

Quantitative analysis, visualization, and modelling of detrital geochronology data

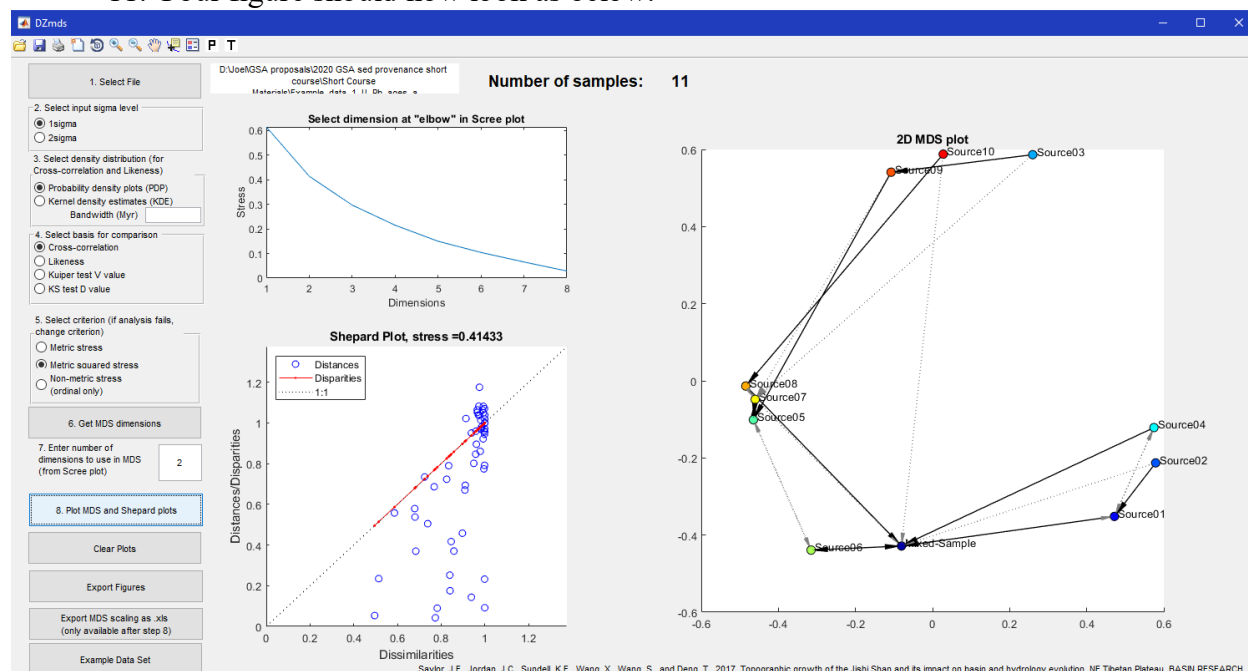
3. APPLICATION 2: DZMDS

Learning goals:

- How to implement metric and non-metric MDS in 2 and 3 dimensions using DZmds
- How to identify high- and low-quality MDS transformations
- How to interpret MDS plots

3.1. 2D metric MDS with squared stress of Cross-correlation

1. Open DZmds by double clicking on the DZmds icon in the folder where you saved the application.
2. Click on “Select File” and navigate to where you saved the example data, “Application_2_U_Pb_ages_and_uncertainties.xlsx”.
3. Select “2sigma” and the input uncertainty level.
4. Select “PDP” as the density distribution
5. Select “Cross-correlation” as the basis for comparison
6. Select “Metric squared stress” as the criterion
7. Click “Get Dimensions”
8. Note the stress versus dimensions relationship in the scree plot. More than 3 dimensions are not particularly useful, because we want to display this data in 2 or 3 dimensions. Given the high stress, we can expect a fairly poor fit in the Shepard plot.
9. Input “2” in the number of dimensions to use in MDS box.
10. Select “Plot MDS and Shepard plots”
11. Your figure should now look as below.

