i}\right) . \qquad (24)$$

For a single-component sample, estimates of z should be approximately Gaussian distributed. The approximately unbiased estimator of z is given by

$$\hat{z} = \ln(\lambda \zeta g \rho_d) + \ln\left(\frac{N_s + 0.5}{N_i + 0.5}\right) \qquad (25)$$

with standard error

$$SE(\hat{z}) = \sqrt{\frac{1}{N_s + 0.5} + \frac{1}{N_i + 0.5}} . \qquad (26)$$

The "$+ 0.5$" terms in equation (25) and equation (26) are continuity corrections which are introduced in the transformation from discrete Poisson distributions to a continuous Gaussian distribution (i.e. N_s and N_i are no longer limited to integer values; see Cox, 1970, p. 33 and McCullagh and Nelder, 1989, p. 107, for a full derivation). A useful feature of z is that $(SE(\hat{z}) \approx SE(\hat{\tau})/\hat{\tau} = RE(\hat{\tau})$ where $RE(\cdot)$ indicates the relative standard error (Galbraith, 1990; Galbraith and Laslett, 1993). Thus, $100 \, SE(\hat{z})$ is approximately equal to the percentage relative standard error of the estimated FT age.

After transformation, the FTGA sample is represented by $\hat{z}_i$ and $SE(\hat{z}_i)$ where i indicates the ith grain among a total sample of n grains. The advantage of this transformation is that each component in a mixed sample is now approximately Gaussian distributed. Furthermore, each observation z_i now has an approximately Gaussian sample distribution with the mean and standard deviation estimated by $\hat{z}_i$ and $SE(\hat{z}_i)$. Thus, the transformed distribution is ideally suited for application of the modified Gaussian-kernel method [equations (19) and (20)].

But how well does the Gaussian approximation hold? This question can be answered using the

binomial method which uses an accurate representation of the expected sample distribution. The binomial method is based on the fact that the sample distribution of the Poisson variables N_s and N_i can be transformed into a univariant binomial distribution. Chapman (1952) and McCullagh and Nelder (1989, pp. 101–102) present the general theory for this transformation. As applied to FT dating, if the sum $(N_s + N_i)$ is fixed, then N_s is a binomial-distributed variable with a distribution parameter $\Theta = \rho_s/(\rho_s + \rho_i)$ and an index equal to $(N_s + N_i)$ (Bardsley, 1983, 1984; Sneyd, 1984; Galbraith and Laslett, 1993). The estimator and associated standard error are given by $\hat{\theta} = N_s/(N_s + N_i)$ and $SE(\hat{\theta}) = \sqrt{\hat{\theta}(1 - \hat{\theta})(N_s + N_i)}$. Note that $\hat{\theta}$ is also binomially distributed. Given observations of N_s and N_i, the estimated sample distribution of $\hat{\theta}$, including its median value and confidence limits, can be calculated using the numerical algorithm of Sneyd (1984).

The approximate form of the sample distribution estimated by the Gaussian method can now be compared with the expected form estimated by the binomial method. Given observations N_s and N_i, the approximate distribution for $\hat{z}$ is a Gaussian PDF with a mean and standard deviation estimated by equations (25) and (26). The expected distribution is determined by calculating the estimated sample distribution of $\hat{\theta}$ using the algorithm of Sneyd (1984) and then transforming this result using

$$z = \ln(\lambda \zeta g \rho_d) + \ln\left(\frac{\theta}{1 - \theta}\right). \tag{27}$$

Figure 3 shows a comparison of the results of the Gaussian and binomial methods for a large range of N_s and N_i values (see figure caption for details). The general conclusion is that the errors introduced by the approximation are roughly inversely correlated with $SE(\hat{z})$. Figure 3(a) shows that for $SE(\hat{z}) < 0.30$, the value of $\hat{z}$ given by equation (25) is nearly identical with the median estimated by the binomial method. Figure 3(b) shows that the relative misfit between the approximate and expected sample distributions is less than 10% for $SE(\hat{z}) < 0.30$. My experience indicates that when estimating FT ages and confidence limits, the Gaussian approximation can be trusted to produce a reasonable result when $SE(\hat{z}) < 0.30$ and an accurate result when $SE(\hat{z}) < 0.15$.

Now consider some measured values of $SE(\hat{z}_i)$, as determined by FT dating of apatite and zircon (Fig. 4). The zircon grains have an average $SE(\hat{z}_i)$ of ~ 0.13 with a range 0.01–0.30, whereas the apatite grains have an average $SE(\hat{z}_i)$ of ~ 0.54 with a range 0.10–1.00. This difference is due to the fact that zircons generally contain more uranium than apatites, which, with all other factors equal, means that zircons will usually have a greater numbers of tracks per grain. The practical result is that the Gaussian approximation usually works well for zircons and less so for apatites.

