FitPDF : a program to calculate and graph probability curves for data measurements with uncertainties

by

Dr Bruce Eglington
Saskatchewan Isotope Laboratory
University of Saskatchewan
Saskatoon
Saskatchewan
S7N 5E2
Canada

email: bruce.eglington@usask.ca

2018

Contents

1.   Introduction

FitPDF is a program to facilitate the calculation and graphing of probability distributions for data with associated analytical uncertainties. It is designed for use on computers using the Windows operating system. Data are imported from Excel spreadsheets, so permitting calculations for large data sets. The program is designed to generate two types of diagram, primarily for use in geology and geochronology, mainly when working with detrital zircon analyses. One can generate a single probability plot and frequency histogram from a set of imported data or one can generate a new form of plot which permits one to illustrate multiple independent probability plots relative to stratigraphic (age) hierarchy. Probability plots utilise the calculation approaches developed for Geodate (Eglington and Harmer 1991) and DateView (Eglington 1999; Eglington, Reddy, and Evans 2009) and provides results equivalent to those available from Isoplot (Ludwig 2003) and AgeDisplay (Sircombe 2004).

2.   Importing Data

Data need to be available in an Excel spreadsheet (both Excel 1997-2003 XLS format and Excel 2007-2017 XLSX formats are supported).

The data need to be arranged in three columns: age, analytical uncertainty and unit (or sample) age. After selecting the spreadsheet containing data to be imported, one needs to specify which columns contain the data as well as start and end rows. One can choose to omit data with negative values and this is a useful technique to distinguish which data are discordant rather than by sorting data in the spreadsheet. If you want to use this approach, simply calculate another column with an appropriate IF statement to check against degree of concordance. Rows that fail the concordance test are given negative ages.

Figure 1. Example dataset. Test data illustrated in this dataset are from Davis et al. ((Davis et al. 2015)) plus other sources (not shown in image).

For the example dataset illustrated in Figure 1, rows 2 to 9203 of the spreadsheet will be imported, taking values from columns D (age), E (uncertainty) and assuming an age for the unit/sample as provided in column F. An extra variable, used in some of the plots, is taken from column G. Uncertainties are 1 sigma.

Once data have been imported, they are shown in a grid and the various unique ages assumed for the different groups (e.g. samples or lithostratigraphic units) are extracted and shown in a second grid.

4

Figure 2. Grid of data values imported to FitPDF and their unique ages.

A graph of the input data (ignoring the individual input uncertainties) provides a simple mechanism to check that the data do not have any extreme incorrect values (Figure 3).
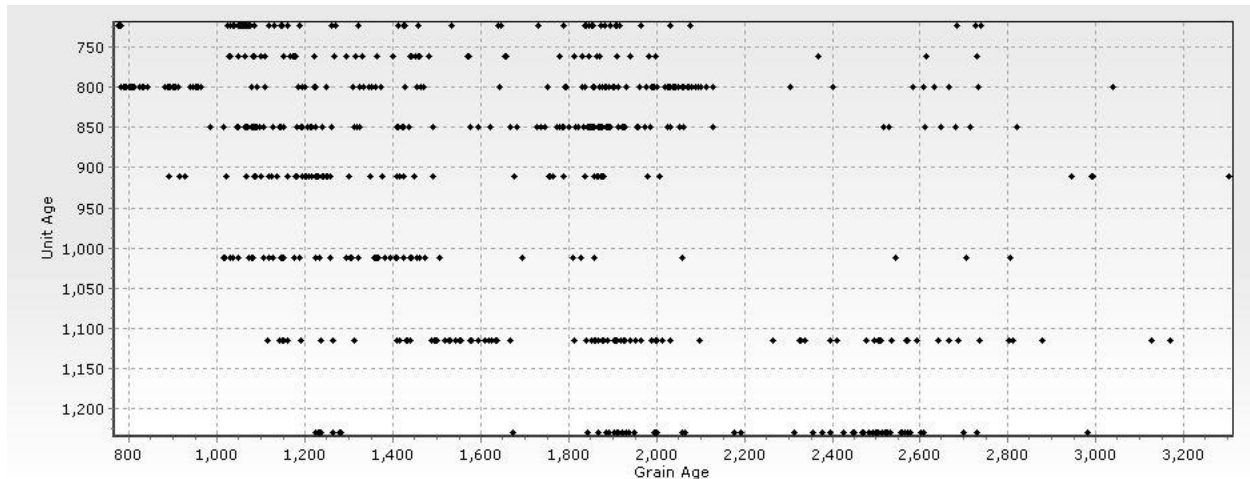

Figure 3. Graphical illustration of input data plotted relative to Unit (Y axis) and Sample (X axis) age.

3.   Defining the Calculation and Graphing Parameters

Here, one can select between a single probability plot for all data imported or a multiple probability plot relative to stratigraphic age, height or some arbitrary level.

Figure 4. Parameters which may be set to create various types of graph. Note that, with the exception of Figure 5, all probability distribution curves are normalised to the maximum peak i.e. the maximum peak is set to 100% and all other values are scaled accordingly.

