

## DZstats2D Version 0.04 User Manual

From “Two-dimensional quantitative comparison in detrital geochronology and geochemistry”  
K. Sundell and J. Saylor (submitted to *Geochemistry, Geophysics, Geosystems*; Nov. 2020)

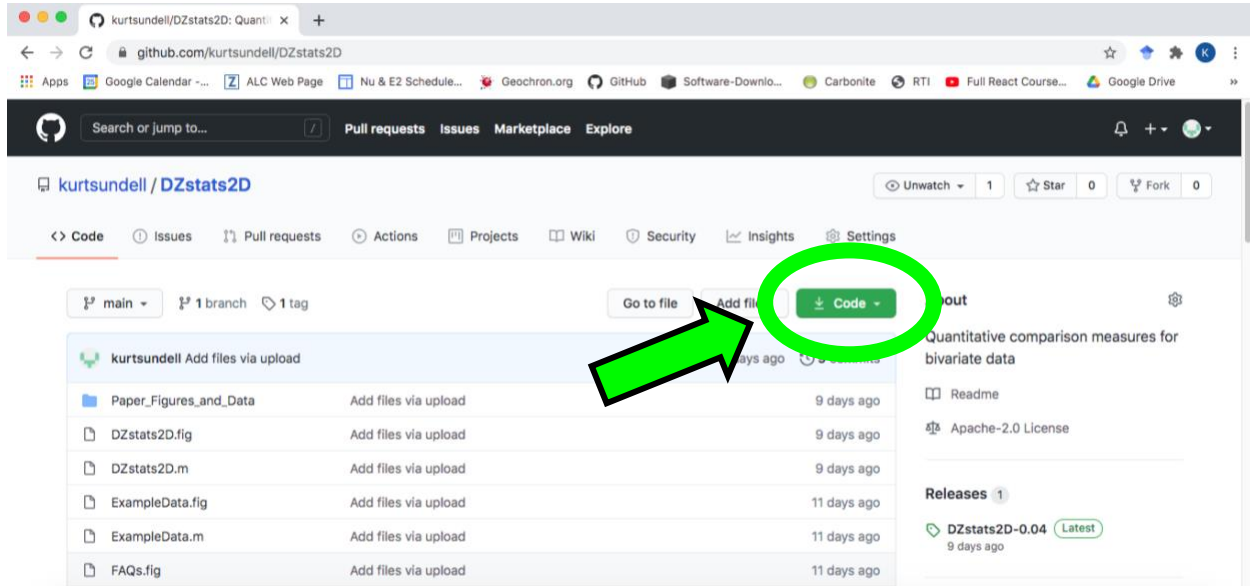
### Table of Contents

<b>Downloading DZstats2D code to run in MATLAB .....</b>	<b>2</b>
<b>Downloading DZstats2D as standalone application to be run without MATLAB and installing the free Runtime Compiler .....</b>	<b>5</b>
<b>Input data setup .....</b>	<b>11</b>
<b>Load Data.....</b>	<b>12</b>
<b>Plot Parameters.....</b>	<b>13</b>
<b>Axes Parameters.....</b>	<b>14</b>
<b>Contour Plot .....</b>	<b>15</b>
<b>Scatter Plot.....</b>	<b>16</b>
<b>Multidimensional Scaling (MDS) .....</b>	<b>17</b>
<b>Monte Carlo Resampling.....</b>	<b>19</b>
<b>Comparison Method .....</b>	<b>22</b>
<b>Conditional Formatting of Results .....</b>	<b>24</b>
<b>View options and exporting and saving plots .....</b>	<b>27</b>

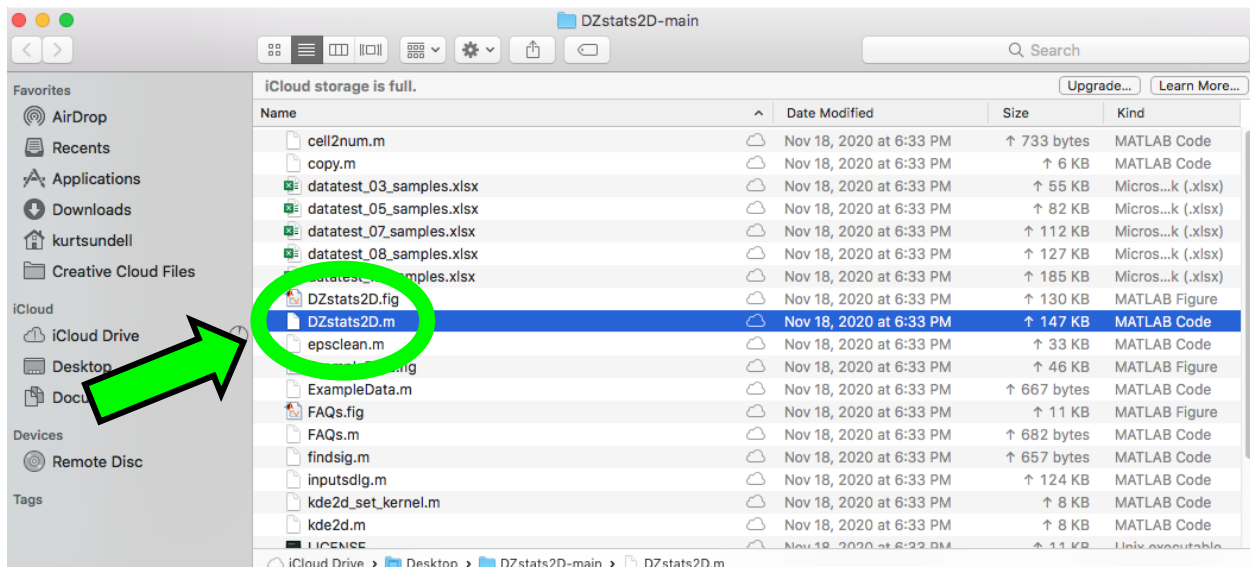
## Downloading DZstats2D code to run in MATLAB

Go to <https://github.com/kurtsundell/DZstats2D>.

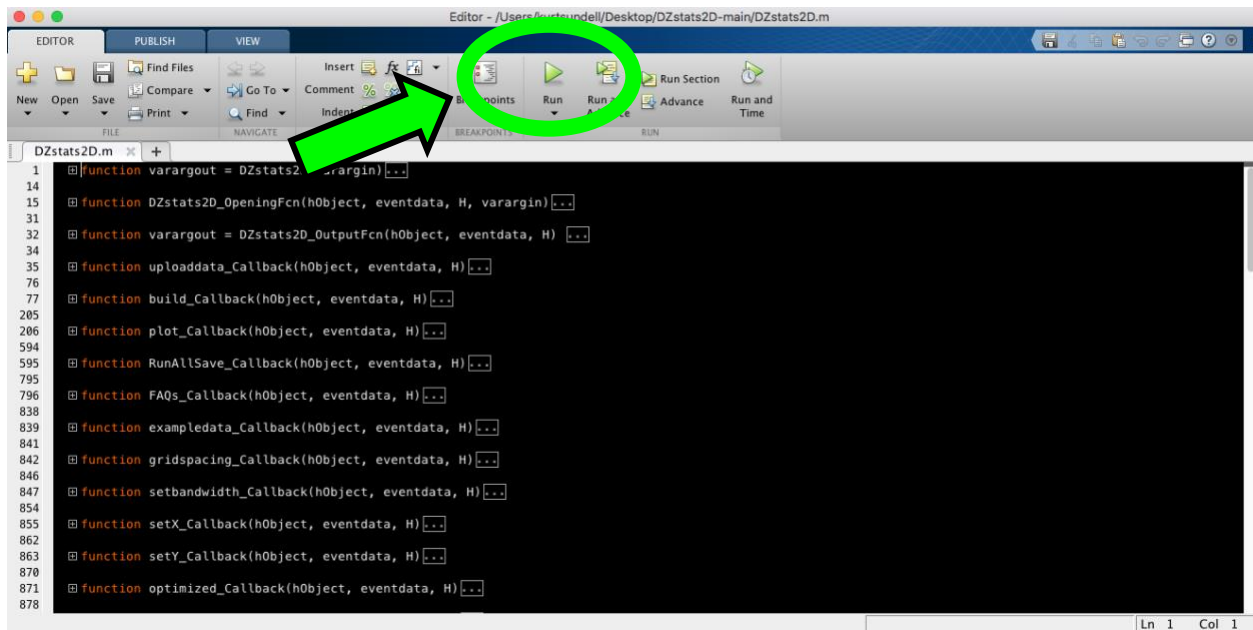
Download zip folder with code.



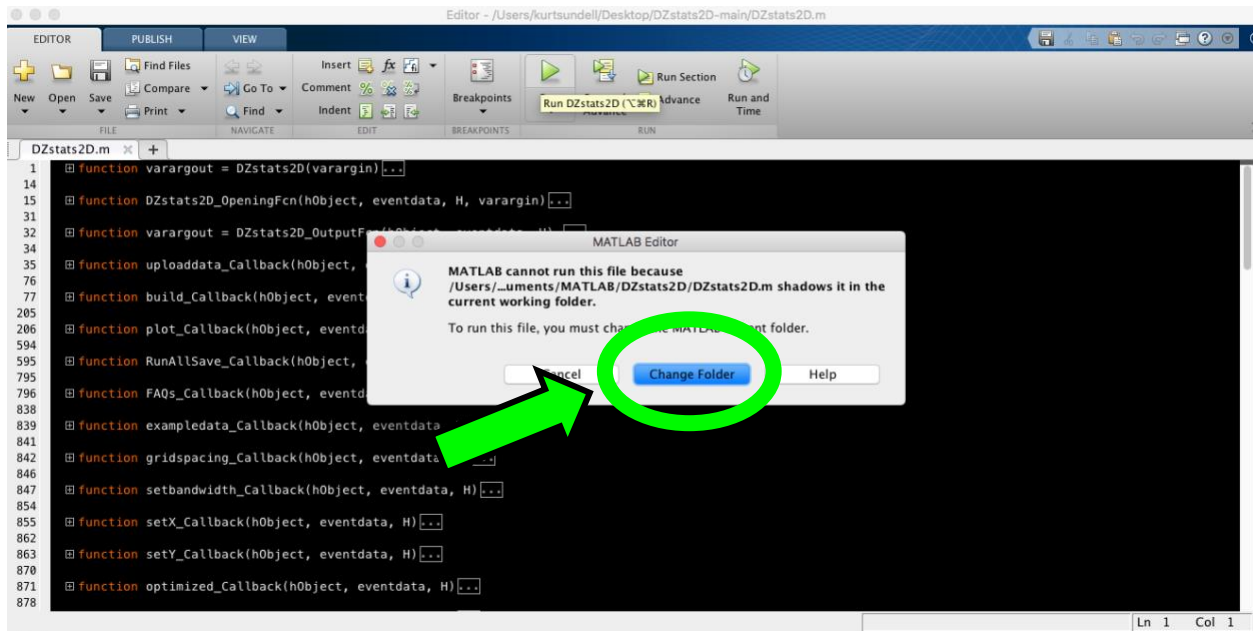
Unzip the .zip file and move the folder to somewhere on your computer (e.g., Desktop, Documents) and open the DZstats2D.m in MATLAB (the program won't work if you try to run in the .zip file).



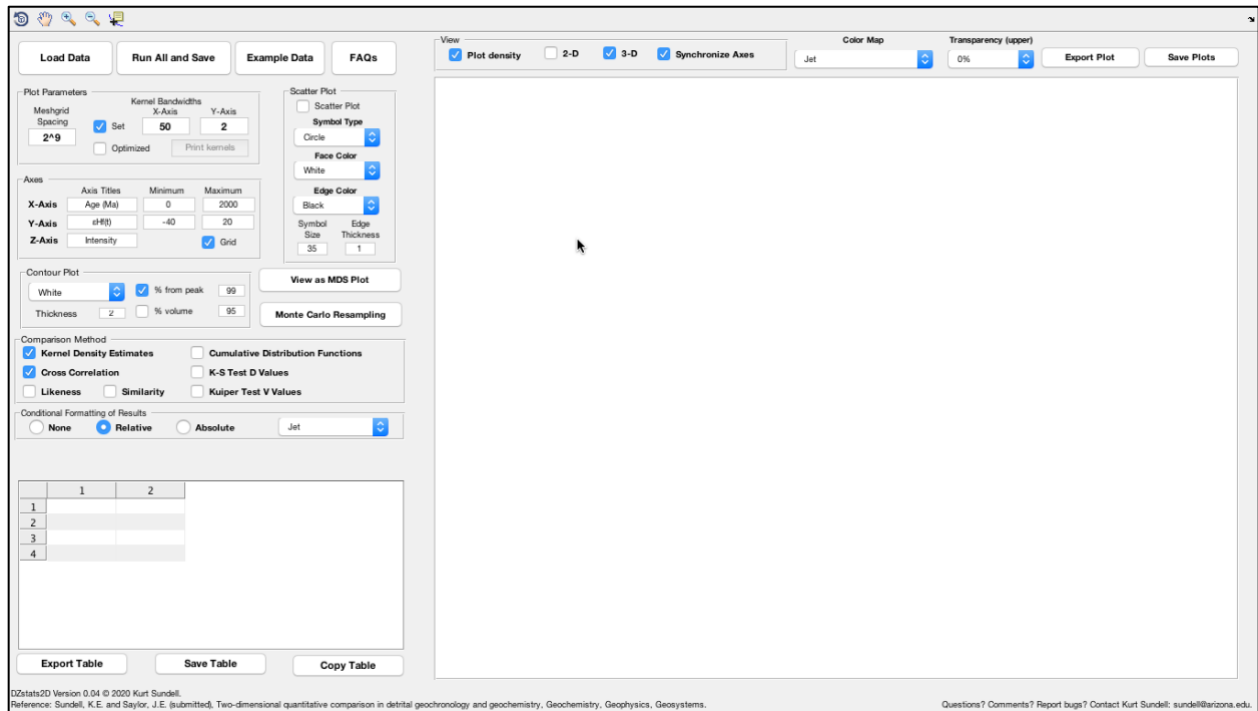
Run the code by pressing the big green Run arrow.



