

DZmds v.1.1.1

J.E. Saylor
 Department of Earth and Atmospheric Sciences
 University of Houston
 Houston TX

Contents

1. INTRODUCTION	2
2. EXECUTION	2
3. DATA ANALYSIS	2
3.1. Data Import	2
3.2. Select the Input Uncertainty Level	2
3.3. Select the Basis for Comparison	2
3.4. Select Criterion	3
3.5. Get MDS Dimensions	3
3.6. Input the Desired Number of Dimensions from Step 5	3
3.7. Plot the MDS and Shepard Plots.....	3
4. ACKNOWLEDGEMENTS	4

1. INTRODUCTION

DZmDs is designed to provide rapid, first-order evaluation of detrital geochronological datasets by visualizing the dissimilarity between detrital geochronological datasets as distance in Cartesian coordinates. The software is designed to maximize flexibility: allowing comparison of MDS based on multiple dissimilarity matrices, using multiple criteria, and in multiple dimensions.

The algorithm executes both metric and non-metric non-classical multidimensional scaling (MDS) of a dissimilarity matrix calculated for a group of samples. Metric MDS can be calculated based on stress or squared stress. Dissimilarity can be selected as the complement of the Cross-correlation of the samples' probability density plots (PDPs) or kernel density estimates (KDEs), the Kuiper test D value, or the Kolmogorov-Smirnov V value. Multidimensional scaling can be conducted in multiple dimensions and displayed in either 2 or 3 dimensions

2. EXECUTION

DZmDs requires the free MATLAB Runtime version 8.3 (R2014a) to be installed before running. If the Runtime is not installed, it can be downloaded and installed from: <http://www.mathworks.com/products/compiler/mcr/index.html>. Following Runtime installation, the DZmDs.exe can be run by directly from the file's location. No further installation is required.

3. DATA ANALYSIS

3.1. Data Import

Data should be organized into a single comma-delimited .csv or .xls file with header rows. In each file, each sample is represented by a pair of columns where the first column is the grain mean ages, and the second column is the uncertainty associated with each age (1σ or 2σ sigma). See the Example Data Set for an example of a correctly formatted sample.

EXAMPLE: To compare four samples, 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 (Table 1). Sample 1 would then occupy columns 1 and 2, Sample 2 would occupy columns 3 and 4, Sample 3 would occupy columns 5 and 6, and Sample 4 would occupy columns 7 and 8.

3.2. Select the Input Uncertainty Level

3.3. Select the Basis for Comparison

Cross-correlation of either PDPs or KDEs calculates the dissimilarity matrix as the complement of the Cross-correlation coefficient ($1-R^2$) based on either PDPs or KDEs, respectively (Saylor et al., 2012; Saylor et al., 2013; Saylor and Sundell, 2016). If the bandwidth window is left blank (Fig. 1), KDEs are calculated using the adaptive bandwidth algorithm of Botev et al. (2010). Otherwise, the bandwidth entered is used for all samples. The Kuiper test D value calculates the dissimilarity matrix as the maximum difference of CDF1-CDF2 plus the maximum difference of CDF2-CDF1 for each sample pair. The Kuiper test D value calculates the dissimilarity matrix as the maximum absolute difference of CDF1-CDF2.

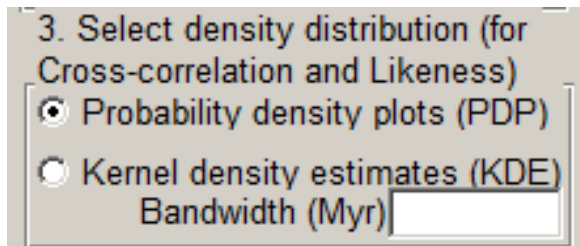


Figure 1. Select density distribution options.

3.4. Select Criterion

Select the criterion to use to calculate and minimize stress. Metric stress (normalized with the sum of squares of the dissimilarities) is the default. Alternatives include metric squared stress, where stress is normalized to the sum of the 4th power of dissimilarities and non-metric stress, where the input dissimilarity matrix is used to order the samples, but stress is normalized to the sum of distances.

3.5. Get MDS Dimensions

This plots the stress versus dimensions for the selected dissimilarity metric. Typically users want to select the either 2- or 3-dimensional analysis in order to minimize both the number of dimensions and the stress (Fig. 2). Therefore the dimension at which the “elbow” occurs is typically the best choice. At this point, adding additional dimensions does not significantly reduce the stress.

Note: if the software returns an error (i.e.: “dings”) when attempting to calculate the Stress versus Dimensions it is unable to complete the calculation for at least one of the dimensions. The preferred solution is to change the criterion (section 3.4). Alternatively, the user can continue to section 3.6, but without the guidance provided by the scree plot. The calculation in section 3.6 may or may not be successful.

