DECOMPOSITION OF MIXED GRAIN AGE DISTRIBUTIONS USING BINOMFIT

Mark Brandon <mark.brandon@yale.edu>,

Department of Geology and Geophysics,

Yale University, P.O. Box 208109, New Haven, CT 06520-8109

Overview

This report describes BINOMFIT, a program for estimating components in a mixed distribution of fission-track (FT) grain ages. BINOMFIT has been around for some 10 years, but we have just finished building a Windows version that is much easier to use and provides extensive options for displaying and recording results. The program and source code (Visual Basic 6) are available, along with other FT programs, at www.geology.yale.edu/~brandon.

Introduction

A common problem in FT dating is the interpretation of discordant fission-track grain -age (FTGA) distributions. Discordance is defined when the variance of the grain ages is greater than expected for analytical error alone. The widely used χ^2 test provides the method for assessing if the distribution is "over-dispersed" relative to the expectation for count statistics for fission processes (Galbraith, 1981).

Most FT-chronologists use the central age (Galbraith and Laslett, 1993) to represent the "estimated age" of a discordant distribution. Implicit in this approach is that over-dispersion is due to random errors by the FT-chronologist and that such errors are approximately symmetrical around the expected count ratio. My experience with results from age standards suggests to me that most skilled FT-chronologists have a relative low error rate compared to the intrinsic variation due to fission processes. The implication is that discordance in FT dating is more commonly an indication of a mixed distribution rather than a poorly dated distribution.

Mixed distributions are expected for samples that have "detrital" FT ages, such as a sandstone with unreset zircon FT ages. In this case, the grain ages preserve information about the cooling history of the source region from which the zircons were eroded, assuming that the sandstone has remained cool enough after deposition to preserve those ages.

Discordance is also encountered in reset samples. In some cases, the discordance is taken as evidence for partial resetting, but the cause is more commonly due to differential resetting of grains with different annealing properties, due to variable composition in apatite or variable alpha-damage for zircon. Heterogeneous annealing is expected for reset sandstones given that the dated apatites and zircons are derived from a variety of sources (Brandon et al., 1998), but this result is also found in some plutonic rocks as well where one might expect that the dated mineral was more homogeneous (O'Sullivan and Parrish, 1995).

The central age is clearly not appropriate for mixed distributions, in that it provides only a simple average for all grains. A better approach is to decompose the FTGA distribution into a set of component distributions, and to interpret the FT results using the average ages for those components. Various methods are available for estimating components in a grain-age distribution (Seward and Rhoades, 1986; Galbraith and Green, 1990; Brandon, 1992; Galbraith and Laslett, 1993; Sambridge and Compston, 1994; Brandon, 1996).

In my experience, the best method for decomposing FT grain ages is the binomial "peak-fitting" method of Galbraith and Green (1990). The only program available for this method has been a DOS program, BINOMFIT, introduced by Brandon (1992). We have used this program extensively over the last 10 years and have found that it works very well for real FTGA distributions. A big advantage of the binomial peak-fitting method is that it provides a one-component solution that is equivalent to the pooled age. As a result, the calculation provides a seamless extension of the estimation method used for concordant FTGA distributions.

In formal terms, a mixed distribution is considered to be a mixture of a finite set of

component distributions, or components. Such components will commonly appear as bumps or peaks in a histogram or probability density plot, but this need not be the case given that overlapping components may appear as only one peak. Nonetheless, the term peak has come to be used as an equivalent for component.

Galbraith and Laslett (1993) introduced the term "minimum age", which can be loosely viewed as the pooled age of the largest concordant fraction of young grain ages in the FTGA distribution. They proposed a two-component mixture model for estimating the minimum age. Binomial peak fitting can also be used to estimate the minimum age of a distribution, while at the same time providing information about older components. The main requirement, however, is that peak fitting must be coupled with a test to find the maximum number of significant components in the distribution, as discussed below.

We have found that FT minimum ages are very useful in a variety of studies. For detrital zircon FT studies of volcanoclastic rocks, the minimum age is commonly a useful proxy for the depositional age of the rock (Brandon and Vance, 1992; Garver et al., 1999, 2000). Likewise, the minimum age for a tuff can remove biases due to older contaminant grains. For reset rocks, the minimum age represents the time of closure for that fraction of grains with the lowest retention for fission tracks, such as fluorapatites in apatite FT dating (e.g., Brandon et al., 1998) and radiation-damaged zircons in zircon FT dating (e.g., Brandon and Vance, 1992).