Choose either the Gaussian 'AND' or 'OR' calculation approach and whether probability values should be expressed as a percentage of the most prominent peak or as true probability values.

Select how many bins are to be used for the frequency histogram. The default value provided by the software is based on methodology outlined in Numerical Recipes (Press et al. 2002).

Minimum and maximum ages may be stipulated for the calculations and plots. Each probability curve is subdivided into 2500 incremental steps between these ages. A minimum 1 sigma uncertainty is provided. Any individual values less than this will be increased to this value. This provides a useful method to smooth the input data and to reduce the influence of overly precise analyses, for instance if combining zircon evaporation data with analyses from other techniques.

3.1. 'AND' and 'OR' Gaussian Calculations

The traditional approach to calculation of probability curves has been to determine the summed Gaussian distribution of all individual analyses, assuming that the uncertainties around the measured data (age) values are normally distributed (hence the term Gaussian). In effect, this

approach determines **the probability of a specific age being found in the specific dataset one is using**, not in the population from which the individual analyses were drawn.
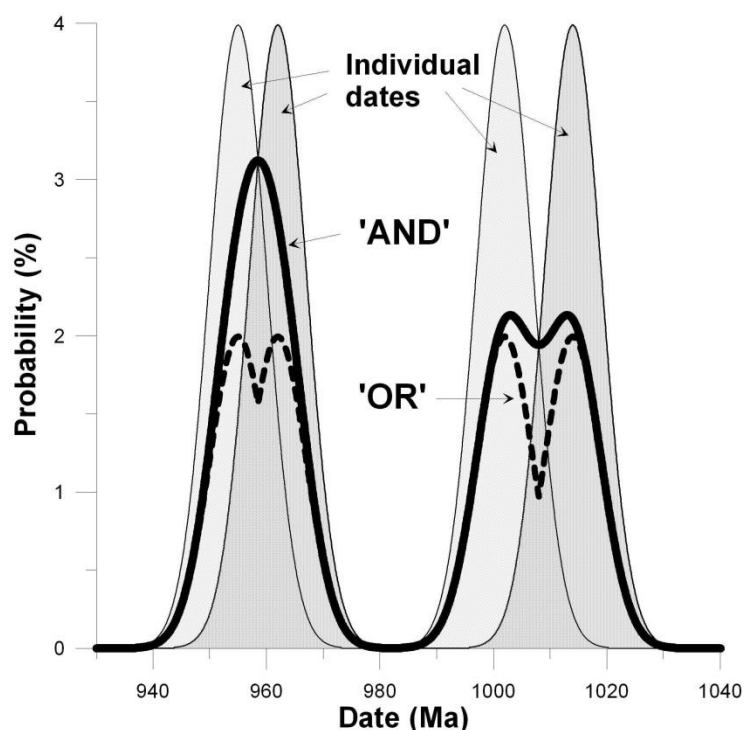
Figure 5. Gaussian 'AND' (solid thick line) and 'OR' (dashed thick line) representation of a simple test dataset comprising four analyses (solid thin line). Note the difference in peak height between the two approaches. Probabilities are presented as 'true probabilities' i.e. without normalisation to the maximum peak. This has the effect of reducing maximum peak height for the combined data since the summed probabilities are divided by the number of samples used to calculate the probability distribution (by 4 in this example).

It is sometime desirable to calculate the **probability of any occurrence of a certain age**. In this case it is appropriate to use a Gaussian 'OR' approach i.e. only the most likely (highest probability) individual age counts in defining the shape of the probability curve. All lesser probabilities at the age are ignored. Figure 5 illustrates the difference between the 'AND' and 'OR' approaches while Figure 6 and Figure 7 illustrate the example dataset used in the other diagrams of this manual.

Figure 6. Gaussian 'AND' representation of example data (red curve) overlaid on a histogram binned according to the settings in Figure 4. Minimum age uncertainty was set as 5 Ma.

Figure 7. Gaussian 'OR' representation of example data (red curve) overlaid on a histogram binned according to the settings in Figure 4. Minimum age uncertainty was set as 5 Ma.

3.2. Influence of Minimum Uncertainty

Changing the minimum age uncertainty permitted for all analyses provides a technique to smooth the probability curve by reducing the influence of very precise ages. Increasing the minimum uncertainty from 5 Ma to 20 Ma is illustrated in Figure 8 and Figure 9.