4. CONSTRUCTING PD PLOTS FOR FTGA SAMPLES

The first step in constructing a PD plot is to use equations (25) and (26) to transform the FT grain estimates to $\hat{z}_i$ and $SE(\hat{z}_i)$ for the $i = 1 - n$ sampled grains. Probability density and standard errors are estimated using equations (19) and (20) with $\alpha = 0.6$. To determine the limits of z for the PD plot, the youngest and oldest grain ages, τ_{min} and τ_{max}, are converted to z_{min} and z_{max} using

$$z = \ln(e^{\lambda \tau} - 1). \tag{28}$$

The PD plot is then extended to $z_{min} - 3W$ and $z_{max} + 3W$ to ensure that the plot includes the tails of the full distribution. This practice is justified by the fact that the expected half width of a component
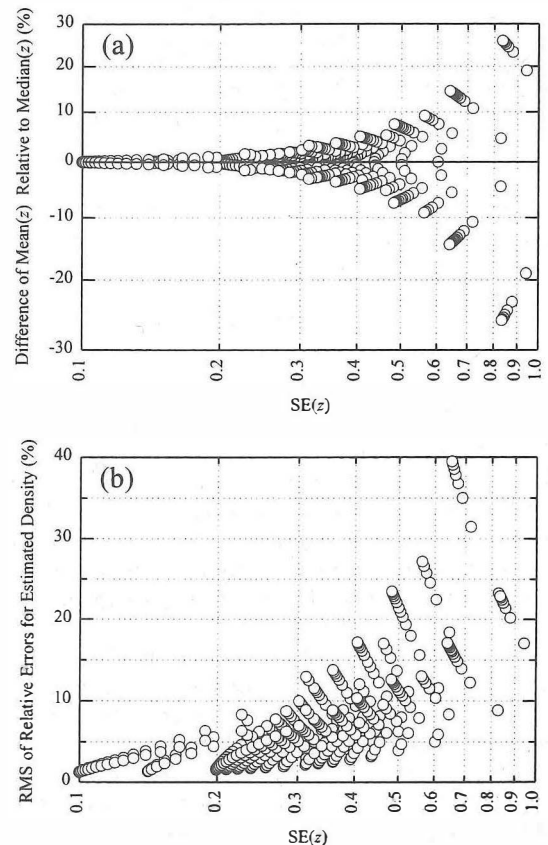

Fig. 3. Comparison of the "approximate" Gaussian method with the "exact" binomial method as a function of $SE(\hat{z})$. Results were calculated for the domains: $(N_s + N_i) = 5$ to 400 and $N_s = 0$ to $(N_s + N_i)$. For counts with $(N_s + N_i) \leqslant 35$, N_s and $(N_s + N_i)$ were incremented by 1 and 5, respectively. Larger increments were used when $(N_s + N_i) > 35$. Part (a) shows the percentage difference between the mean $\hat{z}$ approximated by the Gaussian method and the median $\hat{z}$ determined by the binomial method. Part (b) shows the root-mean-square (r.m.s.) average of the relative misfit of the probability densities calculated by the two methods. Both examples demonstrate that the Gaussian method produces reasonable results when $SE(\hat{z}) < 0.30$ and fairly precise results when $SE(\hat{z}) < 0.15$.
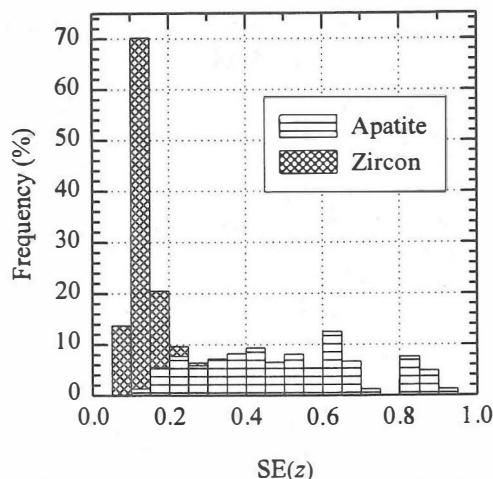
distribution in the PD plot is $\sim 3W$. W is approximated by

$$W_{ave}^2 = \frac{(1+\alpha)^2}{n} \sum_{i=1}^{n} SE(\hat{z}_i)^2 \qquad (29)$$

which is the estimated average of the standard deviations W_j of the component distributions in the PD plot [see equations (15), (17) and (21)]. The plot is calculated in even increments δz to ensure that the peaks in the distribution are equally represented. A useful value for $\delta z = (6/50)W$ which ensures that each component distribution is represented by about 50 points, given that the full width of a component is $\sim 6W$.

When constructing the actual PD plot, we need some easy way to show FT age along the z axis but at the same time, we need to preserve the relative scaling of $\hat{f}(z)$ to z so that the component distributions will retain their Gaussian form. The recommended procedure is to convert from z to τ using equation (23) and to plot τ on a logarithmic scale and $\hat{f}$ on a linear scale (Galbraith, 1990). The resulting plot will have a nearly identical form to a linear–linear plot of $\hat{f}$ vs z. The reason is that z is approximately proportional to the logarithm of τ and $f(z)$ is approximately equal to $f(\ln \tau)$. (This result is based on an approximation of equation (23) $\ln \tau = (1+\epsilon)z - \ln\lambda$, where the absolute value of the approximation error $|\epsilon| < 0.01$ where $\tau < \sim 365$ Ma.)

The current version of my Gaussian peak-fit program provides best-fit estimates of the mean age $\hat{z}_j$, standard deviation $\hat{W}_j$, and number of grains $\hat{n}_j$ for

each of the $j = 1$ to m component distributions. The estimated PDF for the jth component can be shown in the PD plot using

$$\hat{f}_j(z) = \frac{\hat{n}_j}{n} G(z;\hat{z}_j,\hat{W}_j^2), \qquad (30)$$

where n is the total number of dated grains in the sample.

