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

Contents

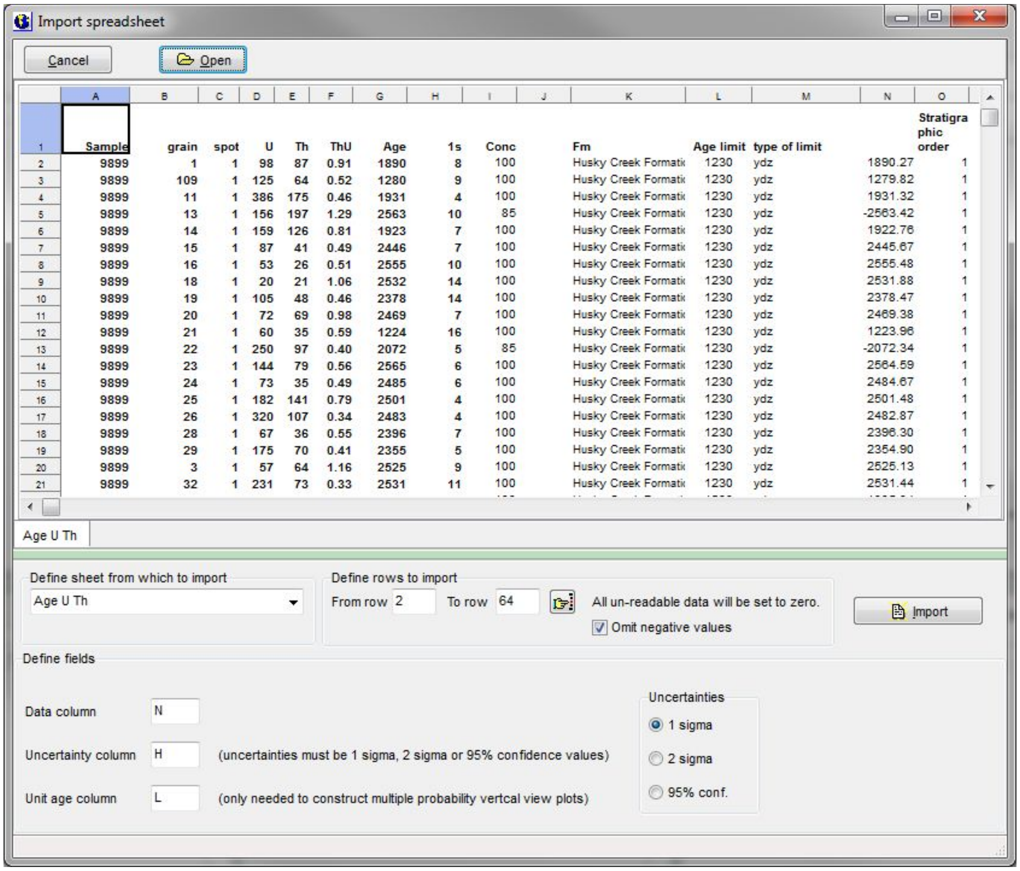
1. Introduction

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, 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 2007-2010 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.