Figure 8. Gaussian 'AND' representation of example data (red curve) overlaid on a histogram binned according to the settings in Figure 4. Minimum age uncertainty was set as 20 Ma.
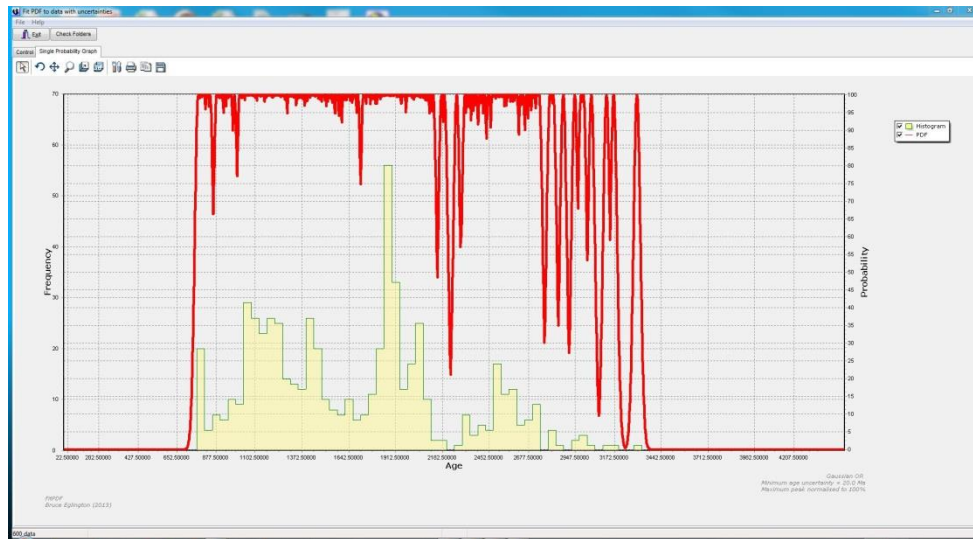
Figure 9. Gaussian 'OR' representation of example data (red curve) overlaid on a histogram binned according to the settings in Figure 4. Minimum age uncertainty was set as 20 Ma.