The binomial peak-fit method of Galbraith (1988) provides best-fit estimates of the binomial parameters
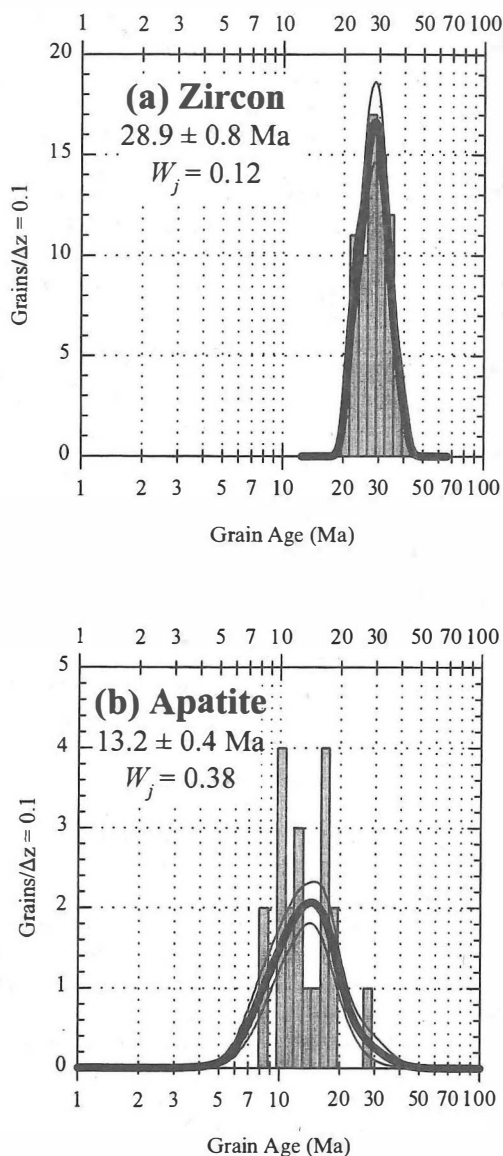

Fig. 5. Logarithmic PD plots for single-component FTGA samples. Density (thick line) and the ± 1 SE envelope (thin lines) were estimated using equations (19) and (20). (a) Volcanic zircons ($n = 72$) from unreset Fish Canyon tuff of Colorado (sample Fl of Brandon and Vance, 1992); (b) detrital apatites ($n = 20$) from a fully reset sandstone of the Olympic subduction complex (sample AR25 of Brandon *et al.*, 1996).

Table 1. Results from the Gaussian peak-fit program

Lab#	Description, number of grains	Age range, W_{ave} from equation (29)	Mean age $\pm$ 1 SE, width and size of best-fit peaks			
Zircon FTGA samples						
Fl	Fish Canyon tuff, CO	21–40 Ma	28.9 $\pm$ 0.8 Ma	—	—	—
	$n = 72$	$W_{ave} = 0.12$	$W_j = 0.12$			
			$n_j = 57.8$			
ZD6	OSC sandstone, WA	14–193 Ma	18.8 $\pm$ 0.4 Ma	43.1 $\pm$ 1.3 Ma	66.8 $\pm$ 2.6 Ma	184 $\pm$ 5.1 Ma
	$n = 50$	$W_{ave} = 0.19$	$W_j = 0.19$	$W_j = 0.16$	$W_j = 0.19$	$W_j = 0.11$
			$n_j = 12.5$	$n_j = 17.0$	$n_j = 16.5$	$n_j = 1.5$
C-1	Siwalik sandstone, Pakistan	22–113 Ma	23.8 $\pm$ 0.7 Ma	36.8 $\pm$ 1.2 Ma	60.8 $\pm$ 1.9 Ma	114 $\pm$ 6.3 Ma
	$n = 80$	$W_{ave} = 0.22$	$W_j = 0.12$	$W_j = 0.20$	$W_j = 0.25$	$W_j = 0.15$
			$n_j = 5.2$	$n_j = 30.9$	$n_j = 41.3$	$n_j = 1.8$
Apatite FTGA samples						
AR25	OSC sandstone, WA	8–27 Ma	13.2 $\pm$ 0.4 Ma	—	—	—
	$n = 20$	$W_{ave} = 0.50$	$W_j = 0.38$			
			$n_j = 19.7$			
AR12	OSC sandstone, WA	5–117 Ma	7.8 $\pm$ 1.3 Ma	34.9 $\pm$ 2.1 Ma	—	—
	$n = 10$	$W_{ave} = 0.93$	$W_j = 0.86$	$W_j = 0.55$		
			$n_j = 2.3$	$n_j = 6.3$		

Notes: n = total number of grains analyzed. W_{ave} was determined using equation (29). For the peak-fit calculation, μ_j, W_j and n_j were all fit as independent free parameters: μ_j is the mean age, W_j is the peak width in z unit and n_j is the number of grains, where j indicates the jth component distribution or peak. OSC = Olympic subduction complex.

for each of the component distributions:

$$\hat{\theta}_j = \left(\frac{N_s}{N_s + N_i} \right)_j, \quad \hat{N}_j = (N_s + N_i)_j \text{ and } \hat{n}_j.$$

Estimated PDFs for these component distributions can be calculated using the algorithm of Sneyd (1984). Alternatively, if the Gaussian approximation is warranted, then the logistic transform can be used, where $\hat{\theta}_j$ is converted to $\hat{z}_j$ using equation (27) and $\hat{W}_j$ is estimated by

$$\hat{W}_j^2 = \frac{1 + \alpha^2}{\theta_j(1 - \theta_j)\hat{N}_j} . \tag{31}$$

Equation (30) is then used to calculate the estimated component PDFs.