General Concepts for the BINOMFIT Program

It is useful to ask why the binomial distribution is important for this problem. The data we measure are spontaneous and induced track counts for each grain, designated as r_i and s_i with the index indicating the ith grain in a sample of $i =$ 1 to N dated grains. The variables r_i and s_i are Poisson distributed, but they can be converted into a single binomial-distributed variable by the transformation $r_i / (r_i + s_i)$ (Galbraith and Green, 1990). This variable can be approximated by a Gaussian distribution when values of r_i and s_i are large, which is commonly the case for zircon, which tends to have a high track density because of greater U content. However, as the cooling age and U content decrease, r_i and s_i get small. Thus,

the Gaussian approximation tends to breaks down for young cooling ages, especially for apatites, which commonly have low U content. In these cases, decomposition methods based on the Gaussian distribution (e.g., GAUSSFIT by Brandon, 1992; MIX by Sambridge and Compston, 1994) will perform poorly. The binomial peak-fitting method provides unbiased estimates of peak ages for track counts of any size.

The binomial peak-fitting method is based on the maximum likelihood method, which means that the best-fit solution is determined by directly comparing the distribution of the grain data to a predicted mixed binomial distribution. This approach is a significant improvement over the convention least-square method, which requires more restrictive assumptions about the distribution of residuals.

A problem with all peak-fitting method is that the calculation requires an initial guess of the number and ages of the peaks in the distribution. The user specifies the number of peaks to be fit, and that number could be as large as the number of grains in the FTGA distribution. So we are left to ask, how many peaks are significant? The solution can also fail to converge given a poor initial set of peak ages used to start the solution.

BIONOMFIT provides an iterative search of peak ages and number of peaks to find an optimal solution. The procedure we use is an automated version of the F-test approach outlined in Brandon (1992). For this search, we consider a large set of trial solutions. The trick is to organize the search and to apply appropriate tests to find the best solution.

The quality of fit for each trial solution is scored using the χ^2 statistic, in a manner similar to the conventional χ^2 test. The strategy for the search is to consider successively larger number of peaks and to use a large number of initial guesses to ensure that for a specified number of peaks, we have found the best-fit solution. The initial guesses for peak ages are generated from the probability density plot for the FTGA distribution. The density plot is estimated using the method in Brandon (1996), and the first and second derivatives are used to find bumps and humps in the plot. The probability density at each candidate age is used to guess the

potential size of the peak. We have tried an alternative approach, using evenly spaced ages, but the computation takes longer and does not provide any improvement in performance.

The next step is to iterate through an increasing number of peaks. The first iteration produces a single component age, with an age identical to the pooled age and a χ^2 value identical to that produced by the conventional χ^2 test. The next iteration finds a best-fit two component solution using all combinations of initial peaks as initial guesses. The best solution is the one with the lowest χ^2 value. We next consider a three-component solution, then a four-component solution, and on.

With each iteration, we use the F test to determine if the introduction of an additional component has produced a significant improvement in the χ^2 statistic (Brandon, 1992). In general, one will find that the χ^2 statistic gets smaller with the introduction of a new component. In part, this is due to the fact that the additional component provides the model with greater flexibility to fit the data. But we need to know if the improvement in fit is large compared with the expected random variability associated with measurements. In other words, we want to be assured that the additional peak is fitting signal and not just noise.

This question is nicely addressed by the F test. To explain, consider two solutions for m and m+1 peaks, and the quality of those fits, indicated by χ^2_m and χ^2_{m+1}. The F statistic is given by $F = (m+1)(\chi^2_m - \chi^2_{m+1})/ \chi^2_m$. When F is large, then the improvement in fit associated with the additional peak is considered significant. The F distribution is used to assign a probability P(F), which is the probability that random variation alone could produce the observed F statistic. We consider P(F) < ~5% to indicate that the improvement in fit is significant. Thus, we can find the optimal number of significant peaks by adding peaks until we get a value of P(F)> ~5%.

Using BINOMFIT

The Windows version of BINOMFIT is written in Visual Basic v. 6.0. Most people will have to deal only with the executable program, but I am distributing the source code as well so that one can inspect the methods used in the program. You are free to modify the source code but please do not distribute the modified version without first consulting with me, in order to avoid confusion about different versions. Also I encourage queries about problems with the program and suggestions for improvements.