You may get an error saying you're in the wrong directory. Select Change Folder.



You should now see DZstats2D



## Downloading DZstats2D as standalone application to be run without MATLAB and installing the free Runtime Compiler

Install the free MATLAB Runtime which can be found here:  
<https://www.mathworks.com/products/compiler/matlab-runtime.html>

For version 1.4x and earlier download R2018b (9.5) 64-bit for Windows or Intel 64-bit for Mac. Version 1.5x will be compiled in R2020a (9.8).

**MATLAB Compiler**

**MATLAB Runtime**

*Run compiled MATLAB applications or components without installing MATLAB*

The MATLAB Runtime is a standalone set of shared libraries that enables the execution of compiled MATLAB applications or components. When used together, MATLAB, MATLAB Compiler, and the MATLAB Runtime enable you to create and distribute numerical applications or software components quickly and securely.

To download and install the MATLAB Runtime:

1. Click the version and platform in the table below that corresponds to the application or component you are using. The version of the MATLAB Runtime is tied to the version of MATLAB.

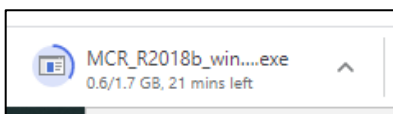
Note: You can find this information in the `readme.txt` file that accompanies the application or component.

2. Save the MATLAB Runtime installer file on the computer on which you plan to run the application or component.
3. Double click the installer and follow the instructions in the installation wizard.

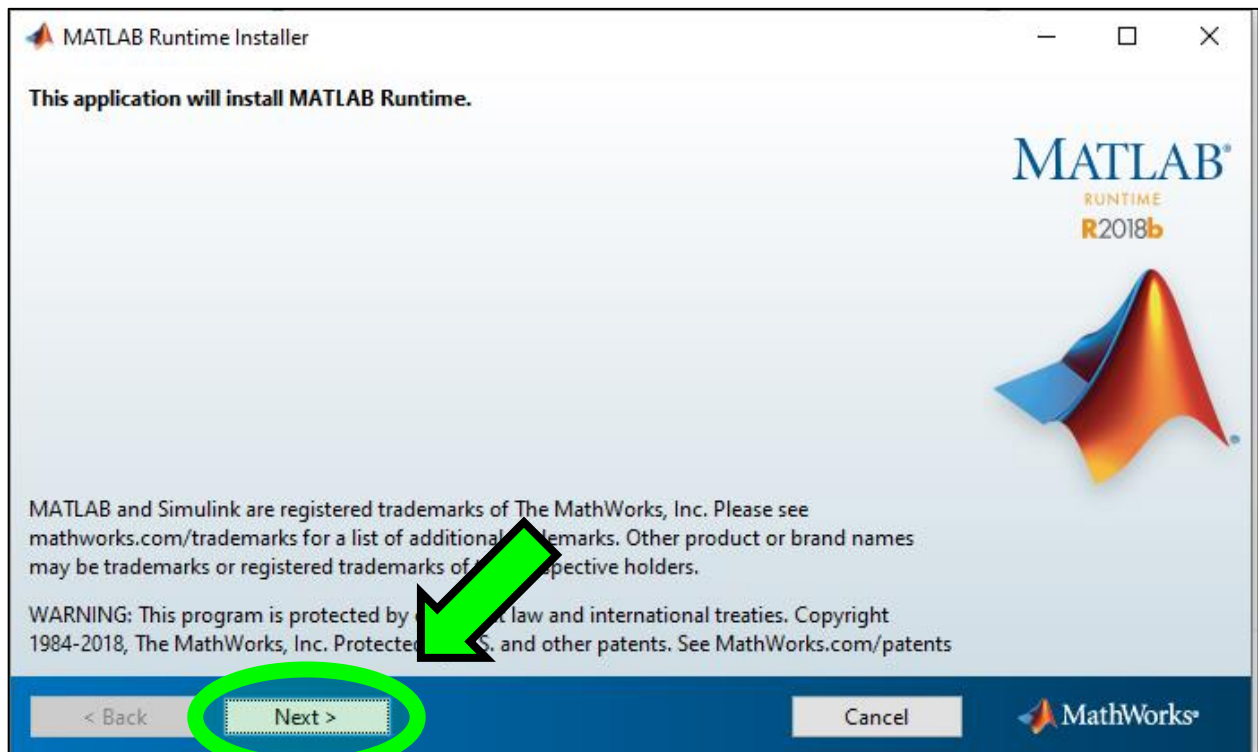
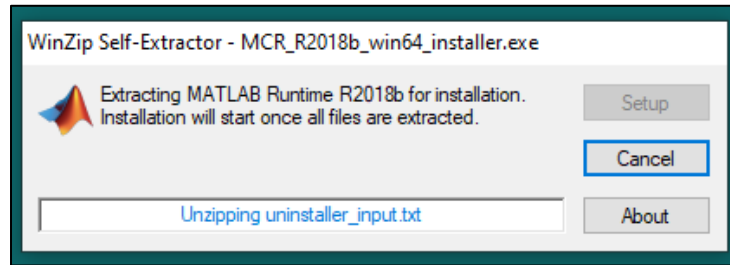
See the [MATLAB Runtime Installer documentation](#) for more information.

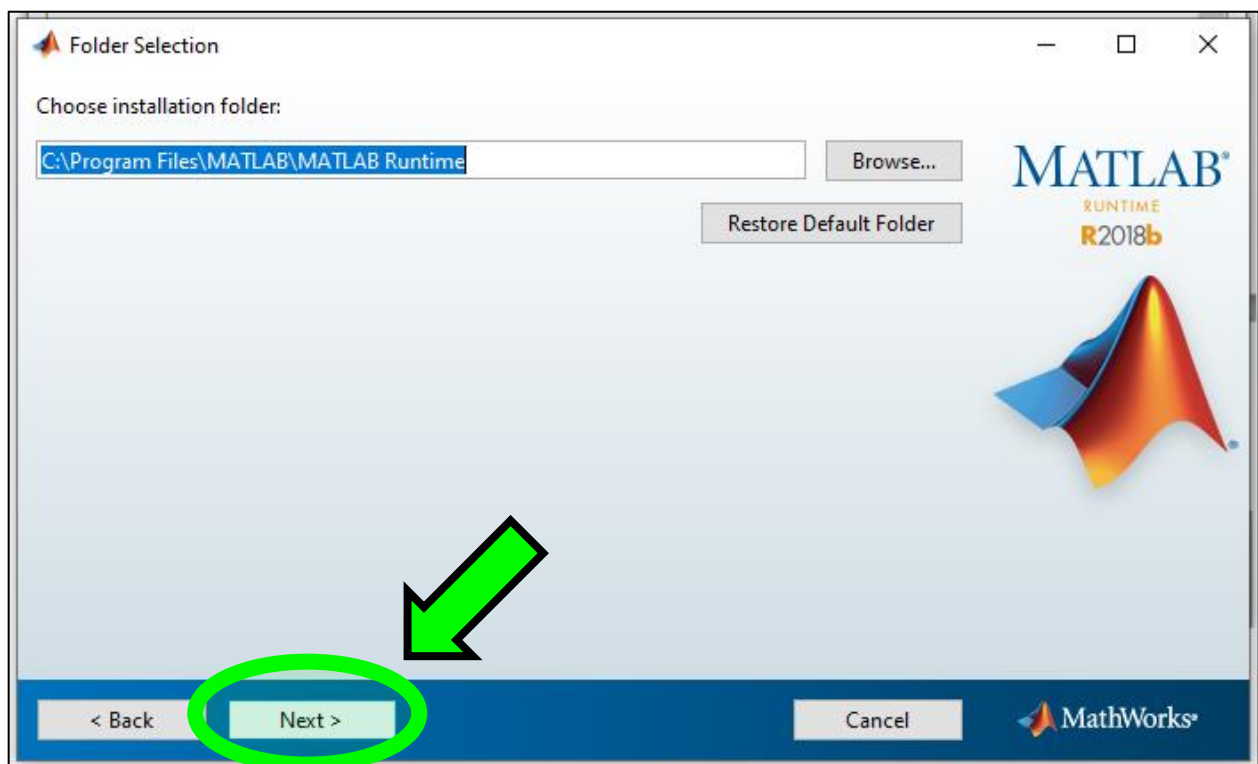
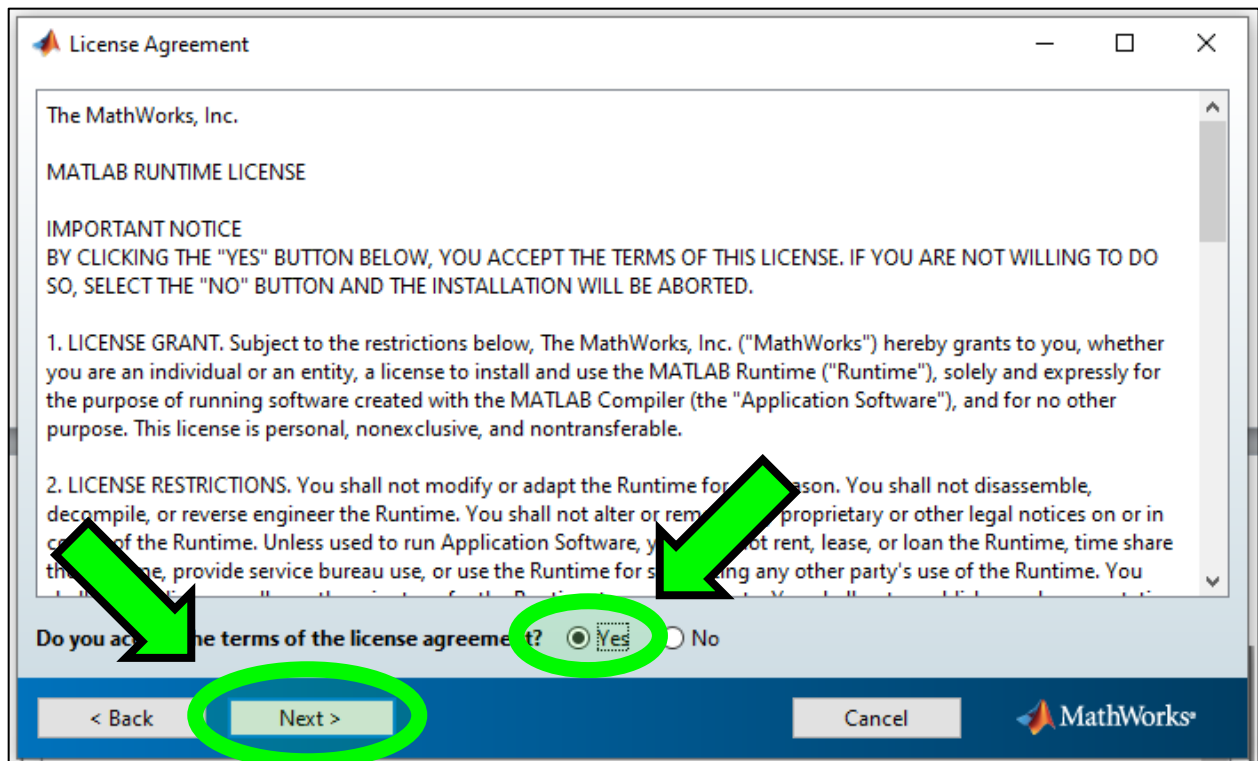
Release (MATLAB Runtime Version#)	Windows	Linux	Mac
R2020a (9.8)	64-bit	64-bit	Intel 64-bit
R2019b (9.7)	64-bit	64-bit	Intel 64-bit
R2019a (9.6)	64-bit	64-bit	Intel 64-bit
R2018b (9.5)	64-bit	64-bit	Intel 64-bit
R2018a (9.4)	64-bit	64-bit	Intel 64-bit
R2017b (9.3)	64-bit	64-bit	Intel 64-bit