4.   Probability Calculations for a Single Probability Plot

The output graph for single probability distributions allows users to select whether to show both probability distribution and frequency histogram (Figure 6), only probability curve (Figure 10) or only the histogram (Figure 11). This is achieved by checking the appropriate checkboxes in the graph Legend.
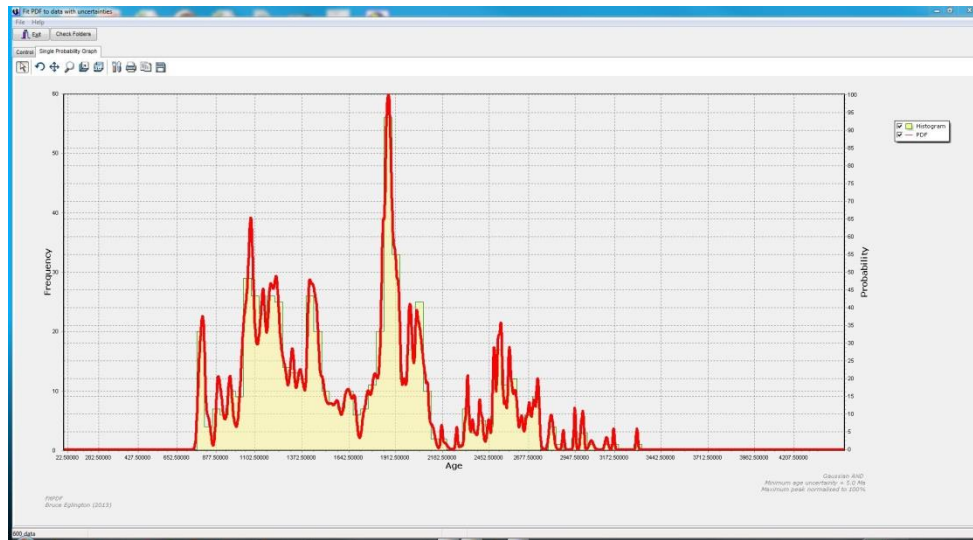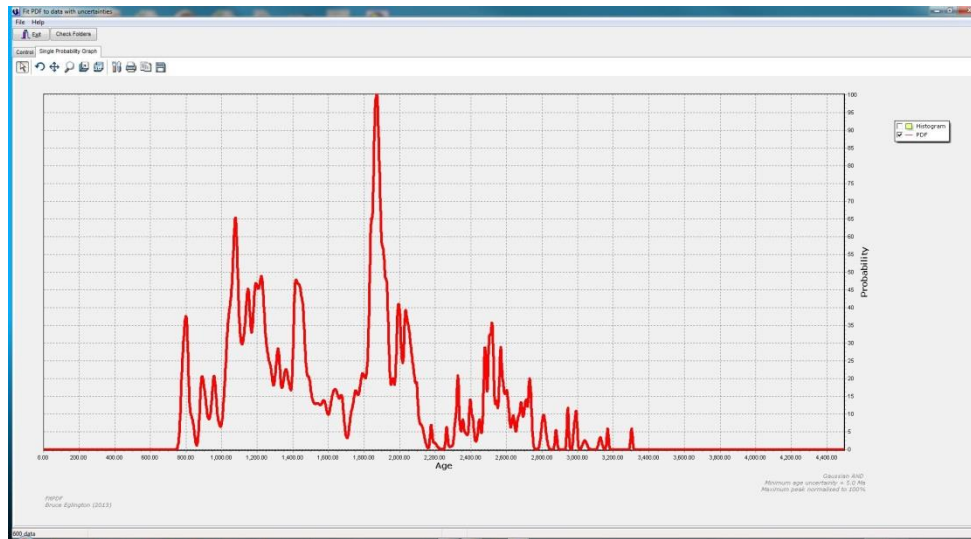
Figure 10. Gaussian 'AND' representation of example data (red curve) overlaid on a histogram binned according to the settings 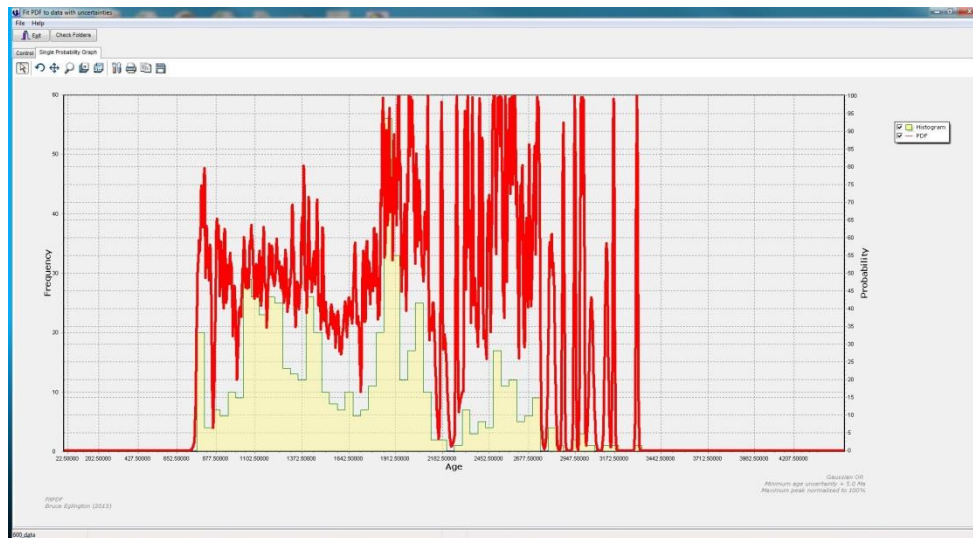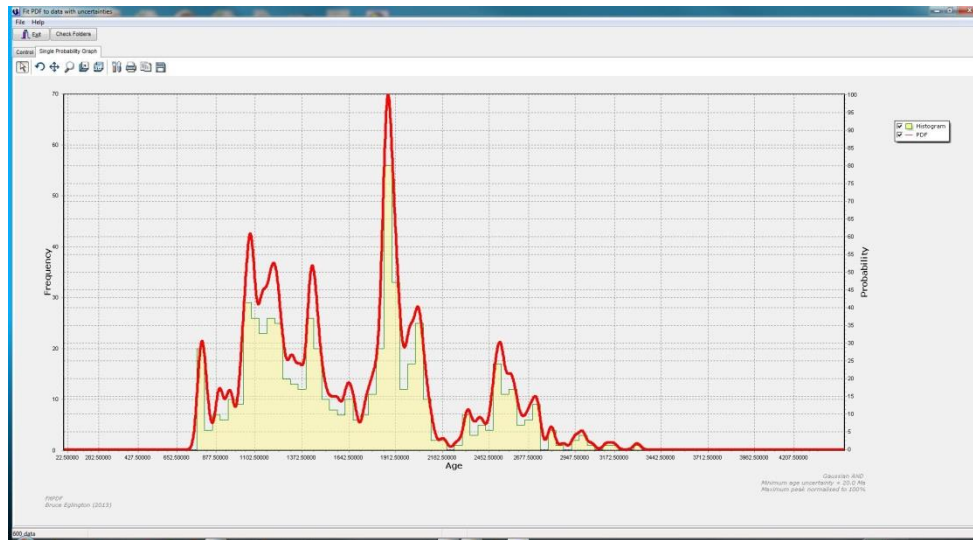in Figure 4. Minimum age uncertainty was set as 5 Ma. Histogram values were switched off by unchecking the appropriate checkbox in the graph Legend.
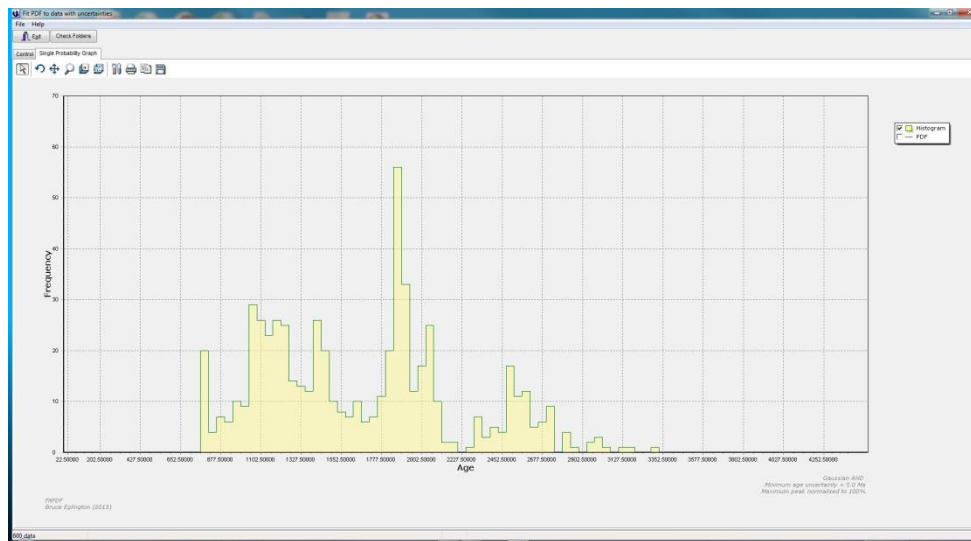

Figure 11. Histogram representation of example data binned according to the settings in Figure 4. Probability curve values were switched off by unchecking the appropriate checkbox in the graph Legend.