I find it useful to superimpose a conventional histogram on the PD plot (e.g. Fig. 5). This result can be accomplished by constructing the histogram on the z scale with the bar width set at a constant Δz. Note that when plotted on the τ scale, the bars of the histogram will show a logarithmic increase in width with increasing age. For instance, the value used here, $\Delta z = 0.1$, is equivalent to a bar width on the τ scale of ~ 0.5 m.y. at $\tau = 5$ Ma and ~ 5 m.y. at $\tau = 50$ Ma.

Up to this point, probability density $\hat{f}(z)$ has been defined in units of fractional probability mass per unit z. An alternative is to express density in the same units used for the histogram which is the number of observations per $\Delta z = 0.10$. This transformation is accomplished by multiplying $\hat{f}(z)$ by the factor $n\Delta z$.

5. EXAMPLES

Five FTGA samples are presented here to illustrate the construction and interpretation of the modified PD plot. Each sample was decomposed using the Gaussian peak-fit method to illustrate the relationship of the PD plot to its estimated peak-fit parameters. Note that $\hat{z}_j$, $\hat{W}_j$ and $\hat{n}_j$ were estimated as independent free parameters (Table 1) and that the maximum number of resolvable component distributions was determined using the F-ratio test (Brandon, 1992).

5.1. *Single-component samples*

Two typical single-component FTGA samples are shown in Fig. 5. The first consists of unreset volcanic zircons ($n = 72$) from the Fish Canyon tuff [Fig. 5(a)] (sample Fl in Brandon and Vance, 1992), and the second, fully reset detrital apatites ($n = 20$) from the Olympic subduction complex of western Washington State [Fig. 5(b)] (sample AR35 in Brandon *et al.*, 1996). Note that the zircon peak is about one-third the width of the apatite peak ($W_j = 0.12$ vs 0.38). Also note that there is fair agreement between W_{ave} determined by equation (29) and $\hat{W}_j$ given by the Gaussian peak-fit method (Table 1).

5.2. *Mixed samples*

Now we turn to mixed FTGA samples. The first and second examples are for detrital zircons from unreset Miocene sandstones of the Olympic subduction complex (Fig. 6; ZD6 in Table 1; Brandon and Vance, 1992) and the Siwalik group from the Himalayan foredeep in Pakistan (Fig. 7; C-1 in Table 1; Cerveny *et al.*, 1988). The third example is

a mixed sample of reset detrital apatites from the Olympic subduction complex (Fig. 7; AR12 in Table 1; Brandon *et al.*, 1996). The bimodal form of this distribution is attributed to variations in annealing and etching properties among the detrital apatites.

Each example is illustrated by three different types of PD plots. Part (a) of each figure shows the original version of the PD plot as calculated from

equation (1). Parts (b), (c) and (d) show the modified PD plot as calculated from equation (19) and equation (20) with $\alpha = 0.6$. Part (b) uses a linear scale for FT age, whereas parts (c) and (d) use a logarithmic scale, as recommended above. Parts (b) and (d) show the best-fit peaks from Table 1 as calculated by equation (30).

The different versions of the PD plot give different impressions of the relative amplitudes and sizes of the