Import spreadsheet

Cancel · Open

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Sample	grain	spot	U	Th	ThU	Age	1s	Conc		Fm	Age limit	type of limit		Stratigraphic order
2	9899	1	1	98	87	0.91	1890	8	100		Husky Creek Formatic	1230	ydz	1890.27	1
3	9899	109	1	125	64	0.52	1280	9	100		Husky Creek Formatic	1230	ydz	1279.82	1
4	9899	11	1	386	175	0.46	1931	4	100		Husky Creek Formatic	1230	ydz	1931.32	1
5	9899	13	1	156	197	1.29	2563	10	85		Husky Creek Formatic	1230	ydz	-2563.42	1
6	9899	14	1	159	126	0.81	1923	7	100		Husky Creek Formatic	1230	ydz	1922.76	1
7	9899	15	1	87	41	0.49	2446	7	100		Husky Creek Formatic	1230	ydz	2445.67	1
8	9899	16	1	53	26	0.51	2555	10	100		Husky Creek Formatic	1230	ydz	2555.48	1
9	9899	18	1	20	21	1.06	2532	14	100		Husky Creek Formatic	1230	ydz	2531.88	1
10	9899	19	1	105	48	0.46	2378	14	100		Husky Creek Formatic	1230	ydz	2378.47	1
11	9899	20	1	72	69	0.98	2469	7	100		Husky Creek Formatic	1230	ydz	2469.38	1
12	9899	21	1	60	35	0.59	1224	16	100		Husky Creek Formatic	1230	ydz	1223.96	1
13	9899	22	1	250	97	0.40	2072	5	85		Husky Creek Formatic	1230	ydz	-2072.34	1
14	9899	23	1	144	79	0.56	2565	6	100		Husky Creek Formatic	1230	ydz	2564.59	1
15	9899	24	1	73	35	0.49	2485	6	100		Husky Creek Formatic	1230	ydz	2484.67	1
16	9899	25	1	182	141	0.79	2501	4	100		Husky Creek Formatic	1230	ydz	2501.48	1
17	9899	26	1	320	107	0.34	2483	4	100		Husky Creek Formatic	1230	ydz	2482.87	1
18	9899	28	1	67	36	0.55	2396	7	100		Husky Creek Formatic	1230	ydz	2396.30	1
19	9899	29	1	175	70	0.41	2355	5	100		Husky Creek Formatic	1230	ydz	2354.90	1
20	9899	3	1	57	64	1.16	2525	9	100		Husky Creek Formatic	1230	ydz	2525.13	1
21	9899	32	1	231	73	0.33	2531	11	100		Husky Creek Formatic	1230	ydz	2531.44	1

Age U Th

Define sheet from which to import: Age U Th

Define rows to import: From row 2 To row 64

All un-readable data will be set to zero.

☑ Omit negative values

Import

Define fields

Data column: N

Uncertainty column: H (uncertainties must be 1 sigma, 2 sigma or 95% confidence values)

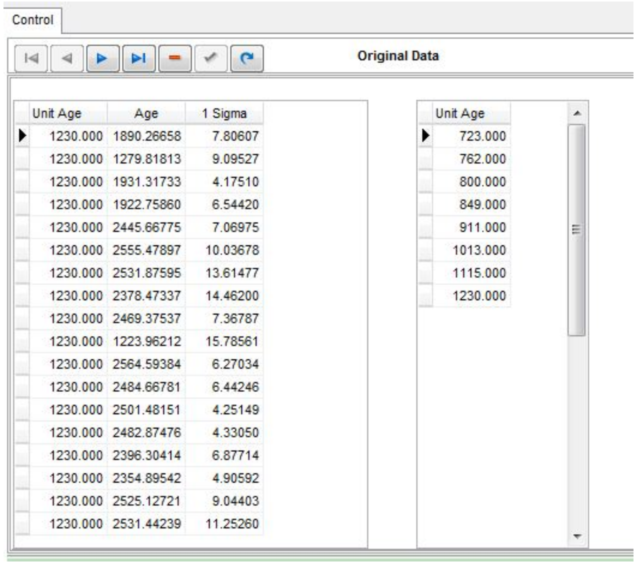
Unit age column: L (only needed to construct multiple probability vertcal view plots)

Uncertainties: ● 1 sigma ○ 2 sigma ○ 95% conf.

Figure 1. Example dataset. Test data illustrated in this dataset were provided by Nicole Rayner and Rob Rainbird (Geological Survey of Canada).

For the example dataset illustrated in Figure **1**, rows 2 to 674 of the spreadsheet will be imported, taking values from columns N (age), H (uncertainty) and assuming an age for the unit/sample as provided in column L. Uncertainties are 1 sigma. All negative values in the data column (column N in the example) will be omitted.

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.



Unit Age	Age	1 Sigma
1230.000	1890.26658	7.80607
1230.000	1279.81813	9.09527
1230.000	1931.31733	4.17510
1230.000	1922.75860	6.54420
1230.000	2445.66775	7.06975
1230.000	2555.47897	10.03678
1230.000	2531.87595	13.61477
1230.000	2378.47337	14.46200
1230.000	2469.37537	7.36787
1230.000	1223.96212	15.78561
1230.000	2564.59384	6.27034
1230.000	2484.66781	6.44246
1230.000	2501.48151	4.25149
1230.000	2482.87476	4.33050
1230.000	2396.30414	6.87714
1230.000	2354.89542	4.90592
1230.000	2525.12721	9.04403
1230.000	2531.44239	11.25260

Unit Age
723.000
762.000
800.000
849.000
911.000
1013.000
1115.000
1230.000

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).