5.   Probability Calculations for Multiple Samples with Different Formation Ages

The traditional approach to illustrating probabilities for multiple samples has been to generate a number of probability curves 'stacked' one above the other. Whilst obviously appropriate in many cases, these plots have the problem that they rapidly become confusing with many samples and there is no way to easily relate them to formation age of the samples e.g. deposition age of the host lithostratigraphic unit. A new form of graph was developed for the StratDB and DateView databases (available online from http://sil.usask.ca/databases.htm) and is included in FitPDF for offline processing of similar data. This graph represents probabilities from individual probability curves by line thickness and symbol size, augmented by colour to emphasise the most prominent peaks. Greater probabilities have thicker lines and bigger symbols with the highest peaks in the probability distributions emphasised in orange and red. Figure 12 illustrates a set of probability curves in a 3-dimensional plot of unit age, grain age and probability. When viewed from above, this plot appears as shown in
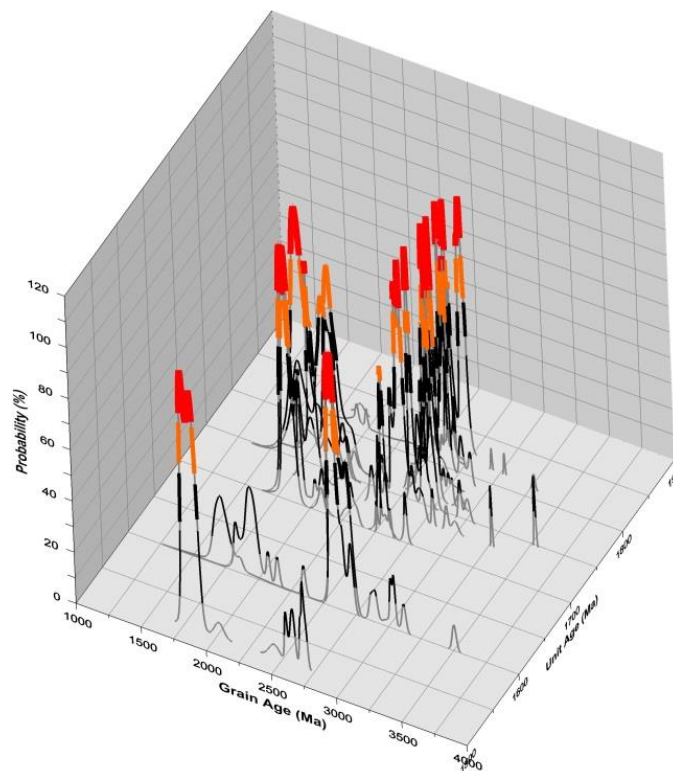
Figure 12. 3D diagram illustrating the layout of probability curves used to produce the 'vertical view' plots.
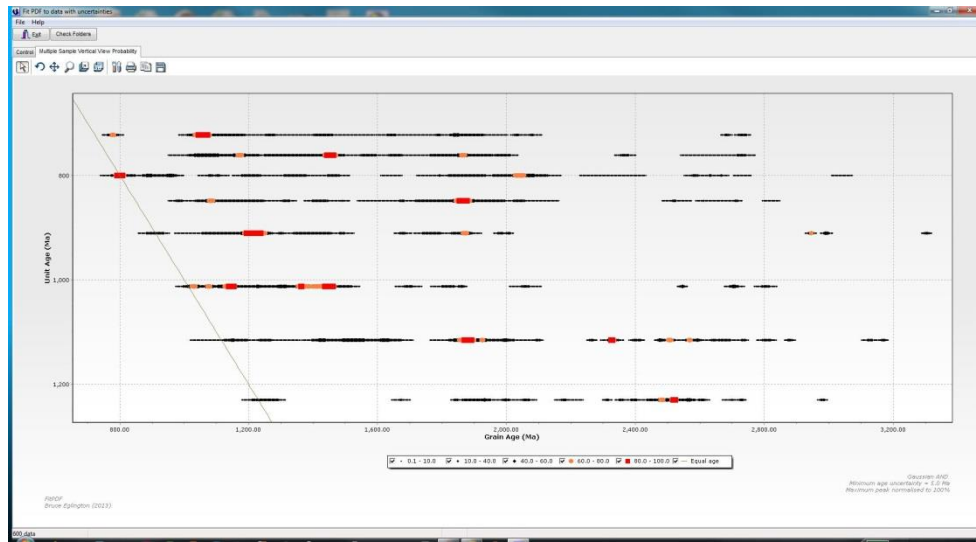
Figure 13. 2D diagram illustrating the probabilities for a set of samples from various lithostratigraphic units arranged according to their formation age. The line sloping from top left towards lower right delineates where grain age and unit age are equal.
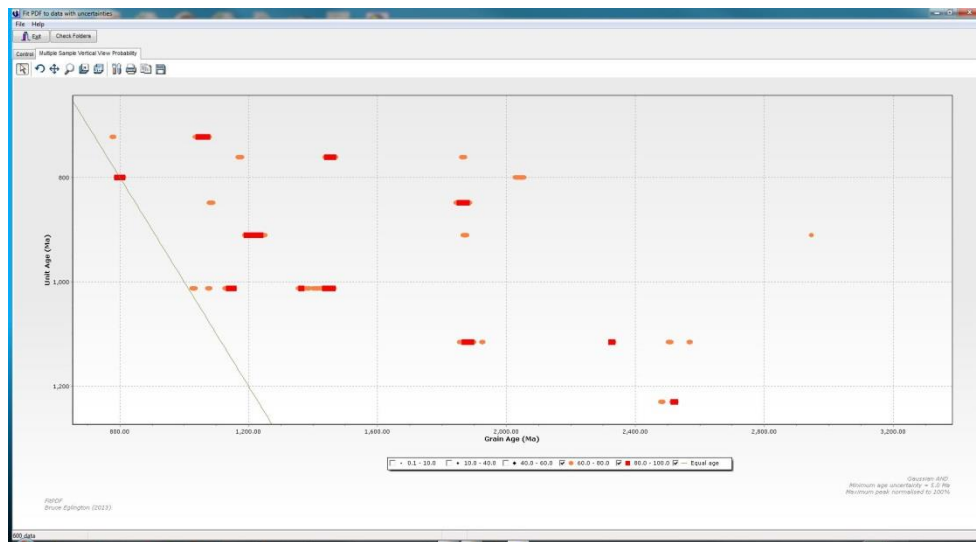
Figure 14. 2D diagram illustrating the probabilities for a set of samples from various lithostratigraphic units arranged according to their formation age. Only the highest probability values are drawn with lower probabilities switched off by unchecking appropriate checkboxes in the Legend. The line sloping from top left towards lower right delineates where grain age and unit age are equal.