The executable version requires some supporting DLL and OCX files. These files are already loaded on many Windows systems since they are used by a variety of other programs. The Readme.txt file in the distribution package provides information for those of you that might be lacking the necessary support files.

Data files are constructed as simple text files. Make sure that the files are standard Windows-based ASCII files. If you are unsure about the provenance of a data file, then open the file in MS Word for Windows, edit out any offending lines or characters, and then save as Plain Text (*.txt). I recommend that the file extension be set to .FTA, .FTZ, etc. depending on the dated mineral. Note that the file extension may not be visible since the default option in newer versions of Windows is to hide the extension. You can change that option or just pay attention to the file type designation to see the extension.

Data files have the following format:

1) Line 1: Title for data set.

2) Line 2: This line is flagged by making the first value negative. All values on this line should be separated by commas.

   a. ZETA METHOD: Effective track density (tr/cm^2) and the relative standard error (%) for the fluence monitor, and the effective uranium concentration (ppm) of the monitor standard.

   b. Z METHOD: Enter line as: -1, 0, 0.

3) Line 3: All values should be separated by commas.

   a. ZETA METHOD: Zeta factor (yr cm^2/tr), standard error (yr cm^2/tr), counter square size (cm^2).

   b. Z METHOD: Z factor (yr), standard error (yr), counter square size (cm^2).

4) Line 4 and on: For each grain, enter the measured spontaneous and induced tracks, and the number of squares for the area counted. Data values can be separated by spaces or commas.

5) To merge another dataset, repeat lines 2 and on.

6) A semicolon in the first space marks the line as a comment. The program skips comments and blank lines.

The methods described here follow Hurford and Green (1983). The zeta method is the most widely used, but there are some cases where the method is needed (such as when an age standard is used as a fluence monitor). The program can handle up to 1000 grain ages, so size limitations should not be problem.

The program comes with some example data files. We focus here on MTTOM.FTZ, a detrital zircon FT sample from Mount Tom in the Olympics Mountains (Brandon and Vance, 1992). We start the program and use the file menu to open the data file. The program comes up with a window showing a list of initial guesses for peak ages. The user is free to remove (i.e., "uncheck") specific peak ages from the list. The user then indicates the mode for the calculation. Automatic mode is as outlined above. Manual mode allows the user to specify a limited set of initial peaks ages, and to solve for that number of peaks only.

The automatic option is followed immediately by the best-fit solution (Figure 1). The best-fit peaks are reported by age, uncertainty, and size. The uncertainty for the peak age is given at 68% and 95% confidence intervals. The size of the individual peaks is reported as a fraction (in percent) and as a count (number of grains). Note that peaks sizes are constrained so that they all sum to 100%.

The "Grain Ages" plot (lower left in Figure 1) shows a vertical line for each grain age, organized according to its position in the file. The best-fit peak ages and 68% confidence intervals are plotted horizontally across the plot. The cursor can be used to find grains according to age and position in the file, by placing the cursor over the grain line and looking at the results in the status bar at the bottom of the program window. Right-clicking the cursor brings up a window where the user can remove the selected grain age or copy the plot to the clipboard, for pasting into another application.
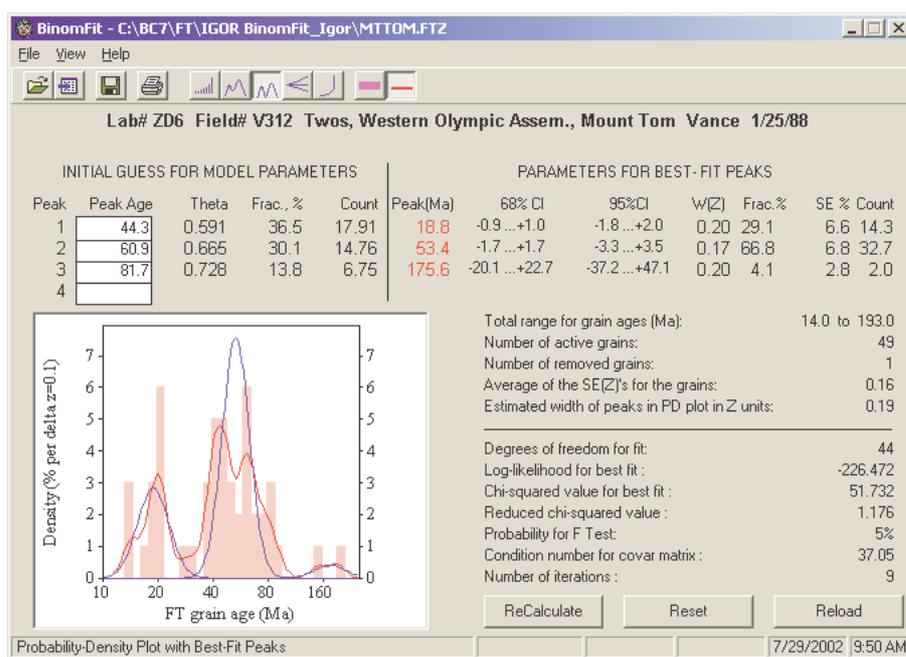