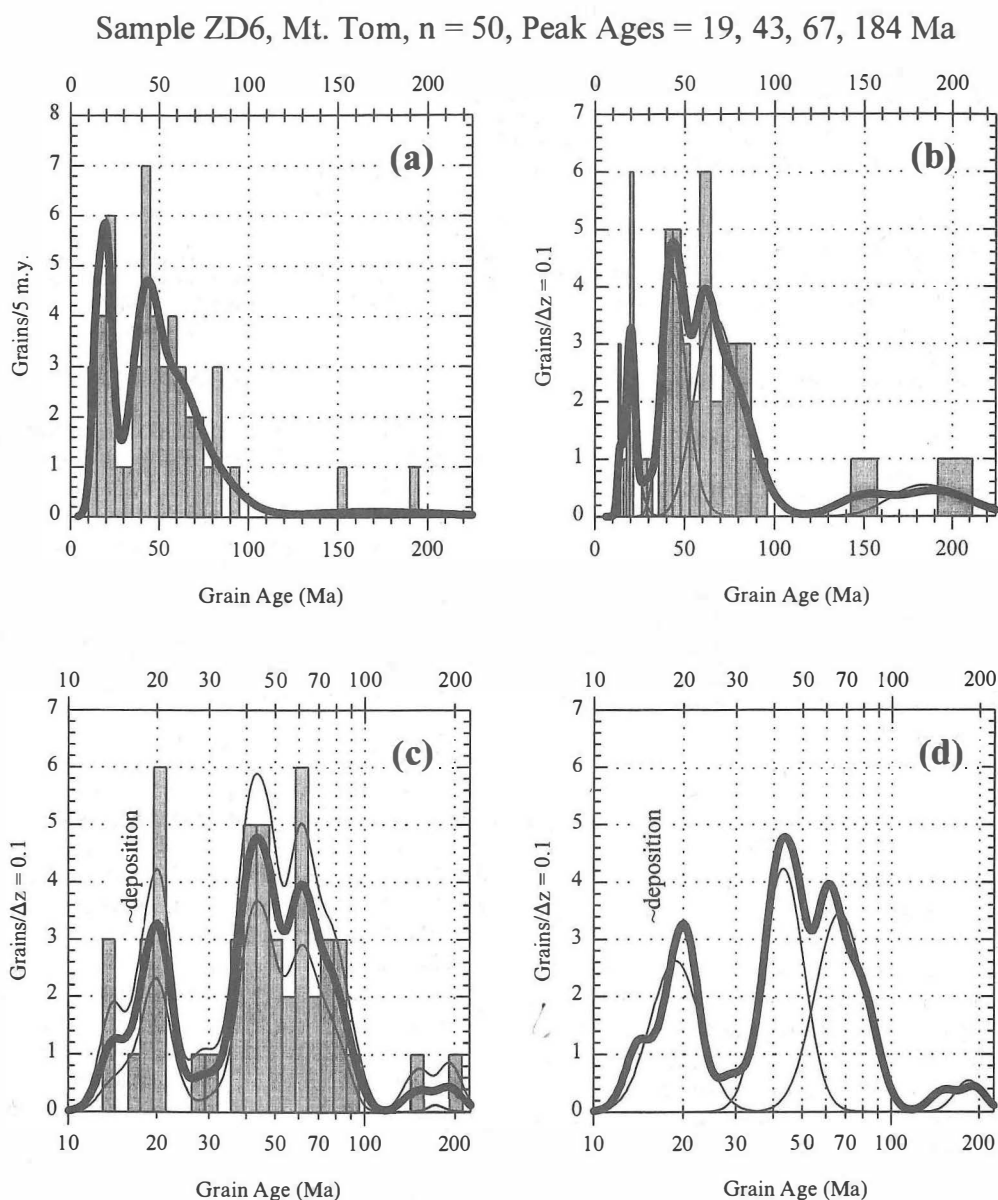## Sample ZD6, Mt. Tom, n = 50, Peak Ages = 19, 43, 67, 184 Ma

Fig. 6. PD plots for a mixed FTGA sample of detrital zircons ($n = 50$) from an unreset sandstone of the Olympic subduction complex (sample ZD6 of Brandon and Vance, 1992). Three versions of PD plots are shown: (a) the original version as proposed by Hurford *et al.* (1984) [see equation (1)], (b) the modified version [equation (19)] with age plotted on a linear scale and (c, d) the modified version [equation (19)] with age plotted on a logarithmic scale, as recommended here. The thin lines in (c) show the ± 1 SE envelope for the estimated density [equation (20)]. The best-fit peaks shown in (b, c) were calculated using equation (30) and the parameters reported in Table 1. Brandon and Vance (1992) conclude that the depositional age of this sample was younger than the youngest peak (< 19 Ma).
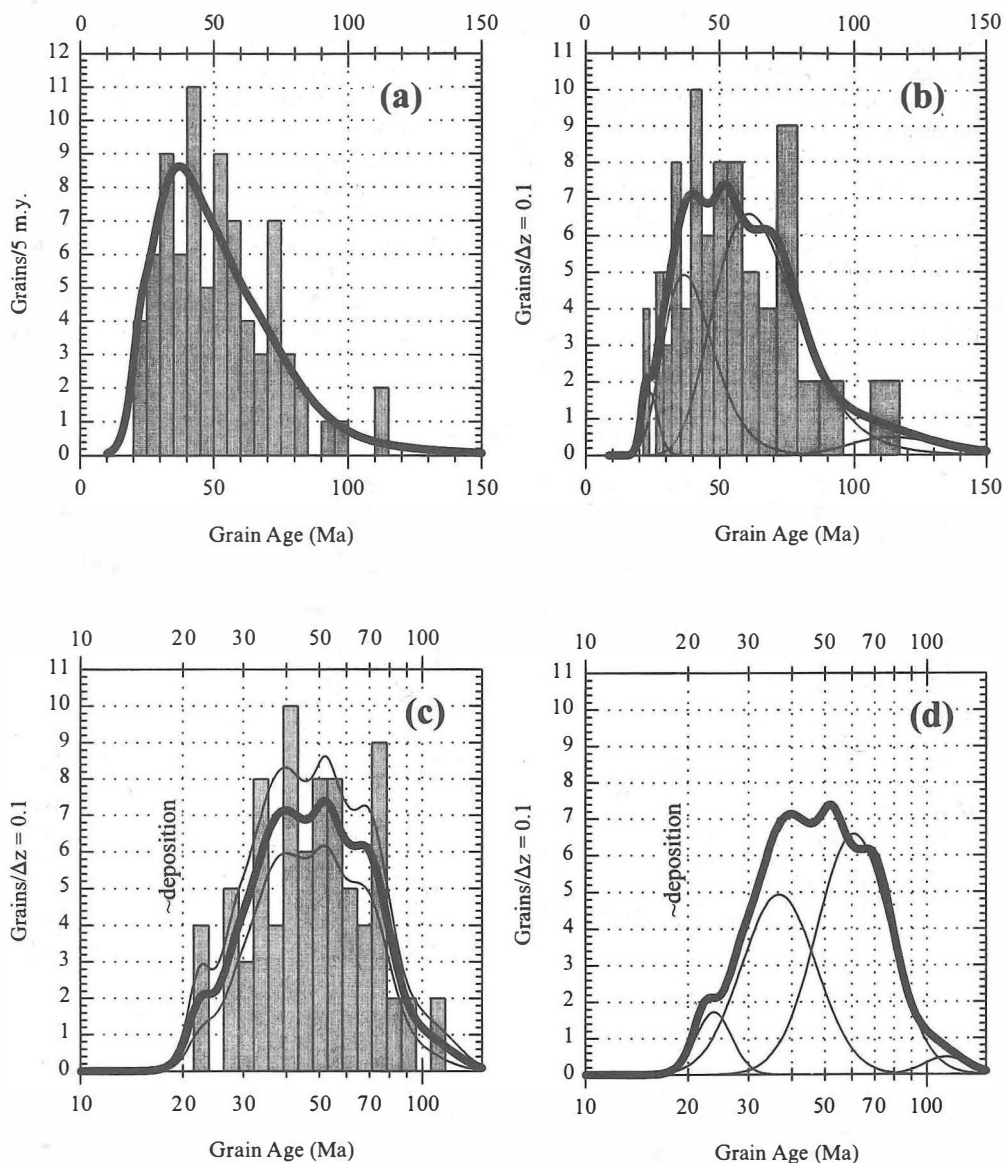
Sample C-1, Siwalik Grp., n = 80, Peak Ages = 24, 37, 61, 114 Ma