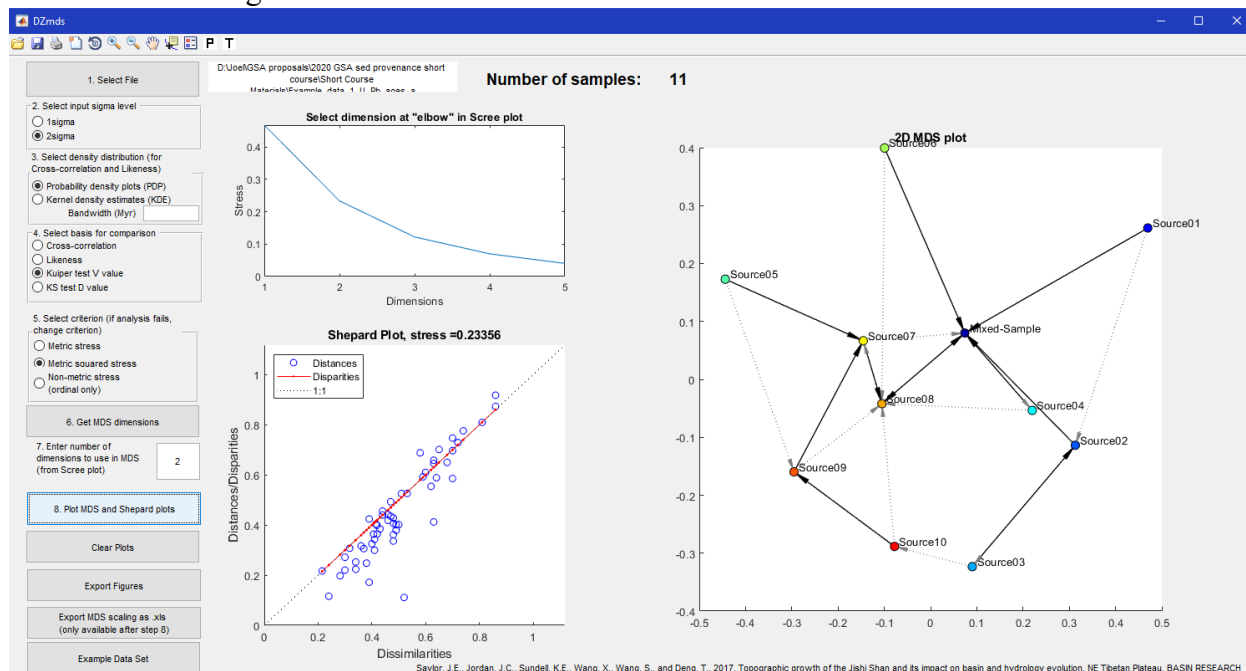


3.2. 2D metric MDS with squared stress of Kuiper V value

1. That is clearly not the MDS that we want to use for this data set. Let's try again.
2. Select “Kuiper test V value” for the basis for comparison

Quantitative analysis, visualization, and modelling of detrital geochronology data

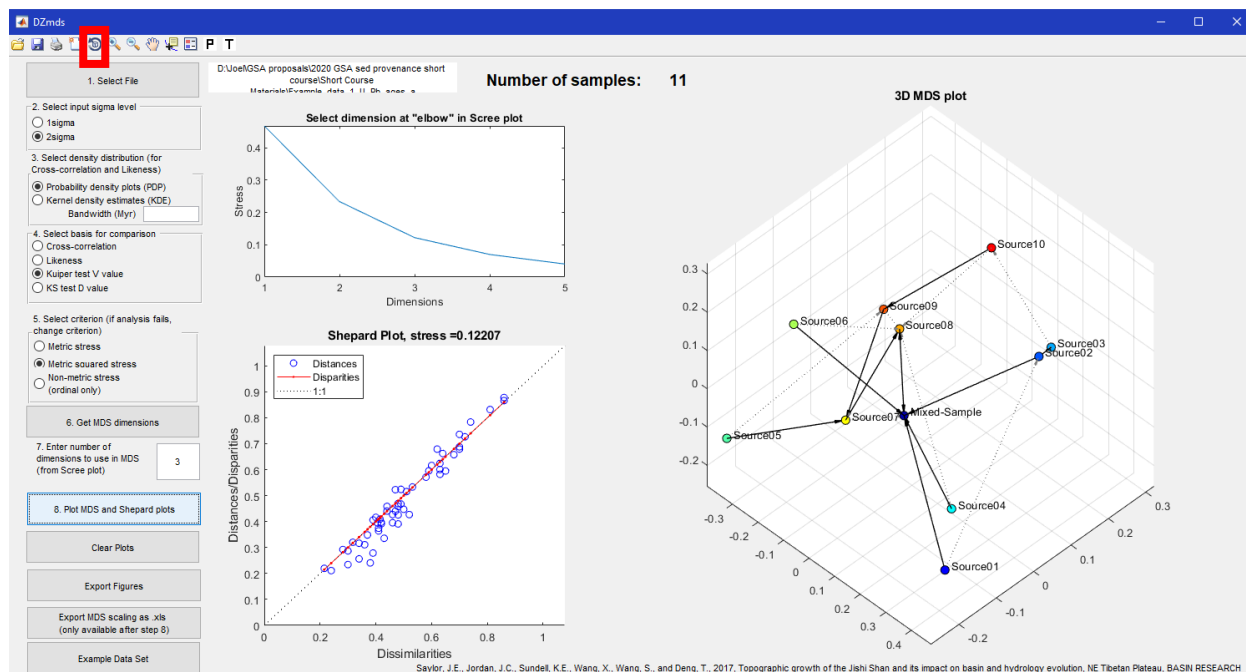
3. Leave the criterion as “Metric squared stress”
4. Select “Get MDS dimensions”
5. Note how much lower the stress is for 2 or 3 dimensions compared to the previous example.
6. Input “2” in the number of dimensions to use in MDS box.
7. Select “Plot MDS and Shepard plots”
8. Your figure should now look as below.