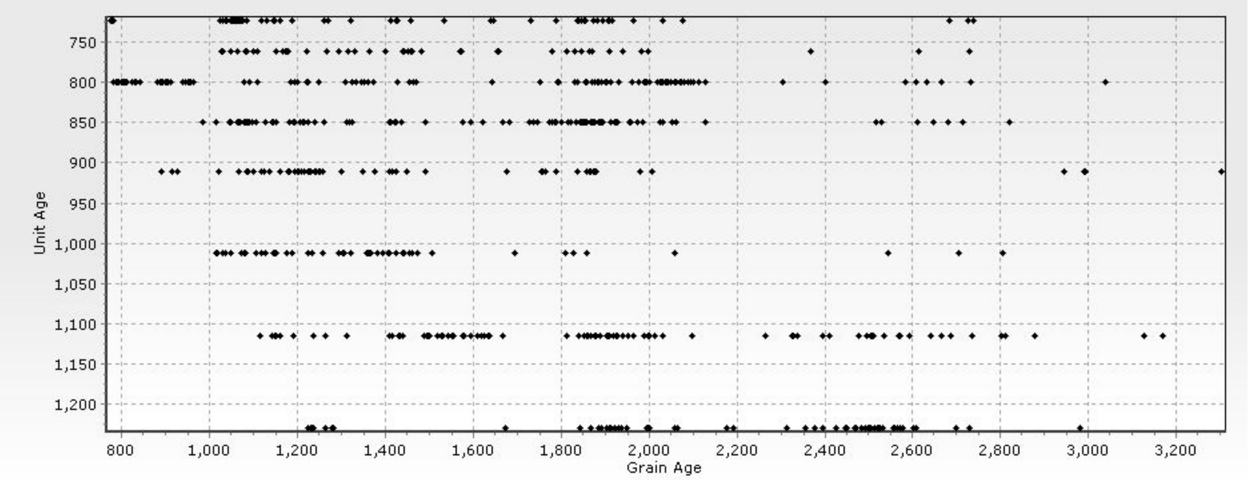
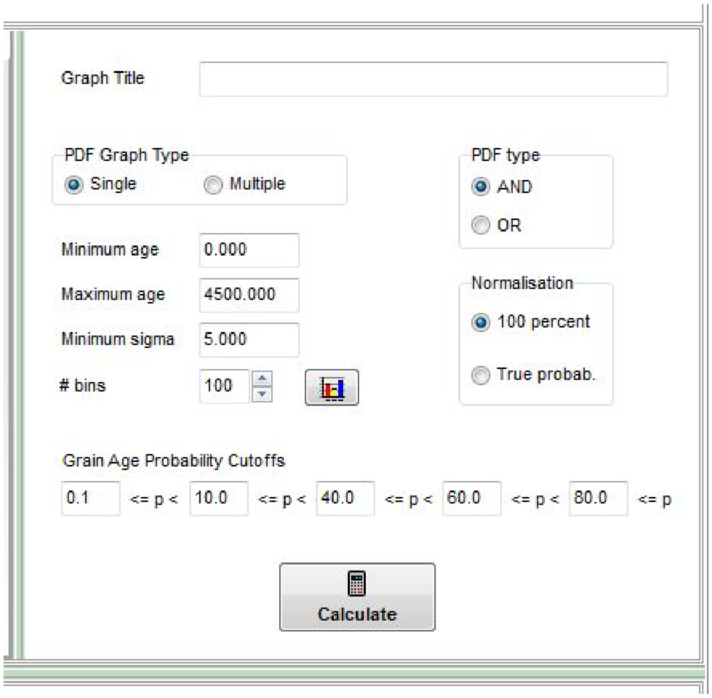


Figure 3. Graphical illustration of the input data plotted relative to Unit (or Sample) age.

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.



