PyQtDZMix V1.0
User Manual

Fangbin Liu
(liufangbin8908@163.com)
Qilu Normal University

October 2025

Content

Indroduction

PyQtDZMix is a comprehensive and user-friendly software tool for zircon U–Pb dating and provenance analysis. Developed in Python with the PyQt5 framework, it provides an intuitive graphical user interface (GUI) that streamlines complex geochronological data workflows.

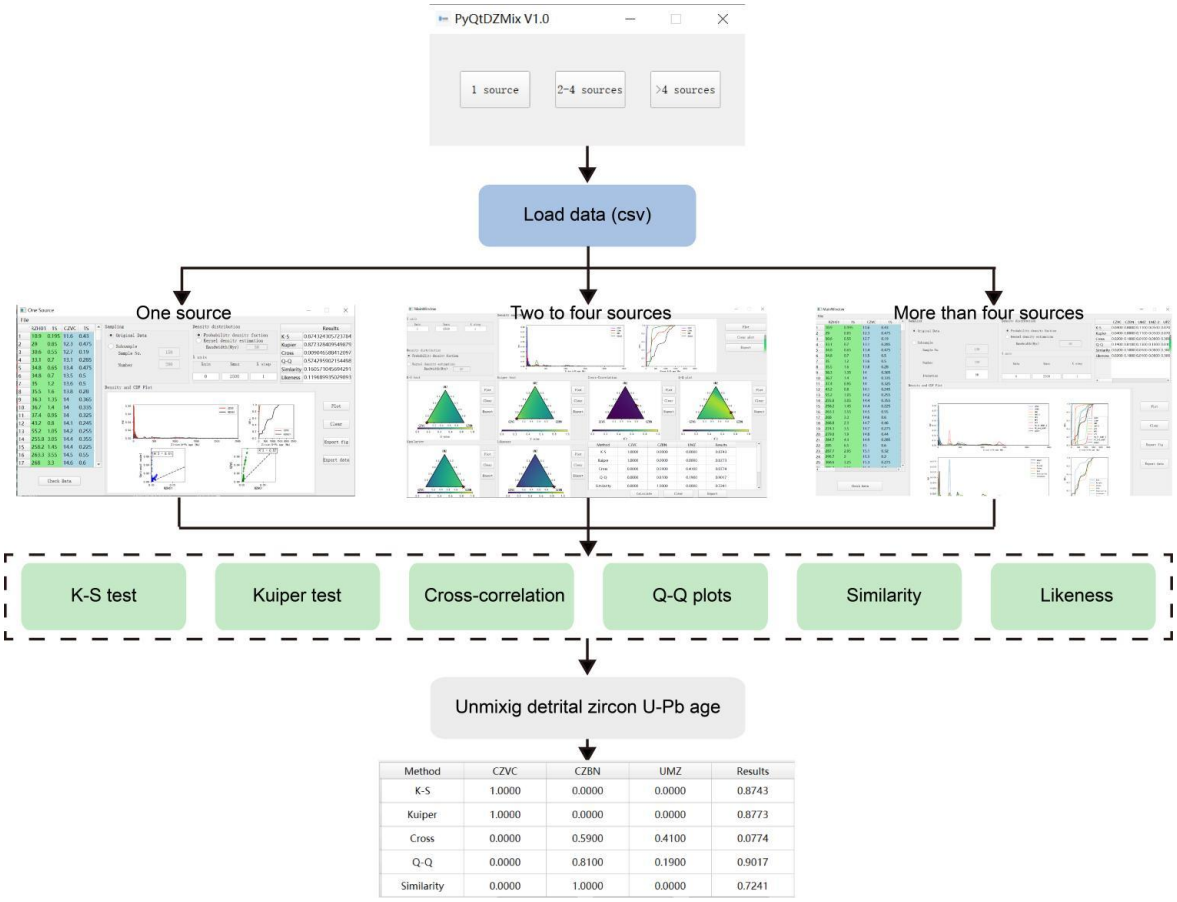
The software integrates key functionalities—data validation, statistical analysis, and dynamic visualization—addressing the increasing complexity of zircon datasets. Its GUI-based design makes it accessible to users with little or no programming experience, offering a visually guided workflow for zircon U–Pb data analysis involving one to multiple source regions (Figure 1).