Fig. 7. PD plot for a mixed FTGA sample of detrital zircons (*n* = 80) from an unreset sandstone of the Siwalik group at Chinji, Pakistan (sample C-1 of Cerveny, 1986; Cerveny *et al.*, 1988). Cerveny *et al.* (1988) report an 18 Ma stratigraphic age for this sample. See Fig. 6 for an explanation of the plots.

component distributions, a point that was first recognized by Galbraith (1990). The original PD plot of Hurford *et al.*, shown as part (a), does a poor job representing the relative sizes of the component distributions. For instance, Fig. 6(a) gives the impression that the youngest peak (19 Ma) for sample ZD6 is the largest in the distribution, but values for n_j in Table 1 shows that the youngest peak is significantly smaller than the peaks located at 43 and 67 Ma. The relative sizes of the component distributions seem to be better represented in part (b) but note that the peaks vary considerably in width and are also slightly skewed. The logarithmic version

of the modified PD plot (c, d) provides the best representation given that component distributions appear as symmetric well defined peaks with the relative area of each peak proportional to the size of the underlying component distribution.

Also note that the histogram bars in (b) become progressively wider with increasing FT age, but those in (c) show a constant width independent of FT age. The reason is that the width of the histogram bars is fixed at $\Delta z = 0.10$ so that the constant width of the bars is only apparent when τ is plotted on a logarithmic scale (as discussed above). This result also indicates that those peaks that have a similar

standard deviation W_j will appear to have a similar width when viewed in the logarithmic version of the PD plot. In this context, note that the peaks for the zircon samples show a more uniform width in logarithmic version of the PD plot. The reason is that for zircon FTGA samples, W_j does not seem to vary much from its average value W_{ave} (see Table 1). This result does not hold for the apatite sample (Fig. 8) because $W_{ave} = 0.93$ is quite large (AR12 in Table 1) indicating that the Gaussian approximation is not applicable. [Note that if $\alpha = 0.6$, then $W_{ave} \approx 1.17 \, SE(\hat{z})$ according to equation (21). Thus, the limiting value for the Gaussian approximation $SE(\hat{z}) < 0.30$ is approximately equal to $W_{ave} \sim 0.35$.]

5.3. Radial plots

The radial plot of Galbraith (1990) provides an alternative method for displaying mixed FTGA distributions. Figure 9 shows a radial plot for the zircon FTGA sample presented in Fig. 6. An important advantage of the radial plot is that it shows the estimated age and precision for each grain. Precision is equal to $1/SE(\hat{z})$ and increases to the right along the horizontal scale. The vertical scale, labeled y on the left side of the plot, is a standardized version of z with $y = (\hat{z}_i - z_{mean})/SE(\hat{z}_i)$ where z_{mean} is the average z for the total distribution. The advantage of this scaling is that the $\pm \, SE(\hat{z}_i)$ for a grain age