Graph Title

PDF Graph Type: ● Single ○ Multiple

PDF type: ● AND ○ OR

Minimum age: 0.000

Maximum age: 4500.000

Minimum sigma: 5.000

bins: 100

Normalisation: ● 100 percent ○ True probab.

Grain Age Probability Cutoffs: 0.1 <= p < 10.0 <= p < 40.0 <= p < 60.0 <= p < 80.0 <= p

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.

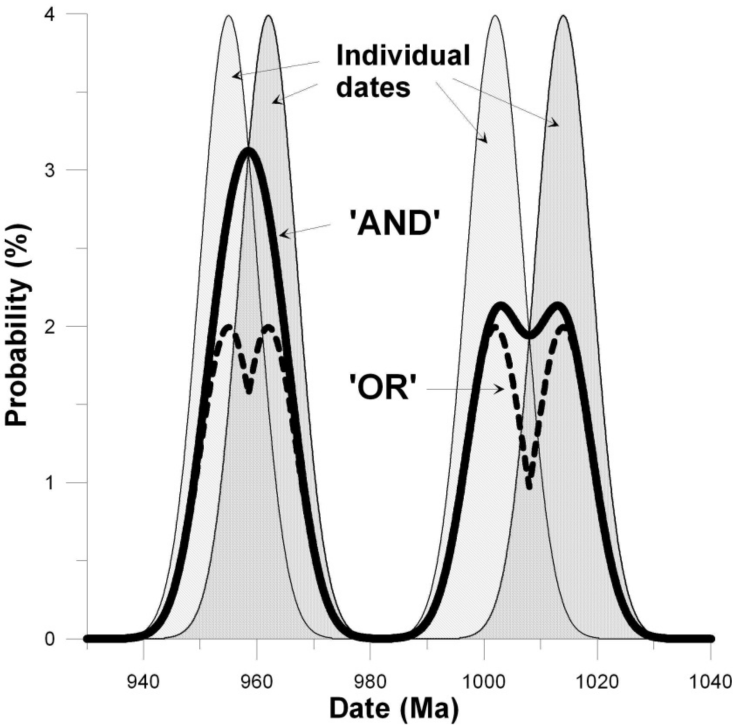


Figure 5. Gaussian ‘AND’ (solid thick line) and ‘OR’ (dashed thick line) representation of a simple test dataset comprising four analyses (solid thin line). Note the difference in peak height between the two approaches. Probabilities are presented as ‘true probabilities’ i.e. without normalisation to 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.

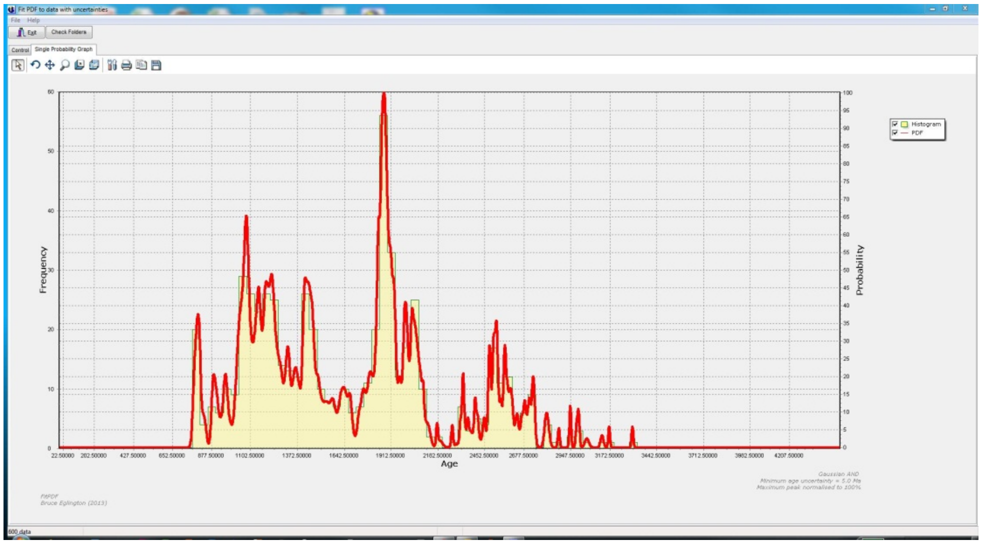


Figure 6. Gaussian ‘AND’ representation of example data (red curve) overlaid on a histogram binned according to the settings in Figure 4. Minimum age uncertainty was set as 5 Ma.