PyQtDZMix incorporates a wide range of advanced statistical tools, including the Kolmogorov–Smirnov (K–S) test, Kuiper test, Q–Q plots, cross-correlation, and similarity and likeness coefficients, enabling rigorous comparison of zircon age distributions and quantitative assessment of sample similarity. It also supports mixture-based provenance analysis using Kernel Density Estimates (KDEs), allowing accurate identification of sediment sources and estimation of contributions from different regions. The software further offers flexibility through both random sampling and non-sampling options, adaptable to datasets of varying size and complexity.

Additional features include simple data import (e.g., CSV format), real-time interactive 2D/3D visualizations, customizable plotting parameters, and versatile export options (JPG, PDF, and data files) for reporting and further analysis.

By combining analytical power with an intuitive interface, PyQtDZMix lowers the technical barrier for researchers and enhances the efficiency, flexibility, and accuracy of zircon U–Pb

geochronological and provenance studies.



Method	CZVC	CZBN	UMZ	Results
K-S	1.0000	0.0000	0.0000	0.8743
Kuiper	1.0000	0.0000	0.0000	0.8773
Cross	0.0000	0.5900	0.4100	0.0774
Q-Q	0.0000	0.8100	0.1900	0.9017
Similarity	0.0000	1.0000	0.0000	0.7241

Figure 1. The general workflow of the PyQtDZMix software

Running

The current PC version was compiled and executed on Windows 11 and is also 64bit only. The software can be launched by simply double-clicking the .exe file. No additional configuration is required. In addition, you can also run the 'PyDesktop.py', but you must install some libraries, like PyQt5, NumPy, SciPy, Matplotlib, Pandas.

Data Import

Data should be organized into a single comma-delimited .csv file with header rows. In each

file, each sample is represented by a pair of columns, where the first two columns correspond to the grain mean ages and the associated uncertainties (1σ), respectively. See the Example Data Set for an example of a correctly formatted sample.

EXAMPLE: To compare one source, the import file should contain four columns with columns 1 (Sample) and 3 (Source), containing the mean grain ages, and columns 2 and 4 containing the grain uncertainty (Figure 2). Sample would then occupy columns 1 and 2, source would occupy columns 3 and 4. If there are more than one source (N⩾2 sources), the import file should contain eight columns with columns 1, 3, 5, and 7 containing the mean grain ages, and columns 2, 4, 6, and 8 containing the grain uncertainty. Among them, the first two columns represent the zircon mean ages and one sigma uncertainty of the sample, respectively. The remaining odd-numbered columns represent the zircon ages of the sources, even followed by its corresponding even-numbered uncertainty column.

A	B	C	D	E	F	G	H
RMDZ02	1S	CZVC	1S	CZBN	1S	UMZ	1S
9.4	0.335	11.6	0.43	11.6	0.43	105.6	2.15
14	0.365	12.3	0.475	12.3	0.475	106.2	2.1
19.2	0.6	12.7	0.19	12.7	0.19	106.8	1.8
45.1	0.75	13.1	0.285	13.1	0.285	107.6	2.9
142.6	2.15	13.4	0.475	13.4	0.475	108.8	2.05
173.2	2.25	13.5	0.5	13.5	0.5	110.1	3.95
235.1	3.75	13.6	0.5	13.6	0.5	117.4	2.8
236.3	2.55	13.8	0.28	13.8	0.28	118.6	3.25
237	3.1	14	0.365	14	0.365	119.2	3.5
238	5.5	14	0.335	14	0.335	122.1	3.35
238.7	2.15	14	0.325	14	0.325	123.2	2.15
241.9	3.5	14.1	0.245	14.1	0.245	128.1	4.65
243.3	3.05	14.2	0.255	14.2	0.255	131.8	3.45

Figure 2. Data input template

GUI

The main GUI includes three parts, such as one source, two to four sources, and more than four

sources (Figure 3).

Figure 3. Main Window

➢ One source

This part focuses on assessing the similarity between the sample and individual source regions. Click it, a sub-interface will pop up, containing a menu bar, a data panel, a parameter panel, a visualization panel, and a results panel for displaying the calculated outputs (Figure 4).