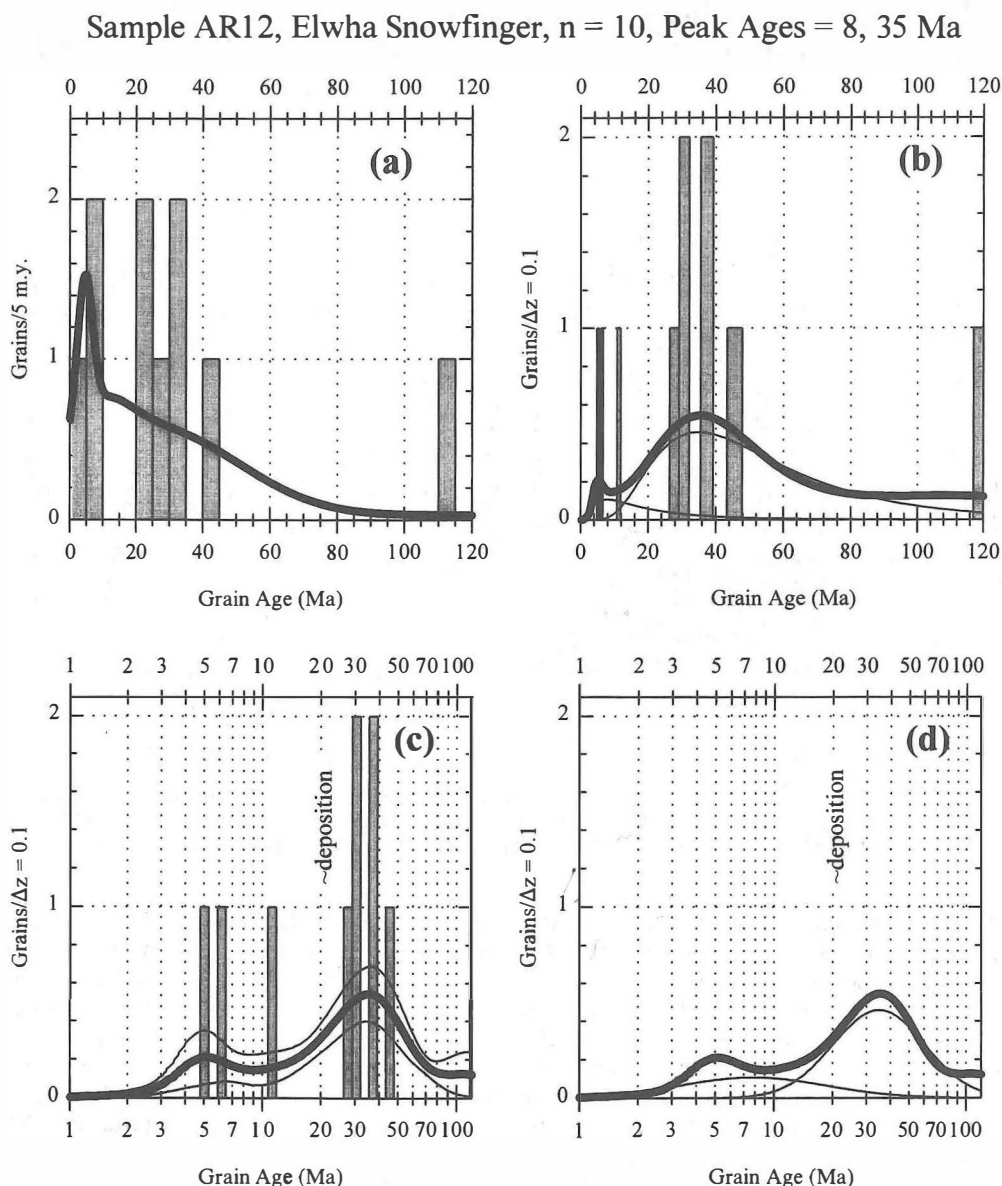Sample AR12, Elwha Snowfinger, n = 10, Peak Ages = 8, 35 Ma

Fig. 8. PD plot for a mixed FTGA sample of detrital apatites ($n = 10$) from a partially reset sandstone from the Olympic subduction complex (sample AR12 of Brandon et al., 1996). This sample has an estimated stratigraphic age of ~ 20 Ma. See Fig. 6 for an explanation of the plots.

Sample ZD6, Mt. Tom, n = 50
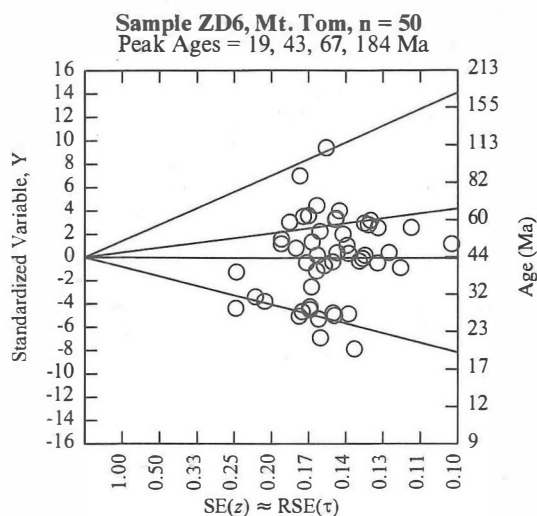Peak Ages = 19, 43, 67, 184 Ma

Fig. 9. Radial plot (Galbraith, 1990) of the FTGA sample displayed in Fig. 6 (sample ZD6 of Brandon and Vance, 1992). The four radial lines mark the mean ages of the component distributions determined by the Gaussian peak-fit method (ZD6 in Table 1). See text for further details.

corresponds to ± 1 y-unit in the vertical around the plotted value for the grain. Lines of constant FT age radiate out from $y = 0$ on the left side of the graph and to the age scale on the right side of the plot. The mean FT ages of the four component distributions determined for this sample (ZD6 in Table 1) are indicated by the radial reference lines in Fig. 9. The radial plot shows a clear clustering of grain ages around the 19 and 184 Ma lines. The other two component distributions at 43 and 67 Ma appear as overlapping clusters and are not well resolved by the radial plot. In contrast, all four components are visible in the modified PD plot [(Fig. 6(c)].