3.3. 3D metric MDS with squared stress of Kuiper V value

1. What if we plot it in 3 dimensions?
2. Input “3” in the number of dimensions to use in MDS box.
3. Select “Plot MDS and Shepard plots”
4. Your figure should now look as below.

Quantitative analysis, visualization, and modelling of detrital geochronology data



5. Select the “Rotate 3D” tool (the icon showing a cube with a circular arrow around it, in the upper left corner). See the red box above).
6. Now you should be able to rotate the MDS plot to examine the relationships between samples.

3.4. Saving figures and MDS

1. Since this is an acceptable stress, let’s save this run.
2. Select “Export Figures”
3. Two figures will pop up showing the Shepard and MDS plots. These figures can be saved in a variety of formats. We will focus on vector formats.
4. 2D plots can be saved in an editable vector format (compatible with Adobe Illustrator for example) by opening the “File” menu on the figure, selecting “Save as”, and saving it as a “.eps” figure. This figure can be opened in Illustrator or other vector art programs.
5. For 3D plots, open the “File” menu as above. Select “Export Setup”. Under “Properties”, highlight “Rendering”. Ensure that the “Custom renderer” box is checked and set the Custom Renderer to “painters (vector format)”. Select “Apply to Figure” and then “Export”. Save as a “.eps” file as before. This file can be opened by Illustrator or other vector art software.
6. “Export MDS scaling” allows you to export the XY coordinates (or XYZ for 3D MDS) to an Excel spreadsheet. This can then be plotted in the program of your choice.

Quantitative analysis, visualization, and modelling of detrital geochronology data

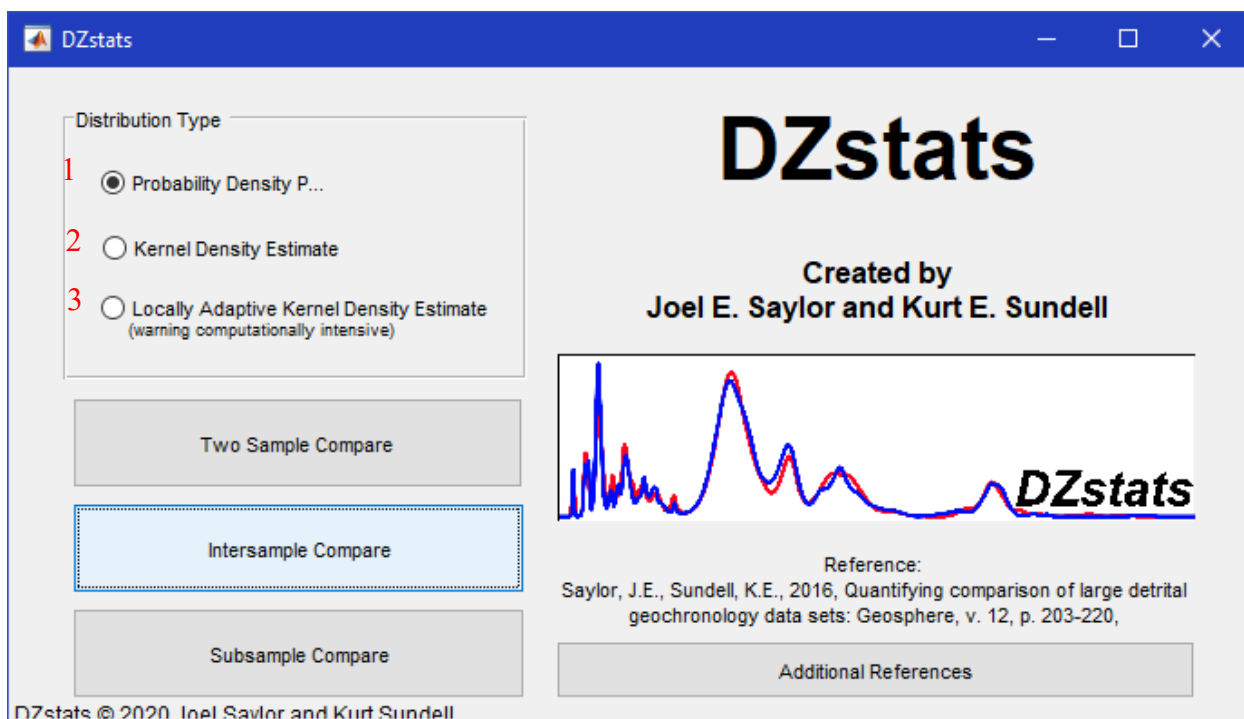
4. APPLICATION 3: DZSTATS

Learning goals:

- Understand how to compare a sample set using DZstats
- Understand the difference between PDPs and KDEs and how to analyze each using DZstats
- How and why to implement bootstrapping (repeated subsampling and comparison) using DZstats
- The dangers of relying on p values for data interpretation

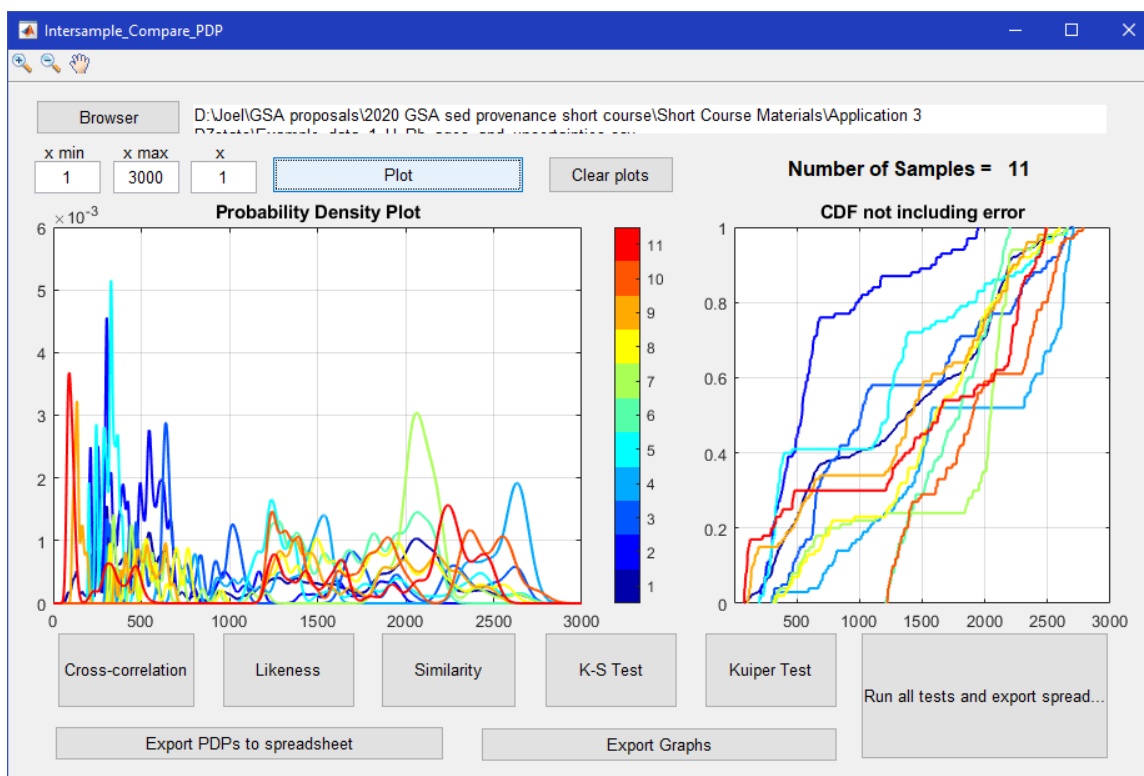
4.1. Comparison using Intersample Compare of PDPs

DZstats has built-in options for sample comparison using PDPs (1, below), adaptive bandwidth KDEs (2, below), or locally adaptive bandwidth (3, below). We will start by comparing 11 samples based on their PDPs and CDFs.