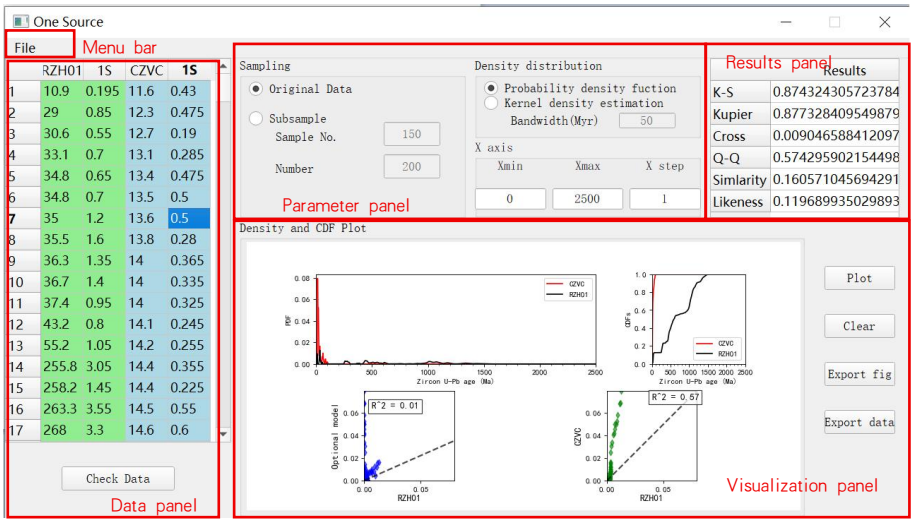


Figure 4. Window of the one source

✧ Menu bar

Click the **File** (Figure 5), the menu option that will appear is shown below.

New: If you click this, the data panel will clear all data and restore default settings.

Open: Select this, we can import the data in the data panel.

Save as: We can save the data updated in the data panel.

Exit: Exit this interface.

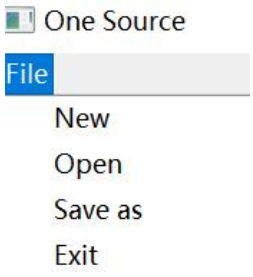


Figure 5. the menu bar

✧ Data panel

If you execute the **"Open"** in the menu bar, you can import the data (.csv) in the data panel. In this part, the table show two colors, the first two columns are grass green (sample), and the other are light blue (source). Click **"Check Data"** button. If the input data pass the validation check, the program automatically displays a dialog box (All column pairs have matching number counts; Figure 6). In contrast (non-numeric entries or NaN values), the program automatically will display a mismatching dialog box (Figure 7). Once all data meet the specified requirements, the program proceeds to the next step — parameter selection.

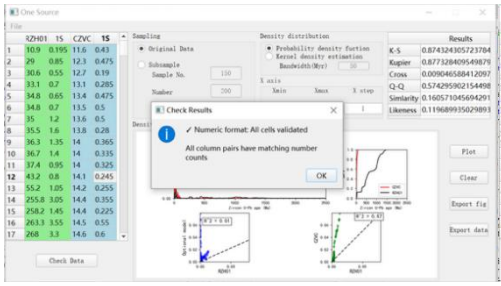


Figure 6. Check data with full data.

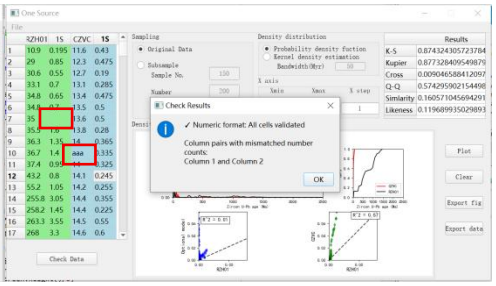


Figure 7. Check data with wrong data.

✧ parameter panel

The parameter panel have three parts, including **Sampling** and **Density distribution** and **X axis** (Figure 4).

Sampling: Original data and Subsample. Select **Subsample,** the **Sample No.** and

Number (Text) will be activated. The default value of **Sample No.** is 150, meaning that a subset is generated by randomly selecting 150 ages from the original dataset. The default value of **Number** is 200, meaning that a total of 200 subsets (including 150 ages) is randomly sampled from the dataset for statistical analysis.

Density distribution: Probability density function and Kernel density estimation. This default option will calculate probability density function based on the uncertainties in the import data file. If you select the Kernel density estimation (KDEs), it will calculate KDEs with the bandwidth (default value is 50).

X axis: Specify the minimum **(Xmin, default is 0)** and maximum **(Xmax, default is 2500)** limits of the x-axis, as well as the step size **(X step, default is 1)** for the calculation.

✧ Visualization panel and results panel

This panel mainly show the plots, including PDFs, KDEs, cumulative distribution functions (CDFs), Cross and Q-Q (Figure 4).