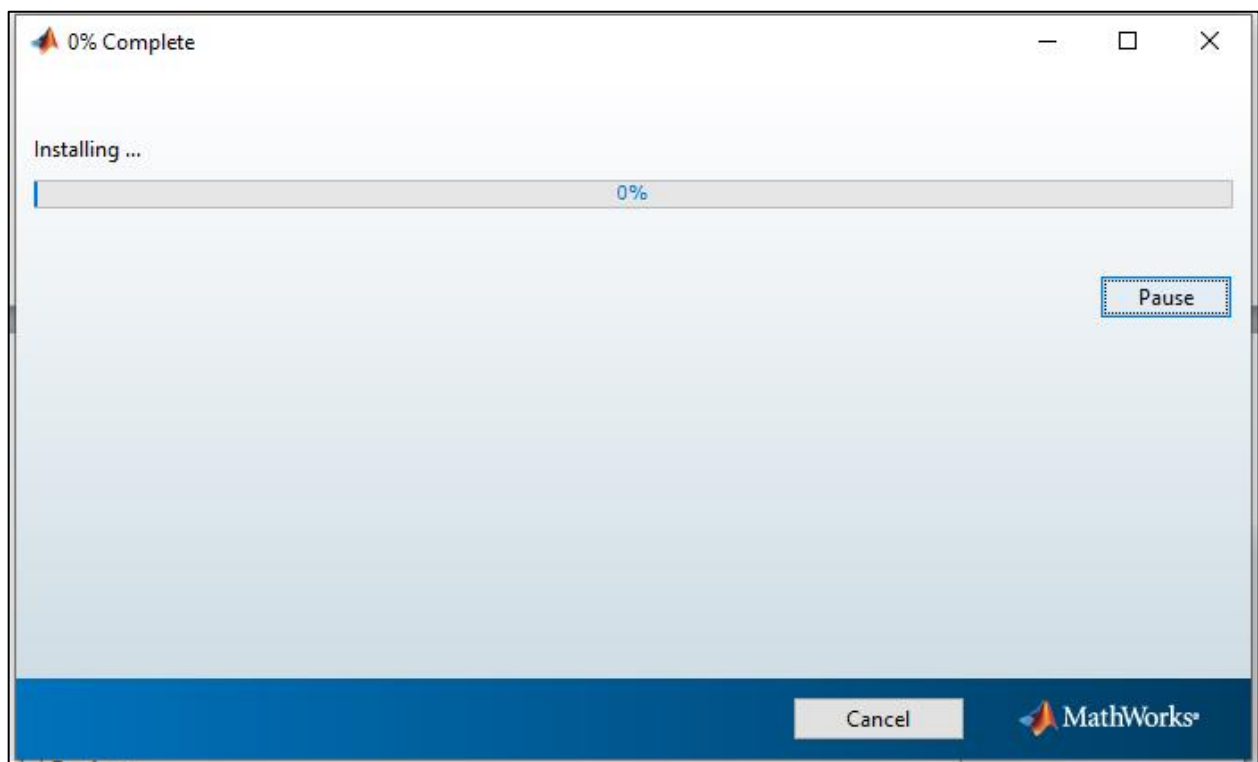
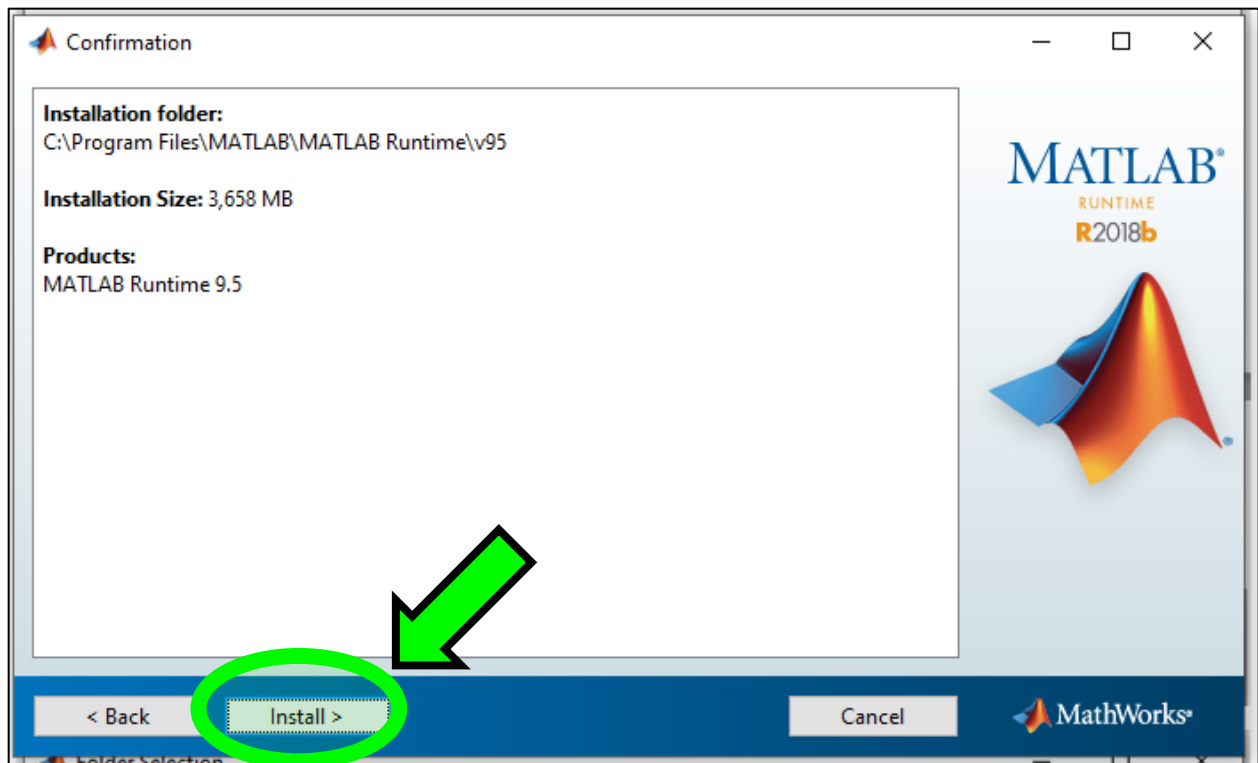
*This file is large, so it may take a while. Once complete, move the installer to your desktop.*



Step 5. Move the installer to your Desktop. Open and follow all default settings. (You can delete the installer afterward)

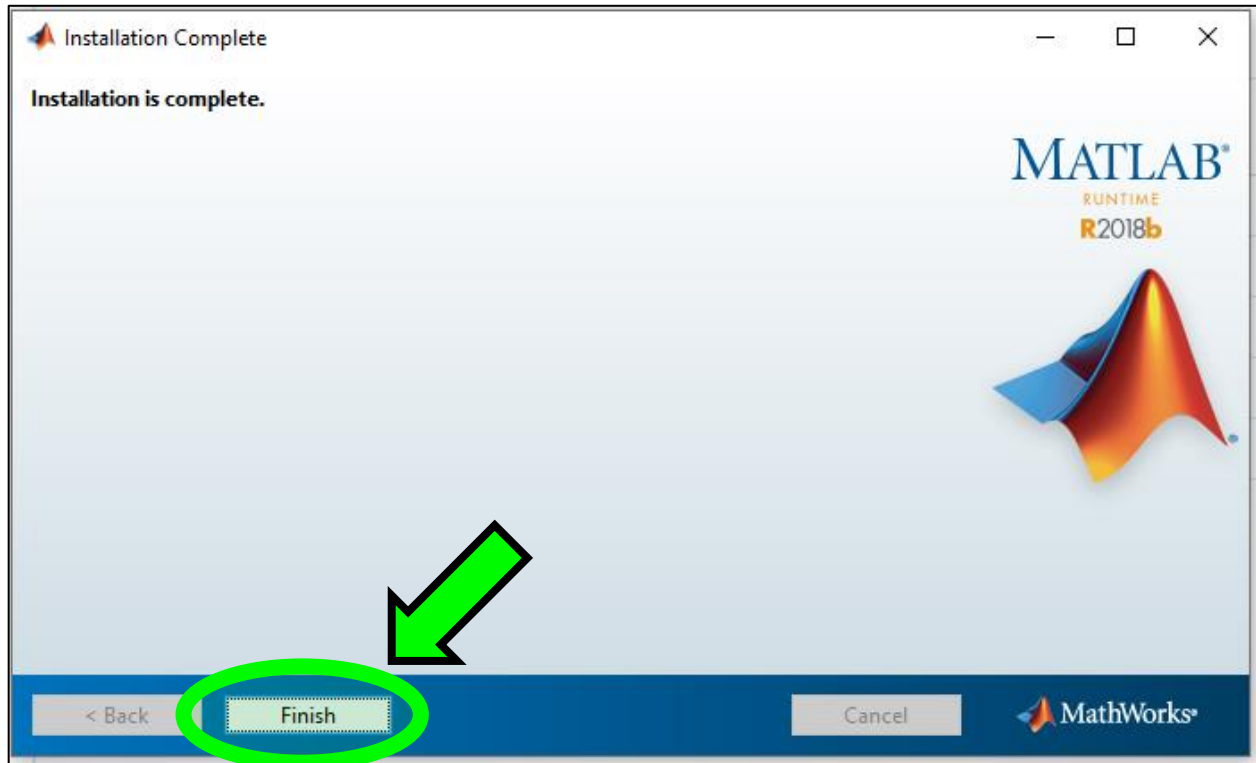




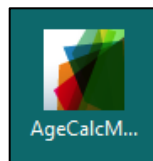




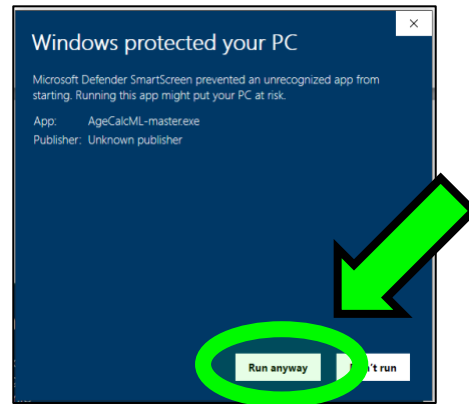
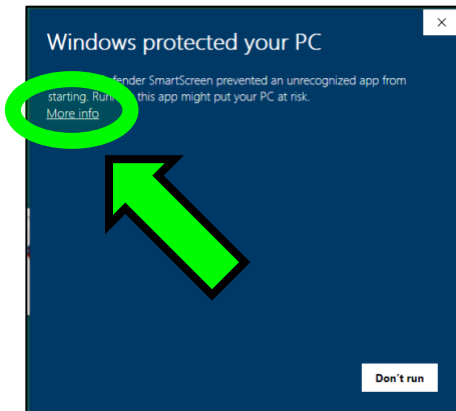
Once the install is complete nothing will happen. *DZstats2D* will automatically access the Runtime. Don't forget to delete the installer.



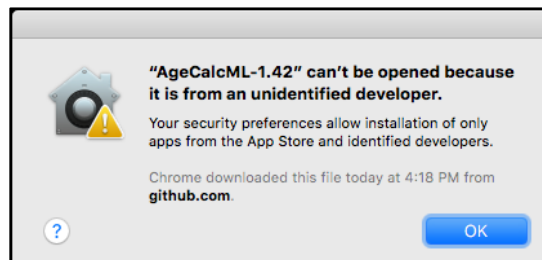
Step 6. Double click to open. Only do this once! The first time opening may take a few minutes.



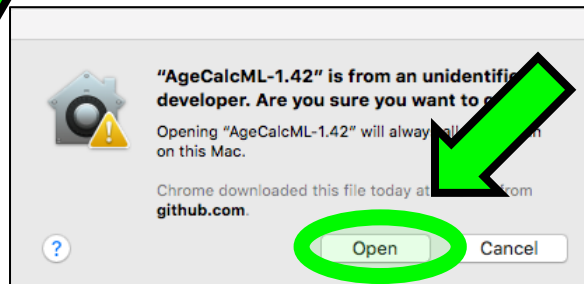
Upon opening you may see a warning message. On Windows select more info and Run anyway.