5.1. Adding an Extra Variable to the Graph

In some cases, it is useful to superimpose the graphical composition of some additional variable, for instance Th/U, over the distribution of detrital zircon data in the multi-sample graphs. In these cases, one needs to select the cut-off value below (or above) which values must be graphed.
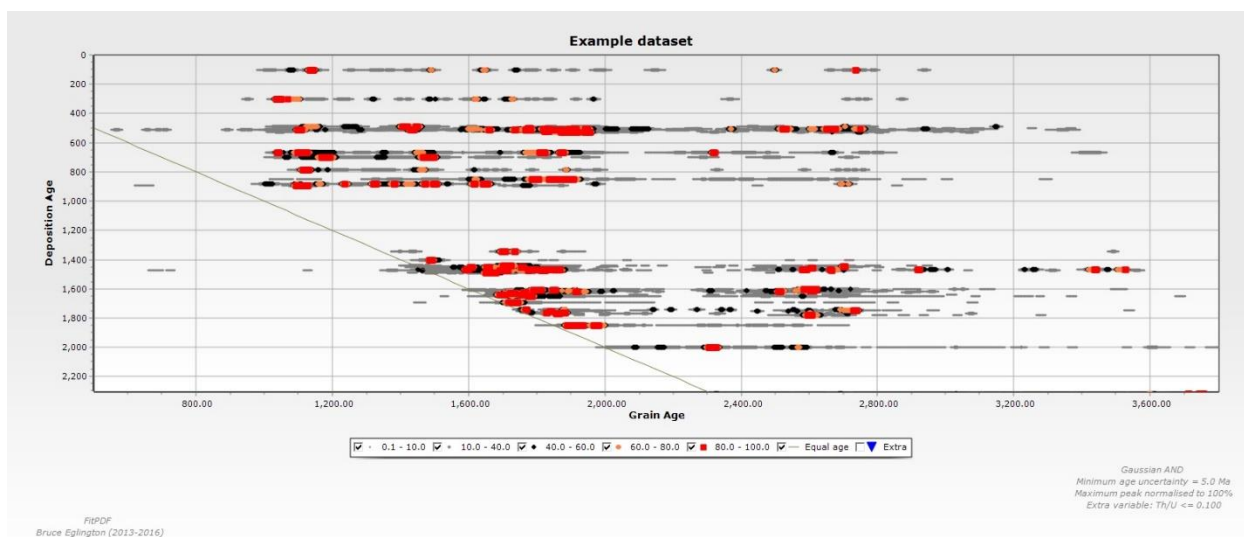

Figure 15. Multi-sample probability distribution relative to deposition age. Extra variable not plotted but see Figure 16 for equivalent dataset with the extra variable plotted.