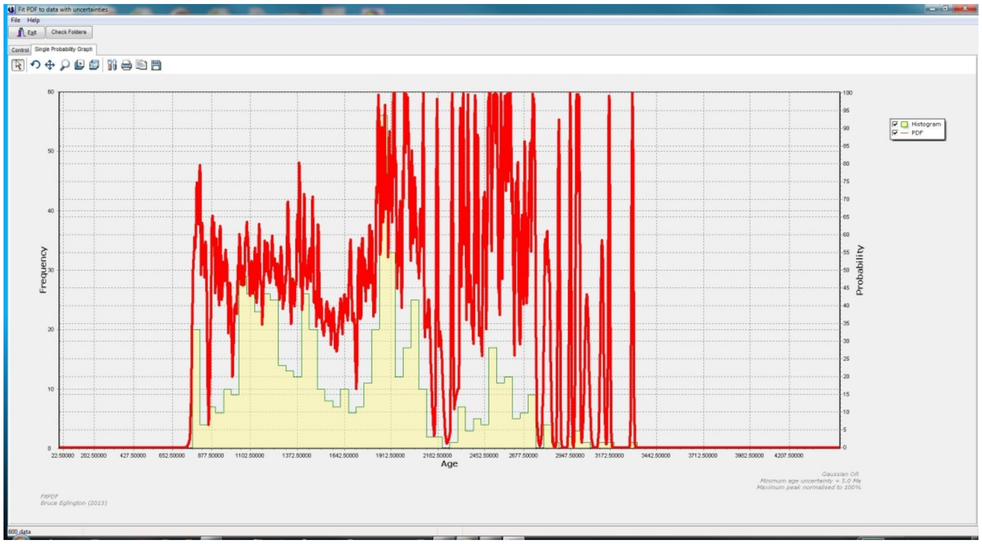


Figure 7. Gaussian ‘OR’ representation of example data (red curve) overlaid on a histogram binned according to the settings in Figure 4. Minimum age uncertainty was set as 5 Ma.

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.

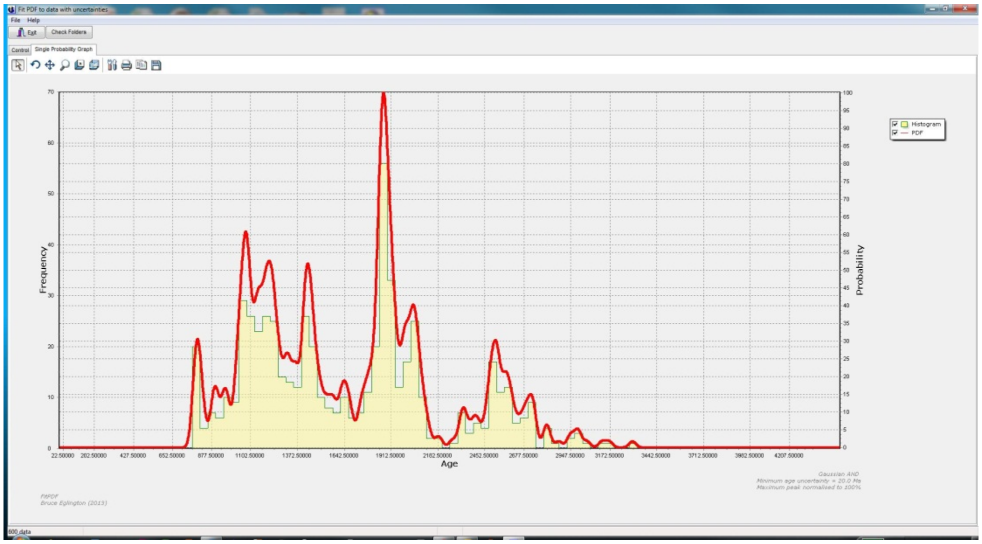


Figure 8. Gaussian ‘AND’ representation of example data (red curve) overlaid on a histogram binned according to the settings in Figure 4. Minimum age uncertainty was set as 20 Ma.