On Mac you might see an error like this:

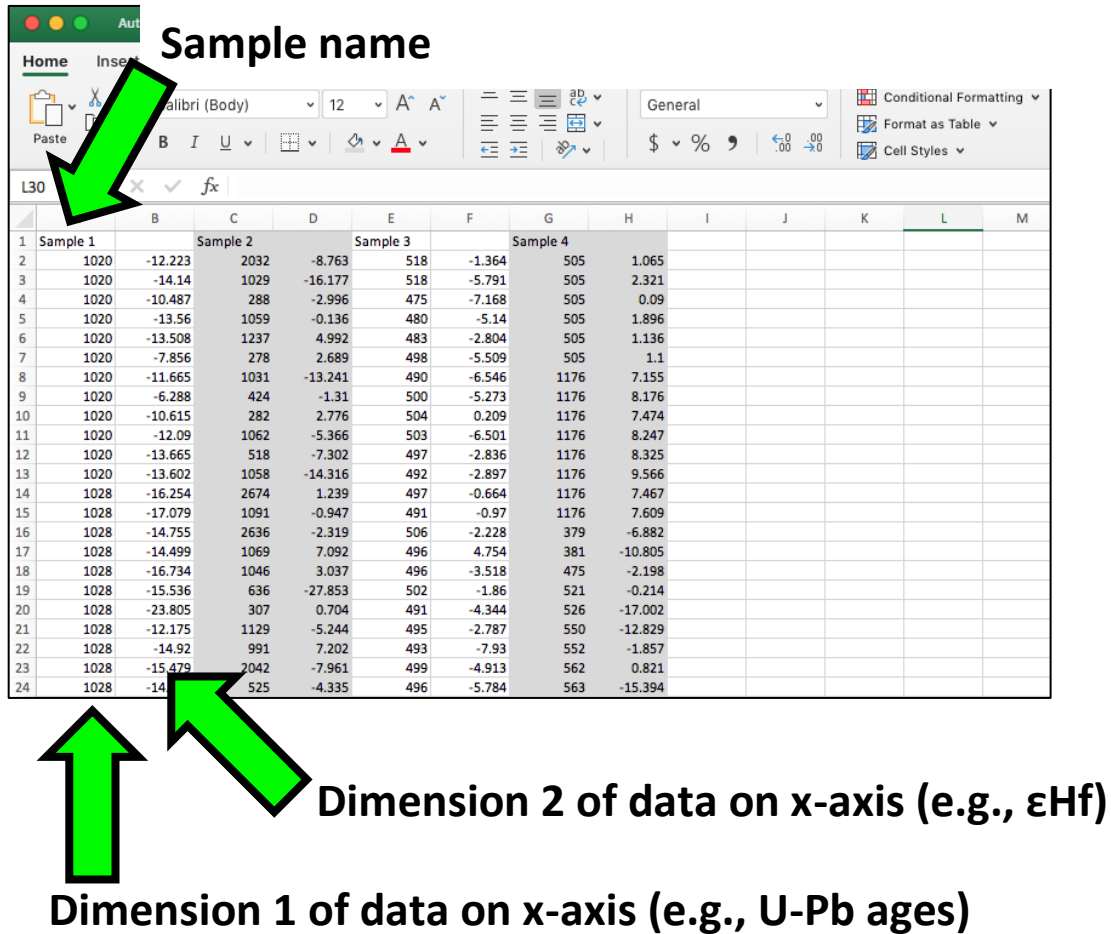


Right click the icon and select open



## Input data setup

The input data is organized as a pair of columns for each sample containing ages and associated uncertainties and sample name. Save as .csv, .xls, or.xlsx file.



**Sample name**

**Dimension 2 of data on x-axis (e.g.,  $\epsilon_{\text{Hf}}$ )**

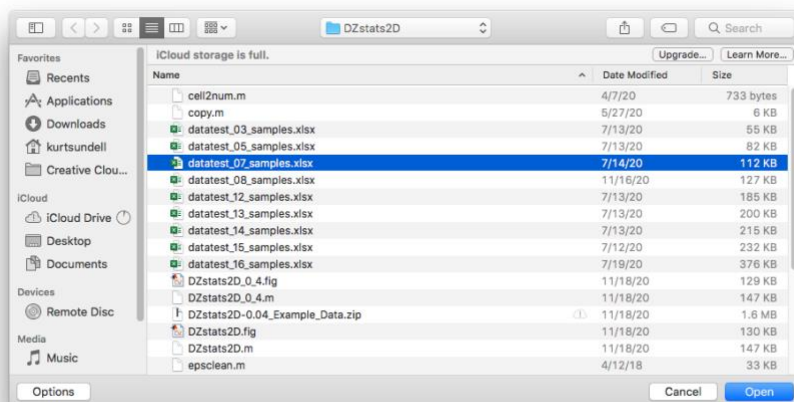
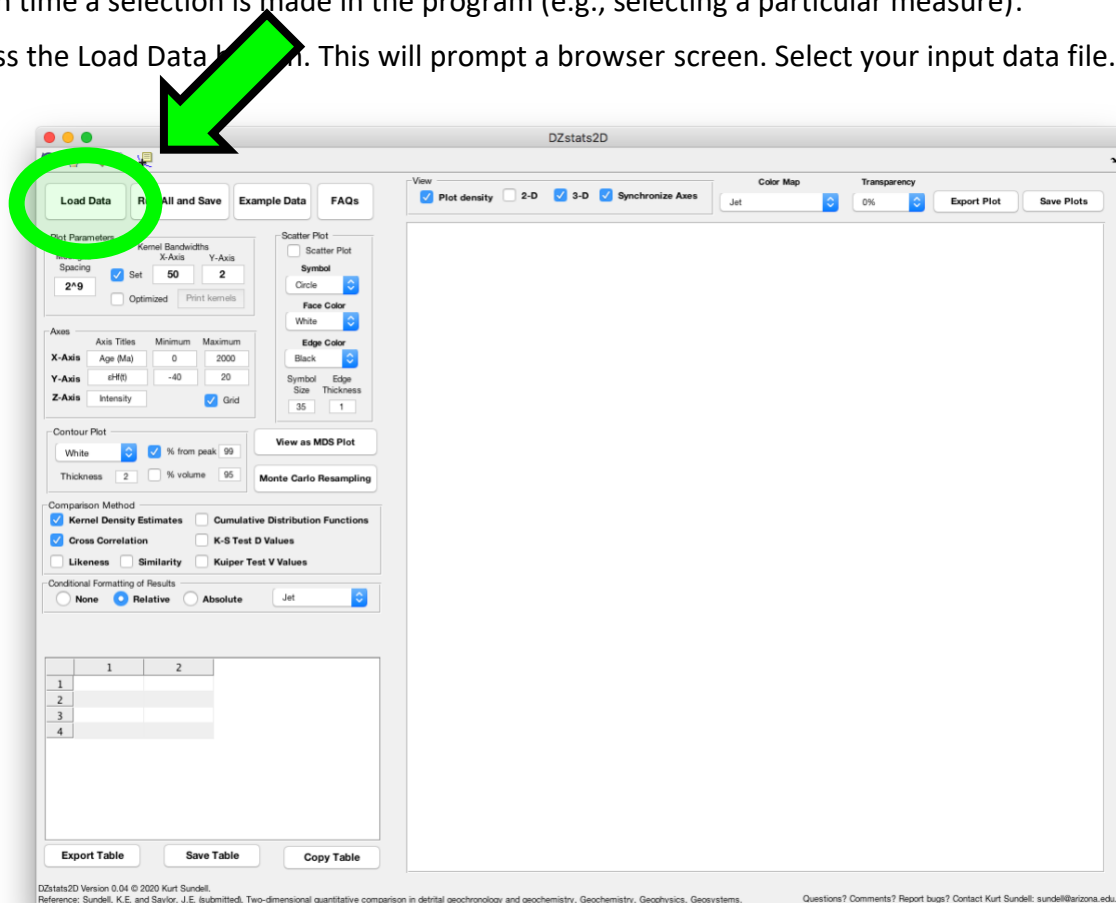
**Dimension 1 of data on x-axis (e.g., U-Pb ages)**

	Sample 1	Sample 2	Sample 3	Sample 4								
1	Sample 1											
2	1020	-12.223	2032	-8.763	518	-1.364	505	1.065				
3	1020	-14.14	1029	-16.177	518	-5.791	505	2.321				
4	1020	-10.487	288	-2.996	475	-7.168	505	0.09				
5	1020	-13.56	1059	-0.136	480	-5.14	505	1.896				
6	1020	-13.508	1237	4.992	483	-2.804	505	1.136				
7	1020	-7.856	278	2.689	498	-5.509	505	1.1				
8	1020	-11.665	1031	-13.241	490	-6.546	1176	7.155				
9	1020	-6.288	424	-1.31	500	-5.273	1176	8.176				
10	1020	-10.615	282	2.776	504	0.209	1176	7.474				
11	1020	-12.09	1062	-5.366	503	-6.501	1176	8.247				
12	1020	-13.665	518	-7.302	497	-2.836	1176	8.325				
13	1020	-13.602	1058	-14.316	492	-2.897	1176	9.566				
14	1028	-16.254	2674	1.239	497	-0.664	1176	7.467				
15	1028	-17.079	1091	-0.947	491	-0.97	1176	7.609				
16	1028	-14.755	2636	-2.319	506	-2.228	379	-6.882				
17	1028	-14.499	1069	7.092	496	4.754	381	-10.805				
18	1028	-16.734	1046	3.037	496	-3.518	475	-2.198				
19	1028	-15.536	636	-27.853	502	-1.86	521	-0.214				
20	1028	-23.805	307	0.704	491	-4.344	526	-17.002				
21	1028	-12.175	1129	-5.244	495	-2.787	550	-12.829				
22	1028	-14.92	991	7.202	493	-7.93	552	-1.857				
23	1028	-15.479	2042	-7.961	499	-4.913	562	0.821				
24	1028	-14.14	525	-4.335	496	-5.784	563	-15.394				

## Load Data

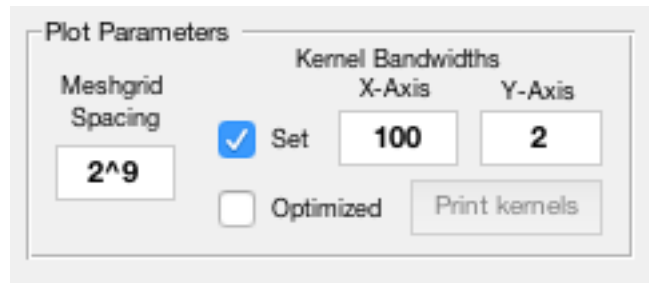
Upon loading, DZstats2D will do two things (1) it will calculate all of the quantitative comparisons and build all of the data it will need to visualize the results, then (2) it will plot the results. It does this in two steps because the 'building' part can be quite time consuming, and it is more efficient to do all of the calculations in one step, rather than recalculate everything each time a selection is made in the program (e.g., selecting a particular measure).

Press the Load Data button. This will prompt a browser screen. Select your input data file.



## Plot Parameters

These options will change the construction of the bivariate density distributions and will require 'rebuilding' the visualization and quantitative comparisons (rebuilding is done automatically).



The image shows a 'Plot Parameters' dialog box. It has a title bar 'Plot Parameters'. Inside, there are two main sections. The first section is 'Meshgrid Spacing' with a text input field containing '2^9'. The second section is 'Kernel Bandwidths', which contains two sub-sections: 'X-Axis' and 'Y-Axis'. The 'X-Axis' has a text input field containing '100'. The 'Y-Axis' has a text input field containing '2'. Below these input fields, there are two checkboxes: 'Set' (checked) and 'Optimized' (unchecked). To the right of the 'Optimized' checkbox is a button labeled 'Print kernels'.

Meshgrid Spacing is the number of points along each axis which are used to define a square grid. The spacing has to be a factor of 2. By default, it is set to 512 (i.e.,  $2^9$ ). Increasing the number of points (e.g.,  $2^{10}$  or  $2^{11}$ ) has little effect on the resulting quantitative comparisons. Decreasing (e.g.,  $2^8$  or  $2^7$ ) will increase the compute speed, but if very coarse (e.g.,  $2^5$  or  $2^4$ ) the quantitative comparison results could change drastically.

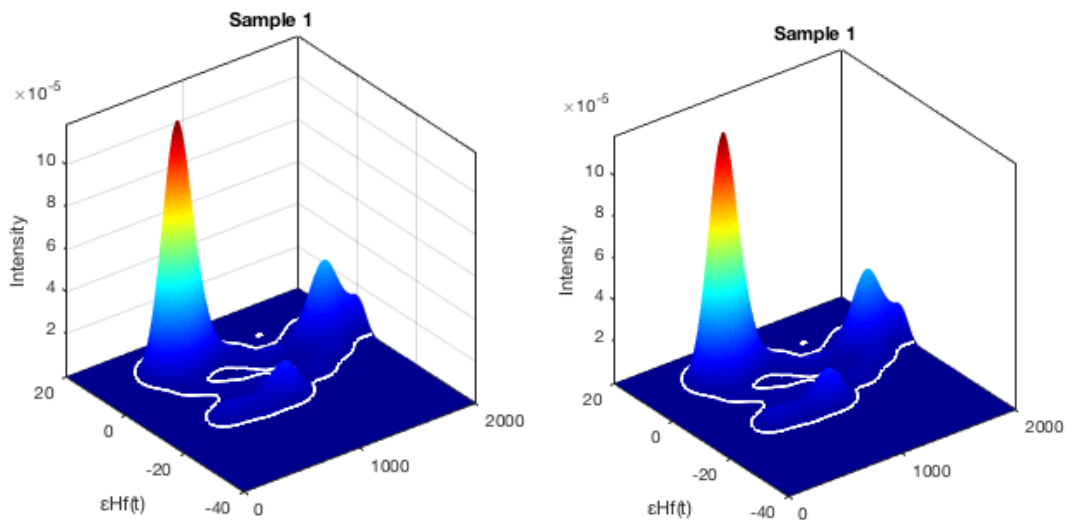
Set and Optimized checkboxes refer to the type of kernel bandwidth to be used. Set uses the x-axis and y-axis values to the right for all input datasets. This is the preferred use. Optimized uses an algorithm from Botev et al. (2010) to determine the optimal bandwidth for each dataset. Using a different kernel for each dataset is not recommended, but rather included for completeness. If an optimized setup is used, the individual kernels can be exported using the Print Kernels button.