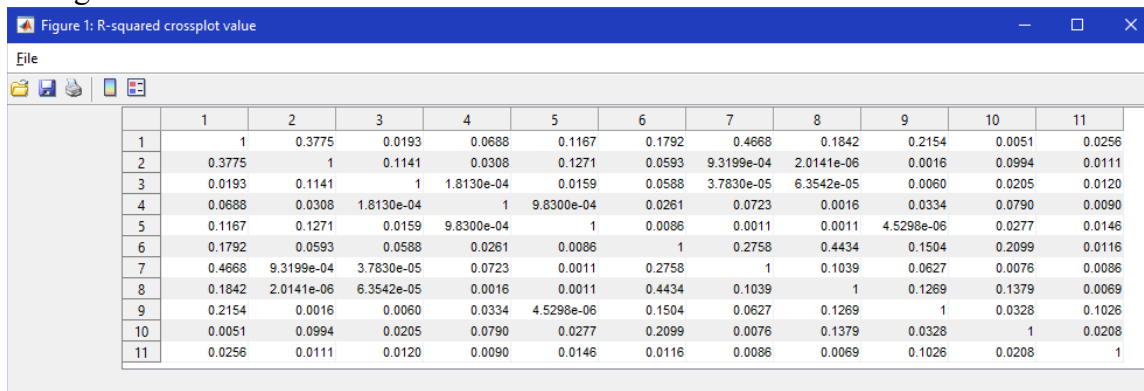


1. Open DZstats by double clicking the icon in in the folder where you saved the application.
2. In the DZstats opening window, select “Probability Density Plot” as the distribution type, and then click on “Intersample Compare”.
3. In the Intersample Compare window, select “Browser” and navigate to “Application_3_U_Pb_ages_and_uncertainties.csv”.
4. Set the x_{min} to 0 Ma, the x_{max} to 3000 Ma, and the x to 1 Myr.
5. Select “Plot”. Your window should look as below.

Quantitative analysis, visualization, and modelling of detrital geochronology data



6. You can now examine the dissimilarity matrix for each of the metrics in buttons below the plots. If you select “Cross-correlation” a new window will open which looks like the figure below.

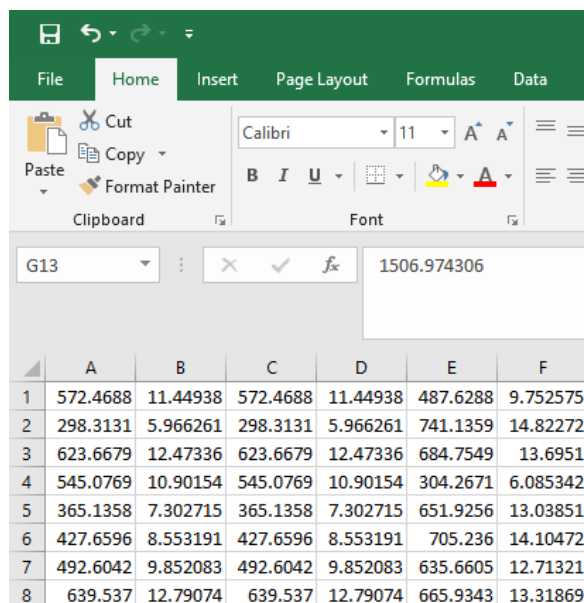


7. You can also run all of the calculations and export the results to a spreadsheet, by selecting the button on the bottom right.
8. This will open a dialog box. Note where you save the file, then navigate to that file and open it. All of the comparisons are saved on sheet 1. The samples are in the same order in the similarity matrix as the input file.

4.2. Comparison using Intersample Compare of KDEs with fixed bandwidths

Quantitative analysis, visualization, and modelling of detrital geochronology data

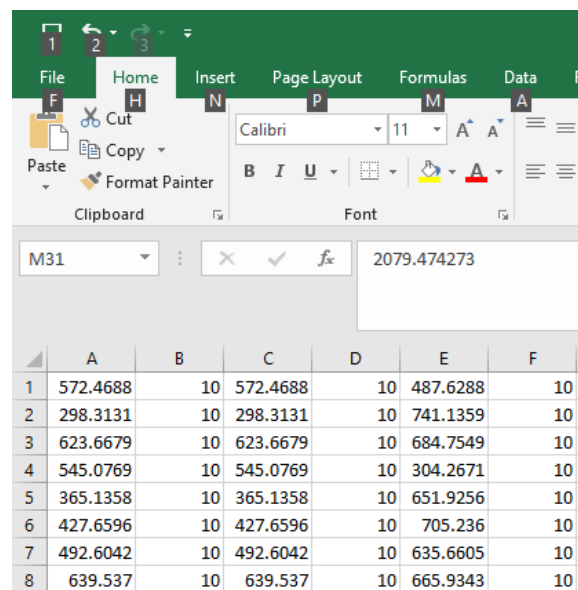
There are times when you might want to compare samples based on their KDEs but with a constant bandwidth, which is applied to all samples.



	A	B	C	D	E	F
1	572.4688	11.44938	572.4688	11.44938	487.6288	9.752575
2	298.3131	5.966261	298.3131	5.966261	741.1359	14.82272
3	623.6679	12.47336	623.6679	12.47336	684.7549	13.6951
4	545.0769	10.90154	545.0769	10.90154	304.2671	6.085342
5	365.1358	7.302715	365.1358	7.302715	651.9256	13.03851
6	427.6596	8.553191	427.6596	8.553191	705.236	14.10472
7	492.6042	9.852083	492.6042	9.852083	635.6605	12.71321
8	639.537	12.79074	639.537	12.79074	665.9343	13.31869

Format of Example_data_1_U_Pb_ages_ and_uncertainties.csv for PDP comparison.

1 σ grain age uncertainties are shown in columns B, D, and F.