Figure 1. Best-fit solution for the Mount Tom zircon FT sample using automatic mode.

The toolbar or view menu can be used to select other reports on the data, including a probability density plot of the data and the best-fit peaks (Figure 2), a radial plot showing data and best-fit peaks (Figure 3), and a P(F) plot (Figure 4), which summarizes the results of the F tests used to determine the number of significant peaks.
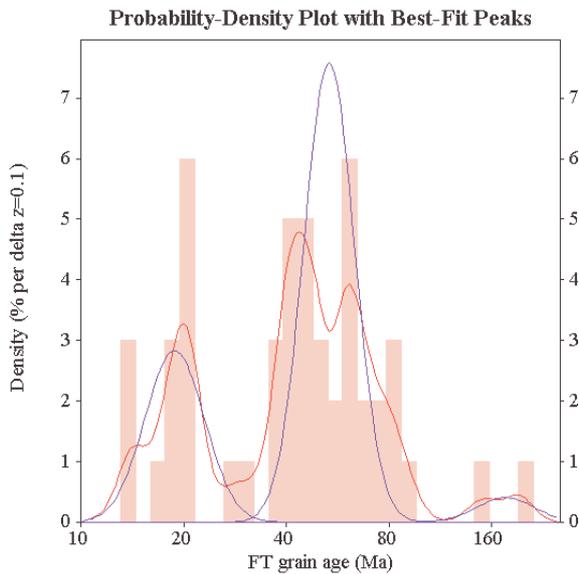

Figure 2. Probability density plot (red line) with best-fit peaks plotted using blue lines for the Mount Tom FT zircon sample. The histogram in the background is scaled to the same density as the density plot.
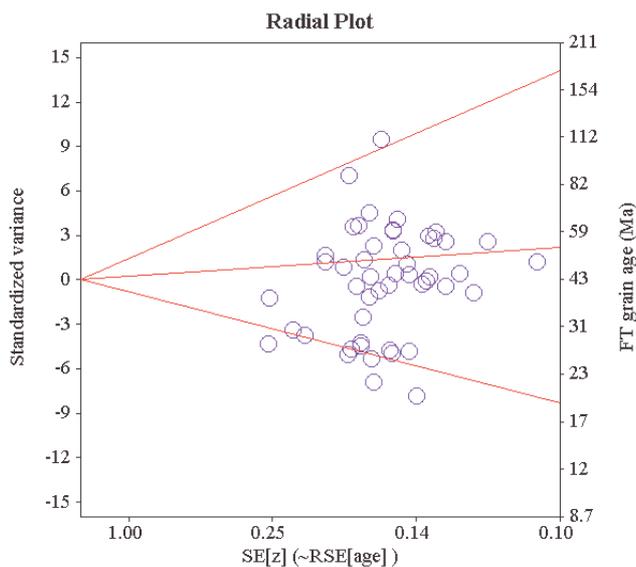