## Axes Parameters

Axes			
	Axis Titles	Minimum	Maximum
X-Axis	Age (Ma)	0	2000
Y-Axis	$\epsilon\text{Hf}(t)$	-40	20
Z-Axis	Intensity		<input checked="" type="checkbox"/> Grid

Axes Parameters controls the titles of the axes and the X-Y range. Changing the axes range will require recalculating the quantitative comparisons. Any input data outside of the X and Y ranges will be excluded before any calculations are made.

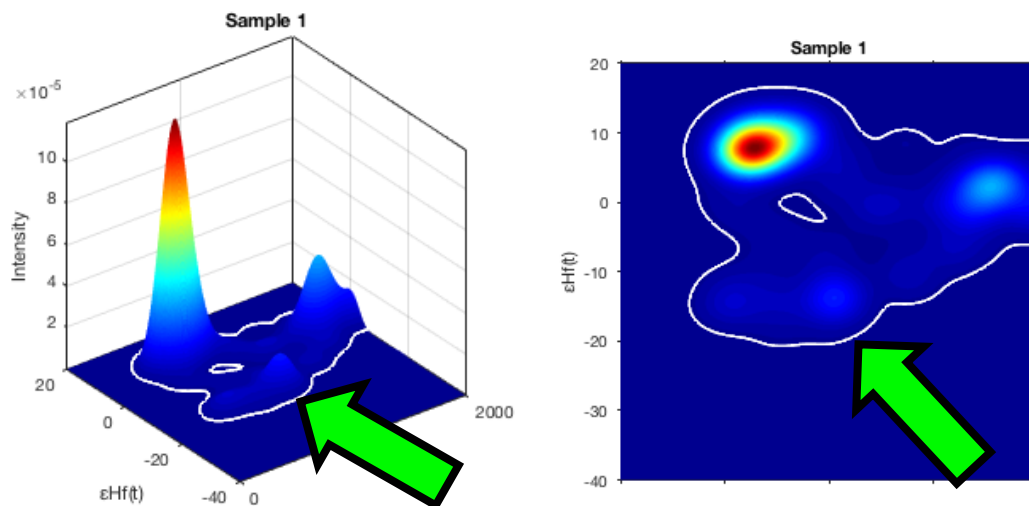
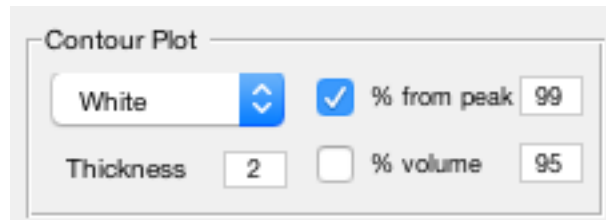
The Grid checkbox will add or remove a grid.



## Contour Plot

These are options that do not require recalculation of the density plots and quantitative comparisons; these data are already calculated and stored in the program during the 'building' step upon loading data or changing parameterization options.

There is a listbox that has different color options (default is white) and thickness specification (default is 2).



Peak density will place the contour a specified percentage away from the highest z-axis value for each plot.

Volume-based density will place the contour at a z-axis value above which the specified density is included.

## Scatter Plot

Scatter Plot is another parameter that is purely for visualization and does not affect the quantitative comparison results. There are a variety of options including symbol type. Color, size, and thickness. Changing any parameter will remake the plots, but will not recalculate the quantitative comparisons.

Scatter Plot

☒ Scatter Plot

Symbol

Circle

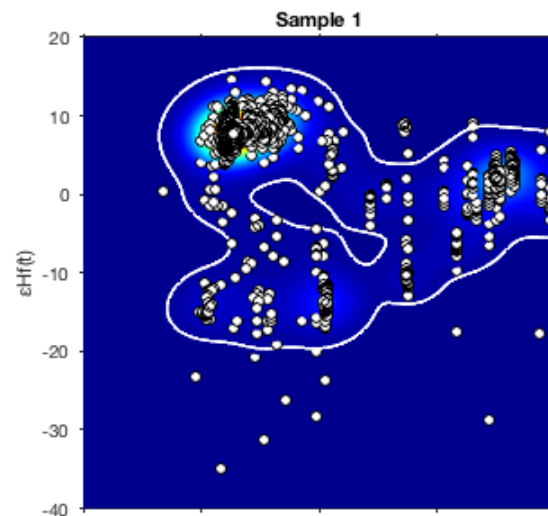
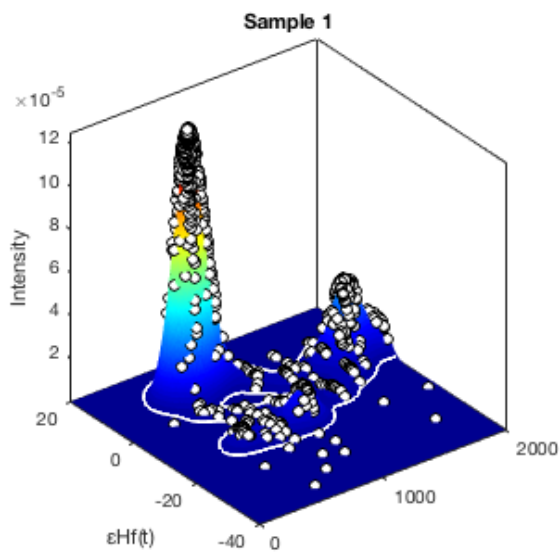
Face Color

White

Edge Color

Black

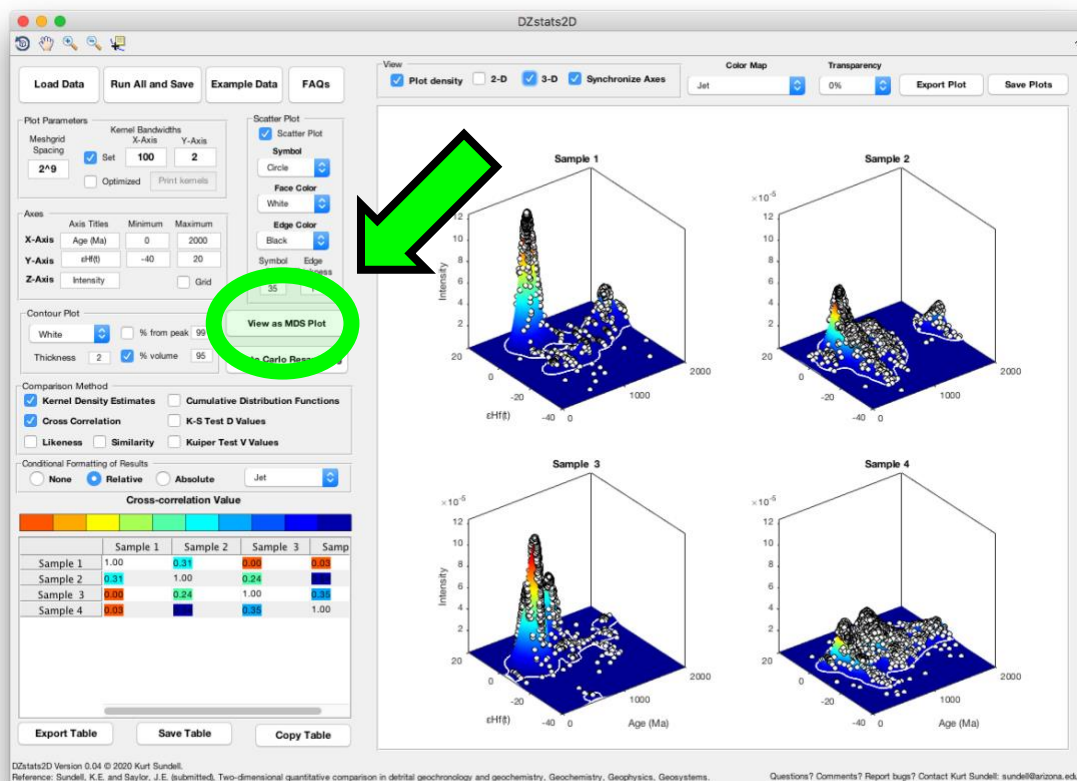
Symbol Size	Edge Thickness
35	1

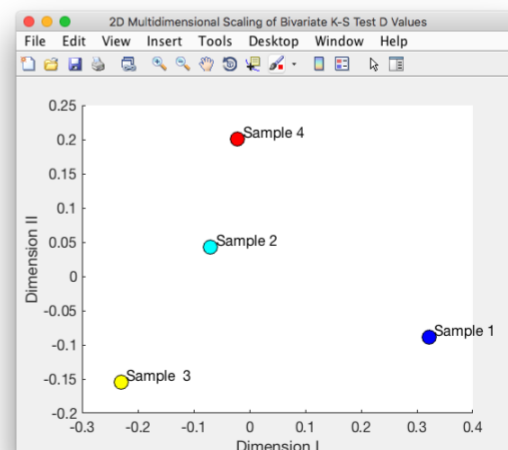
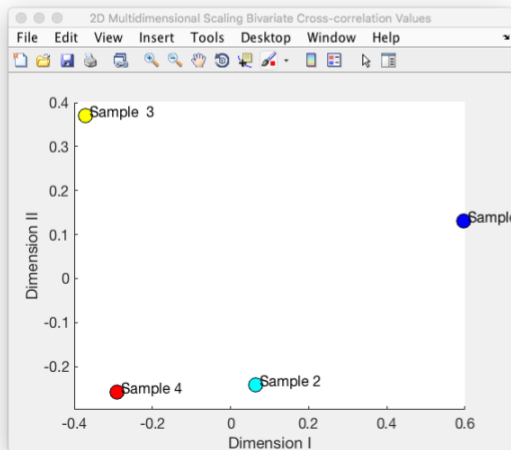
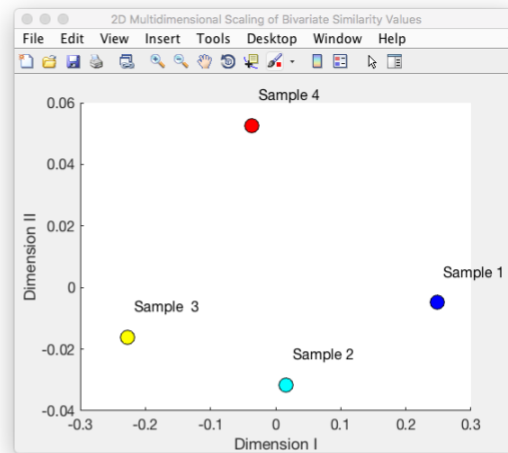
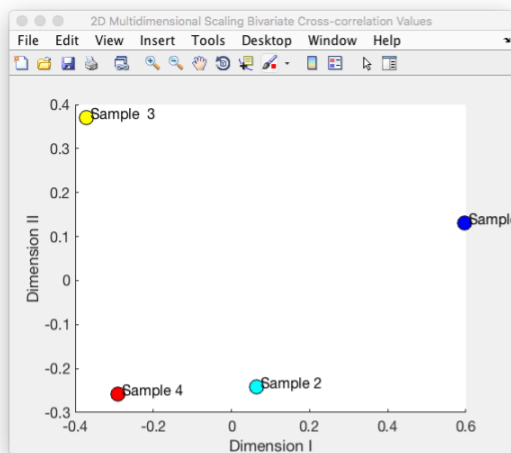




## Multidimensional Scaling (MDS)

DZstats2D implements Kruskal's method of non-metric multidimensional scaling (MDS) in two dimensions. MDS facilitates visualization of pairwise comparison matrices by transforming sample dissimilarity of into distance on a Cartesian plot in N dimensions. Similarity, Likeness, and Cross-correlation all yield 'similarity' matrices, which can be converted into sample dissimilarity by subtracting from 1 (e.g.,  $1 - S2D$ ). Both the K-S Test and Kuiper Test produce sample dissimilarity. Conversion of dissimilarity to distance in MDS is accomplished through linear transformation and iterative rearrangement in Cartesian space, seeking to minimize the misfit (termed 'stress') between distance and disparity. Stress is normalized by the sum of the squares of the interpoint distances. Low stress (e.g.,  $< 0.1$ ) indicates a reasonable transformation (Kruskal, 1961; Vermeesch, 2013). One of the most commonly used methods to accomplish this task is Kruskal's method of nonmetric MDS (Kruskal, 1961). Here, a pairwise dissimilarity matrix is converted to distance through isotonic regression, i.e., fitting a monotonically increasing line through the ranked matrix values, such that the Euclidean distances have approximately the same rank order as the corresponding dissimilarities and approximates a monotonic transformation; the transformed values are disparities. Results are represented as points in Cartesian space with greater distance corresponding to greater dissimilarity. There are other methods of calculating MDS which can be used if they have an option to input a dissimilarity matrix (which can be exported from DZstats2D).