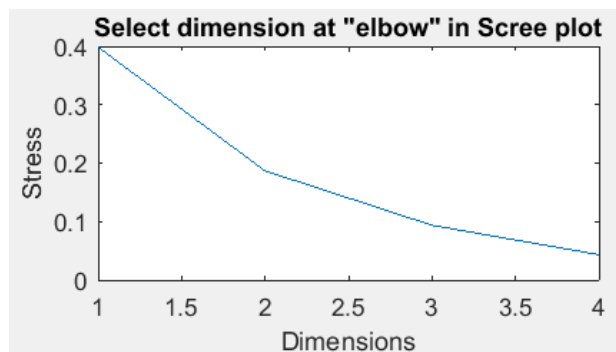


Figure 2. Example of a Scree plot produced by DZmds.

3.6. Input the Desired Number of Dimensions from Step 5

Dimensions must be entered as integers.

3.7. Plot the MDS and Shepard Plots

This will plot the MDS plot in the desired dimension (if the number of dimensions selected in section 3.6 is greater than 3, the results are displayed only in 3 dimensions) as well as the Shepard plot which can be used to evaluate the goodness of fit of the MDS transformation. In the multi-dimensional scaling plot, black arrows point from each sample to its closest

neighbor (the sample most similar to it based on the selected matrix). Grey arrows point from each sample to its second closest neighbor.

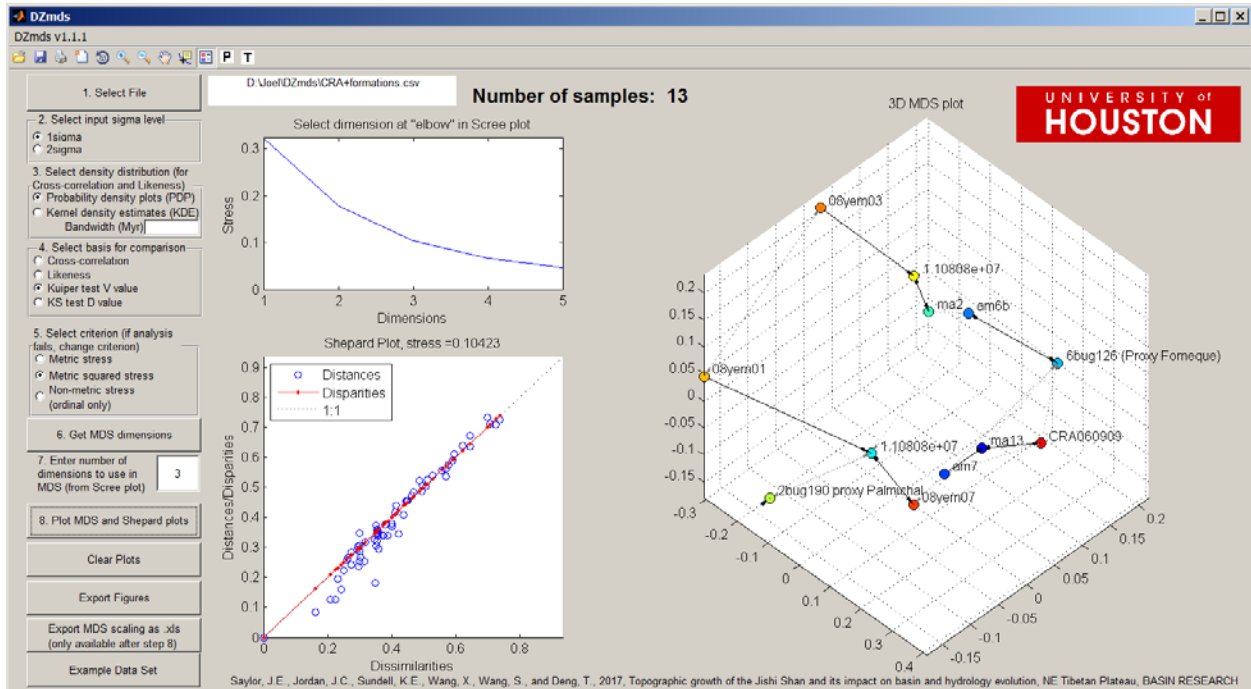


Figure 3. Results of 3-dimensional MDS transformation of the Example Data Set.

4. ACKNOWLEDGEMENTS

This work is only made possible by the research indicated below.

- Botev, Z. I., Grotowski, J. F., and Kroese, D. P., 2010, Kernel density estimation via diffusion: *Annals of Statistics*, v. 38, no. 5, p. 2916-2957.**
- Saylor, J. E., Knowles, J. N., Horton, B. K., Nie, J. S., and Mora, A., 2013, Mixing of Source Populations Recorded in Detrital Zircon U-Pb Age Spectra of Modern River Sands: *Journal of Geology*, v. 121, no. 1, p. 17-33.**
- Saylor, J. E., Stockli, D. F., Horton, B. K., Nie, J., and Mora, A., 2012, Discriminating rapid exhumation from syndepositional volcanism using detrital zircon double dating: Implications for the tectonic history of the Eastern Cordillera, Colombia: *Geological Society of America Bulletin*, v. 124, no. 5-6, p. 762-779.**
- Saylor, J. E., and Sundell, K. E., 2016, Quantifying comparison of large detrital geochronology data sets: *Geosphere*, v. 12, p. 203-220.**
- Vermeesch, P., 2013, Multi-sample comparison of detrital age distributions: *Chemical Geology*, v. 341, p. 140-146.**