Click **"Plot"** button, all plot will be displayed based on the parameter options and the results in the **results panel** will also be shown, including Kolmogorov-Smirnov (K-S) test, Kuiper test, cross-correlation, Q-Q plots, similarity and likeness coefficients.

Click **"Clear"** button, all plot and results will be cleared.

Click **"Export fig"** button, the plot will be saved as jpg or pdf format.

Click **"Export data"** button, the results in the **results panel** will be stored as .csv file.

➢ Two to four sources

This part focuses on checking the data, including the menu bar, a data panel, and the button panel (Figure 8). The functionalities of the first two modules are consistent with those of the

single-source analysis. The button panel has "**Check data", "Subsample", and "Original Data"** buttons. Among them, the "**Check data"** button is the similar with that in the one source. Click the "**Check data"**, if the data is true, the **"Subsample"**, and **"Original Data"** buttons will be activated.

✧ Subsample

Click **"Subsample"** button, a new window will be shown. This window contains a parameter panel, a "Density and CDF Plot" panel, and six statistical analysis plotting panels and results panel for displaying the calculated outputs (Figure 9). Except statistical analysis plotting panels, others are similar with that in one source.

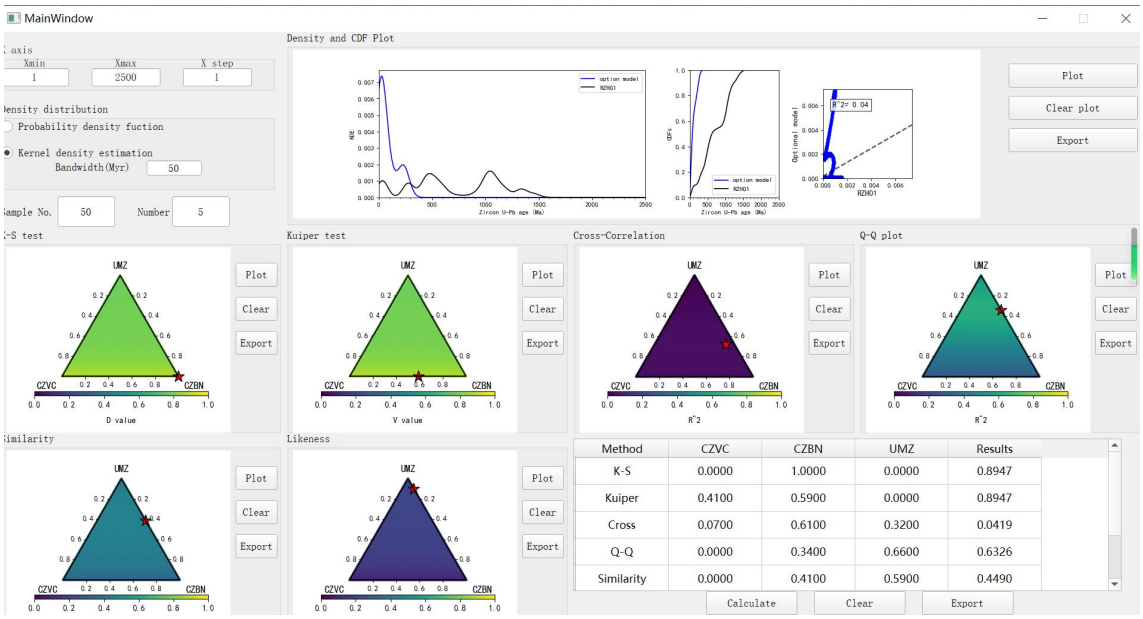


Figure 9. The window of the two to four sources

When you select the parameters that you need, you can click the **"Plot"** button in the **"Density and CDF Plot"** panel. The plots will be shown. Then click **"Plot"** button in the six statistical analysis plotting panels, the 2D or 3D quantitative source plot will be displayed. Meanwhile, the synthetic density and CDF plot, which is based on the different statistical analysis, will be shown in the **"Density and CDF Plot"** panel. The software enables the

construction of band plots (two sources), ternary diagrams (three sources) and quaternary (tetrahedral) diagrams (four sources) to illustrate source contributions under varying quantitative proportions (Figure 10). Click **"Clear"** button, all plot will be cleared. Click **"Export"** button, the plot will be stored as jpg or pdf file.