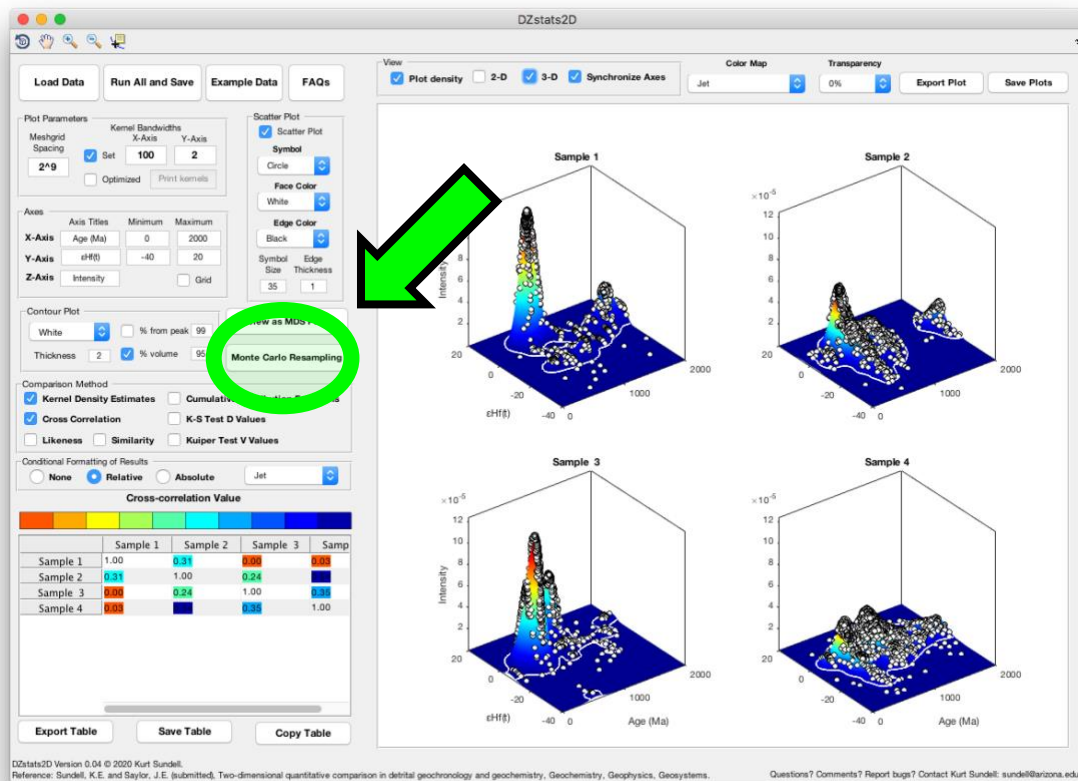


Each MDS plot is a visual representation of the sampli (dis)similarity matrix results. The closer the points are to one another, the more similar the density distributions. The farther apaer, the more dissimilar.

The MDS plots will be slightly different depending on which quantitative comparison measure is used (this is normal). The goal is to look for consistency among the different measures.

## Monte Carlo Resampling

This routine is similar to the 'Subsample Compare' routine in DZstats (Saylor and Sundell, 2016). The idea is to resample the input data sets many times and calculate the mean and standard deviation of the results. This is useful if the input datasets are very different in sample size, and also to test the robustness of results.



**Monte Carlo Resampling**

Monte Carlo resampling routine. If resampling without replacement then the maximum number of subsamples is the smallest sample size. Number of trials is the number of times the input data set is resampled. Results are reported as mean and 1 standard deviation for each metric.

Number of Subsamples (smallest sample size = 840)

Number of Trials

Type of Resampling ☒ With Replacement ☐ Without Replacement

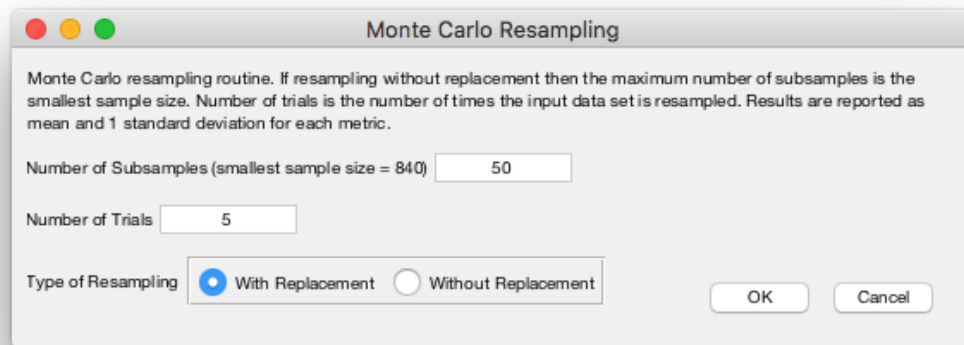
OK Cancel

There are a few options in the Monte Carlo resampling pop-up window. The main option at the bottom of the pop-up window controls what you are capable of doing in the text edit boxes.

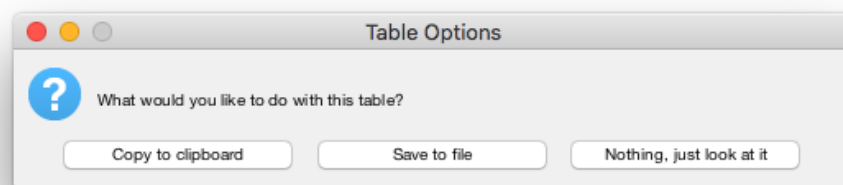
Sampling with replacement means sample selections are equally likely for every selection (including resampling the same value more than once). This also means the sampled values are always independent, i.e., what we randomly select the first time doesn't affect what we select the second time, and so on. Mathematically, this results in zero covariance among sample selections because subsequent samples are not dependent on previous sample selections.

Sampling without replacement means that every time a sample is selected it cannot be selected again because it is removed from the sample. This also means the sample values are not independent because the probability of selecting each sample is dependent on which sample were selected before it (i.e., the probability of selecting a sample is higher if samples have been removed before it); thus, subsequent samples will covary.

Number of subsamples is the number of random selections from each input sample you would like to make. If resampling with replacement, this can be any value. If sampling without replacement, this is limited to the size of your smallest sample, which is shown in parentheses.



Here is an example of selecting 50 data pairs from each data set 5 times, with replacement. You will be asked what you'd like to do with the results:



Monte Carlo Resampling Comparison Matrices with Relative Conditional Format

	1	2	3	4	5	6	7	8	9	10	11
1	DZstats2D Re...										
2											
3	Mean Cross-c...						St.Dev. Cross...				
4		Sample 1	Sample 2	Sample 3	Sample 4			Sample 1	Sample 2	Sample 3	Sample 4
5	Sample 1	1.00	0.32	0.00	0.04		Sample 1	0	0.0974	0.0029	0.0420
6	Sample 2	0.32	1.00	0.15	0.22		Sample 2	0.0974	0	0.0285	0.0495
7	Sample 3	0.00	0.15	1.00	0.22		Sample 3	0.0029	0.0285	0	0.0569
8	Sample 4	0.04	0.22	0.22	1.00		Sample 4	0.0420	0.0495	0.0569	0
9											
10	Mean Likeness						St.Dev. Liken...				
11		Sample 1	Sample 2	Sample 3	Sample 4			Sample 1	Sample 2	Sample 3	Sample 4
12	Sample 1	1.00	0.41	0.21	0.32		Sample 1	0	0.0622	0.0285	0.0460
13	Sample 2	0.41	1.00	0.39	0.22		Sample 2	0.0622	0	0.0299	0.0414
14	Sample 3	0.21	0.39	1.00	0.42		Sample 3	0.0285	0.0299	0	0.0184
15	Sample 4	0.32	0.22	0.42	1.00		Sample 4	0.0460	0.0414	0.0184	0
16											
17	Mean Similarity						St.Dev. Simila...				
18		Sample 1	Sample 2	Sample 3	Sample 4			Sample 1	Sample 2	Sample 3	Sample 4
19	Sample 1	1.00	0.67	0.43	0.60		Sample 1	0	0.0623	0.0622	0.0662
20	Sample 2	0.67	1.00	0.66	0.22		Sample 2	0.0623	0	0.0377	0.0550
21	Sample 3	0.43	0.66	1.00	0.71		Sample 3	0.0622	0.0377	0	0.0238
22	Sample 4	0.60	0.22	0.71	1.00		Sample 4	0.0662	0.0550	0.0238	0
23											
24	Mean K-S Te...						St.Dev. K-S T...				
25		Sample 1	Sample 2	Sample 3	Sample 4			Sample 1	Sample 2	Sample 3	Sample 4
26	Sample 1	0.00	0.43	0.55	0.46		Sample 1	0	0.0346	0.0759	0.0552
27	Sample 2	0.43	0.00	0.37	0.22		Sample 2	0.0346	0	0.0746	0.0958
28	Sample 3	0.55	0.37	0.00	0.46		Sample 3	0.0759	0.0746	0	0.0394
29	Sample 4	0.46	0.22	0.46	0.00		Sample 4	0.0552	0.0958	0.0394	0
30											
31	Mean Kuiper ...						St.Dev. Kuipe...				
32		Sample 1	Sample 2	Sample 3	Sample 4			Sample 1	Sample 2	Sample 3	Sample 4
33	Sample 1	0.00	0.43	0.61	0.47		Sample 1	0	0.0346	0.0394	0.0628
34	Sample 2	0.43	0.00	0.22	0.22		Sample 2	0.0346	0	0.0682	0.0944
35	Sample 3	0.61	0.22	0.00	0.47		Sample 3	0.0394	0.0682	0	0.0387
36	Sample 4	0.47	0.22	0.47	0.00		Sample 4	0.0628	0.0944	0.0387	0

Results are shown as mean values of the 5 comparisons on the left with whatever conditional color formatting was selected before running the resampling routine. The standard deviation of the results is shown on the right. High standard deviation typically means the comparison results are dependent on sample size.

## Comparison Method

Comparison Method

☒ Kernel Density Estimates

☐ Cumulative Distribution Functions

☒ Cross Correlation

☐ K-S Test D Values

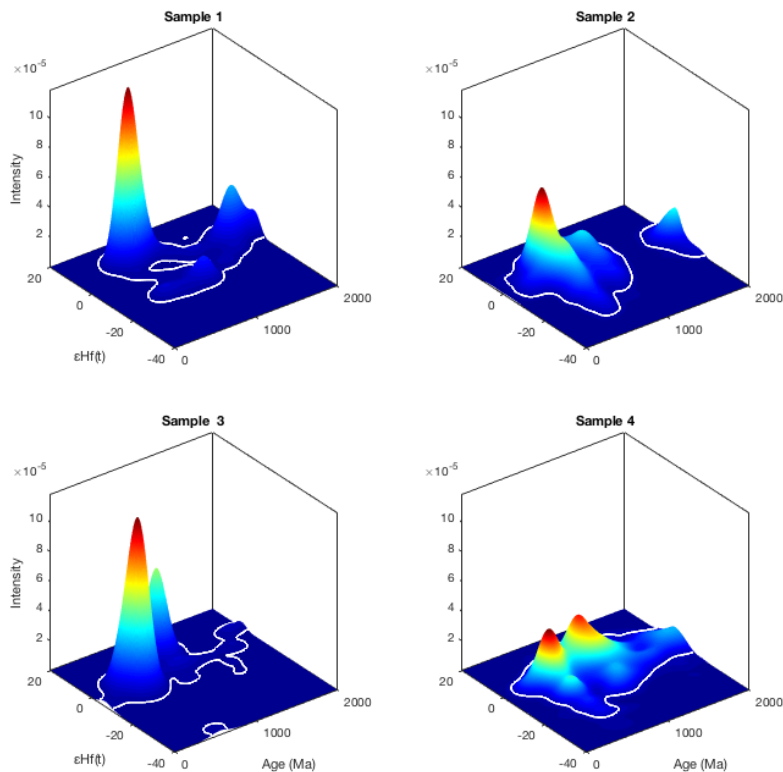
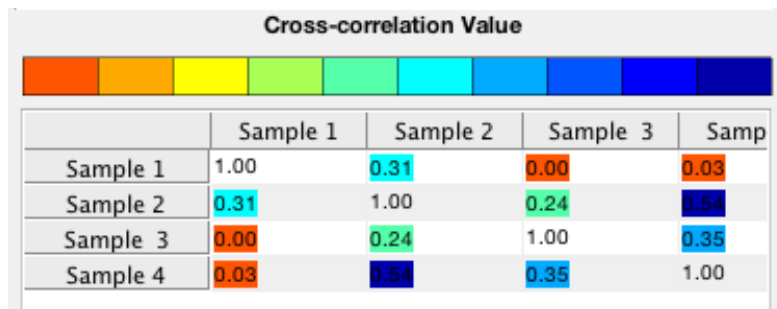
☐ Likeness

☐ Similarity

☐ Kuiper Test V Values

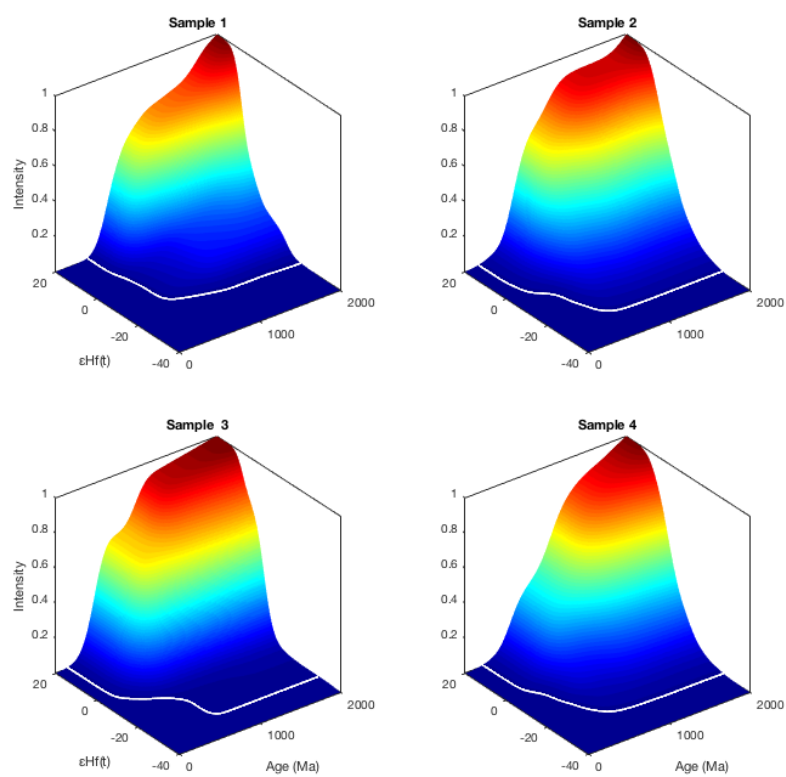
Comparison method selection controls what is currently displayed in DZstats2D both for the comparison matrix, and for the type of density distribution displayed. Selecting a different comparison measure will not make DZstats2D recalculate anything because all comparisons are calculated in the 'building' step of the program.