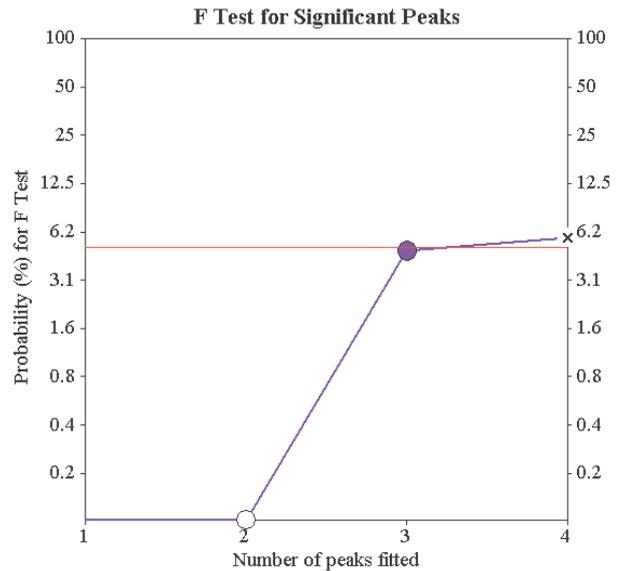**Figure 3.** Radial plot for the Mount Tom FT zircon sample.

Figure 4. F test for Mount Tom zircon FT sample.

The P(F) plot is useful as a summary of the search, the details of which can be found in the Log File, available in the View menu. The peak search determined that the 3 peak solution was optimal, but note that the 4 peak solution failed by only a faction of a percent, with P(F) = 5.7%. Adding additional peaks manually indicates that the 5 peak solution has P(F) = 100%. Thus, we can say with confidence that the number of significant peaks is definitely more than 2 and less than 5. Note that this ambiguous result is unusual. More commonly, we find that P(F) rises rapidly across the 5% threshold.

The Print option in the File menu will send a full report of grain ages, best-fit solution and plots to a designated printer. The Save option in the File menu allows the user to save plot files, which can be imported into a graphics program (e.g., Excel, SigmaPlot) to prepare formal plots for publication. Plot files are available for probability density plots, radial plots, and P(F) plots. Data file can be saved as well, to account for when the file is modified in the program, due to removal of grain ages or merging multiple data files.

In closing, I note that the program presently lacks a Help file. I am hoping to have this available within the next couple of months. The program should be fairly intuitive for the average user.

References

Brandon, M.T., 1992, Decomposition of fission-track grain-age distributions: American Journal of Science, v. 292, p. 535-564.

—, 1996, Probability density plots for fission-track grain age distributions: Radiation Measurements, v. 26, p. 663-676.

Brandon, M.T., Roden, T.M.K., and Garver, J.I., 1998, Late Cenozoic exhumation of the Cascadia accretionary wedge in the Olympic Mountains, Northwest Washington State: Geological Society of America Bulletin, v. 110, p. 985-1009.

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: American Journal of Science, v. 292, p. 565-636.

Galbraith, R.F., 1981, On statistical models for fission-track counts: Journal of Mathematical Geology, v. 13, p. 471-478.

Galbraith, R.F., and Green, P.F., 1990, Estimating the component ages in a finite mixture: Nuclear Tracks and Radiation Measurements, v. 17, p. 197-206.

Galbraith, R.F., and Laslett, G.M., 1993, Statistical models for mixed fission track ages: Nuclear Tracks and Radiation Measurements, v. 21, p. 459-470.

Garver, J.I., Brandon, M.T., Roden, T.M.K., and Kamp, P.J.J., 1999, Exhumation history of orogenic highlands determined by detrital fission-track thermochronology, *in* Ring, U., Brandon Mark, T., Lister Gordon, S., and Willett Sean, D., eds., Exhumation processes; normal faulting, ductile flow and erosion., Volume 154: Geological Society Special Publications: London, United Kingdom, Geological Society of London, p. 283-304.

Garver, J.I., Soloviev, A.V., Bullen, M.E., and Brandon, M.T., 2000, Towards a more complete record of magmatism and exhumation in continental arcs using detrital fission track thermochronometry: Physics and Chemistry of the Earth, Part A, v. 25, p. 565-570.

Hurford, A.J., and Green, P.F., 1983, The zeta age calibration of fission-track dating: Chemical Geology (Isotope Geoscience Section), v. 1, p. 285-317.

O'Sullivan, P.B., and Parrish, R.R., 1995, The importance of apatite composition and single-grain ages when interpreting fission track data from plutonic rocks: a case study from the Coast Ranges, British Columbia: Earth and Planetary Science Letters, v. 132, p. 213-224.

Sambridge, M.S., and Compston, W., 1994, Mixture modelling of multi-component data sets with application to ion-probe zircon ages: Earth and Planetary Science Letters, v. 128, p. 373-390.

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: Nuclear Tracks and Radiation Measurements, v. 11, p. 259-268.