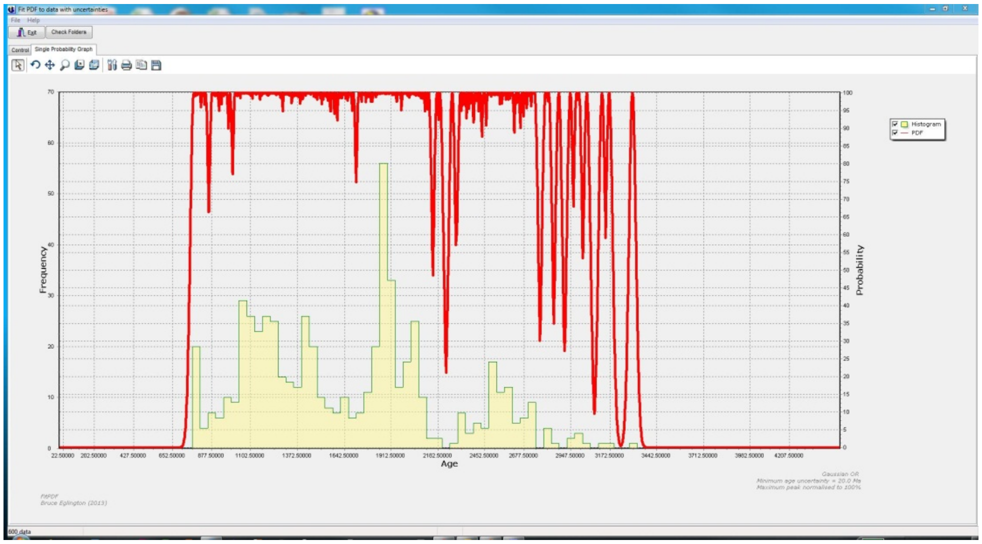


Figure 9. Gaussian ‘OR’ representation of example data (red curve) overlaid on a histogram binned according to the settings in Figure 4. Minimum age uncertainty was set as 20 Ma.

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.

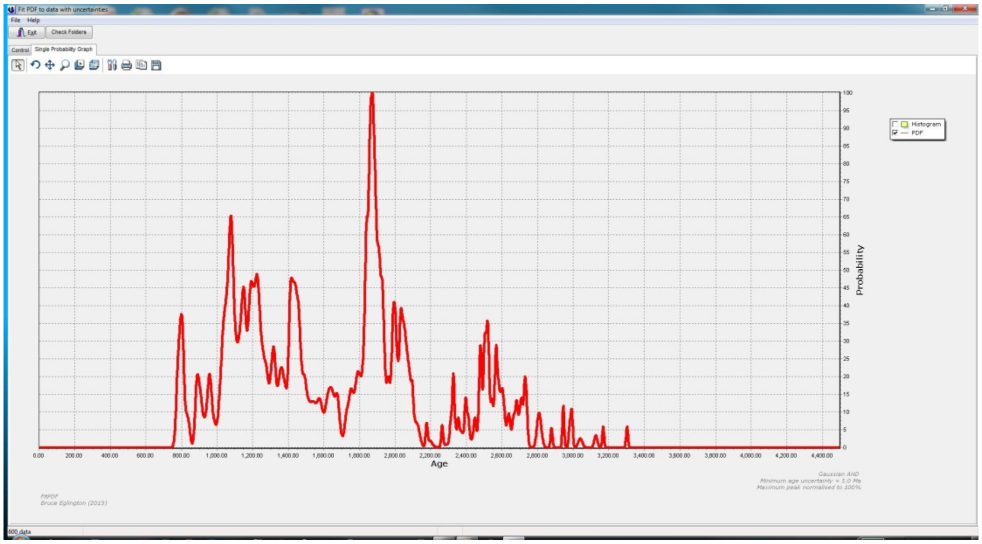


Figure 10. Gaussian ‘AND’ representation of example data (red curve) overlaid on a histogram binned according to the settings in Figure 4. Minimum age uncertainty was set as 5 Ma. Histogram values were switched off by unchecking the appropriate checkbox in the graph Legend.