Figure 15 illustrates a multi-sample graph without plotting of the extra variable and Figure 16 shows the same data with the extra variable plotted. In this case, the extra variable is the Th/U ratio for zircon analyses, with only those analyses having values <= 0.1 being plotted. This allows one to identify which intervals of time contain zircons likely to have been sourced from metamorphic terrains. Other similar situations may be envisaged based on oxygen isotope data or rare earth element data in the case of detrital zircons.
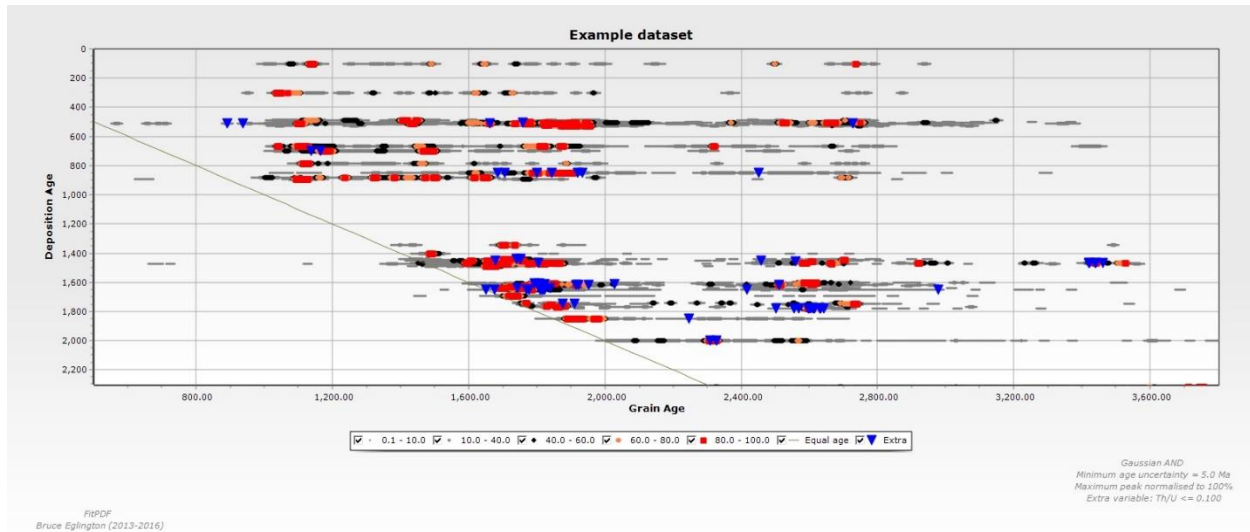
Figure 16. Multi-sample probability distribution relative to deposition age. Extra variable values which are less than the cut-off value of 0.1 specified in Figure 4 are plotted as blue inverted triangles. This provides a mechanism to identify analyses with unique signatures, in this case for low Th/U which may reflect a metamorphic origin for the spots analysed (Hoskin and Schaltegger 2003).

6. Assessing Possible Geotectonic Environment

Cawood et al. (Cawood, Hawkesworth, and Dhuime 2012) proposed a methodology to identify possible geotectonic environment from the cumulative distribution function shapes of detrital zircon records for sedimentary units. FitPDF can colour code each uniquely aged (in terms of deposition age) set of data from multi-sample distributions and create a graph of cumulative age distribution. The colours used for curves in the graph are consistent with those used in the original Cawood et al. publication, as shown in Figure 17.

A menu to the right of the graph allows one to step through the set of cumulative distribution curves from oldest deposition age to youngest to visually identify changing patterns in geotectonic environment.

Colours associated with each unique deposition age record may be exported, via one of the main menu items, so as to link geotectonic interpretation to the original data records for use in other packages such as GIS.
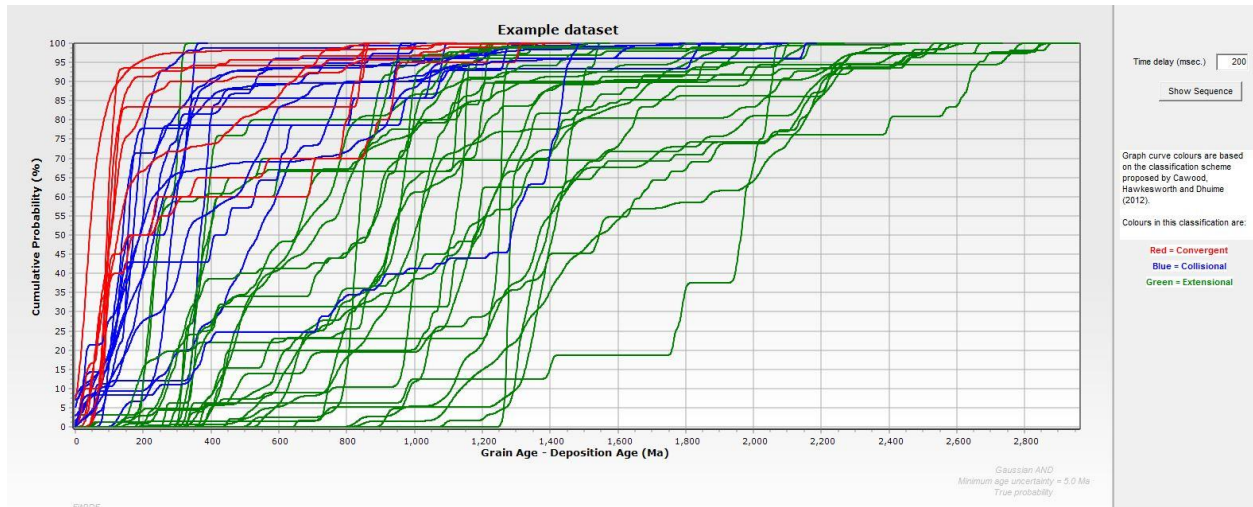
Figure 17. Multi-sample cumulative probability distributions for each unique deposition age in the dataset. Colours used and criteria follow the original publication by Cawood et al. (Cawood, Hawkesworth, and Dhuime 2012).