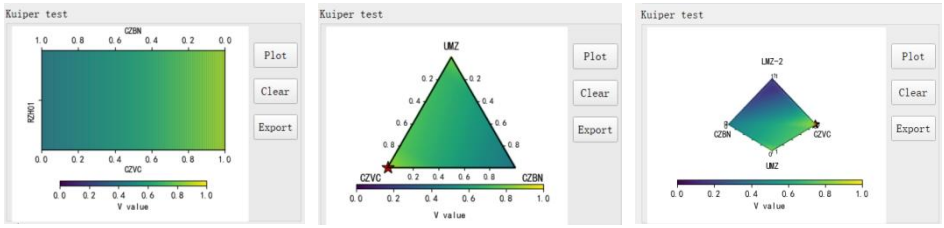


Figure 10. Band plots (two sources), ternary diagrams (three sources) and quaternary (tetrahedral) diagrams (four sources)

Once all the plots have been shown, click **"Calculate"** button, the results in the **results panel** will also be shown, including Kolmogorov-Smirnov (K-S) test, Kuiper test, cross-correlation, Q-Q plots, similarity and likeness coefficients (Figure 11).

Method	CZVC	CZBN	UMZ	Results
K-S	0.0000	1.0000	0.0000	0.8947
Kuiper	0.4100	0.5900	0.0000	0.8947
Cross	0.0700	0.6100	0.3200	0.0419
Q-Q	0.0000	0.3400	0.6600	0.6326
Similarity	0.0000	0.4100	0.5900	0.4490

Calculate Clear Export

Figure 11. The results of Kolmogorov-Smirnov (K-S) test, Kuiper test, cross-correlation, Q-Q plots, similarity and likeness coefficients

Click **"Clear"** button, all plot and results will be cleared.

Click **"Export"** button, the results in the **results panel** will be stored as .csv file.

✧ Original Data

Click **"Original Data"** button, a new window will be shown. This window (including all functions) is nearly identical to the resampling window (**Subsample**), except that the resampling

parameter section is omitted.

➢ More than four sources

This window is almost identical to the one source (Figure 4), the difference is that this one includes an additional iteration option in the **Sampling** of the **parameter panel** (Figure 12). This is because when the number of source regions is too large, it becomes impractical to perform quantitative calculations for each source individually. Therefore, an iterative approach is adopted to prevent the program from crashing due to excessive computational load. In most cases, 10,000 iterations are sufficient, but users can adjust this number as needed. The default value is set to 50.

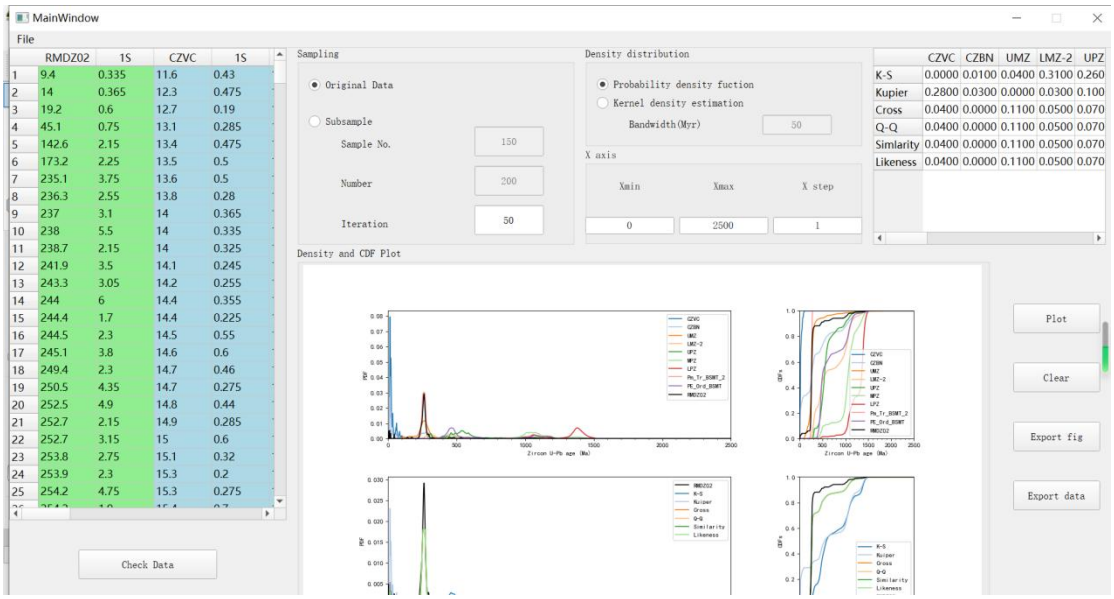


Figure 12. Window of the more than four sources