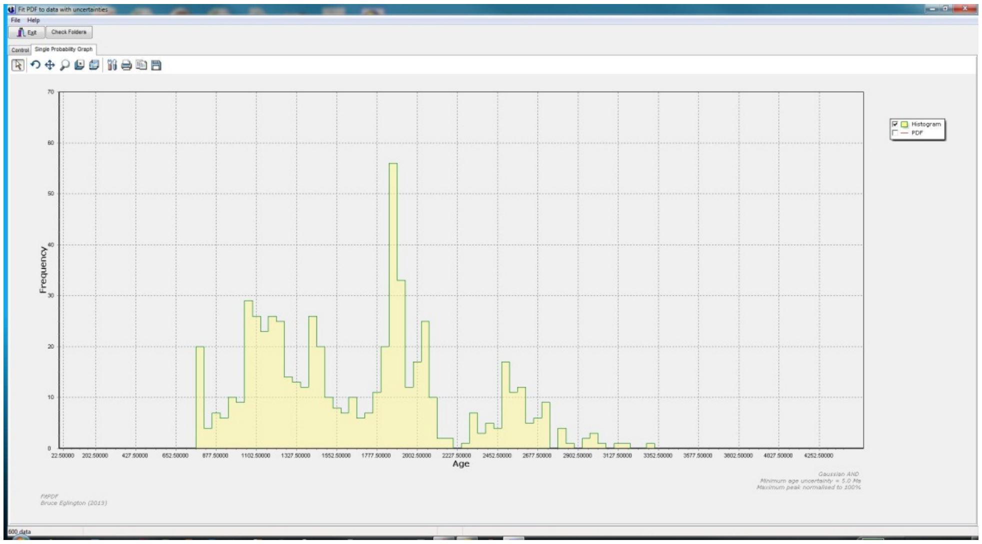


Figure 11. Histogram representation of example data binned according to the settings in Figure 4. Probability curve values were switched off by unchecking the appropriate checkbox in the graph Legend.

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 some 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