Here is an example for Cross-correlation based on bivariate KDEs:

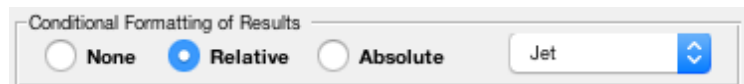


Here is an example of K-S Test D values based on bivariate CDFs:

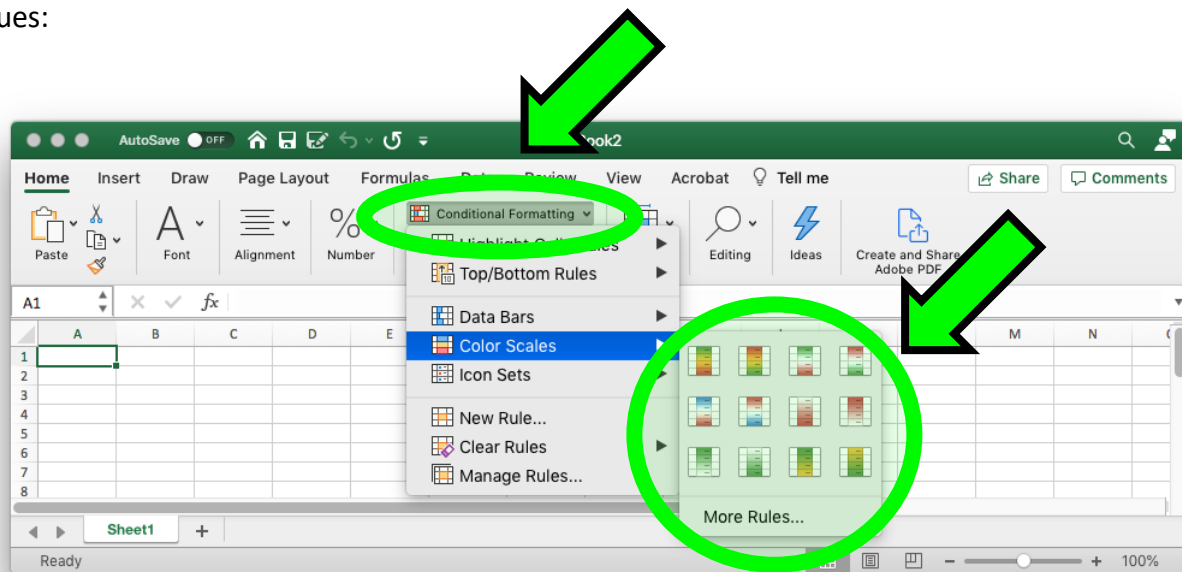
K-S Test D Value				
	Sample 1	Sample 2	Sample 3	Samp
Sample 1	0.00	0.44	0.56	0.45
Sample 2	0.44	0.00	0.32	0.27
Sample 3	0.56	0.32	0.00	0.41
Sample 4	0.45	0.27	0.41	0.00



## Conditional Formatting of Results




Conditional formatting is similar to that implemented in Excel for color-coding numerical values:



Conditional formatting colors the (dis)similarity matrices based on value between zero and one. The color ramp (e.g., Jet) is inverted for CDF-based measures (K-S and Kuiper tests) because they produce dissimilarity matrices whereas KDE-based measures produce similarity matrices. This way the results are comparable. The goal is to look for common patterns among the different quantitative measures as this typically points to a robust assessment that is not dependent on the measure used.

Selecting None will remove all conditional formatting.



Conditional Formatting of Results

☒ None ☐ Relative ☐ Absolute

Jet

**K-S Test D Value**

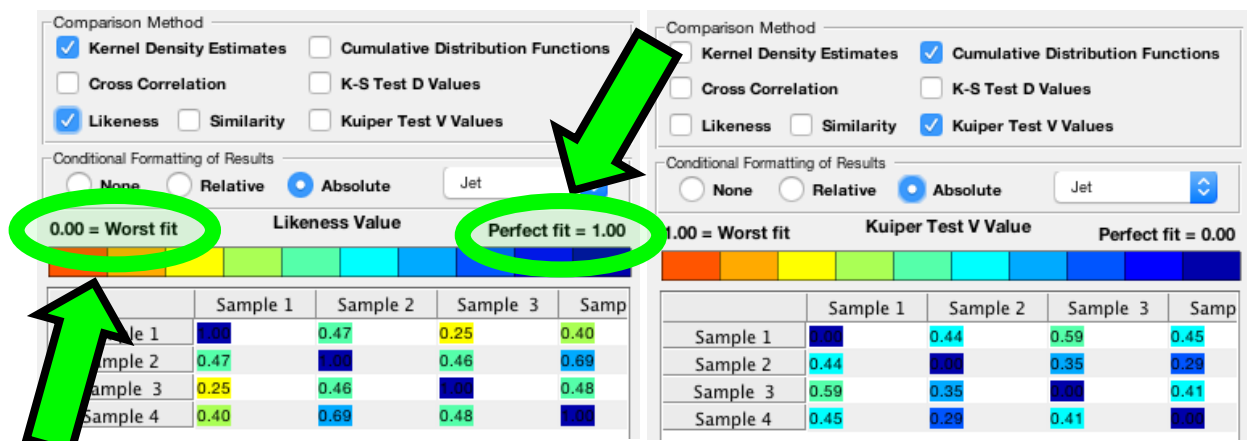
	Sample 1	Sample 2	Sample 3	Samp
Sample 1	0.00	0.44	0.56	0.45
Sample 2	0.44	0.00	0.32	0.27
Sample 3	0.56	0.32	0.00	0.41
Sample 4	0.45	0.27	0.41	0.00



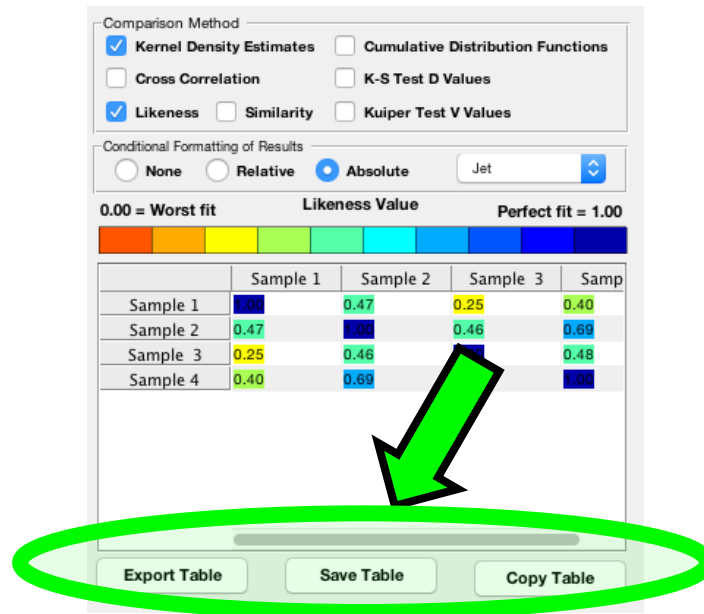
Selecting Relative will set the colormap (e.g., Jet) based on the minimum and maximum results for each measure. Here are Likeness and Kuiper Test, notice how the scale max and min are different:



Selecting Absolute will set the colormap min and max to zero and one for KDE-based measures and set the colormap to one and zero for CDF-based measures.



Results in the table can be exported into a separate window, saved as a spreadsheet table, or copied:

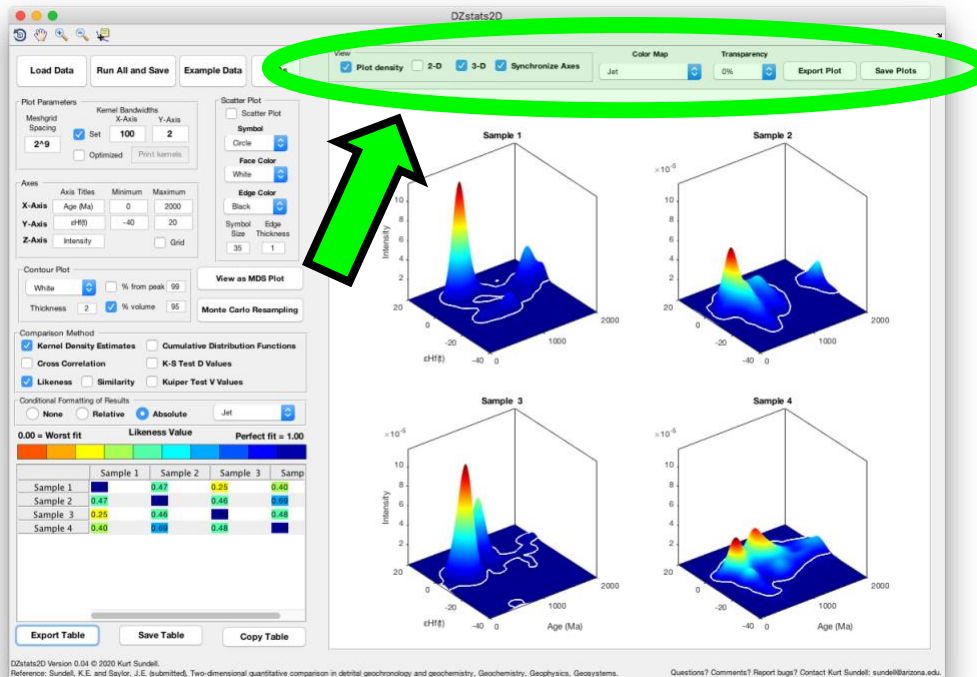


Exporting the table is a good way to quickly compare results. Using relative conditional formatting is a good way to look for similar patterns among the different measures.