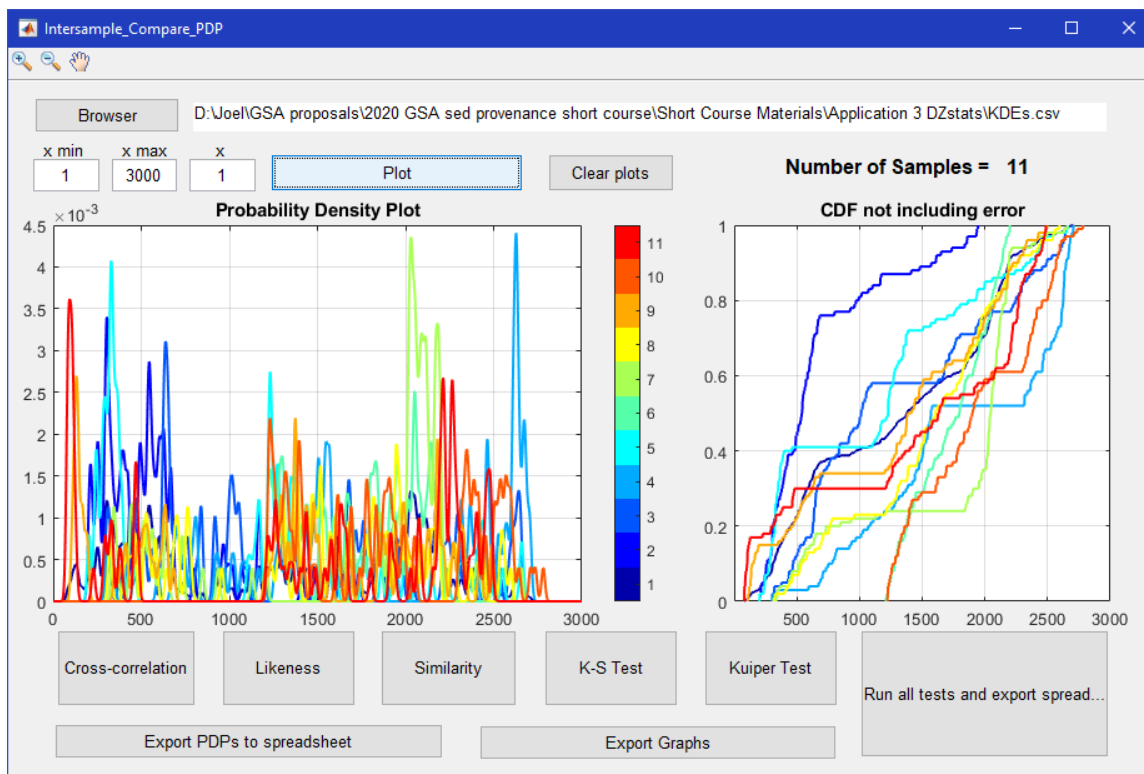
	A	B	C	D	E	F
1	572.4688	10	572.4688	10	487.6288	10
2	298.3131	10	298.3131	10	741.1359	10
3	623.6679	10	623.6679	10	684.7549	10
4	545.0769	10	545.0769	10	304.2671	10
5	365.1358	10	365.1358	10	651.9256	10
6	427.6596	10	427.6596	10	705.236	10
7	492.6042	10	492.6042	10	635.6605	10
8	639.537	10	639.537	10	665.9343	10

Format of Example_data_1_U_Pb_ages_ and_uncertainties.csv for KDE comparison.

1 σ bandwidths (10 Myr) are shown in columns B, D, and F.

- To do this with DZstats, open the file that we used previously (Application_2_U_Pb_ages_and_uncertainties.csv) in Excel or another spreadsheet editing software package. In the file, columns B, D, F, H, J, L, N, P, R, T, and V have the bandwidth for the ages in the preceding column (A, C, etc). Currently the bandwidth is the analytical uncertainty (1 σ) for each age. In order to analyze fixed bandwidth KDEs, replace the uncertainties with a constant value, as shown above.
- Save as a new .csv file.
- In the DZstats opening window, select “Probability Density Plot” as the distribution type, and then click on “Intersample Compare”.
- In the Intersample Compare window, select “Browser” and navigate to your new file and upload it.
- Set the x_min to 0 Ma, the x_max to 3000 Ma, and the x to 1 Myr.
- Select “Plot”. Your window should look as below.