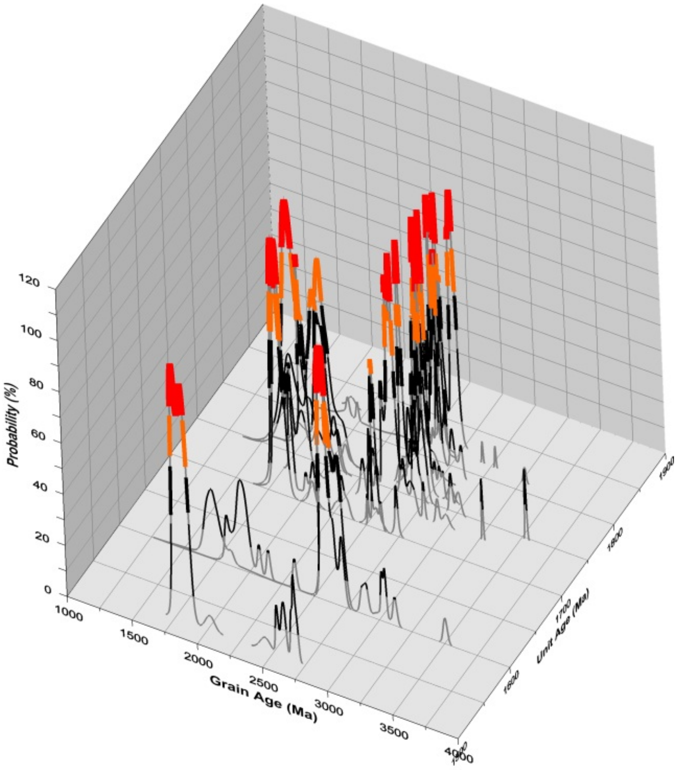
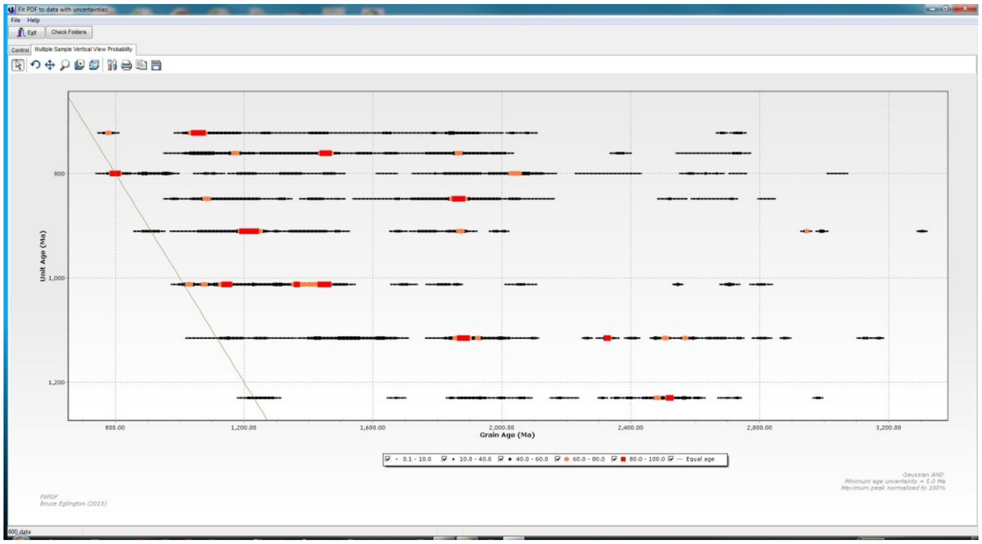
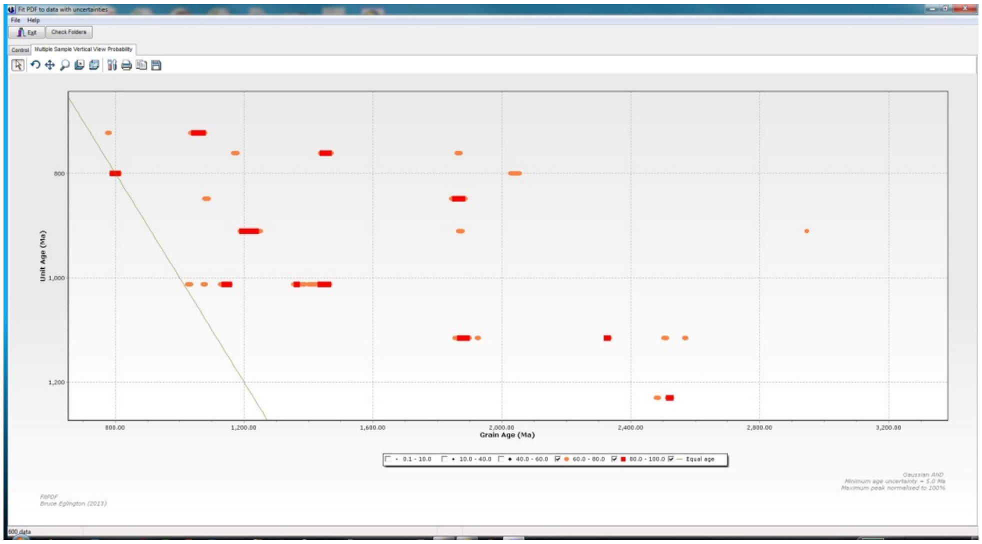


Figure 12. 3D diagram illustrating the layout of probability curves used to produce the ‘vertical view’ plots.



☑ · 0.1 - 10.0 ☑ • 10.0 - 40.0 ☑ ◆ 40.0 - 60.0 ☑ ● 60.0 - 80.0 ☑ ■ 80.0 - 100.0 ☑ — Equal age

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.



□ · 0.1 - 10.0 □ • 10.0 - 40.0 □ ◆ 40.0 - 60.0 ☑ ● 60.0 - 80.0 ☑ ■ 80.0 - 100.0 ☑ — Equal age

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.

6. 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.

7. 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.

8. 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.

9. References

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. 2009. The IGCP 509 and DateView Database Systems: Lessons Learned. American Geophysical Union, Zircon Analysis Workshop

Eglington, B. M., Harmer, R.E. 1991. GEODATE version 2.2: a program for the processing and regression of isotope data using IBM-compatible microcomputers. CSIR manual EMA-H 9102, 1-70.

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., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P. 2002. Numerical recipes in C++ : the art of scientific computing. 2nd ed., Cambridge University Press, Cambridge, 1004 pp.

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.