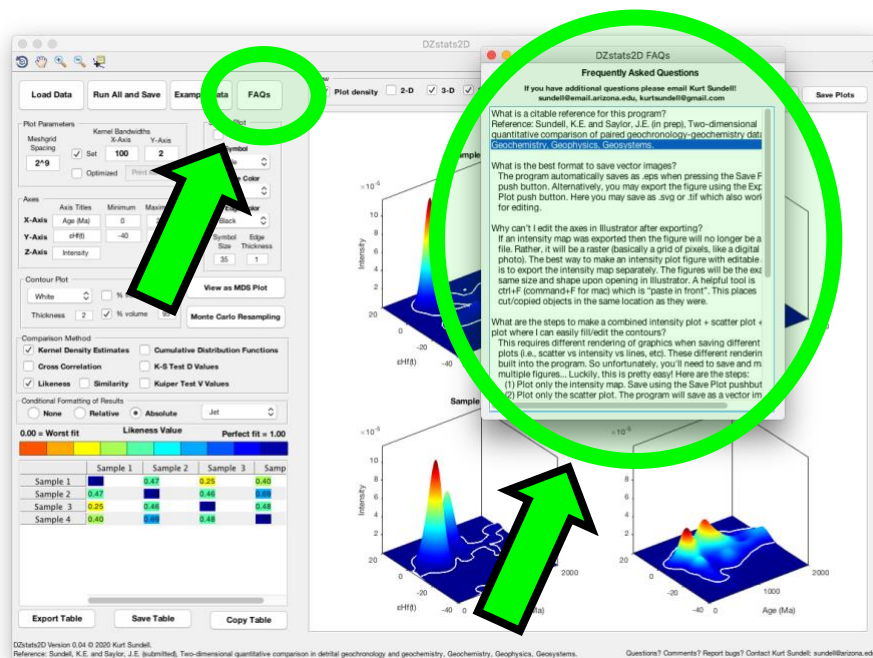
Comparison Matrices with Absolute Conditional Format

	1	2	3	4	5
1	DZstats2D Re...				
2					
3	Cross-correlat...				
4		Sample 1	Sample 2	Sample 3	Sample 4
5	Sample 1	0.00	0.31	0.00	0.03
6	Sample 2	0.31	0.00	0.24	0.54
7	Sample 3	0.00	0.24	0.00	0.35
8	Sample 4	0.03	0.54	0.35	0.00
9					
10	Likeness				
11		Sample 1	Sample 2	Sample 3	Sample 4
12	Sample 1	0.00	0.47	0.25	0.40
13	Sample 2	0.47	0.00	0.46	0.69
14	Sample 3	0.25	0.46	0.00	0.48
15	Sample 4	0.40	0.69	0.48	0.00
16					
17	Similarity				
18		Sample 1	Sample 2	Sample 3	Sample 4
19	Sample 1	0.00	0.78	0.53	0.71
20	Sample 2	0.78	0.00	0.77	0.00
21	Sample 3	0.53	0.77	0.00	0.00
22	Sample 4	0.71	0.00	0.00	0.00
23					
24	K-S Test D				
25		Sample 1	Sample 2	Sample 3	Sample 4
26	Sample 1	0.00	0.44	0.56	0.45
27	Sample 2	0.44	0.00	0.32	0.27
28	Sample 3	0.56	0.32	0.00	0.41
29	Sample 4	0.45	0.27	0.41	0.00
30					
31	Kuiper Test V				
32		Sample 1	Sample 2	Sample 3	Sample 4
33	Sample 1	0.00	0.44	0.59	0.45
34	Sample 2	0.44	0.00	0.35	0.29
35	Sample 3	0.59	0.35	0.00	0.41
36	Sample 4	0.45	0.29	0.41	0.00

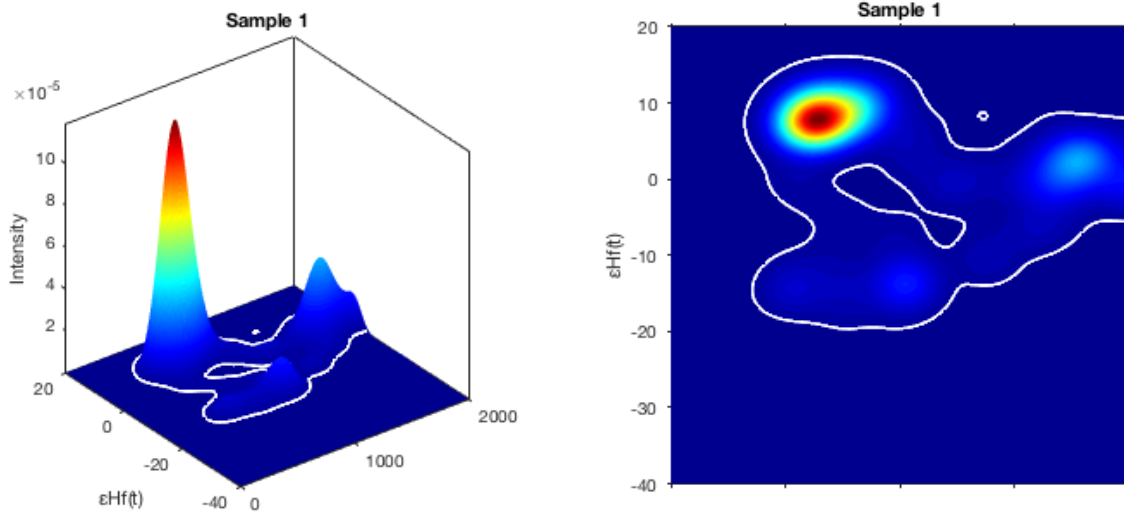
## View options and exporting and saving plots



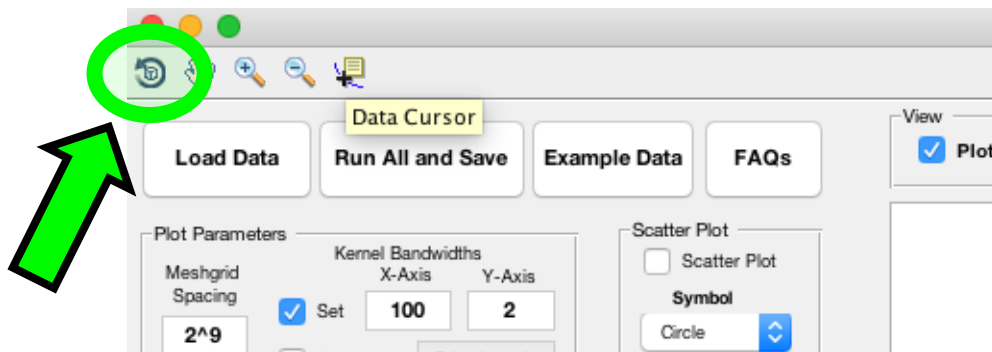
Plot Density controls whether or not the colored density plots is shown. This is important for making complex figures in Adobe Illustrator or a similar drafting program because the density plots are best exported and saved as raster images and the axes and contour and scatter plots (if included) are best saved as vector images (see more details below or in the FAQ button in the top left of the program).



The 2D or 3D options control if the plots are in perspective view (3D) or plan view (2D, i.e., parallel to the z-axis).

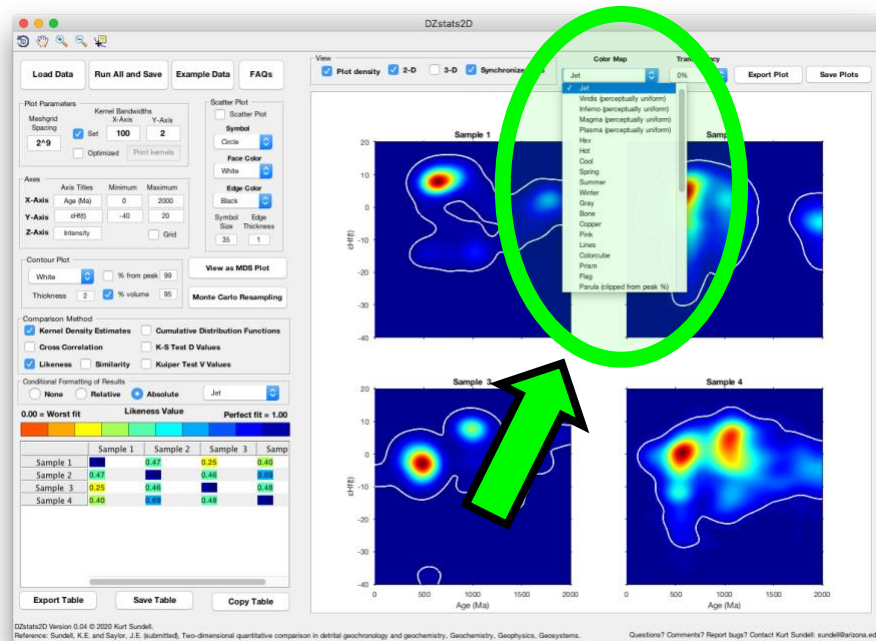


Synchronize Axes will make all axes have the same view when rotating using the rotation tool. To use the rotation tool, select the tool and click and move the plots.

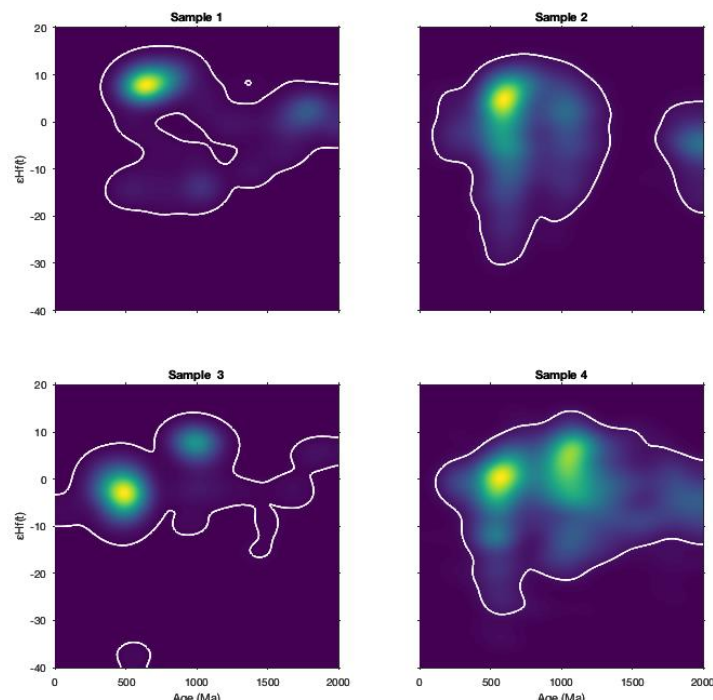


There are a variety of colormaps built into DZstats2D. by default, Jet is used. The Jet colormap is not good for persons with colorblindness. DZstats2D includes perceptually uniform colormaps from Ander Biguri, who improved them from Python:

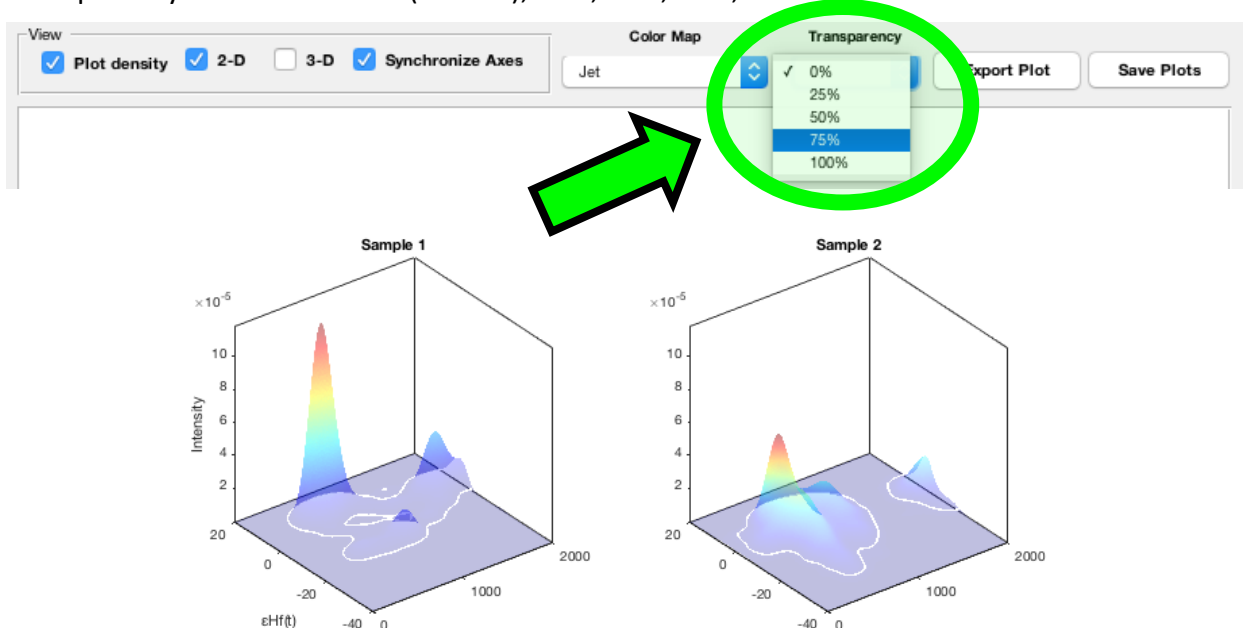
<https://www.mathworks.com/matlabcentral/fileexchange/51986-perceptually-uniform-colormaps>.



Viridus colormap:



Transparency can be set to 0% (default), 25%, 50%, 75%, or 100%:



Export plot will produce a new, resizable figure. This can be saved in a variety of formats.

Save plot will save the current figure in .eps format. If Plot Density is checked, the figure will no longer be a vector file. Rather, it will be a raster (basically a grid of pixels, like a digital photo). The best way to make a figure with editable axes is to export the density map separately. The figures will be the exact same size and shape upon opening in Adobe Illustrator (AI). A helpful tool in AI ctrl+F (command+F for mac) which is “paste in front”. This places cut/copied objects in the same location as they were.

How to make a figure in AI that combines a density map, scatter plot, contours, and editable axes? This requires different rendering of graphics when saving different types of plots (i.e., scatter vs intensity vs lines, etc). These different renderings are built into the program. Here are the steps:

- (1) Plot only the intensity map. Save using the Save Plot pushbutton.
- (2) Plot only the scatter plot. The program will save as a vector image.
- (3) Plot only the contours. The program will save as a vector image, then simplify the polygons for ease of editing. You will thank yourself for doing this separately! Otherwise it's a pain to fill them in Illustrator.
- (4) Open the intensity plot in Illustrator. Add a clipping mask or white boxes to cover the uneditable axes.
- (5) Open the Contour Plot. Copy and "paste in front" (ctrl+f or command+f) in the density plot file. Do this in a new layer in Illustrator if you've used a clipping mask.
- (6) Repeat part 5 for the Scatter Plot. I recommend placing the different plots in different layers. Organization is key!