Quantitative analysis, visualization, and modelling of detrital geochronology data



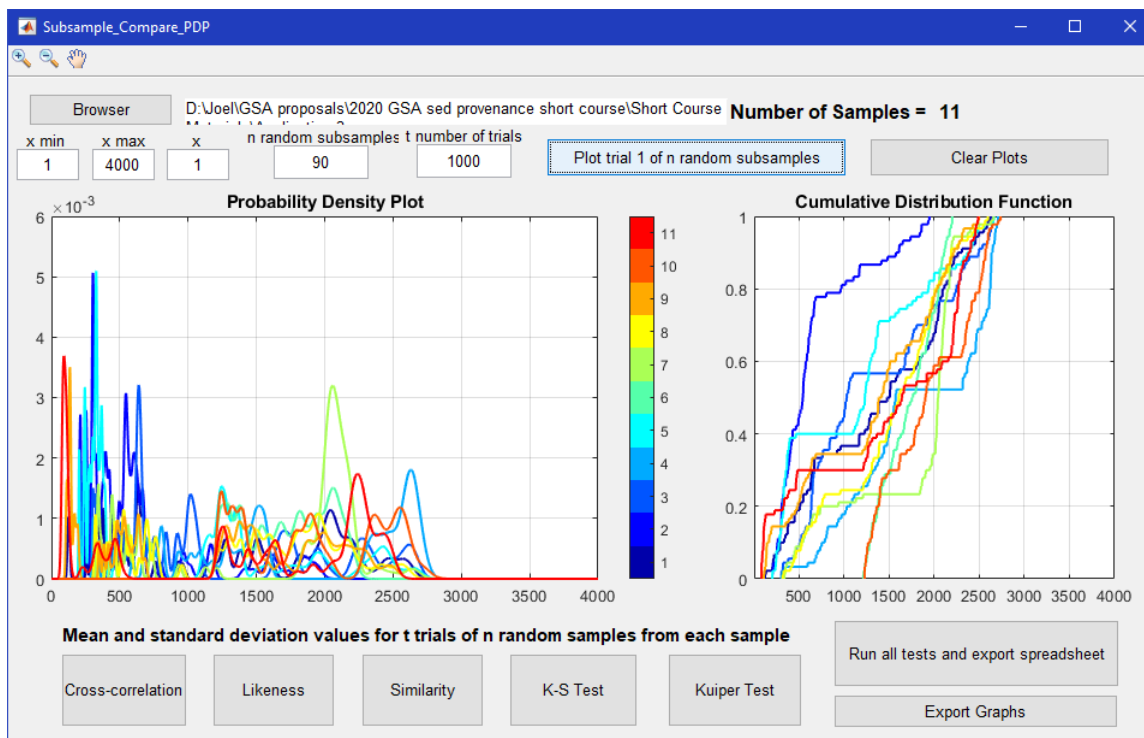
7. The rest of the analysis can proceed following the steps for “Comparison using Intersample Compare of PDPs” above

4.3. Comparison using Subsample Compare of PDPs

Analysis of sample sets with widely differing sample sizes can lead to erroneous results because the comparison does not account for the differing accuracy of the age distribution based on fewer or more grain ages. Repeated analysis using subsamples with a fixed grain count drawn from the samples with replacement is one way to mitigate these errors. This method also provides an estimate of the precision of the metrics. In DZstats, this can be implemented using the Subsample Compare routine.

1. Open DZstats by double clicking the icon in in the folder where you saved the application.
2. In the DZstats opening window, select “Probability Density Plot” as the distribution type, and then click on “Subsample Compare”.
3. In the Intersample Compare window, select “Browser” and navigate to “Application_3_U_Pb_ages_and_uncertainties.csv”.
4. Leave x_min, x_max, and x at their default values.
5. “n” determines how many grains will be selected from each sample. Since the smallest sample size in this example is 100 grains, set n for 90. This will ensure that there is variation even in the smallest sample.
6. “t” determines how many times we will drawn and compare grain ages. Set this for 1,000.
7. Select “Plot trial 1 of n random subsamples”. Your window should look as below.

Quantitative analysis, visualization, and modelling of detrital geochronology data



8. You can now view the similarity matrix for each of the metrics by selecting the appropriate button on the bottom.
9. Selecting “Cross-correlation” will result in a pop-up window that shows the selected n (= 90) and t (= 1000) as well as the mean Cross-correlation from the 1,000 trials, and the associated standard deviation based.