An additional graph is provided at the bottom right of this window, illustrating the inferred geotectonic setting for each unique deposition age, as illustrated in Figure **18**.

Figure 18. Inferred geotectonic setting relative to deposition age. Colours used follow the original publication by Cawood et al. (Cawood, Hawkesworth, and Dhuime 2012).

7.   Modifying Graph Images

Graphs may be directly modified by clicking on the 'spanners' icon above the graph. Users are able to change line and symbol colours, sizes, shapes; axis parameters; titles, etc and to export the graphs to both vector and raster format files.
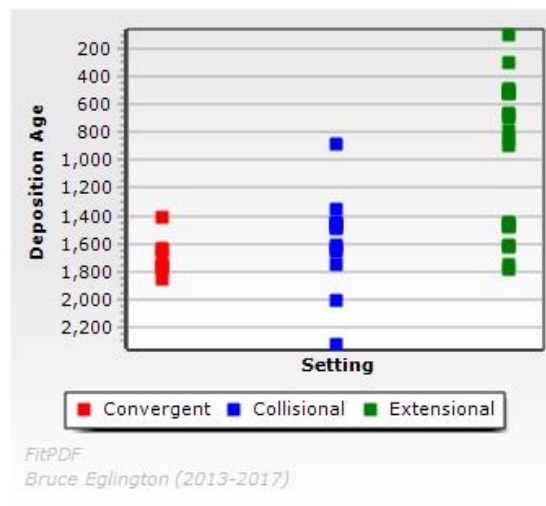
With the 'arrow' icon selected, one can zoom into images by left-mouse dragging from top left to bottom right or zoom out by left-mouse dragging in the reverse direction. The graph may be scrolled by right-mouse dragging.

8.   Exporting Graph Images

The graphs may be exported to either vector or raster format files by clicking on the appropriate icon above each graph or by selecting the 'spanners' icon and choosing the 'export' option within the menu which pops up.

9.   Exporting Data to Reproduce Graphs with Offline Software

The data values used to produce the curves in the graphs may be exported to Excel spreadsheets by selecting the appropriate menu item from the main menu.

10.   References

Cawood, P. A., C. J. Hawkesworth, and B. Dhuime. 2012. "Detrital Zircon Record and Tectonic Setting." Geology 40 (10): 875–78. doi:10.1130/G32945.1.

Davis, William J., Luke Ootes, Lucy Newton, Valerie Jackson, and Richard A. Stern. 2015. "Characterization of the Paleoproterozoic Hottah Terrane, Wopmay Orogen Using Multi-Isotopic (U-Pb, Hf and O) Detrital Zircon Analyses: An Evaluation of Linkages to Northwest Laurentian Paleoproterozoic Domains." Precambrian Research 269 (Supplement C): 296–310. doi:10.1016/j.precamres.2015.08.012.

Eglington, B. M. 1999. "DateView: A 32-Bit Windows Geochronology Database." Council for Geoscience Open File Report 1999–0207–O: 1–12.

Eglington, B. M., and R. E. Harmer. 1991. "GEODATE Version 2.2: A Program for the Processing and Regression of Isotope Data Using IBM-Compatible Microcomputers." CSIR manual EMA-H 9102.

Eglington, B. M., S. M. Reddy, and D. A. D. Evans. 2009. The IGCP 509 Database System: Design and Application of a Tool to Capture and Illustrate Litho- and Chrono-Stratigraphic Information for

Palaeoproterozoic Tectonic Domains, Large Igneous Provinces and Ore Deposits; with Examples from Southern Africa. 323rd ed. London: Geological Society.

Hoskin, Paul W. O., and Urs Schaltegger. 2003. "The Composition of Zircon and Igneous and Metamorphic Petrogenesis." Reviews in Mineralogy and Geochemistry 53 (1): 27–62. doi:10.2113/0530027.

Ludwig, K. R. 2003. "User's Manual for Isoplot 3.00: A Geochronological Toolkit for Microsoft Excel." Berkeley Geochronology Center Special Publication 4: 1–74.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. 2002. Numerical Recipes in C++ : The Art of Scientific Computing. 2nd ed. Cambridge: Cambridge University Press.

Sircombe, K. N. 2004. "AgeDisplay: An Excel Workbook to Evaluate and Display Univariate Geochronological Data Using Binned Frequency Histograms and Probability Density Distributions." Computers and Geosciences 30: 21–31.