6. CONCLUSIONS

This paper demonstrates the advantages of the modified PD plot for presentation and analysis of FTGA samples. The proposed modifications are: (1) density estimation should use a Gaussian kernel with the kernel width scaled by $\alpha = 0.6$ to achieve the best compromise between resolution and smoothness of the PD plot; (2) density should be estimated using the transform variable z to ensure the that component distributions are approximately Gaussian distributed; (3) the calculated probability density should be transformed from z to τ and plotted using a logarithmically-scaled axis for FT age. The result is a PD plot where component distributions appear as symmetric Gaussian-shaped peaks, with the relative area of each peak proportional to the relative size of the component distribution. This procedure works well for zircon FTGA samples, which generally have large numbers of tracks per grain. It is less successful for apatite FTGA samples because the lower number

of tracks per grain typical of apatite mean that the Gaussian approximation is more commonly not applicable.

REFERENCES

Bardsley W. E. (1983) Confidence limits for fission-track dating. *Mathl Geol.* **15,** 649–658.
Bardsley W. E. (1984) Note on confidence bounds for fission-track dating with low track counts. *Nucl. Tracks* **9,** 69–70.
Brandon M. T. (1992) Decomposition of fission-track grain-age distributions. *Am. J. Sci.* **292,** 535–564.
Brandon M. T. and Vance J. A. (1992) Tectonic evolution of the Cenozoic Olympic subduction complex, Washington State, as deduced from fission-track ages for detrital zircons. *Am. J. Sci.* **292,** 565–636.
Brandon M. T., Roden-Tice M. K. Garver J. I. (1996) Late Cenozoic exhumation of the Cascadia accretionary wedge in the Olympic Mountains, northwest Washington State. *Geological Society of America Bulletin*, in press.
Brigham E. O. (1988) *The Fast-Fourier Transform and its Applications*, pp. 1-448. Prentice–Hall, Englewood Cliffs, NJ.
Cerveny P. F. (1986) Uplift and erosion of the Himalaya over the past 18 million years: evidence from fission track dating of detrital zircons and heavy mineral analysis. M.Sc. thesis, Dartmouth College, Hanover, New Hampshire.
Cerveny P. F., Naeser N. D., Zeitler P. K., Naeser C. W. and Johnson N. M. (1988) History of uplift and relief of the Himalaya during the past 18 million years: evidence from fission-track ages of detrital zircons from sandstones of the Siwalik group. In: *New Perspectives in Basin Analysis* (eds Kleinspehn K. L. and Paola C.), pp. 43-61. Springer, New York.
Chapman D. G. (1952) On tests and estimates for the ratio of Poisson means. *Ann. Inst. Statist. Math. (Tokyo)* **4,** 45-49.
Cox D. R. (1970) *The Analysis of Binary Data*, pp. 1-142. Methuen, London.
Galbraith R. F. (1988) Graphical display of estimates having differing standard errors. *Technometrics* **30,** 271–281.
Galbraith R. F. (1990) The radial plot: graphical assessment of spread in ages. *Nucl. Tracks Radiat. Meas.* **17,** 207–214.
Galbraith R. F. and Green P. F. (1990) Estimating the component ages in a finite mixture. *Nucl. Tracks Radiat. Meas.* **17,** 197–206.
Galbraith R. F. and Laslett G. M. (1985) Some remarks on statistical estimation in fission-track dating. *Nucl. Tracks Radiat. Meas.* **10,** 361–363.
Galbraith R. F. and Laslett G. M. (1993) Statistical models for mixed fission track ages. *Nucl. Tracks Radiat. Meas.* **21,** 459–470.
Garver J. I. and Brandon M. T. (1994a) Fission-track ages of detrital zircon from Cretaceous strata, southern

British Columbia: implications for the Baja BC hypothesis. *Tectonics* **13**, 401–420.

Garver J. I. and Brandon M. T. (1994b) Erosional denudation of the British Columbia Coast Ranges as determined from fission-track ages of detrital zircon from the Tofino basin, Olympic Peninsula. *Washington, Geol. Soc. Am. Bull.* **106**, 1398–1412.

Hurford A. J. and Green P. F. (1983) The zeta age calibration of fission-track dating. *Isotope Geosci.* 285-317.

Hurford A. J., Fitch F. J. and Clarke A. (1984) Resolution of the age structure of the detrital zircon populations of two Lower Cretaceous sandstones from the Weald of England by fission track dating. *Geol. Mag.* **121**, 269–277.

McCullagh P. and Nelder J. A. (1989) *Generalized Linear Models*, 2nd edn, pp. 1–511. Chapman & Hall, London.

Seward D. and Rhoades D. A. (1986) A clustering technique for fission-track dating of fully to partially annealed minerals and other non-unique populations. *Nucl. Tracks Radiat. Meas.* **11**, 259–268.

Silverman B. W. (1986) *Density Estimation for Statistics and Data Analysis*, pp. 1-175. Chapman & Hall, London.

Sneyd A. D. (1984) A computer program for calculating exact confidence intervals for age in fission-track dating. *Computers Geosci.* **10**, 339–345.

Wagner G. A. and Van Den Haute P. (1992) *Fission-Track Dating*, pp. 1–285. Kluwer Academic, Amsterdam.

Whittle P. (1958) On the smoothing of probability density functions. *J. R. Statist. Soc. B* **20**, 334–343.