Figure 1: Mean and standard deviation of Cross Correlation Coefficient

	1	2	3	4	5	6	7	8	9	10	11
1	n										
2	90										
3	trials										
4	1000										
5											
6	Mean Cross ...										
7	1	0.2695	0.0627	0.0046	0.1191	0.1793	0.3233	0.2092	0.2109	0.0172	0.0597
8	0.2695	1	0.1530	0.0049	0.1557	0.0181	0.0014	0.0152	0.0183	0.0335	4.9120e-04
9	0.0627	0.1530	1	0.0208	7.7764e-04	0.0079	0.0076	0.0334	0.0070	7.9421e-04	0.0011
10	0.0046	0.0049	0.0208	1	0.0045	0.0012	0.0241	0.0400	8.3308e-04	0.1413	0.0015
11	0.1191	0.1557	7.7764e-04	0.0045	1	0.0290	0.0015	0.0236	0.0104	0.0593	0.0385
12	0.1793	0.0181	0.0079	0.0012	0.0290	1	0.3094	0.4842	0.2107	0.2695	0.0414
13	0.3233	0.0014	0.0076	0.0241	0.0015	0.3094	1	0.1492	0.1002	6.4787e-04	0.0275
14	0.2092	0.0152	0.0334	0.0400	0.0236	0.4842	0.1492	1	0.2143	0.2226	0.0495
15	0.2109	0.0183	0.0070	8.3308e-04	0.0104	0.2107	0.1002	0.2143	1	0.0879	0.1591
16	0.0172	0.0335	7.9421e-04	0.1413	0.0593	0.2695	6.4787e-04	0.2226	0.0879	1	0.0616
17	0.0597	4.9120e-04	0.0011	0.0015	0.0385	0.0414	0.0275	0.0495	0.1591	0.0616	1
18											
19	Standard de...										
20	0	0.0988	0.0416	0.0057	0.0659	0.0772	0.1085	0.0664	0.0682	0.0195	0.0320
21	0.0988	0	0.0248	0.0033	0.0252	0.0024	0.0011	0.0071	0.0075	0.0032	7.7639e-04
22	0.0416	0.0248	0	0.0079	0.0013	0.0036	0.0035	0.0109	0.0047	0.0011	0.0012
23	0.0057	0.0033	0.0079	0	0.0027	0.0015	0.0030	0.0133	0.0012	0.0227	0.0014
24	0.0659	0.0252	0.0013	0.0027	0	0.0082	0.0012	0.0088	0.0046	0.0131	0.0104
25	0.0772	0.0024	0.0036	0.0015	0.0082	0	0.0335	0.0393	0.0309	0.0334	0.0101
26	0.1085	0.0011	0.0035	0.0030	0.0012	0.0335	0	0.0253	0.0207	7.1147e-04	0.0077
27	0.0664	0.0071	0.0109	0.0133	0.0088	0.0393	0.0253	0	0.0328	0.0320	0.0122
28	0.0682	0.0075	0.0047	0.0012	0.0046	0.0309	0.0207	0.0328	0	0.0191	0.0248
29	0.0195	0.0032	0.0011	0.0227	0.0131	0.0334	7.1147e-04	0.0320	0.0191	0	0.0136
30	0.0320	7.7639e-04	0.0012	0.0014	0.0104	0.0101	0.0077	0.0122	0.0248	0.0136	0

Quantitative analysis, visualization, and modelling of detrital geochronology data

10. As an example of the dangers of using p values, select “K-S Test” or “Kuiper Test”. In order to interpret the result, you need to know that sample 1 is a mixture of samples 2, 4–9, and 11. So, those p values in row 1 should be >0.05 . However, that is not the case with either the Kuiper or K-S tests (see below). For the Kuiper test, one sample (9) yielded a p value > 0.05 . For the K-S test, two samples (3 and 9) have p values >0.05 . One of those samples (3) was not mixed into sample 1.

Figure 2: Mean and standard deviation of Kuiper Test p values

	1	2	3	4	5	6	7	8	9	10	11	
1	n											
2		90										
3	trials											
4		1000										
5												
6	Mean Kuiper ...											
7		1	1.2038e-05	0.0028	2.4408e-05	0.0099	7.3296e-08	1.2146e-04	0.0152	0.1954	5.3773e-05	5.9666e-05
8		1.2038e-05	1	3.2736e-08	1.2731e-18	1.0029e-08	4.4434e-29	1.1684e-18	8.3989e-15	7.6460e-17	4.4434e-29	2.8098e-20

Figure 1: Mean and standard deviation of K-S Test p values

	1	2	3	4	5	6	7	8	9	10	11	
1	n											
2		90										
3	trials											
4		1000										
5												
6	Mean K-S te...											
7		1	4.8754e-06	0.0623	2.3942e-06	0.0200	6.7262e-06	7.4827e-05	0.0372	0.3584	4.3415e-06	0.0085
8		4.8754e-06	1	1.1373e-09	2.8999e-20	3.9144e-07	9.5987e-31	2.6604e-20	2.0815e-16	3.1370e-11	9.5987e-31	5.5963e-13

11. As above, you can export all of the comparison metrics as a spreadsheet by selecting the “Run all tests and export spreadsheet” button on the bottom right. This will open a dialog box. Note where you save the file, then navigate to that file and open it. In addition to the mean and standard deviation, the software will export the minimum and maximum for each metric.